CONFIDENTIAL

Statistical Analysis Plan

OPN-375 (fluticasone propionate delivered by exhalation delivery system)

OPN-FLU-CS-3205

A 24-Week Randomized, Double-blind, Placebo-controlled, Parallel-group, Multicenter Study Evaluating the Efficacy and Safety of Intranasal Administration of 186 and 372 µg of OPN-375 Twice a Day (BID) in Subjects with Chronic Rhinosinusitis With or Without the Presence of Nasal Polyps

IND Number:	110089
EudraCT Number:	2019-000368-12
Date of Plan:	14 February 2022
Based on:	Protocol Amendment 6, Date: 07 January 2022
CRF Version 1.0 Date:	10 March 2021

Author: Thomas Dobbins

The undersigned have reviewed and agree on the content of this Statistical Analyses Plan.

Review and Approval Signatures

Name: **John Messina**

Title: **SVP. Clinical Research and Medical Affairs**

Signature: _____

Date: _____

Name: **Thomas Dobbins**

Title: **Author, Statistician, The Griesser Group**

Signature:

Date: 14 February 2022

Table of Contents

1 ABBREVIATIONS

Abbreviation or Term	Definition
AE	Adverse event
AM	Morning
ANCOVA	Analysis of covariance
APOV	Average percent of opacified volume
APOV-E	Average percent of opacified volume in the ethmoid sinuses
APOV-M	Average percent of opacified volume in the maxillary sinuses
BID	Two times a day
BOCF	Baseline Observation Carried Forward
CFR	Code of Federal Regulations
CRF	Case Report Form
CRS	Chronic Rhinosinusitus
CSNS	Composite score of nasal symptoms
CI	Confidence Interval
CT	Computed tomography
DB	Double-blind
DCF	Data Clarification Form
EDS	Exhalation Delivery System
EOS	End of Study
EQ-5D	EuroQol-5
EQ VAS	EQ Visual Analogue Scale
EVA	Electronic visual acuity
ET	Early termination
FAS	Full Analysis Set
GCP	Good Clinical Practice
GLM	Generalized linear model
GLMM	Generalized linear mixed model
HEOR	Health Economics & Outcomes Research
HPQ	Health and Work Performance Questionnaire
IA	Interim analysis
IASAP	Interim analysis Statistical Analysis Plan
ICH	International Conference on Harmonization
INS	Intranasal Steroid
IOP	Intraocular pressure
ITT	Intent-to-Treat
IWRS	Interactive web response system
J2C	Jump-to-Control
LS	Least Squares
LM	Lund-Mackay Staging
MAR	Missing at Random
MCS	Mental component score
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed-effect model for repeated measures
NB	Negative Binomial
PASS 15	Power Analysis and Sample Size Software
PBO	Placebo
PCS	Physical component score
PGIC	Subject Global Impression of Change
PHE	Public Health Emergency
PM	Evening
PMM	Pattern Mixture Model
POV	Percent of Opacified Volume

PRO	Patient Reported Outcome
PSQI	Pittsburgh Sleep Quality Index
QIDS	Quick Inventory of Depressive Symptomatology
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard Deviation
SE	Standard Error
SF-36v2	36-Item Short Form Health Survey Version 2
SF-6D	Short Form-6 dimension
SIT	Smell Identification Test
SNOT-22	22-item Sinonasal Outcomes Test
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
US	United States
WHO	World Health Organization
WHODrug	World Health Organization Drug Dictionary
WPOV-E	Worst Percentage of Opacified Volume in the Ethmoid Sinuses
WPOV-M	Worst Percentage of Opacified Volume in the Maxillary Sinuses
WPOV	Worst Percentage of Opacified Volume among the Maxillary and Ethmoid Sinuses
ZLM	Zinreich Modification of Lund-Mackay Staging
WZLM	Worst ZLM score for the worst sinus among Maxillary and Ethmoid Sinuses

2 INTRODUCTION

This Statistical Analysis Plan (SAP) has been developed after review of Sponsor Protocol OPN-FLU-CS-3205 (Amendment 6.0, dated 07 January 2022) and the corresponding electronic case report form (eCRF). This SAP describes the statistical methods to be used for the analysis and reporting of all efficacy and safety data collected during the conduct of Protocol OPN-FLU-CS-3205.

This SAP supersedes the statistical considerations identified in original Protocol OPN-FLU-CS-3205, as well as the SAP dated 19 April 2021, which is based on Protocol Amendment 4.1 and is the basis for the IASAP dated 8 June 2021. This SAP contains all changes incorporated into Protocol Amendment 6.0, and there are no deviations from this amendment presented in this SAP. If additional analyses are performed to supplement the planned analyses described in this SAP they will be identified as such in the Clinical Study Report (CSR).

This SAP is being written with consideration of the recommendations outlined in the following International Conference on Harmonization (ICH) Guidance for Industry documents:

- E9: Statistical Principles for Clinical Trials
- E9-R1 Addendum: Estimands and Sensitivity Analysis
- ICH E3: Structure and Content of Clinical Study Reports

This SAP also incorporates the recent FDA Guidance regarding the COVID-19 Public Health Emergency (PHE):

- FDA Guidance on the Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency: Guidance for Industry, Investigators, and Institutional Review Boards, (11 May 2020)

3 STUDY OVERVIEW

3.1 General Study Design and Study Schema

This is a 24-week, randomized, double-blind, placebo-controlled, parallel-group, multicenter study designed to assess the efficacy and safety of intranasal administration of OPN-375 (186 and 372 μg twice daily) in subjects with CRS with or without the presence of nasal polyps. In addition, a sub-study will be conducted (in select sites in the US and Canada only) to evaluate clinical biomarkers to characterize the microbiome and cytokine profiles in a subset of subjects at selected study centers as detailed in Protocol Attachment 6.0.

Subjects must sign an informed consent before any study-related procedure is performed. Subjects who meet eligibility criteria at Visit 1 (Screening) will enter a 7- to 21-day single-blind placebo run-in period to determine disease status and to ensure that they can comply with study procedures. At Visit 1 (Screening), nasoendoscopy-related (nasal examination and assessment of nasal cavity for presence of edema, polyps, or mucopurulence) procedures will be performed for all subjects. During the single-blind run-in period, subjects will administer morning and evening doses of the study drug (placebo) and will complete a daily diary (electronic) immediately before the morning and evening doses including recording of both instantaneous (evaluation of symptom severity immediately preceding the time of scoring) and reflective (evaluation of symptom severity over the past 12 hours) scores for nasal symptoms.

Subjects determined to meet diary eligibility criteria over a 7-day period during the first 14 days of the single-blind run-in period, will have a sinus computed tomography (CT) scan performed. The CT scan should be performed after diary eligibility is confirmed and prior to Day 21 of the screening run in period. CT scans performed prior to Day 21 and within 2 weeks prior to Visit 1, as part of routine clinical care, may be utilized if the scan meets the CT scan standards outlined in the Imaging Charter for this protocol. The resultant scan will be reviewed centrally to confirm study eligibility prior to randomization.

Those subjects meeting disease severity eligibility criteria (symptom severity and sinus opacification) will be randomly assigned through an interactive web response system (IWRS) to receive 1 of 2 active treatments or placebo: 186 or 372 μg of OPN-375 or placebo BID for the 24-week double-blind treatment phase. When the number of subjects with polyps reaches 50% of the total population, no additional subjects with polyps will be enrolled.

During the first 12 weeks of the double-blind phase, electronic diaries will be completed twice daily by the subject to capture symptom scores for nasal congestion, nasal discharge, (anterior and/or posterior), facial pain or pressure sensation, and sense of smell, as well as use of approved rescue medication after the Week 4 visit. Subject-completed questionnaires will be used to measure the symptoms and social/emotional consequences of the subject's nasal disorder (SNOT- 22), health- related quality of life (EQ-5D, SF-36v2, and SF-6D), sleep quality and disturbances (PSQI), depressive symptoms (QID), work productivity, and objective smell test (SIT). Subjects will assess their global impression of change since starting the study drug using the PGIC scale.

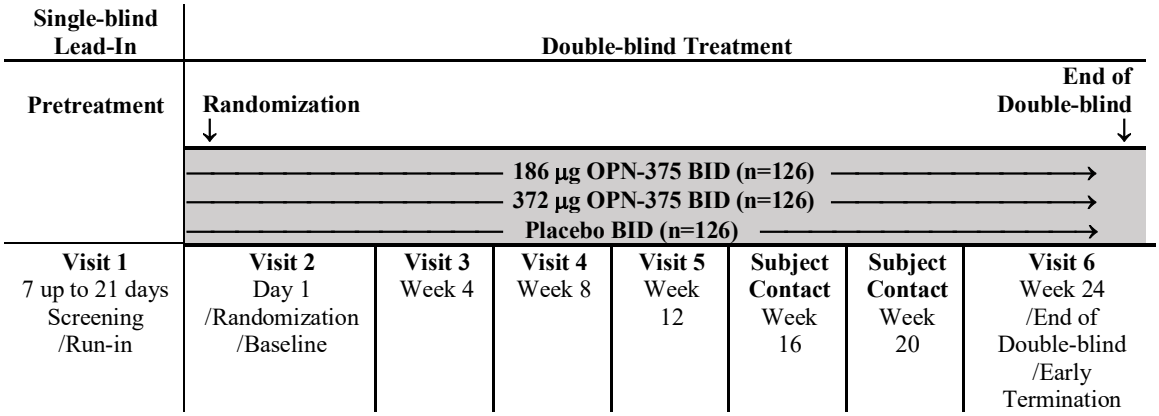
Health economic information related to CRS will also be collected (HPQ). Objective changes in disease severity will be assessed via sinus CT scan.

Safety will be assessed by monitoring of AEs, performing nasal examination, measuring vital signs (i.e., blood pressure, pulse) and weight, and through collection of information for concomitant medications.

Refer to the Schedule of Study Procedures and Evaluations in Table 3.1 for the timing of all tests, procedures, and assessments that will be performed during the study.

A study flow diagram is shown in Figure 3.1

Figure 3.1: Study Flow Diagram



3.2 Study Procedures and Assessments

The study procedures and assessments are outlined in Table 3.1 below.

Table 3.1: Schedule of Study Procedures and Evaluations

Period	Pretreatment (Screening/ Run-in)	Double-blind Treatment						
Visit	1	2	3	4	5	Subject Contact		6**
Procedures and assessments	Single-blind Placebo Run-in (-7 up to -21 days)	Baseline (Randomization) Day 1	Week 4[a] ±7 days	Week 8[a] ±7 days	Week 12[a] ±7 days	Week 16 ±7 days	Week 20 ±7 days	Week 24[a]/ET ±7 days
Informed consent	X							
Demographics	X							
Medical/surgical/smoking history	X							
Inclusion and exclusion criteria	X	X						
Confirm ability to use EDS[b]	X							
Serum chemistry, hematology, urinalysis	X							
Urine drug screen	X							
Blood Sample for SARS-CoV-2 Serology Testing[j]								X
Brief physical examination	X							
Urine pregnancy test[c]	X	X	X	X	X			X
Vital signs and weight	X	X	X	X	X			X
Nasal and sinus endoscopic assessment	X							
CRS treatment and history	X							
Nasal examination[d]	X							X
Nasal polyp grading	X							X
CT with Lund-Mackay scoring	X[e]							X[f]
Exacerbation assessment			X	X	X	X	X	X
Surgical intervention assessment		X						X
SNOT-22		X	X	X	X			X
EQ-5D		X						X
SF-36v2		X						X
PSQI		X			X			X
QIDS		X						X
HPQ		X	X	X	X			X
SIT		X						X
PGIC			X					X
Dispense/collect study drug (OPN-375)	X[g]	X	X	X	X			X
Review proper use of EDS	X	X	X	X	X			
Treatment compliance		X	X	X	X	X	X	X
Provide/collect subject diary	X				X			
Review subject's diary entries		X	X	X	X			

Period	Pretreatment (Screening/ Run-in)	Double-blind Treatment						
Visit	1	2	3	4	5	Subject Contact		6**
Procedures and assessments	Single-blind Placebo Run-in (-7 up to -21 days)	Baseline (Randomization) Day 1	Week 4[a] ±7 days	Week 8[a] ±7 days	Week 12[a] ±7 days	Week 16 ±7 days	Week 20 ±7 days	Week 24[a]/ET ±7 days
Contact IWRS[h]	X	X	X	X	X			X
AE collection[i]	X	X	X	X	X	X	X	X
Prior/concomitant medication/procedures/nondrug therapy	X	X	X	X	X	X	X	X
Contact with subject						X	X	

AE=adverse event; CRF=case report form; DB=double-blind; EDS=Exhalation Delivery System; EOS=End of Study, ET=early termination; HPQ=Health and Work Performance Questionnaire; IWRS=interactive web response system; PGIC=Patient Global Impression of Change; EQ-5D=EuroQol-5; SF-36v2=36-Item Short Form Health Survey version 2; SNOT-22=Sinonasal Outcome Test – 22; PSQI = Pittsburgh Sleep Quality Index; QIDS = Quick Inventory of Depressive Symptomatology, SIT = Smell Identification Test.

**If subject cannot complete EOS CT scan at the Week 24 visit, subject should continue treatment until an EOS CT scan can be completed. If subject continues treatment:

- Complete Subject Contacts approximately every 4 weeks to assess safety until subject can complete the EOS CT scan
- When EOS CT scan has been completed and subject returns to study site, perform additional SNOT-22, and urine pregnancy test (women of childbearing potential only)

[a] All visits up to Visit 6 (Week 24) should be scheduled based on the date of the Day 1 visit. If necessary, visits may be performed within the time window shown; however, subsequent visits should be scheduled based on the date of the Day 1 visit.

[b] The assignment of a kit number for the demonstrator model EDS is not performed by IWRS, rather, the study staff will dispense a demonstrator model EDS kit and record the kit number in the drug dispensing log. This practice EDS is only used at the site at Visit 1 (Screening) and is not to be sent home with the subject.

[c] Women of child-bearing potential only; urine pregnancy tests will be performed at each visit, not including subject contacts. Subjects continuing treatment past Week 24/Visit 6 will need an additional pregnancy test completed upon returning to site.

[d] Nasoendoscopy will be performed using a rigid or flexible endoscope. The examiner must be able to visualize the middle meatus with the scope. Decongestants and/or local anesthetics may be used for the nasoendoscopic procedure.

[e] The CT scan should be performed after diary eligibility is confirmed and prior to Day 21 of the screening run in period.

[f] For the Week 24 study visit, the CT examination should be performed prior to the on-site visit.

[g] Demonstrator model EDS kit (see footnote b) and single-blind placebo dispensation only.

[h] At Visit 1 (Screening), the IWRS will be contacted to obtain the subject ID number, the identifier for the single-blind kit to be dispensed and to track enrollment. IWRS will be contacted at Visit 2 to randomly assign subjects to treatment. At Visit 2 through – Visit 5, IWRS will be used to assign the study drug kit(s) to be dispensed. If a subject cannot complete the EOS CT scan at Visit 6 (Week 24), study staff should not complete Visit 6 in IWRS. Subjects should continue study treatment until an EOS CT scan can be completed. Study staff will dispense additional study medication via the "Unscheduled Visit' functionality in IWRS. IWRS will be contacted for each 'Unscheduled Visit,' at which time one treatment kit will be dispensed.

[I] The period of observation for collection of AEs extends from the time the subject gives informed consent until completion of the double-blind treatment period or an early termination visit. Serious adverse events will be reported through 30 days after the last dose of study drug administration.

[j] Blood sample for SARS-CoV-2 serology test at Visit 6 should be completed unless the subject declines testing.

3.3 Study Objectives

3.3.1 Primary Objective

The primary objective of this study is to compare the efficacy of intranasal administration of twice-daily doses of 186 and 372 µg of OPN-375 (fluticasone propionate) with placebo in subjects with CRS using the following co-primary endpoints:

- change from baseline in AM instantaneous symptoms as measured by a composite score of nasal symptoms (CSNS): congestion, facial pain or pressure sensation, and nasal discharge (anterior and/or posterior) at the end of Week 4

and

- change from baseline to Week 24/ET in the average percent of opacified volume (APOV) in the ethmoid and maxillary sinuses

3.3.2 Key Secondary Objectives

The following key secondary objectives will compare the efficacy of twice daily doses of 186 and 372 µg of OPN-375 with placebo by pooling the data from studies OPN-FLU-CS-3205 and OPN-FLU-CS-3206, which has identical study objectives and study design but only includes subjects with CRS without nasal polyps):

- change from baseline to Week 4 in CSNS in subjects who were symptomatic at trial entry despite reported use of an intranasal steroid for treatment of CRS within 30 days of Visit 1
- frequency of acute exacerbations of CRS over the 24-week treatment period, defined as a worsening of symptoms that requires an escalation of treatment

3.3.3 Other Secondary Objectives

Secondary objectives of this study are to compare the effect of twice-daily doses of 186 and 372 µg of OPN-375 with placebo in:

- change from baseline in AM/PM instantaneous and AM/PM reflective symptoms as measured by a CSNS: congestion, facial pain or pressure sensation, and nasal discharge (anterior and/or posterior) at the end of Week 4, 8, and 12 for the total population, for the CRS with NP and without NP sub-groups, and in subjects with and without previous sinus surgery

- change from baseline to Week 4, 8, and 12 for the total population, for CRS with and without NP, and subjects with and without previous sinus surgery, on facial pain or pressure sensation, nasal discharge (anterior and/or posterior), nasal congestion, and

sense of smell as measured by average AM/PM instantaneous symptom scores and AM/PM reflective symptom scores

- change from baseline to Week 24/ET in APOV in the ethmoid sinuses (APOV-E)
- change from baseline to Week 24/ET in APOV in the maxillary sinuses (APOV-M)
- change from baseline to Week 24/ET in the APOV in the ethmoid and maxillary sinuses for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery
- change from baseline to Week 24/ET in APOV in the ethmoid sinuses (APOV-E) for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery
- change from baseline to Week 24/ET in APOV in the maxillary sinuses (APOV-M) for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery
- change from baseline in the volume of the nasal cavity opacified
- change from baseline to defined time points in subject symptoms and functioning, as measured by SNOT-22 total score and sub-domain scores for the total population and for the CRS with NP and CRS without NP sub-groups
- change from baseline to Week 24/ET in the Lund-Mackay Staging System: total score, scores for the maxillary and ethmoid sinuses, and scores for each sinus pair
- change from baseline to Week 24/ET in percent of sinus volume occupied by disease in the worst maxillary sinus, in the worst ethmoid sinus, and in the worst sinus between maxillary and ethmoid sinuses
- change from baseline to Week 24/ET in percent of sinus volume occupied by disease in the worst maxillary sinus, in the worst ethmoid sinus, and in the worst sinus between maxillary and ethmoid sinuses, in subjects with NP and without NP sub-groups and in subjects with and without previous sinus surgery
- change from baseline to Week 24/ET in the Zinreich modification of Lund-Mackay Staging System total score, scores for the maxillary and ethmoid sinuses combined, and scores for each sinus pair
- change from baseline to Week 24/ET in Zinreich modification of the Lund Mackay Staging System score for the worst sinus between maxillary and ethmoid sinuses
- change from baseline to Week 24/ET in Zinreich modification of the Lund Mackay Staging System score for the worst sinus between maxillary and ethmoid sinuses in subjects with NP and without NP sub-groups and in subjects with and without previous sinus surgery

- time to first acute exacerbation of CRS, defined as a worsening of symptoms that requires escalation of treatment
- change from baseline to Week 24/ET in the 36-Item Short Form Health Survey version 2 (SF36v2) mental component score (MCS)
- change from baseline to Week 24/ET in the SF36v2 physical component score (PCS)
- change in patient-reported outcomes from baseline to Week 24/ET as measured by:
  - change in sleep quality from baseline to Weeks 12 and 24/ET, using the PSQI global and component scores
  - percent of subjects improved as indicated in the Patient Global Impact of Change (PGIC) at Week 4 and Week 24/ET
  - SF-36v2, 8 individual, physically and emotionally based domains (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health) standardized scores
  - change in the severity of depression from baseline to Week 24/ET as measured by the Quick Inventory of Depression Symptomatology (QIDS)
  - change in severity of olfactory symptoms from baseline to Week 24/ET as measured by the Smell Identification Test (SIT)
  - EuroQol-5 (EQ-5D)
  - Short Form-6 dimension (SF-6D) derived scores from the SF-36v2
  - assessment of health economic measures during the double-blind treatment phase related to CRS (eg, criteria for surgical intervention for CRS is met [independent of actual surgery performed], subjects who entered the study approved for surgery [with a scheduled surgery date] who no longer elect to undergo surgery, missed work or school days/lost productivity) by Health and Work Performance Questionnaire (HPQ)

The following secondary objectives will compare the efficacy of twice daily doses of 186 and 372 µg of OPN-375 with placebo by pooling the data from Study OPN-FLU-CS-3205 with Study OPN-FLU-CS-3206, which has identical study objectives and study design but only includes subjects with CRS without nasal polyps):

- change from baseline to Week 24/ET in subject symptoms and functioning, as measured by Sinonasal Outcome Test-22 (SNOT-22) total score in subjects who report using an intranasal, topically acting nasal steroid for the treatment of CRS within 30 days of Visit 1
- change in sleep quality from baseline to Week 24/ET, using the Global Pittsburgh Sleep Quality Index (PSQI) Score

3.3.4 Safety Objectives

To evaluate the safety of OPN-375 by monitoring AEs throughout the study; results of nasal examination, vital signs measurements (i.e., blood pressure, pulse), and weight; and monitoring concomitant medication usage.

3.4 Study Endpoints

3.4.1 Primary Endpoint

Co-primary Endpoints:

- change from baseline to the end of Week 4 in average total instantaneous AM scores (evaluation of symptom severity immediately preceding the time of scoring) of:
  - nasal congestion
  - nasal discharge (anterior and/or posterior)
  - facial pain/pressure sensation

  The baseline CSNS is the average of the total instantaneous AM scores over the last 7 days of the single-blind run-in period, and the end of Week 4, scores are averaged over the 7 days before Week 4.

- change from baseline to Week 24/ET in the APOV in the ethmoid and maxillary sinuses

3.4.2 Key Secondary Endpoints

Key Secondary Endpoints include:

- change from baseline to Week 4 in CSNS in subjects who were symptomatic at trial entry despite reported use of an intranasal steroid for treatment of CRS within 30 days of Visit 1 from studies OPN-FLU-CS-3205 and OPN-FLU-CS-3206

- frequency of acute exacerbations of CRS over the 24-week treatment period, defined as a worsening of symptoms that requires escalation of treatment, from studies OPN-FLU-CS-3205 and OPN-FLU-CS-3206

3.4.3 Other Secondary Endpoints

- change from baseline in AM/PM instantaneous and AM/PM reflective symptoms as measured by a CSNS: congestion, facial pain or pressure sensation, and nasal discharge (anterior and/or posterior) at the end of Week 4, 8, and 12 for the total population, for the CRS with NP and without NP sub-groups, and in subjects with and without previous sinus surgery

- change from baseline to Weeks 4, 8, and 12 for the total population, for the CRS with NP and without NP sub-groups, and in subjects with and without previous sinus surgery:

- in average AM/PM instantaneous (evaluation of symptom severity immediately preceding the time of scoring) and AM/PM reflective (evaluation of symptom severity over the past 12 hours). Averages are based on scores recorded in the diary for the 7 days before each time point for:
    - nasal congestion
    - nasal discharge (anterior and/or posterior) score
    - facial pain or pressure sensation score
    - sense of smell score

- change from baseline to Week 24/ET in APOV in the ethmoid sinuses (APOV-E)

- change from baseline to Week 24/ET in APOV in the maxillary sinuses (APOV-M)

- change from baseline to Week 24/ET in the APOV in the ethmoid and maxillary sinuses for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery

- change from baseline to Week 24/ET in APOV in the ethmoid sinuses (APOV-E) for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery

- change from baseline to Week 24/ET in APOV in the maxillary sinuses (APOV-M) for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery

- change from baseline to Week 4, 8, 12 and 24/ET in the SNOT-22 total and sub-domain scores for the total population and for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery

- change from baseline to the Week 24/ET for CT related assessments:
    - Lund-Mackay Staging System total score, scores for ethmoids and maxillary sinuses combined, and scores for each sinus pair (eg, ethmoid, maxillaries, frontal, and sphenoid, and ostiomeatal complex),
    - percent of sinus volume occupied by disease in the worst maxillary sinus (WPOV-M), percent of sinus volume occupied by disease in the worst ethmoid sinus (WPOV-E), and percent of sinus volume occupied by disease in the worst sinus between maxillary and ethmoid sinuses (WPOV)
    - WPOV-M, WPOV-E, WPOV in subjects with NP and without NP sub-groups and in subjects with and without previous sinus surgery

  - Zinreich modification of the Lund-Mackay Staging System total score, scores for ethmoids and maxillary sinuses combined, and for each sinus pair (eg, ethmoids, maxillary, frontal, and sphenoid)
  - Zinreich modification of the Lund-Mackay Staging System score for the worst sinus between maxillary and ethmoid sinuses in the total population and in subjects with NP and without NP sub-groups, and in subjects with and without previous sinus surgery
  - Change from baseline in the volume of the nasal cavity opacified

- time to first acute exacerbation of CRS, defined as a worsening of symptoms that requires escalation of treatment
- change from baseline to Week 12 and Week 24/ET in the PSQI global and component scores.
- percent of subjects indicating improvement on the PGIC at Week 4 and Week 24/ET
- QIDS Severity of Depression at Week 24/ET
- change from baseline to Week 24/ET for:
  - the MCS of the SF36v2
  - the PCS of the SF36v2
  - SF-36v2 individual domains
  - QIDS total score
  - SIT scores
  - EQ-5D VAS and dimensions
  - SF-6D derived scores

The following secondary endpoints are defined by pooling the data from Study OPN-FLU-CS-3205 and Study OPN-FLU-CS-3206:

- change from baseline to Week 24/ET in SNOT-22 total score in subjects who report using an intranasal, topically acting nasal steroid for the treatment of CRS within 30 days of Visit 1
- change from baseline to Week 24/ET the Global PSQI Score

3.4.4 Safety Endpoints

Safety Endpoints include:

- Prior-concomitant and concomitant medications
- Vital signs and weight
- Nasal and endoscopic assessments
- Nasal examinations
- Adverse Events

3.4.5 HEOR Endpoints and Assessments

Health economic information related to chronic rhinosinusitis will be collected during the study and will include information on criteria for eligibility for escalation of care to surgical intervention (independent of actual surgery performed) for subjects approved for surgery, who no longer elect to undergo surgery, and missed work or school days/lost productivity. The investigator will ask the subject if he/she would consider surgery and the subject will complete the Health Productivity Questionnaire (HPQ).

Objective criteria for being eligible for surgical intervention are listed below for subgroups of subjects with and without polyps.

Subjects **with polyps** present at baseline who meet the following criteria:

- used a topical intranasal corticosteroid (≥8 weeks duration)
- used a course of systemic corticosteroid (1- to 3-week duration)
- SNOT-22 total score ≥20
- Lund-Mackay CT score ≥1

Subjects **without polyps** present at baseline who have:

- used topical intranasal corticosteroid (≥8 weeks duration)
- used either a short course of broad-spectrum/culture-directed systemic antibiotic (2 to 3 weeks duration) or a prolonged course of systemic low-dose anti-inflammatory antibiotic (i.e., macrolide or trimethoprim/sulfamethoxazole) (≥12 weeks duration)
- SNOT-22 total score ≥20
- Lund-Mackay CT score ≥1

The World Health Organization (WHO) Health and Performance Questionnaire (HPQ) is a work productivity questionnaire which asks about employment status, work absences, and on-the-job productivity. Based on HPQ work productivity questions, a comparison of health economic measures during the double-blind treatment phase related to CRS will include:

- percent of subjects indicating that they are willing to consider sinus surgery at baseline and Week 24/ET

- percent of subjects who meet the minimal objective criteria for surgical intervention (as defined in Section 7.6.1) at baseline and Week 24/ET

- percent of subjects approved for surgery who no longer elect to undergo a surgery Week 24/ET

- number and percentage of missed work days and the percentage of productive hours lost while at work from the HPQ

3.5 Investigational Product, Dosage and Mode of Administration

For the single-blind, run-in period, subjects will be dispensed a single-blind kit at Visit 1 (Screening). Eligible subjects will receive a double-blind study drug kit at Visit 2, Day 1 (Baseline). Thereafter, at scheduled visits (every 4 weeks through Week 8 during the double-blind treatment phase) subjects will receive 1 kit. At Week 12, subjects will receive 3 kits.

If subject cannot complete the EOS CT scan at Week 24, study staff should not complete Visit 6 in IWRS. Subjects should continue study treatment until an EOS CT scan can be completed, study staff will dispense additional study treatment via the "Unscheduled Visit" functionality in IWRS. IWRS will be contacted for each "Unscheduled Visit," at which time one treatment kit will be dispensed.

The IWRS will assign study drug kit numbers as described in the Protocol Section 3.3. Study drug packaging is described in Protocol Section 5.3.

As shown in Table 3.2, the study drug dosage during the double-blind treatment phase will be 186 µg or 372 µg of OPN-375 BID or placebo BID.

Table 3.2: Study Drug Dosing

	OPN-375 Dosage											
	186 µg Twice a Day				372 µg Twice a Day				Placebo			
Time	AM		PM		AM		PM		AM		PM	
Bottle	1	2	1	2	1	2	1	2	1	2	1	2
Left nostril	PBO	93	PBO	93	93	93	93	93	PBO	PBO	PBO	PBO
Right nostril	PBO	93	PBO	93	93	93	93	93	PBO	PBO	PBO	PBO

BID=twice a day AM=morning; PBO=placebo; PM=evening.

3.6 Method of Treatment Assignment or Randomization

Subjects will be randomly assigned to treatment according to a computer-generated randomization code. Randomization will be coordinated centrally through an IWRS. The system will provide subject identification numbers at Visit 1 (Screening), which are subsequently linked to the treatment assignments at randomization. Following completion of all baseline evaluations on Day 1, subjects who meet all eligibility requirements will be randomly allocated to 1 of the 3 treatment groups using a 1:1:1 ratio:

- OPN-375 186 µg BID

- OPN-375 372 µg BID
- Placebo BID

Randomization will be stratified by:

- presence of nasal polyps of grade 1 or larger at baseline (Present vs Absent)
- previous sinus surgery (Yes vs No) (previous sinus surgery is defined as evidence of ethmoidectomy [partial or total], or maxillary antrostomy).

An IWRS will be used in this study to track enrollment, randomly assign subjects to treatment groups, and manage study drug supplies. At Visit 1 (Screening) and at each double-blind treatment phase study visit through Week 24, site staff will contact the IWRS and provide requested information to uniquely identify the subject. The IWRS will then provide the identifier (kit number) for the single-blind placebo kit at Visit 1 (Screening). The IWRS will then provide the identifier (kit number) for the double-blind study drug kit(s) to be dispensed at the respective double-blind visit.

4 GENERAL CONSIDERATIONS FOR STATISTICAL ANALYSIS

4.1 Unblinding and Responsibility for Analyses

This trial is conducted as a double-blind study. During the conduct of the study, the subject, investigator, and study personnel at each center, and the Sponsor or its designated personnel directly involved in the clinical study will remain blinded to study treatment. The Investigator will not be provided with the randomization scheme. The randomization scheme will be maintained within the Interactive Web Response System (IWRS).

The official, final database is the responsibility of OptiNose, US, Inc. and its designated personnel, and will not be unblinded until medical review has been performed, protocol deviations and per-protocol population have been identified, and data have been declared final and complete.

If a medical emergency occurs and a decision regarding the subject's condition/treatment requires knowledge of the treatment assignment, the study blind may be broken for the specific subject via the IWRS; the Investigator will immediately notify the Medical Monitor of the situation. The date, time, and reason for un-blinding will be documented in the source document and in the appropriate section of the CRF. Additionally, the documentation received from the IWRS indicating the blind break will be retained in a secure manner, in the subject's source documents.

A partial unblinding of specified data and personnel will occur for an interim analysis of this study. The plan, procedures, and methods for this interim analysis are the subject and scope of a separate IASAP, and are summarized briefly in Section 4.5.

The statistical analysis of data from this trial, as specified in this SAP, will be the responsibility of OptiNose US, Inc. and its designated personnel.

4.2 Design considerations for Sample Size and Power

Sample size assumptions for CSNS were based on previous Phase 3 studies of OPN 375 in subjects with nasal polyposis. Sample size assumptions for APOV were based on data from a study evaluating the change in maxillary and ethmoid opacification in chronic sinusitis patients treated with an intramuscular dose of triamcinolone (Pallanch 2013).

Based on the results for changes from baseline in morning symptoms among subjects with nasal polyps in the previous studies conducted with OPN 375, a sample of 111 subjects per group (333 total subjects) is sufficient to detect a -0.8 scale difference in morning symptoms score between treatments (OPN 375 vs placebo) using a 2-sided test at the 5% significance level with 88% power, and assuming a standard deviation of 1.9.

This sample size is also adequate to detect an -11.3% difference between treatments in average percent opacification using a 2-sided test at the 5% significance level with 88% power, and assuming a standard deviation of 27 percent.

The sample size was estimated for the co primary efficacy endpoints using Power Analysis and Sample Size Software (PASS 15) (NCSS, 2017).

4.3 Multi-Centers

Although this is a multi-center study, a central randomization is used with IWRS; therefore, there are no planned analyses by study center, and study site will not be included as a factor in statistical models.

4.4 Randomization Stratification Factors

There are two stratification factors for randomization: 1) presence of nasal polyps of grade 1 or larger at baseline (Present vs Absent) and 2) previous sinus surgery (Yes vs No).

Both factors will be included in the analysis model as categorical covariates. In the event that a stratification error occurs, i.e., a subject is incorrectly stratified, in the primary analyses the subject will be analyzed as randomized (stratified), consistent with the ITT principle. The error will be documented as a major protocol violation and these subjects will be removed from the per-protocol analyses.

4.5 Interim Analysis (IA)

The Interim Analysis Statistical Analysis Plan (IASAP), dated 19 March 2021 (original) and June 8, 2021 (revised) describes the statistical methods to be used for the interim analysis (IA) of the co-primary CSNS and APOV endpoints, respectively, and these IASAPs are considered an addendum to the SAP Version 4.0 dated 19 April 2021. All statistical methods, data handling conventions and rules defined in the SAP Version 4.0 apply to the original and revised IASAP.

Original IASAP (19 March 2021) is the reference for the blinded IA CSNS:

1. A blinded, pooled IA of co-primary endpoint Composite Score of Nasal Symptoms (CSNS) to assess the variance of change from baseline to Week 4 in CSNS. This IA is planned to occur when approximately 50% of randomized subjects have reached Week 4 CSNS endpoint.

Revised IASAP (8 June 2021) is the reference for the IA of APOV: The revised IASAP added a blinded IA of APOV endpoint and made the unblinded IA of APOV conditional on the outcome of this analysis as described below.

2. A blinded, pooled IA of co-primary endpoint change from baseline to Week 24/ET in Average of Percentage Opacified Volume (APOV) to assess variance. This IA is planned to occur after approximately one-third of randomized subjects have completed the study (Week 24/ET). Based on the results of this blinded IA of APOV, an unblinded IA of APOV to assess the conditional probability of rejecting the null hypothesis at the end of the study, given the observed interim result (Conditional Power), will be considered according to the following decision criteria:

    a) If SD <= 29.7%, (i.e. within 10% of target), then no unblinded IA will be performed.

    b) If observed SD > 29.7%, then perform unblinded IA at the earliest feasible time post the blinded IA. This unblinded IA will be performed by an independent, unblinded statistician not involved in the conduct of the study.

As of the date of this final SAP, the blinded IA of both CSNS and APOV co-primary endpoints have occurred and resulted in a decision to not increase sample size based on variability; consequently, the planned unblinded IA of APOV did not occur.

4.5.1 Alpha Spending for the IA and final analyses

Assuming a single unblinded interim analysis of APOV at 1/3 of the planned enrollment with a target sample size of 111 per group, and a futility bound, the Type 1 error probability for the

interim analysis is $\alpha<0.0001$, and the Type 1 error probability for the final analysis is $\alpha=0.0499$ (two-sided test).

Since the unblinded interim analysis of APOV did not occur, no Type 1 error probability was spent; therefore, the Type 1 error probability for the final analysis is $\alpha=0.05$ (two-sided test).

4.6 Considerations for statistical analysis resulting from the COVID-19 PHE

4.6.1 Co-Primary Endpoints

Since the co-primary efficacy variable CSNS is based on electronic diaries used by subjects during this study, as described in Section 9.3.1, there is no anticipated impact of COVID-19 on the subject's ability to complete the diaries per the protocol. Therefore, there is no anticipated impact of COVID-19 on this co-primary endpoint.

For the co-primary endpoint APOV, the endpoint CT scan will occur at the Week 24 visit, or the ET visit if the subjects discontinues the study prior to Week 24. Since subjects must visit the clinic for the CT scan, subjects may be impacted by COVID-19 in a manner similar to other acute illnesses as described in the Protocol, Amendment 6, Section 6. In this circumstance a subject must be symptom-free for at least 28 days prior to undergoing the Week 24/ET CT scan, and consequently the Week 24/ET visit may be delayed. Subjects are instructed to continue on treatment until the EOS CT scan can be completed. There is no anticipated impact of such a delay on the APOV co-primary endpoint.

Thus, there is no practical impact of the COVID-19 PHE on either of the co-primary endpoints.

4.6.2 Secondary Endpoints

The following secondary efficacy variables are PROs which are self-administered by subjects during study visits:

- Sinonasal Outcome Test (SNOT-22) – a subject-completed questionnaire that consists of 22 symptoms and social emotional consequences of their nasal disorder.
- EuroQol-5 (EQ-5D) – a subject-completed 2-page questionnaire including the EQ-5D descriptive system and the EQ Visual Analogue Scale (EQ VAS).
- 36-Item Short Form Health Survey Version 2 (SF-36v2) – a subject-completed, multipurpose, 36-item subject-completed validated questionnaire that measures 8 domains of health.
- Short Form-6 Dimension (SF-6D) – a subject-completed single health state index derived from the 8 domains from the SF-36v2.
- Pittsburgh Sleep Quality Index (PSQI) – a self-rated, validated questionnaire, which assesses sleep quality and disturbances over a 1-month time interval.
- Quick Inventory of Depressive Symptomatology (QIDS) – a subject-completed 16-item questionnaire to assess the severity of depressive symptoms.

- Health and Work Performance Questionnaire (HPQ) – a subject-completed Health economic information questionnaire.
- Smell Identification Test (SIT) – a subject-completed test comprised of 4 booklets each containing 10 microencapsulated (scratch and sniff) odors.
- Patient Global Impression of Change (PGIC) – a subject-completed global impression of change scale assessment questionnaire.

4.6.2.1 Remote Study Visits

A COVID-19 Supplemental Guidance Document for Protocol CS-3205 has been provided to Investigators which includes detailed instructions for handling the PROs where remote visits are necessary.

For any subjects participating in study visits remotely, all PROs required for the visit will be sent by email to the subject or by courier directly to the subject's home for completion prior to the study visit. Site staff are instructed to notify study subjects to complete the self-administered PROs within two days prior to the remote study visit. At the time of this remote visit, the site staff are instructed to collect and document the information provided in the PROs in the patient's chart. Subject-completed hard copies of all PROs will be collected at the subject's next in-person visit or sent to the site by courier.

As subjects will complete the PROs before the virtual study visit, it will prevent the risk of study staff influencing subject responses. The site staff will verify the PROs were completed by the subjects at the virtual visit and the subjects will share their completed PRO responses. The subject's responses will be documented in the subject's chart.

Site staff are instructed to review the PRO responses provided during the virtual study visit with the hard copies of the PROs submitted by the subject after their visit. If any discrepancies are noted, site staff are instructed to update the Electronic Data Capture (EDC) system with the data from the hard copy. The hard copy subject-completed PROs are considered the source record and data from these will be reported in the EDC. Additionally, the study Monitoring Plan will include a requirement for monitors to confirm discrepant data during monitoring visits.

4.6.2.2 Sensitivity Analyses

The number and percentage of virtual visits will be summarized by study visit, overall and by treatment group, in order to assess the extent of the virtual visits in the study. Also, a tabulated listing of discrepancies between verbal responses reported at the virtual visit and reported on the PRO hard copies will be provided for each PRO.

A sensitivity analysis of the impact of virtual clinic visits will depend on the extent of the virtual visits at Baseline and Week 24/ET. If there are not sufficient virtual visits to include a term in the statistical model, then a statistical analysis excluding subjects with a virtual visit at Week 24/ET will be performed. The treatment difference and 95% confidence interval will be reported to assess consistency with the pre-specified analysis result.

If there are a sufficient number of virtual visits to include in a statistical model then a categorical variable for visit type (in-clinic, virtual) and treatment-by-visit type interaction will be included in the statistical model. The treatment difference and 95% confidence interval in each subgroup will be reported to assess consistency with the pre-specified analysis result.

It is assumed that both the number and percentage of data discrepancies between the verbal report and the hard copy report will be small. A statistical analysis excluding subjects with discrepant results at Week 24/ET will be performed. The treatment difference and 95% confidence interval will be reported to assess consistency with the pre-specified analysis result.

Modification of these analyses or other sensitivity analyses of secondary endpoints may be considered, as appropriate, once the extent of the virtual visits is determined.

5 ANALYSIS POPULATION SETS

5.1 Enrolled Subjects

This set will include all subjects screened and who have signed the informed consent.

5.2 Single-Blind Run-In Set

The single-blind run-in set includes all subjects who received at least one dose of placebo EDS in the placebo run-in period.

5.3 Intent-to-treat (ITT) Analysis Set

The ITT analysis set will include all subjects who are randomized regardless of whether they received double-blind treatment.

5.4 Safety Analyses Set

The safety analysis set will include all randomized subjects who receive at least one dose of randomized study drug. The treatment group assignment in this population will be defined by the treatment actually received. The safety population will be used for the analysis of safety.

5.5 Full Analysis Set (FAS)

The FAS will include all subjects in the ITT set who receive at least one dose of randomized study treatment during the double-blinded treatment phase and have a baseline efficacy measurement for APOV endpoint, and also have a baseline value and at least one post-baseline value (7-day average AM instantaneous CSNS score) on or before Week 4 (i.e., Week 2 and/or Week 4) for the CSNS endpoint. Subjects will be included in the treatment group to which they were randomized.

The FAS population will be the primary study population for all efficacy analyses in this study.

5.6 Per Protocol Analysis Set (PPS)

The PPS analysis set excludes subjects from the FAS population due to major protocol deviations with potential to substantively affect the results of the primary efficacy analysis. The PPS population will be used for a sensitivity analysis for primary and key secondary efficacy endpoints.

Determination of the major protocol deviations will be made prior to unblinding of the database and documented in a separate report. Examples may include but are not necessarily limited to:

- Subjects who did not use study medication for more than 15 consecutive days in a given month
- Subjects who used a prohibited medication (e.g., oral steroids)
- Subjects who have sinus surgery during the study
- Subjects with an acute sinus exacerbation, for example as evidenced by taking antibiotics for sinus infection, within 28 days of a study-directed CT scan (screening or Week 24/ET)

6 STUDY POPULATION SUMMARY

All study population summaries will be based on the Safety Analysis Set, (reference Section 5.4), and summarized by treatment group and total population, unless otherwise indicated.

Descriptive statistics for continuous variables include n, mean, standard deviation (SD), standard error (SE) of the mean, median, minimum, and maximum. Descriptive statistics for categorical variables include subject counts and percentages. Percentages are based on the number of subjects in the analysis set for whom there are non-missing data at the specified visit, unless otherwise specified. Counts of "missing" will be presented if at least one treatment group has missing data.

6.1 Subject Disposition

The number of subjects in the Enrolled Set, Single-Blind Analysis Set, ITT Analysis Set, Safety Analysis Set, FAS, PPS will be summarized using descriptive statistics by treatment group and overall.

Subjects who completed the study, and subjects who discontinued the study prior to week 24, will be summarized by treatment group and overall, and reasons for study discontinuation will be summarized.

The denominator for calculating percentages in the Subject Disposition displays will be the ITT Analysis Set.

6.1.1 Subject Disposition with respect to Intercurrent Events

Subject disposition with respect to Intercurrent Events will be summarized for both the primary and supplemental estimands defined in Sections 7.1.3 and 7.2.1, respectively. The efficacy FAS population will be used for the Subject Disposition with respect to Intercurrent Events.

A subject listing of intercurrent events including subject outcomes under each estimand definition will be provided.

6.2 Demographics

The continuous variables of subject age, weight (kg), height (cm), and body mass index (BMI) (kg/m^2), and the categorical variables of subject sex, ethnicity, and race will be summarized.

6.3 Subject Characteristics

The following subject characteristics will be summarized: nasal polyp status at baseline, prior sinus surgery status.

6.4 CRS History

Information pertaining to subjects' chronic sinusitis history will be summarized.

6.5 Prior Medications

All prior medications will be coded using the World Health Organization dictionary of medical codes (WHO Drug, Version December 2018 or later). The incidence of prior medications will be summarized by therapeutic class and preferred term. Subjects are counted only once in each

therapeutic class category, and only once in each preferred term category. Prior medications will include all medications taken on or before the first dose of double-blind treatment. Further details on the definition of prior medications is provided in Section 9.5.1.

6.6 Medical History

A complete medical history will be obtained before randomization to ensure subjects qualify for the study. Medical, surgical and smoking history will be summarized.

6.7 Chronic Rhinosinusitis (CRS) Treatment History

Information pertaining to subjects' CRS treatment history will be summarized.

6.8 Nasal and Sinus Endoscopic Assessment

Results of the nasal and sinus endoscopic assessment performed during the placebo run-in period will be summarized.

6.9 Urine Drug Screening

Drug screen will be performed at the screening/placebo run-in period. Positive findings indicating drug abuse will lead to screen failure. Urine drug screen findings will be listed.

6.10 Protocol Deviations/Violations

Protocol violations and deviations considered to be major will be listed by treatment group. Reference Section 5.6 for definition of Per Protocol Analysis Set.

7 STATISTICAL METHODS FOR EFFICACY ANALYSES

Hypotheses testing and statistical inference for efficacy endpoints and the strategy for control of Type I error probability is described in Section 7.4. All statistical tests are 2-sided tests at the 5% level of significance. Statistical significance is defined as $p < 0.05$ (two-sided test). Nominal p-values will be provided for all other efficacy endpoints not included in the formal testing strategy.

7.1 Primary Estimand, Intercurrent Events, and Estimation

Primary Objective and Population: The primary objective is to compare the efficacy of intranasal administration of twice-daily doses of 186 µg and 372 µg of OPN-375 (fluticasone propionate) with placebo in subjects with documented CRS with or without nasal polyps.

7.1.1 Co-primary Endpoints

1. Change from baseline to the end of Week 4 in the 7-day average AM, instantaneous CSNS; CSNS includes congestion, facial pain or pressure sensation, and nasal discharge (anterior and/or posterior). Further details on the daily nasal symptoms diary and the derivation of the CSNS and the 7-day average are provided in Section 9.3.1.

AND

2. The change from baseline to Week 24/ET in the APOV in the ethmoid and maxillary sinuses.

   The APOV is the average percentage of volume opacified across both maxillary and ethmoid sinuses (all 4 sinuses). Further details on derivation of APOV is provided in Section 9.3.2.

7.1.2 Intercurrent Events

The potential intercurrent events are as follows:

1. Need for systemic corticosteroids for CRS exacerbation or worsening of CRS nasal/sinus symptoms
2. Need for nasal surgery; i.e., study discontinuation for nasal surgery
3. Discontinuation from study treatment

7.1.3 Primary Estimand

The primary efficacy estimand is the difference between each active treatment of twice-daily doses of 186 µg and 372 µg of OPN-375 and placebo for each of the co-primary endpoints.

For the primary estimand, intercurrent events will be addressed as follows:

1. Systemic corticosteroids use for CRS exacerbation or worsening of CRS nasal/sinus symptoms: a composite strategy will be employed with a poor score assigned to this event.

   For CSNS, a value of '9' will be assigned; further detail regarding the assignment of a worst score to the diary data are provided in Section 9.3.1. For APOV, a poor score is defined as follows: for a subject with a baseline APOV ≥ 75%, a value of '100%' is assigned; if baseline APOV < 75%, a poor score is defined as baseline APOV + 25%.

The rationale for assignment of a poor score is provided in Section 7.1.4 below. Further detail regarding the handling of a missing POV in individual sinuses is provided in Section 9.3.2.

Systemic corticosteroid use for other indications will be considered a treatment and study discontinuation event (reference criteria 3c).

2. Nasal surgery: composite strategy with poor scores assigned as above.
3. Treatment Discontinuation: treatment policy strategy; subjects may continue in the study until one of the following three mutually exclusive outcomes:
   a) Subject completes Week 24/ET CT scan. In the absence of an intercurrent event (systemic corticosteroids, nasal surgery) use the endpoint value;
   b) An intercurrent event (systemic corticosteroids for CRS exacerbation or worsening of CRS nasal/sinus symptoms, nasal surgery) occurs: use composite strategy and assign poor score as described above (reference criteria 1, 2);
   c) Subject discontinues the study. Missing data due to study discontinuation following treatment discontinuation will be imputed; further detail on the imputation method is provided in Section 7.1.5 for each of the co-primary endpoints.

7.1.4 Rationale and Definition of Poor Scores for the Primary Estimand

The majority of the clinical endpoints in this study, including the co-primary endpoints, are defined for a range of values which have a defined minimum and maximum value. For these endpoints, a worst score can be defined based on the maximum, or minimum, score in the range depending on which value represents a worst score. The analyses of interest are based on a comparison of the change from baseline values for these endpoints. Given the assignment of a worst score, i.e., the most extreme value possible for an endpoint, the magnitude of the change score depends directly on the baseline value. For subjects with lower baseline scores, reflective of mild or moderate CRS, the resulting change score may be too extreme and not reflective of a change that is observed in the clinic. Consequently, a statistical analysis using these extreme values can result in subjects with unrealistic change scores; even a few such subjects can have a large, unintended influence on the assessment of treatment effect. In this sense, assignment of a worst score can produce a change score that is artificial and arbitrary, and this approach is not recommended (O'Kelly & Ratitch, 2014).

Instead, a 'poor score' (O'Kelly & Ratitch, 2014) is assigned to subjects with intercurrent events (1) and (2) defined above based on the upper 97.5th percentile of the distribution of the change scores observed in the clinic. The definition of a poor score and documentation of pre-specified poor scores assigned to each clinical endpoint is provided in the SAP Addendum, *Blinded Review of Clinical Endpoints in CRS for Assignment of Poor Scores* (26 October 2021).

7.1.5 Estimator and Estimation Method - Primary Efficacy Analysis

Estimator and estimation method (CSNS)**:** For the co-primary endpoint CSNS, the estimator is the mean difference between treatments (active versus placebo) in change from baseline to the end of Week 4 in the 7-day average instantaneous AM CSNS. Estimation is based on a mixed-effects model for repeated measures (MMRM) where changes from baseline to Week 2 and Week 4, in the 7-day average instantaneous AM CSNS, respectively, are the repeated measures. The 7-day windows used to define the 7-day average instantaneous AM CSNS from baseline through Day 84 (Week 12) are provided in Section 9.1.2.1, Table 9.1. The MMRM model will employ Restricted Maximum Likelihood (REML) for parameter estimation and the Kenward-Roger method for calculating the denominator degrees of freedom. An unstructured covariance matrix will be used to estimate within-subject error. For the treatment policy estimand, (reference Section 7.1.3, criteria 3,c), missing data due to study discontinuation will be estimated within the MMRM model under the assumption that the missing data are missing at random (MAR) and that early discontinuation from the study (prior to Week 4) is unlikely to reflect a treatment effect. Contingencies in the event of an issue with MMRM model convergence are described in Section 10.5.

The MMRM model will include categorical effects for previous sinus surgery (Y,N), nasal polyp status (present, absent), treatment (OPN-375 186 µg, OPN-375 372 µg, placebo), Week (2, 4), treatment-by-day interaction, and the continuous covariate: baseline 7-day average total instantaneous AM CSNS, with baseline-by-day interaction. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and p-value will be displayed.

Estimator and estimation method (APOV)**:** For the co-primary endpoint APOV, the estimator is the mean difference between treatments (active versus placebo) in the change from baseline to Week 24/ET APOV. Estimation is based on an ANCOVA model including categorical effects for previous sinus surgery (Y, N), nasal polyp status (present, absent), treatment (OPN-375 186 µg, OPN-375 372 µg, placebo), and baseline APOV. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and p-value will be displayed.

For the treatment policy estimand, (reference Section 7.1.3, criteria 3, c), missing data due to study discontinuation will be imputed with a pattern mixture model (PMM) multiple imputation (MI) using the Jump–to-Control (J2C) method. The PMM assumes that the missing data at Week 24/ET on the control arm (placebo) is MAR, and the missing data on the active treatment arms have the profile of the control arm at Week 24/ET.

Multiple imputation using a PMM with J2C method will be conducted using the sequential modeling method as described in (O'Kelly & Ratitch, 2014). Further detail on the MI implementation, and statistical analysis and inference is provided in Section 10.

7.2 Sensitivity Analyses

7.2.1 Supplementary Estimand

A sensitivity analysis of the co-primary endpoints based on an alternative estimand will be performed.

Supplementary estimand: the difference between each active treatment of twice-daily doses of 186 µg and 372 µg of OPN-375 and placebo in each of the co-primary endpoints, irrespective of systemic corticosteroid use.

That is, for the supplementary estimand, intercurrent events will be addressed as follows:

1. Systemic corticosteroids use for all indications: treatment policy strategy; consider outcomes irrespective of systemic corticosteroid use; subjects may continue in the study until one of the following two mutually exclusive outcomes:

   a) Subject completes Week 24/ET scan, use the endpoint value;

   b) Subject discontinues the study. Missing data due to study discontinuation following treatment discontinuation will be imputed; further detail on the imputation method is provided in Section 7.1.5 for each of the co-primary endpoints.

2. Nasal surgery; i.e., study discontinuation for nasal surgery: composite strategy with poor scores assigned.

3. Treatment discontinuation: composite strategy with poor scores assigned.

For the supplementary estimand, the estimator and estimation methods for the co-primary endpoints, as well as the composite strategy with poor score assignments, are the same as for the primary estimand described above.

7.2.2 ANCOVA on Ranked APOV Values

A sensitivity analysis of the primary estimand and APOV co-primary endpoint will be performed based on ranked APOV values after modification of the primary estimand as follows. For the composite strategy, a worst score of 100% is assigned; for the treatment policy strategy, baseline APOV is assigned (BOCF). Baseline and endpoint APOV values are then ranked. Ranked APOV values will be analyzed with an ANCOVA model including categorical effects for previous sinus surgery and treatment. The method described by Koch, et. al. (1982) using extended Mantel-Haenszel statistics will be used to perform nonparametric treatment comparisons, adjusting for the baseline covariate. Implementation of this method is described in Stokes, et. al., Chapter 7 (2012).

7.2.3 Observed Cases Analysis

A sensitivity analysis will be performed for both co-primary endpoints based on the FAS, observed cases population. For each co-primary endpoint, the observed cases population is defined to include all subjects who have an observed value for that endpoint, irrespective of an intercurrent event, and the observed value is used in the analysis. For both co-primary endpoints, the estimator and estimation methods are the same as described in Section 7.1.5, with the exception that in the observed cases population there are no missing data; consequently, no imputation is required.

7.2.4 Tipping Point Analysis

If the analyses of both co-primary endpoints are statistically significant, then an additional sensitivity analysis of the primary estimand will be performed to determine the smallest constant, δ, applied to the imputed missing data in the active treatment groups, that will yield a mean treatment difference such that the result using the primary analysis model described in Section 7.1.5 is no longer statistically significant. The constant δ represents a worsening of treatment values and the smallest value, say δ^*, that yields a non-significant result for the primary analysis is called the "Tipping Point."

For the co-primary APOV endpoint, a worsening of treatment values means that APOV values in the active treatment groups would increase at Week 24/ET; therefore, $\delta > 0$. The value δ^* provides a measure of the robustness of the primary analysis result. The larger the value of δ^* the more robust are the results of the primary analysis to the impact of missing data, and, conversely, the smaller the value of δ^* the more sensitive the results of the primary analysis are to missing data. Ultimately however, a Tipping Point analysis is a numerical search procedure. A value of the Tipping Point, if reached, needs to be interpreted in the context of its clinical plausibility.

For the treatment policy estimand, (reference Section 7.1.3, criteria 3, c), missing data due to study discontinuation for APOV endpoint will be imputed under MAR assumption in both active treatment and placebo groups, then the constant δ is applied (added) to the imputed values in the active treatment groups. Further detail on the implementation is provided in Section 10.

For the CSNS co-primary endpoint, it is anticipated that there will be an insufficient number of early study discontinuations (discontinuations prior to Week 4) to warrant a formal Tipping Point analysis. If this is not the case, then a Tipping Point analysis will be performed for the CSNS co-primary endpoint using the same approach as described above.

7.3 Estimation - Key Secondary Efficacy Endpoints

The following key secondary endpoints (defined in Section 3.4.2) are included in the strategy for strong control of the Type I error probability in the trial. For the key secondary endpoints, intercurrent events will be addressed as described in Section 7.1.3 for the primary estimand.

The analyses of these key secondary endpoints are defined by pooling the data from Study OPN-FLU-CS-3205 and Study OPN-FLU-CS-3206.

7.3.1 CSNS in Subjects on INS at Baseline – Pooled Analysis

Change from baseline to Week 4 in CSNS in subjects who were symptomatic at trial entry despite reported use of an intranasal steroid (INS) for treatment of CRS within 30 days of Visit 1 will be analyzed using the same MMRM model described for the co-primary endpoint CSNS and the model will include a categorical effect term for protocol (CS-3205, CS-3206).

7.3.2 Acute CRS Exacerbations – Pooled Analysis

How an acute CRS exacerbation is defined and captured in this study is described in Section 9.3.5. The frequency of acute sinus exacerbations over 24 weeks will be compared between OPN-375

dose groups and placebo using generalized linear model (GLM) for data from the negative binomial distribution that is commonly referred to as the negative binomial (NB) regression model. The NB model will include the same model terms as the primary models; in addition, the model will include a categorical effect for protocol (CS-3205, CS-3206), and the logarithm of exposure time as an offset variable.

Regarding intercurrent events, use of systemic corticosteroids is included in the definition of acute exacerbation; subjects who have surgery or study discontinuation events will be censored at the time of the intercurrent event. Subjects who discontinue treatment are followed to study completion or intercurrent event as described in Section 7.1.3.

7.4 Multiplicity - Type I Error Control

The study hypothesis encompasses the primary and key secondary endpoints; therefore, hypothesis tests for these endpoints are included in the strategy for strong control of the Type I error in the study. The familywise Type 1 error probability for the test of the primary hypothesis is α=0.05, since the planned unblinded interim analysis of APOV endpoint did not occur (reference Section 4.5.1).

The strategy for strong, familywise control of the Type I error is based on closed testing procedure as follows.

1. Test each co-primary endpoint for OPN-375 high dose (372 µg) vs placebo: If
   - $p < 0.05$ for both endpoints, then proceed to test low dose comparison; otherwise, stop.
2. Test each co-primary endpoint for OPN-375 low dose (186 µg) vs placebo: If
   - $p < 0.05$ for both endpoints, then proceed to test key secondary endpoints; otherwise, stop.

There are two comparisons for each dose: OPN-375 high dose (372 µg) vs placebo and OPN-375 low dose (186 µg) vs placebo. The order of the testing is as follows:

1. OPN-375 372 µg versus placebo in the change from baseline in the 7-day average instantaneous AM CSNS at the end of Week 4
2. OPN-375 372 µg versus placebo in the APOV in the ethmoid and maxillary sinuses change from baseline to Week 24/ET
3. OPN-375 186 µg versus placebo in the change from baseline in the 7-day average instantaneous AM CSNS at the end of Week 4
4. OPN-375 186 µg versus placebo in the APOV in the ethmoid and maxillary sinuses change from baseline to Week 24/ET

7.4.1 Type I Error Control for Key Secondary Endpoints

There are four planned comparisons among the two key secondary efficacy endpoints. These comparisons are defined based on pooled data from Study OPN-FLU-CS-3205 and Study OPN-FLU-CS-3206. Testing will continue based on closed testing procedure with α=0.05, if both primary hypotheses are met, as follows:

1. OPN-375 372 µg versus placebo in change from baseline to Week 4 in CSNS in subjects who were symptomatic at trial entry despite reported use of an intranasal steroid for treatment of CRS within 30 days of Visit 1

2. OPN-375 186 µg versus placebo in change from baseline to Week 4 in CSNS in subjects who were symptomatic at trial entry despite reported use of an intranasal steroid for treatment of CRS within 30 days of Visit 1

3. OPN-375 372 µg versus placebo in the frequency of acute exacerbations of CRS over 24 weeks

4. OPN-375 186 µg versus placebo in the frequency of acute exacerbations of CRS over 24 weeks

7.5 Estimation - Other Secondary Efficacy Endpoints

All statistical analyses in this section that reference the primary MMRM and ANCOVA models described in Section 7.1.5 will include the same factors as their respective primary models, except for the baseline covariate. The baseline covariate for the specific endpoint described in this section will be used, unless otherwise specified.

In addition, a combined dose group (OPN-375 186 µg, OPN-375 372 µg) is defined in all models of primary and secondary endpoints by including appropriate contrasts, assigning equal weight to each dose, to obtain the inferential statistics for combined doses and for the comparison: combined doses versus placebo.

Intercurrent events for secondary endpoints will be addressed as described in Section 7.1.3 for the primary estimand, unless otherwise indicated. For the treatment policy estimand, (reference Section 7.1.3, criteria 3,c), missing data due to study discontinuation will be handled as described in Section 7.1.5 for the primary MMRM and ANCOVA models, respectively. Further details on the MI method for endpoints based on the ANCOVA model are described in Section 10.

7.5.1 Averages of Daily Nasal Symptom Diary – Individual & CSNS

Details on the daily nasal symptoms diary and the derivation of the CSNS is provided in Section 9.3.1. For the 7-day averages of AM, PM, and instantaneous, reflective individual nasal symptoms scores, the worst score of '3' will be applied to the individual symptom scores for the composite strategy, per the primary estimand. For the 7-day average of AM, PM, and instantaneous, reflective CSNS scores, the worst score of '9' will be applied; further details regarding the assignment of a worst score to the diary data are provided in Section 9.3.1. The 7-day windows used to define the 7-day averages from baseline though Day 84 (Week 12) are provided in Section 9.1.2.1, Table 9.1.

The following individual and CSNS endpoints will be analyzed:

- Changes from baseline to Weeks 4, 8 and 12 in the 7-day average of AM, PM, and instantaneous, reflective individual nasal symptoms scores (16 endpoints at each time point: four individual nasal symptom scores each at four endpoints).
- Changes from baseline to Weeks 4, 8 and 12 in the 7-day average of AM, PM, and instantaneous, reflective CSNS scores (4 endpoints at each time point).

These endpoints will be analyzed using the MMRM model described in Section 7.1.5 with the repeated measures including Week 4, 8, 12. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed for each analysis time point.

7.5.2 SNOT-22 Total Score and Domain Scores

The SNOT-22 instrument is described in Section 9.3.3. For SNOT-22 total score, a poor score is defined as follows: for a subject with a baseline SNOT-22 total score ≥ 92, a value of '110' is assigned; if baseline SNOT-22 total score < 92, a poor score is defined as baseline SNOT-22 total score + 18 for the composite strategy, per the primary estimand. For the individual Domain SNOT-22 scores, the worst value for the respective domain (Table 9.4) will be applied.

Changes from baseline to Weeks 4, 8, 12 and 24/ET in SNOT-22 total scores and each of the 5 Domain scores will be analyzed using the MMRM model described in Section 7.1.5 where the repeated measures include change from baseline in SNOT-22 total scores at Week 4, 8, 12, and 24/ET. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed for each analysis time point.

7.5.2.1 SNOT-22 in Subjects on INS at Baseline – Pooled Analysis

The following analysis is defined by pooling the data from Study OPN-FLU-CS-3205 and Study OPN-FLU-CS-3206. Change from baseline to Week 24/ET in SNOT-22 total score in subjects who report using an intranasal, topically acting nasal steroid (INS) for the treatment of CRS within 30 days of Visit 1 will be analyzed using the same MMRM model described above, and the model will include a categorical effect term for protocol (CS-3205, CS-3206).

7.5.3 Percentage of Opacified Volume

Percentage of opacified volume, and average (APOV) and worst POV (WPOV) are defined in Section 9.3.2. For APOV-E, APOV-M, WPOV, WPOV-E, WPOV-M the following endpoints are analyzed.

7.5.3.1 Average Percentage of Opacified Volume in the Individual Sinuses

Change from baseline in APOV-E and APOV-M will be analyzed with the ANCOVA model described in Section 7.1.5. A poor score is assigned as defined in Section 7.1.3 for the APOV co-primary endpoint. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

7.5.3.2 Percentage of Opacified Volume (POV) in the Worst Sinus

i. Change from baseline to Week 24/ET in the WPOV sinus at baseline.

   For WPOV, a poor score is defined as follows: for a subject with a baseline WPOV ≥ 71%, a value of '100%' is assigned; if baseline WPOV < 71%, a poor score is defined as baseline WPOV + 29%.

ii. Change from baseline to Week 24/ET in the WPOV-E sinus at baseline.

   For WPOV-E, a poor score is defined as follows: for a subject with a baseline WPOV-E ≥ 67%, a value of '100%' is assigned; if baseline WPOV-E < 67%, a poor score is defined as baseline WPOV-E + 33%.

iii. Change from baseline to Week 24/ET in the WPOV-M sinus at baseline.

   For WPOV-M, a poor score is defined as follows: for a subject with a baseline WPOV-M ≥ 69%, a value of '100%' is assigned; if baseline WPOV-M < 69%, a poor score is defined as baseline WPOV-M + 31%.

Changes from baseline to Week 24/ET for (i) WPOV (ii) WPOV-E, and (iii) WPOV-M will be analyzed with the ANCOVA model described in Section 7.1.5. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

7.5.4 Qualitative CT Scan Assessment

The protocol for qualitative CT scan imaging assessments is based on the Image Review Charter (IRC) – Revision C, (29 January 2020). Qualitative CT scan assessments will be performed at baseline and Week 24/ET in accordance with the IRC for Lund-Mackay Staging (LM) and Zinreich Modification of the Lund-Mackay Staging (ZLM). Further detail on LM and ZLM image scoring metrics and adjudication procedure, and the definitions of the analysis variables derived from these scoring metrics is provided in Section 9.3.6.

For LM total score, a poor score is defined as follows: for a subject with a baseline LM total score ≥ 18, a value of '24' is assigned; if baseline LM total score < 18, a poor score is defined as baseline LM total score + 6 for the composite strategy, per the primary estimand.

For ZLM total score, a poor score is defined as follows: for a subject with a baseline ZLM total score ≥ 39, a value of '50' is assigned; if baseline ZLM total score < 39, a poor score is defined as baseline ZLM total score + 11 for the composite strategy, per the primary estimand.

Other LM and ZLM efficacy endpoints will be assigned scores for the composite strategy, per the primary estimand, as listed in Table 7.1.

Table 7.1
Worst Score Assignment for LM and ZLM efficacy variables

Efficacy Variable	LM worst score	ZLM worst score
Ethmoid + Maxillary	12	30
Sinus Pair	4	10

7.5.4.1 Lund-Mackay (LM) Staging

Changes from baseline to Week 24/ET in the LM Total Score, LM scores for ethmoids and maxillary sinuses combined, and the LM score for each sinus pair will be analyzed using the ANCOVA model described in Section 7.1.5. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

A sensitivity analysis of reviewer agreement will be provided as a '3X3' cross-classification table of Reviewer 1 versus Reviewer 2 assigned LM scores.

7.5.4.2 Zinreich Modification of the Lund-Mackay Staging (ZLM)

Changes from baseline to Week 24/ET in the ZLM Total Score, ZLM scores for ethmoid (L, R) and maxillary (L, R) sinuses combined, the ZLM score for each sinus pair (ethmoid, maxillary) and the ZLM score for the worst sinus among maxillary (L, R) and ethmoid (L, R) sinuses (WZLM) will be analyzed using the ANCOVA model described in Section 7.1.5. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

A sensitivity analysis of reviewer agreement will be provided as a '6X6' cross-classification table of Reviewer 1 versus Reviewer 2 assigned ZLM scores.

7.5.5 PSQI Global and Component Scores

The Pittsburgh Sleep Quality Index (PSQI) is a validated, self-rated questionnaire, which assesses sleep quality and disturbances over a 1-month time interval; further details on the PSQI instrument are provided in Section 9.3.7. PSQI is measured at baseline, Week 12, and Week 24/ET. For PSQI Global score, a poor score is defined as follows: for a subject with a baseline PSQI Global score ≥ 17, a value of '21' is assigned; if baseline PSQI Global score < 17, a poor score is defined as baseline PSQI Global score + 4 for the composite strategy, per the primary estimand. For PSQI component scores, the worst score for the component will be assigned.

Changes from baseline to Week 12 and Week 24/ET in PSQI Global score and component scores (reference Appendix, Section 13.1) will be analyzed using the MMRM model described in Section 7.1.5, where the repeated measures include change from baseline scores at Week 12 and 24/ET. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed for each analysis time point.

7.5.5.1 PSQI Global Score – Pooled Analysis

This analysis is defined by pooling the data from Study OPN-FLU-CS-3205 and Study OPN-FLU-CS-3206.

Change from baseline to Week 24/ET in PSQI Global score will be analyzed using the MMRM model described above and the model will include a categorical effect term for protocol (CS-3205, CS-3206).

7.5.6 SF36v2 – MCS and PCS Domains

The SF36v2 instrument is described in Section 9.3.4. For SF36v2 MCS domain score, a poor score is defined as follows: for a subject with a baseline MCS Domain score ≤ 18, a value of '0' is assigned; if baseline MCS Domain score > 18, a poor score is defined as baseline MCS Domain score −18 for the composite strategy, per the primary estimand.

For SF36v2 PCS domain score, a poor score is defined as follows: for a subject with a baseline PCS Domain score ≤ 11, a value of '0' is assigned; if baseline PCS Domain score > 11, a poor score is defined as baseline PCS Domain score −11 for the composite strategy, per the primary estimand.

The change from baseline to Week 24/ET in SF-36v2 PCS and MCS t-scores, respectively, will be analyzed using the ANCOVA model described in Section 7.1.5. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

7.5.7 SF36v2 – Individual Health Domains

The SF36v2 instrument is described in Section 9.3.4. For SF36v2 Individual Health Domain scores, the worst SF36v2 Individual Health Domain score of '0' will be assigned for the composite strategy, per the primary estimand.

The change from baseline to Week 24/ET in SF-36v2 Health Domain scores will be analyzed using the ANCOVA model described in Section 7.1.5. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

7.5.8 PGIC - "Subject" Global Impression of Change

PGIC and endpoints "Improved" and "Much Improved" are defined in Section 9.3.8. PGIC is measured at Weeks 4 and 24/ET. For PGIC endpoints, the worst score of "Not Improved" will be assigned for each endpoint (Table 9.6) for the composite strategy, per the primary estimand.

The proportion of subjects "Improved" or "Much Improved" will be compared between treatments (each active versus placebo) based on a logistic regression MMRM using the generalized linear mixed model (GLMM). The logistic regression MMRM model will include categorical effects for previous sinus surgery (Y,N), nasal polyp status (present, absent), treatment (OPN-375 186 µg, OPN-375 372 µg, placebo), week (4, 24/ET) and treatment-by-week interaction,. Treatment group comparisons will be estimated from the least squares mean difference between treatments from treatment-by-visit interaction and will be presented as odds ratios with accompanying p-

values and 95% confidence intervals. The model-based LS Mean estimate of the percentage response in each treatment group will also be presented.

Model parameters will be estimated using restricted pseudo-likelihood with Newton-Raphson ridging optimization and the Kenward-Roger method for calculating the denominator degrees of freedom. An unstructured covariance matrix will be used to estimate within-subject error. The logistic regression MMRM will be implemented in SAS PROC GLIMMIX.

7.5.9 QIDS-SR Total Score and Severity of Depression Categories

The QIDS-SR Total Score and Severity of Depression categories are described in Section 9.3.9. For QIDS-SR total score, a poor score is defined as follows: for a subject with a baseline QIDS-SR total score ≥ 22, a value of '27' is assigned; if baseline QIDS-SR total score < 22, a poor score is defined as baseline QIDS-SR total score + 5 for the composite strategy, per the primary estimand.

Change from baseline to Week 24/ET for QIDS-SR total scores will be analyzed using the ANCOVA model described in Section 7.1.5. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

The proportion of subjects reporting "None" or "Mild" depression (Table 9.7) will be compared between treatments (each active versus placebo) based on a logistic regression at Week 24/ET. The logistic model will include categorical effects for previous sinus surgery (Y, N), nasal polyp status (present, absent), treatment (OPN-375 186 µg, OPN-375 372 µg, placebo), and baseline QIDS-SR total score. Model-based odds ratios between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

The statistical analyses of the proportion of subjects reporting "None" or "Mild" depression will include all subjects with a baseline and Week 24/ET QIDS-SR total score; no imputation of QIDS-SR depression categories will be performed.

7.5.10 Time to First CRS Exacerbation Event

How an acute CRS exacerbation is defined and captured in this study, as well as the definition of time to first acute CRS exacerbation is described in Section 9.3.5. The analyses of the frequency of acute CRS exacerbation events is described in Section 7.3.2

Subjects who complete the study without an acute CRS event will be censored at Week 24/ET. Regarding intercurrent events, use of systemic corticosteroids is included in the definition of acute exacerbation; subjects who have surgery or study discontinuation events will be censored at the time of the intercurrent event. Subjects who discontinue treatment are followed to study completion or intercurrent event as described in Section 7.1.3.

The Kaplan-Meier method will be used to estimate the time to first acute CRS exacerbation event, and the median, 25th and 75th percentiles of time to first acute CRS exacerbation events (in weeks). A figure for Kaplan-Meier estimates by treatment group and 95% CI for the median time (in weeks) will be presented.

The stratified log-rank test will be used to compare the time to first acute CRS exacerbation event between each active treatment group and placebo including stratification factors sinus surgery

status (Y vs N) and nasal polyp status (present vs absent) at baseline. The nominal p-value from the stratified log-rank test will be presented.

7.5.11 SIT Scores

The Smell Identification Test (SIT) is described in Section 9.3.10. The SIT is administered at baseline and Week 24/ET. The SIT score represents the number of correct answers to the 40 response items. For SIT scores, a worst score of '10' will be assigned, per the primary estimand. A SIT score of '10' is the expected number of correct responses due to chance alone; i.e., a SIT score of '10' is consistent with total anosmia.

Change from baseline to Week 24/ET in SIT scores will be analyzed using the ANCOVA model described in Section 7.1.5. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed.

7.6 Health Economic Outcomes Research (HEOR) Endpoints

7.6.1 Surgical Intervention Assessment (SIA)

The surgical intervention assessment (SIA) is performed at baseline and Week 24/ET. This assessment includes a question to the subject as well as an objective assessment.

For the question, subjects are asked, "If symptoms of CRS were to continue at the current severity level would subject consider undergoing sinus surgery if it were recommended to subject today by subject's physician?"

The Surgical Intervention Assessment (SIA) is based on the following pre-defined, objective criteria used to create a standardized determination of whether or not a subject is eligible for sinus surgery.

a) Subjects **with** polyps present at baseline will have to have all of the following:
   - used a topical intranasal corticosteroid (≥8 weeks duration)
   - used a short-course of systemic corticosteroid (1- to 3-week duration)
   - SNOT-22 total score ≥20
   - Lund-Mackay CT score ≥1

b) Subjects **without** polyps present at baseline will have to have all of the following:
   - used topical intranasal corticosteroid (≥8 weeks duration)
   - used either a short course of broad-spectrum/culture-directed systemic antibiotic (2 to 3 weeks duration) or a prolonged course of systemic low-dose anti-inflammatory antibiotic (i.e., macrolide or trimethoprim/sulfamethoxazole) (≥12 weeks duration)
   - SNOT-22 total score ≥20
   - Lund-Mackay CT score ≥1

A subject who satisfies the criteria in either (a) or (b) above is defined to have the event: eligible for sinus surgery. Only subjects who have SIA completed at both baseline and Week 24/ET will be included in the analysis.

The number and percentage of subjects answering Yes/No to the surgery question, and the number and percentage of subjects who are eligible for surgery (Yes/No) based on SIA within each active treatment group and placebo will be summarized in a '2x2' contingency table representing baseline versus Week 24/ET. McNemar's test will be used to assess the extent to which subjects improve ('Yes' at baseline to 'No' at Week 24/ET) versus worsen ('No' at baseline versus 'Yes' at Week 24/ET) within each treatment group, and the odds ratio (improved versus worsened) and 95% confidence interval will be presented for each treatment group.

The proportion of subjects who meet the SIA eligibility criteria for surgery at Week 24/ET will be compared between treatments with a logistic regression model including categorical effects for previous sinus surgery (Y, N), nasal polyp status (present, absent), treatment (OPN-375 186 µg, OPN-375 372 µg, placebo), and baseline SIA eligibility (Yes, No) as a covariate. Treatment group comparisons will be estimated from the least squares mean difference between treatments and will be presented as odds ratios with accompanying p-values and 95% confidence intervals.

The same logistic model will be used for the proportion of subjects answering 'yes' to the surgery question at week 24/ET, with baseline response as a covariate.

7.6.2 EQ-5D-5L

The EQ-5D-5L is described in Section 9.4.1; subjects complete the EQ-5D-5L at baseline and Week 24/ET. For each of the 5 dimensions, the number and percentage of subjects in each of the 5 levels will be summarized within each active treatment group and placebo in a '5x5' contingency table representing baseline versus Week 24/ET. Only subjects who have completed the EQ-5D-5L at both baseline and Week 24/ET and have complete data for each dimension will be included in the summary.

For the EQ-5D-5L Visual Analogue Scale (VAS), the change from baseline to Week 24/ET VAS will be analyzed using the ANCOVA model described in Section 7.1.5. The model-based LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed. A worst VAS score of '0' will be assigned for the composite strategy, per the primary estimand.

7.6.3 Health and Performance Questionnaire (HPQ)

The World Health Organization (WHO) Health and Performance Questionnaire (HPQ) is a work productivity questionnaire which asks about employment status, work absences, and on-the-job productivity. The HPQ is measured at baseline, and at weeks 4, 8, 12, and 24/ET. HPQ variables will be defined and summarized outside of this analysis plan.

7.6.4 SF-6D

The SF-6D provides a means for using the SF-36 in economic evaluation by estimating a preference-based single index measure for health from these data using general population values. The SF-6D can be used to obtain quality adjusted life years (QALYs) from the SF-36 for use in cost utility analysis. SF-6D will be scored and summarized outside of this analysis plan.

7.7 Per Protocol Analysis

As a sensitivity analysis, the co-primary endpoints will be analyzed using the Per Protocol Set, defined in Section 5.6, according to the primary estimand and estimation methods used in the primary analyses and described in Sections 7.1.1−7.1.5.

7.8 Subgroup Analyses

The following endpoints will be analyzed according to the primary estimand for the following stratification subgroups:

1. Subjects with versus without polyps at baseline
2. Subjects with versus without prior sinus surgery at baseline

These subgroups are defined based on a subject's actual polyp status and actual nasal surgery status at baseline.

7.8.1 Subgroup Analysis of Co-Primary Endpoints

a) Change from baseline to the end of Week 4 in the 7-day average AM, instantaneous CSNS
b) Change from baseline to Week 24/ET in the APOV in the ethmoid and maxillary sinuses.

7.8.2 Subgroup Analysis of Daily Nasal Symptom Scores

a) Changes from baseline to Weeks 4, 8 and 12 in the 7-day average of AM, PM, and instantaneous, reflective individual nasal symptoms scores (16 endpoints at each time point: four individual nasal symptom scores each at four endpoints).
b) Changes from baseline to Weeks 4, 8 and 12 in the 7-day average of AM, PM, and instantaneous, reflective CSNS scores (4 endpoints at each time point) – except primary endpoint covered above.

7.8.3 Subgroup Analysis of SNOT-22 Total and Domain Scores

a) Changes from baseline to Weeks 4, 8, 12 and 24/ET in SNOT-22 total scores and each of the 5 Domain scores.

7.8.4 Subgroup Analysis of APOV in Individual Sinuses

a) change from baseline to Week 24/ET in APOV-E
b) change from baseline to Week 24/ET in APOV-M

7.8.5 Subgroup Analysis of POV in the Worst Sinus

c) change from baseline to Week 24/ET in WPOV
d) change from baseline to Week 24/ET in WPOV-E
e) change from baseline to Week 24/ET in WPOV-M
f) change from baseline to Week 24/ET in WZLM

7.8.6 Models for Subgroup Analyses

The primary estimand and models described in described in Sections 7.1.3-7.1.5 will also include the appropriate terms for treatment-by-subgroup interaction as well as the appropriate contrasts to obtain the within-subgroup inferential statistics. The model-based, within-subgroup LS mean difference between each active treatment group and placebo (active-placebo), 95% CI, and nominal p-value will be displayed, as well as the treatment-by-subgroup interaction nominal p-value.

For the APOV endpoint, the subgroup analysis will be based on the same MI dataset created for the primary analysis.

8 SAFETY ANALYSIS

The Safety Analysis Set defined in Section 5.4 will be used for all safety analyses specified in this section unless otherwise noted. Safety and tolerability will be assessed by clinical review of all relevant parameters including adverse events, laboratory tests, and physical and vital sign measurements.

All safety summaries will be descriptive and will be displayed by treatment group in the following order (left to right): Placebo, OPN-375 186 µg, OPN-375 372 µg, and OPN-375 combined doses.

8.1 Study Drug Exposure and Administration

Study drug exposure will be measured by study duration which is defined in Section 9.1.1. Summary statistics for the duration of drug exposure will be calculated and presented using the Kaplan-Meier method which will display in a figure the proportion of subjects with continuing drug exposure over the duration of the study, by treatment group. A corresponding table will include the summary statistics for duration of exposure, including minimum, maximum, and quartiles, median duration and 95% confidence interval for the median duration.

Compliance is calculated as follows.

$$Compliance = \frac{number\ of\ doses\ of\ study\ medication}{Study\ Duration\ \cdot 4} \cdot 100$$

8.2 Adverse Events

All adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 21.0 or greater. Treatment emergent adverse events (TEAEs) are defined as adverse events with onset dates on or after the first dose of DB study drug. Further details regarding the definition of TEAEs are provided in Section 9.5.2.

All AE summaries will include total number of TEAEs and total number of subjects with TEAEs unless otherwise stated. The incidence of adverse events will be summarized using descriptive statistics by system organ class and preferred term. Subjects are counted only once in each system organ class category, and only once in each preferred term category. For the summaries by severity, subjects are counted at the greatest severity and will not have the number of events included. Additional summaries by severity will provide all adverse events for overall TEAEs and treatment related TEAEs.

An overall summary of adverse events will include number (%) of subjects with TEAEs (and number of AEs), treatment-related TEAEs, SAEs, severe TEAEs, TEAEs that resulted in treatment discontinuation, and fatal AEs; this overall AE summary will be presented for the Safety Analysis set.

Additionally, summaries will be presented for all TEAEs by system organ class and preferred term (overall and by severity). TEAEs determined by the investigator to be treatment-related (overall and by severity) will be summarized. Serious AEs will also be summarized. TEAEs causing permanent discontinuation from treatment will be summarized, and AEs resulting in death will be summarized.

A summary of adverse events that started during the placebo run-in period will be presented for both Safety Analysis set and the Enrolled Subjects set.

Listings for pre-treatment AEs, fatal AEs, serious AEs, and TEAEs leading to discontinuation will also be presented.

8.3 Clinical Laboratory Tests

Blood and urine samples will be collected at Screening (Visit 1) only. Screening laboratory values will be provided in the listings. No summary statistics will be created.

8.4 Urine Pregnancy Test

For women of childbearing potential, urine pregnancy tests will be performed at the site during every visit. Urine pregnancy test results will be provided in a listing.

8.5 Nasal Examinations

The nasal examination will be performed during the placebo run-in period and again at the Week 24/ET visit. The nasal examination will be performed via nasoendoscopy. Sites will be provided with a nasal examination worksheet and a polyp grading sheet. At the nasal examination, the subject is evaluated for epistaxis and non-active bleeding, septal erosion/ulceration/perforation, ulcerations in areas of the nose other than septum, and mucosal candidiasis. A summary of the key safety assessments from the nasal examination is provided in the Protocol, Attachment 3.

Findings of the nasal examination will be collected on the nasal examination CRF. Findings from the nasal examination worksheet that are deemed to be clinically significant by the Investigator will also be recorded on the AE CRF. The nasal examination findings will be summarized with descriptive statistics.

8.6 Ocular Examinations

An ocular examination is an unscheduled visit and be must promptly be performed if a subject experiences any unexplained worsening in vision during the study (e.g. difficulty reading or seeing traffic signs from a distance). The examiner must be a health care provider with expertise in examination and diagnosis of conditions of the eye. Reports will be forwarded to the investigator.

Results of Ocular Examinations will be presented in the Subject Listings.

8.7 Vital Signs

Vitals signs at each visit and changes from baseline to each visit will be summarized using descriptive statistics.

8.8 Physical Examination

A brief physical examination will be performed on all subjects at Screening (Visit 1). Height will be collected. Results of the brief physical examination will be presented in the Subject Listings.

8.9 Concomitant Medications

Concomitant medications include all medications and other treatments taken by the subject during the study, including those treatments initiated prior to the start of the study. Specifically, concomitant medications are defined as non-study medications with a start date on or after the first dose of study medication inclusive. Medications that started prior to the first dose of study medication but continued during treatment will also be defined as concomitant. Further details regarding the definition of concomitant medications are provided in Section 9.5.1.

Concomitant medications will be coded using the World Health Organization (WHO) Drug Dictionary (WHODrug), Version December 2010 or later, and summarized descriptively by treatment group and drug class with subject counts.

Concomitant medications will be summarized by therapeutic class and preferred term for the Safety Analysis Dataset. For each subject, multiple records of the same concomitant medication will be counted only once within each therapeutic class and preferred term.

All concomitant medications will be provided in a listing for the placebo run-in period based on the Safety Analysis set and Enrolled Subjects set.

8.10 Rescue Medications

The incidence of rescue medication use during the placebo-run period will be presented in the Subject Listings.

9 DATA HANDLING RULES AND DEFINITIONS

9.1 Data Handling Conventions

9.1.1 Study Days and Duration

Study Day 1 is the date of first dose of randomized treatment with study drug after randomization; it is typically the date of randomization.

Study days will be numbered relative to the first day of randomized study drug administration.

- Study days will be numbered relative to study start (i.e., ..., –2, –1, 1, 2, ...; with **Day 1** being the start of study drug and **Day –1** being the day before the start of study drug).
- The elapsed time since Day 1 for an event (e.g. onset of an adverse event) will be calculated as [date of event – date of Day 1 + 1].
- For the purpose of converting days to years or months, one year = 365.25 days, and one month = 30.44 days.

Study Duration is defined as follows:

$$(Date\ of\ Double\ Blind\ Treatment\ completion\ or\ discontinuation - Date\ of\ First\ Day\ of\ Study\ Drug) + 1$$

9.1.2 Visit Windows for Repeated Measures Efficacy Analyses

All efficacy measures that are repeated (repeated measures) over the duration of the study and analyzed with statistical models for repeated measures (i.e., linear and logistic MMRM models) will have efficacy windows defined for these analyses.

For efficacy measures that occur only once post-baseline during the study at Week 24/ET, the nominal study visit (Week 24/ET) will be used for the efficacy analyses.

All safety analyses will be based upon nominal visit days.

9.1.2.1 Efficacy Windows for 7-Day Average Daily Nasal Symptom Diaries and CSNS

As discussed in Section 9.3.1 below, electronic diaries will be used by subjects during the single-blind placebo run-in period and during the first 12-weeks of the 24-week double-blind treatment period in this study to capture daily nasal symptoms and the CSNS. Table 9.1 displays the efficacy windows defined for the 7-day average Daily Nasal Symptoms diary and CSNS for each study week from baseline (Visit 2) through Week 12 (Visit 5). Scheduled study visits corresponding to study endpoints are included and highlighted in the table for reference.

The efficacy windows defined in Table 9.1 may yield intermittent missing data due to a missing 7-day average from the daily nasal symptom diary, as described in Section 9.3.1. Missing data will be estimated within the MMRM models under the assumption they are Missing at Random (MAR), conditional on the observed values and the statistical model (Mallinckrodt, Lane, Schnell, Peng, & Mancuso, 2008).

Table 9.1: Windows for 7-Day Average Individual Daily Nasal Symptom Diary and CSNS

Study Visit	Study Week	Nominal Visit Day	Interval (Days)
2	**Baseline**	**1**	**[-7, -1]**
	1	7	[2, 6]
	2	14	[7, 13]
	3	21	[14, 20]
3	**4**	**28**	**[21, 27]**
	5	35	[28, 34]
	6	42	[35, 41]
	7	49	[42, 48]
4	**8**	**56**	**[49, 55]**
	9	63	[56, 62]
	10	70	[63, 69]
	11	77	[70, 76]
5	**12**	**84**	**[77, 83]**

9.1.2.2 Efficacy Windows for Repeated Measures based on Clinic Visits

Several secondary efficacy endpoints in this study are based on efficacy measures that are repeated over several clinical visits; these include the SNOT-22 Total and Domain Scores, PSQI Global and Component Scores, PGIC, and HPQ. Since these endpoints include repeated measures, they are analyzed using MMRM models.

Table 9.2 displays the windows for the efficacy measures at each scheduled clinic visit. If multiple observations for a subject occur within the same visit window, the observation closest to the nominal visit day will be used; in the case of two equidistant observations the later one will be used. The efficacy windows defined in Table 9.2 may yield intermittent missing data due to a missed clinic visit. Missing data will be estimated within the MMRM models under the assumption they are MAR, conditional on the observed values and the statistical model (Mallinckrodt, Lane, Schnell, Peng, & Mancuso, 2008).

Subjects who discontinue from double-blind study prematurely are instructed to fill out the assessments in the Week 24/ET visit form. For efficacy endpoints described above, these data will be windowed as defined in Table 9.2 and used at the defined visit. For example, if a subject has an ET visit on Day 50 the value will be used as the Week 8 visit value, and the subsequent visits Week 12 and Week 24/ET will be missing. If a subject has an ET visit on Day 120, the value will be used as the Week 24/ET visit value and there is no monotone missing data for this subject.

Table 9.2: Windows for Repeated Measures based on Clinic Visits

Visit Number	Nominal Visit Day	Interval (Days)
2: Baseline	1	[-21 , 1]
3: Week 4	28	[2, 42]
4: Week 8	56	[43, 70]
5: Week 12	84	[71, 98]
6: Week 24/ET	168	[≥ 99]

9.1.3 Baseline Value

Baseline is defined as the last measurement prior to receiving the first dose of double-blind treatment. For nasoendoscopy nasal examination, the assessment at Screening (Visit 1) will be used as the baseline assessment; for CT scan, the date of the procedure during the placebo run-in period will be considered the baseline visit. For all other procedures, the assessment at Visit 2 (Day 1) is considered the baseline assessment.

For the individual daily nasal symptom scores and CSNS, the 7-day average of the last 7 days in the run-in period immediately prior to randomization is considered the baseline value. Reference Section 9.3.1 and Table 9.1 above for further details.

9.1.4 Enrollment Date

The enrollment date is the date the subject signed informed consent to be enrolled into the study. i.e., screening visit date (Visit 1).

9.1.5 Week 24/End of Treatment (ET)

Week 24/ET value is the last non-missing post-baseline value for each subject. This value will occur either at Week 24 or earlier if, for example, the subject discontinues the study prior to Week 24.

9.2 Demographic Variable Definitions

9.2.1 Age

Subject age will be calculated as the integer part of the number of years from date of birth to the date of signing the informed consent form, using the following formula:

Age = integer part of [(date of informed consent – date of birth + 1)/365.25].

9.2.2 Body Mass Index (BMI)

Body mass index (BMI) calculation: BMI (kg/m^2) = [weight (kg)] / [height (m)]^2

9.3 Efficacy Endpoint Variable Definitions

9.3.1 Daily Nasal Symptoms Diary Assessments

Electronic diaries will be used by subjects during the single-blind placebo run-in period and during the first 12-weeks of the 24-week double-blind treatment phase in this study to capture daily nasal symptoms (i.e., symptom scores for nasal congestion, nasal discharge [anterior and/or posterior], facial pain or pressure sensation, and sense of smell) and the use of approved rescue medication after the Week 4 visit through Week 12 of the double-blind period. Subjects will be instructed to complete the electronic diary twice daily immediately before dosing (AM and PM).

Each morning and evening subjects will record their nasal symptom severity scores using the nasal symptoms scoring scale (Table 9.3) to evaluate their symptoms both immediately preceding the time of scoring (instantaneous), and based on reflection of symptoms over the past 12 hours, with regard to nasal congestion, nasal discharge (anterior and/or posterior), facial pain or pressure sensation, and sense of smell.

AM and PM assessment scores, and instantaneous and reflective scores, are all considered to be separate and distinct measures of nasal symptoms and will be analyzed separately.

Composite Score of Nasal Symptoms (CSNS): The CSNS is defined as the sum of the nasal congestion, nasal discharge and facial pain/pressure for each type of assessment (instantaneous or reflective) at each time point (AM or PM), resulting in four CSNS per day.

Table 9.3: Nasal Symptoms Scoring Scale

Score	Description
0	None
1	Mild – symptoms clearly present, but minimal awareness, and easily tolerated
2	Moderate – definite awareness of symptoms that is bothersome but tolerable
3	Severe – symptoms that are hard to tolerate, cause interference with activities or daily living

Note: Sense of smell, scored as 0=normal, 1=slightly impaired, 2=moderately impaired, 3=absent.

For statistical analyses, the daily AM and PM, instantaneous and reflective, individual nasal symptoms scores, and the CSNS will be averaged weekly, defined as the average of the daily scores for the 7 study days prior to the nominal day associated with each 7-day week; for example, for Week 2, Day 14 is the nominal day and the 7-day average is calculated based on study days 7 to 13.

Table 9.1 displays the efficacy windows defined for the 7-day average Daily Nasal Symptoms diary and CSNS for each study week from baseline (Visit 2) through Week 12 (Visit 5).

The 7-day average diary score will be derived for each nasal symptom (range 0-3) and CSNS (range 0-9) by time point (AM or PM) and measure (instantaneous or reflective). Thus, there are sixteen 7-day average scores representing the individual nasal symptoms scores: 2 (AM, PM) * 2 (instantaneous, reflective) * 4 nasal symptoms (congestion, discharge, facial pressure, sense of smell), and there are four 7-day average CSNS scores: 2 (AM, PM) * 2 (instantaneous, reflective).

The 7-day average AM, instantaneous CSNS score is the primary efficacy variable in this study.

Missing Diary Data

On any given study day, for time (AM, PM), or measure (instantaneous, reflective), if any of the 3 nasal symptom scores is missing then the CSNS score will be missing for that study day, time, and measure. For both individual symptoms and the CSNS, the average 7-day value will be missing only if there are missing values for all 7 days. For the first week (Study Days 2-6, Table 9.1), the average 5-day value will be missing only if there are missing values for all 5 days.

Intercurrent Events (IE)

If an intercurrent event occurs, under the composite strategy per the primary estimand, a worst score of '3' will be assigned to each of the daily nasal symptoms scores recorded on the day of the IE and for the remaining days in the week in which the IE occurs; that is, the 7-day average score will include actual scores for days up to the IE and then a worst score for diary days recorded on or after the day of the IE. Since CSNS is a composite score as defined above, a worst score of '9' is assigned. Subsequent 7-day average scores will be assigned a worst score for the remainder of the CSNS study period.

9.3.2 Quantitative (Volumetric) Measurement of CT Scan Image

The protocol for quantitative (volumetric) measurement of CT scan imaging is based on the Image Review Charter (IRC) – Revision C, (29 January 2020). Quantitative measurement of Sinus Volume and Opacification Volume will be performed by a trained analyst under the supervision of a Biomedical Engineer using a validated software system as described in the IRC.

The following quantitative variables are defined for each of the four sinuses: Maxillary (Left, Right), Ethmoid (Left, Right).

Sinus Volume (SV): the total volume of the specified sinus cavity regardless of the amount of mucosal inflammation that may be present. Total Sinus Volume will consist of both the soft tissue and air signal within the specified sinus. Sinus Volume will be reported in cubic millimeters.

Opacification Volume (OV): the total volume of the soft tissue and mucosal inflammation as visualized in the specified sinus. Opacification Volume will be reported in cubic millimeters.

Percentage of Opacified Volume (POV): $POV = OV/SV \cdot 100\%$

Missing Volumetric Measurements

Volumetric measurements (SV, OV) of a sinus may be missing due to technical issues that are unrelated to disease measurement; these measurements are reported as 'Unable to Assess.'

If the volumetric measurements in any sinus are missing at baseline, then the subject will be excluded from the FAS population. If either of the volumetric measurements in any sinus is missing at Week 24/ET, then the volumetric measurement for that sinus will be replaced by the baseline volumetric measurement.

Average Percentage of Opacified Volume (APOV)

APOV is the average percentage of opacified volume across both maxillary (L, R) and ethmoid (L, R) sinuses (all 4 sinuses) at each time point (Baseline, Week 24/ET). Larger values of APOV,

i.e., higher percentages of opacified volume, indicate greater severity of CRS; therefore, negative values for change from baseline APOV indicate improvement from baseline.

APOV-E is the average percentage of opacified volume in the ethmoid (L, R) sinuses at each time point (Baseline, Week 24/ET).

APOV-M is the average percentage of opacified volume in the maxillary (L, R) sinuses at each time point (Baseline, Week 24/ET).

Worst Percentage of Opacified Volume (APOV)

WPOV is the worst sinus across the maxillary (L, R) and ethmoid (L, R) sinuses (all 4 sinuses) at baseline. In the event of a tie, the rule is: use the right sinus, then use the ethmoid sinus.

WPOV-E is the worst ethmoid (L, R) sinus at baseline. In the event of a tie, the rule is: use the right sinus.

WPOV-M is the worst maxillary (L, R) sinus at baseline. In the event of a tie, the rule is: use the right sinus.

The endpoint analyzed is the change from baseline to Week 24/ET in the WPOV sinus; this is similarly defined for WPOV-E, WPOV-M.

9.3.3 SNOT-22 Total and Domain Scores

SNOT-22 is a subject-completed questionnaire that consists of 22 symptoms and social/emotional consequences of their nasal disorder. Each item is rated as follows: 0=no problem, 1=very mild problem, 2=mild or slight problem, 3=moderate problem, 4=severe problem, 5=problem as bad as it can be. The SNOT-22 is validated in CRS (Hopkins 2009). The recall period is the past 2 weeks.

Five domain scores will be derived from the SNOT-22 for efficacy evaluation (Table 9.4): Rhinologic Symptoms; Extra-Nasal Rhinologic symptoms, Ear and Facial Symptoms; Sleep Function; Psychological Issues; and the total score (sum of all items).

Table 9.4: SNOT-22 Domain and Total Scores

Domain	Sum of Items	Range
Rhinologic Symptoms	1, 2, 3, 6, 21, 22	0-30
Extra-Nasal Rhinologic Symptoms	4, 5, 6	0-15
Ear/Facial Symptoms	2, 7, 8, 9, 10	0-25
Psychological Dysfunction	14, 15, 16, 17, 18, 19, 20	0-35
Sleep Dysfunction	11, 12, 13, 14, 15	0-25
TOTAL SCORE	**Sum of all 22 items**	**0-110**

9.3.4 SF36v2 and PCS Domain Scores

The SF-36 is the most widely used general health-related quality of life instrument in clinical trials, with a rich and diverse body of psychometric supporting data. The SF-36v2 is a multipurpose, 36-item subject-completed validated questionnaire that measures 8 domains of health: physical functioning, role limitations due to physical health, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, and mental health. The SF-36v2 survey with a 4-week recall will be used. It yields scale scores for each of these 8 health domains, and two summary measures of physical and mental health, respectively: the PCS and MCS.

The SF-36v2 includes algorithms for interval level scoring for all eight scales ranging from 0 (for worse health) to 100 (best possible health as measured by the questionnaire), as well as the same standardized scoring (mean = 50, standard deviation = 10) for the SF-36 summary scores (PCS and MCS).

The SF-36v2 scoring will be performed according to the current SF-36v2 algorithm and specifications for missing data using proprietary software provided by Optum PRO CoRE. The resulting t-scores will be analyzed as described in Sections 7.5.6, 7.5.7.

9.3.5 Acute CRS Exacerbations

An acute exacerbation of chronic rhinosinusitis event is defined as a worsening of symptoms requiring antibiotic or corticosteroid use, specifically:

- An acute worsening of one or more of the cardinal diagnostic symptoms of chronic rhinosinusitis (listed below), lasting at least 3 days, that causes the subject to seek medical care:
  - facial pain or pressure
  - nasal congestion/blockage
  - rhinorrhea
  - reduction in sense of smell

AND

- An escalation of treatment, defined as either initiation of treatment with antibiotics or systemic steroids, or the escalation of treatment involving an unscheduled acute care visit (e.g., emergency room or out subject clinic) or in-subject care, for increased symptoms related to sinusitis.

Date of onset of a CRS exacerbation event is defined by on the earliest onset date reported on the Exacerbation CRF; differentiation between CRS exacerbation events will also be determined from the onset dates reported on the Exacerbation CRF. If CRS exacerbation event start date is less than 7 days after the end date of a previous exacerbation, the entire period will be treated as a single exacerbation.

Duration of double blind period for each subject will be derived for this analysis as:

Duration (days) = (Week 24/ET visit date – Date of Day 1) + 1.

This duration of DB period will be the follow-up (exposure) time for analysis; subjects who did not experience any CRS exacerbation events during the DB period will be assigned '0' for frequency.

The time (weeks) to first acute CRS exacerbation event is defined as:

$$\frac{start\ date\ of\ first\ CS\ exacerbation - date\ of\ randomization + 1}{7}$$

9.3.6 Qualitative CT Scan Image Assessment

The protocol for qualitative CT scan imaging assessments is based on the Image Review Charter (IRC) – Revision C, (29 January 2020). Qualitative CT scan assessments will be performed at baseline and Week 24/ET in accordance with the IRC for Lund-Mackay Staging and Zinreich Modification of the Lund-Mackay Staging.

As described in the IRC, two radiologist reviewers and a third adjudicating radiologist reviewer will independently review and score the CT scan images. When there are discrepant scores between the two primary reviewers, the adjudicating reviewer will review the image and choose a score recorded by one of the primary reviewers.

Specifically, the adjudicated value is assigned as follows, where R3 represents the adjudicating radiologist, and R1, R2 represent the Radiologist reviewers:

- If R1 and R2 agree, then the adjudicated value is the value of R1 and R2.
- If R1 and R2 both have an integer value and they disagree, then R3 reviews the CT scan image and chooses one of those integer values and that is also the adjudicated value.
- If one of R1 or R2 is 'Unable to Assess,' then R3 reviews the CT scan image and records either the integer value of R1 or R2 or 'Unable to Assess.'

Lund-Mackay (LM) Staging

The LM sinus (Lund and Mackay, 1993) score will be graded separately for all 10 sinuses listed in Table 9.5, as well as the ostiomeatal complex, (12 total LM scores). The sinuses are scored: 0 (no abnormality), 1 (partial opacification), or 2 (total opacification), and the ostiomeatal complex is scored: 0 (not obstructed), or 2 (obstructed).

LM Scores for the sinuses will be derived from the Zinreich Modification of the Lund-Mackay Staging System (ZLM) scores, which are defined below.

- ZLM scores of 0 will = 0 on LM
- ZLM scores of 1, 2, 3, or 4 will = 1 on LM
- ZLM scores of 5 will = 2 on LM
- The LM score for the OMC is captured as defined above.

The following variables are defined based on LM scores:

- LM Total Score: sum of 12 LM scores. Therefore, the total LM score for a CT scan ranges from 0-24.
- LM scores for ethmoids and maxillary sinuses combined: The sum of the two ethmoid and one maxillary sinuses listed in Table 9.5, including the left and right sinus for each. The score for the total of six sinuses ranges from 0 to 12.
- LM Score for each Sinus Pair: For each of the 5 sinuses listed in Table 9.5, the left and right sinus pair is summed. Each sinus pair listed in Table 9.5 will be analyzed separately. These variables range from 0 to 4.

Zinreich Modification of the Lund-Mackay Staging (ZLM)

The Zinreich Modification of the Lund-Mackay Staging System (Zinreich 2004), the ZLM score, is also graded separately for all 10 sinuses listed in Table 9.5. The scoring is a follows.

Score 0: 0% opacification of the sinus

Score 1: 1% to 25% opacification of the sinus

Score 2: 26% to 50% opacification of the sinus

Score 3: 51% to 75% opacification of the sinus

Score 4: 76% to 99% opacification of the sinus

Score 5: 100% opacification of the sinus

The following variables are defined based on ZLM scores:

- ZLM Total Score: sum of 10 ZLM scores. Therefore, the total ZLM score for a CT scan ranges from 0-50.
- ZLM scores for ethmoids and maxillary sinuses combined: The sum of the two ethmoid and one maxillary sinuses listed in Table 9.5, including the left and right sinus for each. The score for the total of six sinuses ranges from 0 to 30.
- ZLM Score for each Sinus Pair: For each of the 5 sinuses listed in Table 9.5, the left and right sinus pair is summed. Each sinus pair listed in Table 9.5 will be analyzed separately. These variables range from 0 to 10.
- WZLM score is the score for the worst sinus among the anterior ethmoid (L, R), posterior ethmoid (L, R), and maxillary (L, R) sinuses (all 6 sinuses) at baseline. In the event of a tie, the rule is: use the right sinus, then use the ethmoid sinus, then use the anterior ethmoid sinus. The change in this sinus at Week 24/ET is the endpoint of interest.

Table 9.5: Sinuses (Left and Right) Assessed and Scored

Sinus (Left, Right)	LM Score	ZLM Score
Ethmoid Anterior	✓	✓
Ethmoid Posterior	✓	✓
Frontal	✓	✓
Maxillary	✓	✓
Sphenoid	✓	✓
Ostiomeatal Complex	✓	

Missing Qualitative CT Scan Image Measurements

Qualitative LM and ZLM measurements of a sinus may be missing due to technical issues that are unrelated to disease measurement; these measurements are reported as 'Unable to Assess' through the adjudication process.

Rules for handling missing qualitative ZLM scores in an individual sinus are as follows: If the qualitative CT scan measurements in any sinus are missing at baseline, then the subject will be excluded from the FAS population. If the qualitative CT scan measurement in any sinus are missing at Week 24/ET, then the measurement for that sinus will be replaced by the corresponding baseline measurement. As described above, LM scores are derived from ZLM scores, except for the LM Ostiomeatal Complex which will use this missing data rule as well.

9.3.7 Pittsburgh Sleep Quality Index (PSQI)

PSQI is a validated, self-rated questionnaire, which assesses sleep quality and disturbances over a 1-month time interval. Nineteen individual items scored 0 (no difficulty) to 3 (severe difficulty) generate 7 component scores: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction. The sum of scores for these 7 components yields a global score (range 0 to 21). Higher scores indicate worse sleep quality. Refer to Section 7.5.5 for analysis and Section 13.1, Appendix 1: PSQI Scoring for the scoring algorithm of the PSQI instrument (Buysse, DJ 1989).

9.3.8 PGIC - "Subject" Global Impression of Change

"Subject" Global Impression of Change (PGIC) is assessed based on a 7-point Likert scale as shown in Table 9.6. Two endpoints are defined based on a dichotomization of the scale. Category 1 (Improved) is defined as any improvement (minimally, much, very much); Category 2 (Much Improved) is defined as improvement (much, very much). Reference Section 7.5.8 for analysis.

Table 9.6: Subject Global Impression of Change (PGIC) Scale

Since starting the study drug, how would you rate the change in your symptoms?			
Score	**Description**	**Category 1**	**Category 2**
1	Very much improved	**Improved**	**Improved**
2	Much improved	**Improved**	**Improved**
3	Minimally improved	**Improved**	**Not improved**
4	No change	**Not improved**	**Not improved**
5	Minimally worse	**Not Improved**	**Not Improved**
6	Much worse	**Not improved**	**Not improved**
7	Very much worse	**Not Improved**	**Not Improved**
Missing	Missing	**Not Improved**	**Not Improved**

9.3.9 Quick Inventory of Depressive Symptomatology – Self Report (QIDS-SR)

The 16-item QIDS-SR (Rush et al, 2003) is designed to assess the severity of depressive symptoms. The QIDS-SR is a self-report version and assesses all the criterion symptom domains designated by the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders – 5th Edition to diagnose a major depressive episode.

Sixteen individual items are scored 0 to 3 with lower scores indicating lower severity of depression. The 7-day period prior to assessment is the usual time frame for assessing symptom severity. Severity of depression will be assessed based on the total score (range 0-27) and categories of the severity of depression as shown in Table 9.7.

If any item Q1-Q5, Q10-Q16 is missing, or if either of ((Q6 and Q7) or (Q8 and Q9)) are missing, then the total QIDS-SR score is missing. If either of Q6 or Q7 is answered and either of Q8 or Q9 is answered the questionnaire can be categorized by severity of depression as described in Table 9.7.

The QIDS-SR total score is the summation of the following items:

- The highest score from Q1-Q4
- Response from Q5
- The highest score from Q6-Q9
- Responses from Q10, Q11, Q12, Q13, Q14
- The highest score from Q15 and Q16

Table 9.7: QIDS-SR Severity of Depression Categories

QIDS Score	Severity of Depression
0 – 5:	None
6 – 10:	Mild
11 – 15:	Moderate
16 – 20:	Severe
21 – 27:	Very Severe

9.3.10 Smell Identification Test (SIT)

SIT is a smell identification test assessing olfaction and is comprised of 4 booklets each containing 10 microencapsulated (scratch and sniff) odors, producing 40 multiple choice response items each with one correct and 3 incorrect answers. SIT scores range from 0-40, with a higher score indicating better olfaction and a score of '10' suggesting total anosmia (Doty, et. al., 1984). Forced choice response alternatives accompany each test item. The test provides an indication of smell loss (anosmia; mild, moderate or severe hypersomnia) as well as an index to detect malingering.

9.4 Health Economic Outcomes Research (HEOR) Endpoints

9.4.1 EQ-5D-5L

The EQ-5D-5L is a standardized measure of health status developed to provide a simple, generic measure of health for clinical and economic appraisal.

The five-level EQ-5D (EQ-5D-5L) consists of 2 pages: the EQ-5D descriptive system and the EQ visual analogue scale (EQ VAS). The descriptive system comprises 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. And each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The subject is asked to indicate his/her health state by ticking the box next to the most appropriate statement in each of the 5 dimensions. This decision results in a 1-digit number that expresses the level selected for that dimension. The digits for the 5 dimensions can be combined into a 5-digit number that describes the subject's health state. EQ-5D-5L dimension is scored using algorithms obtained from the EuroQol Group authors.

Each dimension is coded according to the level selected by the subject as follows: 1= no problems, 2=slight problems, 3=moderate problems, 4=severe problems, 5=extreme problems. Missing dimension values and ambiguous values (2 values selected for the same domain) coded as '9' will be changed to missing in the analysis dataset.

The EQ VAS records the subject's self-rated present health on a vertical visual analogue scale, in response to the statement, *"We would like to know how good or bad is your health TODAY."* The endpoints on the VAS are labelled 100: 'The best health you can imagine' and 0: 'The worst health you can imagine'. The VAS can be used as a quantitative measure of health outcome that reflect the subject's own judgement. The VAS is scored as collected and should have a number between

0-100. Missing values coded as '999' for this instrument will be changed to missing in the analysis dataset.

9.5 Therapy and Safety Endpoint Variable Definitions

9.5.1 Prior and Concomitant Medication

A prior medication is defined as any non-study medication that started before the date of first dose of study drug.

Concomitant medication is defined as any non-study medication that:

- Started before the date of first dose of study drug and is ongoing throughout the study or ends on/after the date of first study medication administration.
- Started on/after the date of first dose of study drug and is ongoing or ends during the course of study medication.

The start/stop dates recorded in the CRF will be used to determine when a medication is considered prior-to or after first dose of study drug. A medication can be considered both prior and concomitant if it was started prior to the first dose of study drug and continues to be taken after the first dose of study medication.

Unresolved missing start dates will be handled as follows for determination of concomitance: partial dates will be imputed in such a way as to consider the non-study medication as prior and/or concomitant if it is possible that the missing information could lead to a prior and/or concomitant outcome.

9.5.2 Treatment Emergent Adverse Events (TEAEs)

A treatment emergent adverse event (TEAE) is defined as an adverse event with onset date on or after the first dose of randomized study drug.

All reports of adverse events should include the severity and relationship to study drug that are determined by the investigators. In the event of missing information, the following rules will be applied for the adverse event summaries:

- AEs with missing onset dates will be included as treatment- emergent (unless end date is prior to the first dose date)
- AEs with missing severity will be counted as severe in severity
- AEs with missing relationship to study drug will be counted as related

10 STATISTICAL ANALYSES IMPLEMENTATION SPECIFICATIONS

This section contains the specifications required to implement the statistical methods for efficacy analyses described in Section 7 with a multiple imputation (MI) method for missing data. These specifications are given in sufficient detail such that an independent reviewer can replicate the efficacy analysis and results based on the methods described in Section 7 and the additional MI implementation specifications contained in this section.

These specifications incorporate the guidance for pre-specifying details of MI provided in (O'Kelly & Ratitch, 2014) Chapter 6. Multiple imputation procedures are implemented in a three step process: Step 1: Multiple Imputation, Step 2: Statistical Analyses, Step 3: Statistical Inference. Each of these steps are described for the MI procedures presented herein.

SAS® (SAS Institute), software Version 9 or higher will be used to implement all statistical methods, and to produce all Tables and Figures containing analyses results and summary descriptive statistics.

10.1 Declarations and Multiple Imputation Considerations

a) Random Number Seed: 07091993

   A random number seed is specified for each invocation of an MI procedure in SAS. Specifying a seed enables the numeric results of the multiple imputation procedures to be reproduced.

b) Number of multiple imputations (M): **40**

   This is the number of imputation datasets that will be generated for all imputation procedures. M=40 is chosen since the fraction of missing information, and the relative increase in variance, are anticipated to be small.

c) Discreteness of the efficacy variables

   For the efficacy measures that occur only once post-baseline during the study at Week 24/ET, the endpoint value is Week 24/ET. For the following efficacy variables the Week 24/ET value is a discrete random variable: WPOV-E, WPOV-M, WPOV, LM, ZLM, and WZLM scores, QIDS total score, SIT scores. The MI method described below imputes the Week 24/ET value as a continuous variable. Consequently, imputed values will be real numbers; these imputed values will not be rounded to nearest integers as this may introduce bias and is not recommended (O'Kelly & Ratitch, 2014). For each of these efficacy variables, and including APOV, SF36v2 t-scores, and EQ-5D-5L VAS scores as well, imputation values obtained from the MI may be outside the defined range of the variable. If a value outside this range is obtained in the imputation, the imputed value will be truncated at the maximum or minimum value depending on which of these values it exceeds.

10.2 Multiple Imputation Method – PMM with J2C

As described in Section 7.1.3 for the treatment policy estimand, missing data due to study discontinuation will be imputed with a multiple imputation method (MI) for the following efficacy

variables: APOV, WPOV-E, WPOV-M, WPOV; LM, ZLM, and WZLM scores; SF36-v2 scores, QIDS total score, SIT score, and EQ-5D-5L VAS.

Multiple imputation with a PMM and J2C method will be conducted using the sequential modeling method as described in (O'Kelly & Ratitch, 2014), Chapter 7. To implement J2C method, first the initial dataset is used to impute missing data in the control group with multiple imputation under MAR assumption. The initial dataset used is the analysis dataset – in this case a dataset with a single post-baseline observation at Week 24/ET; this dataset is referred to as *EFF*. Therefore, MI is performed on a single post-baseline observation with missing data.

The statistical model used to estimate parameters in the imputation of missing data for analysis variable in the control group under MAR assumption includes the following terms: treatment, baseline covariate, nasal polyp status (present, absent), pervious sinus surgery (Y, N). This model is implemented in PROC MI with MONOTONE REGRESSION statement. The output dataset is referred to as *EFF_MI_MAR*. This dataset, after some data manipulation to restore the missing data in the active active treatment arms, is the input dataset for implementation of J2C method in SAS, which involves three steps for imputing a single time point.

Step 1: Remove subjects in the active treatment groups with complete data at Week 24/ET (OUTPUT REST_V6). Keep remaining subjects (OUPUT IMPUTE_V6):

a) Subjects in active treatment groups with data to be imputed at Week 24/ET

b) Control subjects with observed and multiply imputed data under MAR at Week 24/ET

Step 2: Multiple imputation of Week 24/ET missing data (INPUT IMPUTE_V6) using a regression of Week 24/ET analysis variable on baseline value of endpoint only (OUTPUT IMPUTED_V6).

Since the regression method uses only observed data to estimate regression parameters for the MI, only control subjects are used for the imputation of missing values at Week 24/ET in the active treatment groups.

Step 3: Reassemble (concatenate) datasets (REST_V6, IMPUTED_V6).

The final, reassembled dataset will include M=40 imputations of all subjects with missing data at Week 24/ET imputed using J2C method. This dataset, labeled *EFF_MI_J2C* for reference, is used in the statistical analysis described in Section 10.4.

10.3 Multiple Imputation Method – Tipping Point

Tipping Point analyses will be conducted using the sequential modeling method as described in, (O'Kelly & Ratitch, 2014), Chapter 7, and specifically the third (sequential) approach described Section 7.4.7 of the book. First, the initial analysis dataset is used to impute missing data in the control group with multiple imputation under MAR assumption; the MI is performed as described above in Section 10.2 and dataset *EFF_MI_MAR* is used as the input dataset to the Tipping Point analysis.

The Tipping Point analysis will then be performed on the *PEFF_MI_MAR* dataset. A constant, δ, will be applied to the imputed values in the active treatment groups; δ will be defined in units of the within-group standard deviation (σ) of change from baseline in APOV; that is, $\delta = c \cdot \sigma$, where the constant c is a fraction. An estimate of σ will be obtained from the root mean square error (RMSE) of the primary analysis model described in Section 7.1.5. To find the tipping point, the sequence of fractions c = 0.25, 0.50, 0.75, 1.0, and so on in 0.25 increments, will be chosen until the tipping point is reached. However, this search will employ an efficient numerical searching algorithm to find the Tipping Point value δ^*. Since the APOV endpoint has a range from 0-100, imputed values with δ added cannot exceed the maximum value of 100. Therefore, any imputed values with δ added that exceed the maximum score of 100 will be assigned the maximum score of 100.

Note that it is possible a Tipping Point does not exist. If the Tipping Point is not reached for value of δ such that all missing data in the active treatment groups are assigned the maximum score, then a Tipping Point does not exist.

The output dataset for the Tipping Point analysis will be labelled *PEFF_ MI_TP_D* for reference, where D will represent the current value of δ in the search.

10.4 Multiple Imputation – Statistical Analysis and Inference

Statistical Analysis

The ANCOVA statistical model for the primary efficacy analysis described in Section 7.1.5 will be used to analyze the MI dataset *EFF_MI_J2C* (primary analysis) and *MI_PEFF_TP_D* (Tipping Point sensitivity analysis). The statistical analysis is performed separately on each of the M=40 complete datasets. The summary statistics resulting from these analyses are output to SAS ODS datasets: DIFFS (least squares mean (LSM) estimates of differences between treatments), LSMEANS (LSM estimates for each treatment group, and SOLUTIONF (model parameter estimates). Note the SAS PROC MI procedure produces a SAS variable *_Imputation_*; this SAS variable identifies each of the M=40 MI datasets for analyses and inference.

Statistical Inference

The statistical method for combining the summary statistics obtained from the statistical analysis of each of the M=40 MI datasets is well established (Rubin's method); it is based on obtaining an overall estimate of mean and variance from these MI summary statistics, forming a test statistic for the hypothesis test and estimating the degrees of freedom, and obtaining a p-value and a confidence interval for formal statistical inference. Rubin's method is described in (O'Kelly & Ratitch, 2014), Chapter 6, Section 6.4.

This method is implemented in SAS PROC MIANALYZE; the procedure is designed to accept the SAS ODS output from the statistical analysis. If the unadjusted degrees of freedom is very large compared to the complete-data degrees of freedom, then adjusted degrees of freedom will be used.

10.5 MMRM Model convergence contingencies

Model convergence is not anticipated to be an issue with the linear and logistic MMRM models used for the analysis of data from this trial; however, in the event a convergence issue arises for any analyses, in each case, both a heterogeneous Toeplitz and Toeplitz covariance structure will be considered in sequence in place of the unstructured covariance. In this case, the empirical sandwich estimator will be used as a sensitivity analysis to correct for potential model misspecification.

11 CHANGES FROM THE PROTOCOL

The following changes were made to this SAP post Protocol Amendment 6.0.

Section 3.3.3 and 3.4.3 (Objectives and Endpoints)

The following objectives and secondary endpoints were added:

- change from baseline to Week 24/ET in APOV in the ethmoid sinuses (APOV-E)
- change from baseline to Week 24/ET in APOV in the maxillary sinuses (APOV-M)
- change from baseline to Week 24/ET in APOV in the ethmoid sinuses (APOV-E) for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery
- change from baseline to Week 24/ET in APOV in the maxillary sinuses (APOV-M) for CRS with NP and without NP sub-groups and in subjects with and without previous sinus surgery
- change from baseline in the volume of the nasal cavity opacified

The following objective and secondary endpoint was removed:

- percent of subjects requiring rescue medication after Week 4

Section 7.2 Sensitivity Analyses

- Section 7.2.3 Added an Observed Cases Sensitivity Analysis

Section 7.5.3 and Section 9.3.2 (Percentage of Opacified Volume)

- Clarified the definition of WPOV, WPOV-E, WPOV-M and the corresponding endpoints
- Added the definition of APOV-E and APOV-M and the corresponding endpoints

Section 7.5.4 and Section 9.3.6 (Qualitative CT Scan Assessment)

- Clarified the definition of WZLM and the corresponding endpoints

12 REFERENCES

Buysse, DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ: The Pittsburgh Sleep Quality Index (PSQI): A new instrument for psychiatric research and practice. Psychiatry Research 1989; 28:193-213.

Doty, RL, Shaman, P, Kimmelman, CP, Dann, MS. University of Pennsylvania Smell Identification Test: A rapid quantitative olfactory function test for the clinic. Larynoscope. 1984; 94:176-178.

DeConde, A, Mace J, Bodner T, et al. SNOT-22 quality of life domains differentially predict treatment modality selection in chronic rhinosinusitis. Allergy Rhinology. 2014 December; 4 (12): 972–979.

Hopkins C, Slack R, Lund V, et al. Long-term outcomes from the English national comparative audit of surgery for nasal polyposis and chronic rhinosinusitis. Laryngoscope. 2009; 119 (12):2459-65.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonized Tripartite Guideline, Statistical Principles for Clinical Trials (E9), 5 February 1998.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonized Tripartite Guideline, Statistical Principles for Clinical Trials Addendum: Estimands and Sensitivity Analysis in Clinical Trials (E9 [R1]), Current Step 2 version 16 June 2017.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonized Tripartite Guideline, Structure and Content of Clinical Study Reports (E3), 30 November 1995.

Koch, GG, Amara, IA, Davis, GW, Gillings, DB. A review of some statistical methods for covariance analysis of categorical data. Biometrics. 1982; 38: 563-595

Lund VJ, Mackay IS. Staging in rhinosinusitis. Rhinology. 1993; 31:183–184. [PubMed: 8140385].

Mallinckrodt, C., Lane, P., Schnell, D., Peng, Y., & Mancuso, J. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. Drug Information Journal, 303-319, 2008.

O'Kelly, M., & Ratitch, B. Clinical Trials with Missing Data: A Guide for Practitioners. Wiley, 2014.

Maruish, M. E. (Ed.). User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric Incorporated

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

NCSS, LLC. PASS 15 Power Analysis and Sample Size Software. 2017. Kaysville, Utah, USA. ncss.com/software/pass.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. K., Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in subjects with chronic major depression. Biological Psychiatry, 54, 573-583.

Soler Z, Rudmik L, Hwang PH, et al. Subject-Centered Decision Making in the Treatment of Chronic Rhinosinusitis. Laryngoscope. 2013 October; 123(10): 2341–2346

Stokes, ME, Davis, CS, Koch, GG. Categorical Data Analysis Using SAS, Third Edition. 2012. Cary, NC: SAS Institute, Inc.

13 APPENDICES

13.1 Appendix 1: PSQI Scoring

Scoring the PSQI

The order of the PSQI items has been modified from the original order in order to fit the first 9 items (which are the only items that contribute to the total score) on a single page. Item 10, which is the second page of the scale, does not contribute to the PSQI score.

In scoring the PSQI, seven component scores are derived, each scored 0 (no difficulty) to 3 (severe difficulty). The component scores are summed to produce a global score (range 0 to 21). Higher scores indicate worse sleep quality.

Component 1: Subjective sleep quality—question 9

Response to Q9	Component 1 score
Very good	0
Fairly good	1
Fairly bad	2
Very bad	3

Component 1 score:_____

Component 2: Sleep latency—questions 2 and 5a

Response to Q2	Component 2/Q2 subscore
≤ 15 minutes	0
16-30 minutes	1
31-60 minutes	2
> 60 minutes	3

Response to Q5a	Component 2/Q5a subscore
Not during past month	0
Less than once a week	1
Once or twice a week	2
Three or more times a week	3

Sum of Q2 and Q5a subscores	Component 2 score
0	0
1-2	1
3-4	2
5-6	3

Component 2 score:_____

Component 3: Sleep duration—question 4

Response to Q4	Component 3 score
> 7 hours	0
6-7 hours	1
5-6 hours	2
< 5 hours	3

Component 3 score:_____

Component 4: Sleep efficiency—questions 1, 3, and 4

Sleep efficiency = (# hours slept/# hours in bed) X 100%

hours slept—question 4

hours in bed—calculated from responses to questions 1 and 3

Sleep efficiency	Component 4 score
> 85%	0
75-84%	1
65-74%	2
< 65%	3

Component 4 score:_____

Component 5: Sleep disturbance—questions 5b-5j

Questions 5b to 5j should be scored as follows:

Not during past month	0
Less than once a week	1
Once or twice a week	2
Three or more times a week	3

Sum of 5b to 5j scores	Component 5 score
0	0
1-9	1
10-18	2
19-27	3

Component 5 score:_____

Component 6: Use of sleep medication—question 6

Response to Q6	Component 6 score
Not during past month	0
Less than once a week	1
Once or twice a week	2
Three or more times a week	3

Component 6 score:_____

Component 7: Daytime dysfunction—questions 7 and 8

Response to Q7	Component 7/Q7 subscore
Not during past month	0
Less than once a week	1
Once or twice a week	2
Three or more times a week	3

Response to Q8	Component 7/Q8 subscore
No problem at all	0
Only a very slight problem	1
Somewhat of a problem	2
A very big problem	3

Sum of Q7 and Q8 subscores	Component 7 score
0	0
1-2	1
3-4	2
5-6	3

Component 7 score:_____

Global PSQI Score: Sum of seven component scores:_____