Official Title:	A PHASE III, RANDOMIZED, MULTICENTER, DOUBLE-BLIND, PLACEBO-CONTROLLED CLINICAL TRIAL OF OMALIZUMAB IN PATIENTS WITH CHRONIC RHINOSINUSITIS WITH NASAL POLYPS
NCT Number:	NCT03280537
Document:	STATISTICAL ANALYSIS PLAN
Version & Date:	Version 2: 15-Apr-2019

STATISTICAL ANALYSIS PLAN

TITLE:	**A PHASE III, RANDOMIZED, MULTICENTER, DOUBLE-BLIND, PLACEBO-CONTROLLED CLINICAL TRIAL OF OMALIZUMAB IN PATIENTS WITH CHRONIC RHINOSINUSITIS WITH NASAL POLYPS**
PROTOCOL NUMBER:	GA39855
STUDY DRUG:	Omalizumab (IGE025) RO5489789
VERSION NUMBER:	2
IND NUMBER:	5369
EUDRACT NUMBER:	2017-003450-28
SPONSOR:	F. Hoffmann-La Roche Ltd.
PLAN PREPARED BY:	███████████ , Ph.D.
DATE FINAL:	24 January 2019
DATE AMENDED:	See electronic date stamp below.

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

Name	**Reason for Signing**	**Date and Time** (UTC)
███████████	Company Signatory	15-Apr-2019 17:01:11

CONFIDENTIAL

STATISTICAL ANALYSIS PLAN AMENDMENT
RATIONALE

Changes to Section 4.2.4 analysis of Daily eDiary Assessments were made to reflect that symptom evaluation would occur only through Day 186 (last day of Week 24 window), the point in time of the primary endpoint, but not through a later date as indicated in the prior version.

Changes in Section 4.3.1 (Intra-Reader Reliability of Nasal Polyp Score Assessments) were made to add the missing data imputation rules.

Changes to Section 4.5.1.1.1 were made to address an inconsistency between the definition of the estimand for co-primary endpoints (Section 4.5.1.1.1) and the estimator (Section 4.5.1.1.3).

Changes to Section 4.5.1.1.3 were made to clarify the data handling for patients with complete baseline and no post-baseline values in the primary analysis.

Changes to Section 4.5.1.1.4.1 were made to add one sensitivity analysis to evaluate the robustness of the primary endpoint results using different imputation rules for data after intercurrent events.

Changes to Section 4.5.3 were made to add four exploratory endpoints.

Changes to Section 4.7.2 were made to clarify that some events identified as risks associated with omalizumab will first be captured by broad search criteria or broad Standardized MedDRA Query and then confirmed by a Sponsor scientist during medical reviews.

Additional minor changes have been made to improve clarity and consistency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

GLOSSARY OF ABBREVIATIONS AND DEFINITION OF TERMS

AAC	anaphylaxis adjudication committee
AE	adverse event
AESI	adverse event of special interest
ARS	anterior rhinorrhea score
ATE	arterial thrombotic events
AQLQ	asthma quality of life questionnaire
BID	twice a day
Comp	completers
CRSwNP	chronic rhinosinusitis with nasal polyps
CS	corticosteroid
CSS	Churg Strauss Syndrome
Day 1	study day one; defined as the day of randomization
Day -7	study day minus seven is defined as seven days prior to randomization
Drop	dropouts
eCRF	electronic Case Report Form
EGPA	Eosinophilic Granulomatosis with Polyangiitis
EMA	European Medicines Agency
EQ-5D-5L	EuroQol 5-Dimension 5-level Questionnaire
FAS	full analysis set
FDA	U.S. Food and Drug Administration
HES	Hyper Eosinophilic Syndrome
HGLT	High Level Group Term
HLT	High Level Term
HRQoL	health-related quality of life
ICC	intraclass correlation coefficient
IWRS	interactive Web-based response system
LOE	lack of efficacy
LS	least squares
M	mometasone
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MID	minimal important difference
MMRM	mixed-effect model repeated measurement
MNAR	missing not at random
NCS	nasal congestion score

GLOSSARY OF ABBREVIATIONS AND DEFINITION OF TERMS

NPS	nasal polyp score
OLE	open label extension
PD	progressive disease
PFAS	pooled full-analysis set
PRS	posterior rhinorrhea score
PT	Preferred Term
QD	once a day
QoL	quality of life
Rx	prescription
SAE	serious adverse event
SAP	Statistical Analysis Plan
SFU	safety follow-up
SSLD	Serum Sickness Like Disease
SmPC	Summary of Product Characteristics
SMQ	Standardized MedDRA Query
SNOT-22	Sino-Nasal Outcome Test-22
SSS	sense of smell score
SOA	schedule of assessments
TNSS	total nasal symptom score
TEAE	treatment emergent adverse event
UPSIT	University of Pennsylvania Smell Identification Test
V1	Visit 1 (see SOA)
V2	Visit 2 (see SOA)
VAS	visual analogue scale
Study drug	placebo or omalizumab

# 1.	BACKGROUND

The purpose of the analyses outlined in this Statistical Analysis Plan (SAP) is to evaluate the efficacy and safety of omalizumab compared with placebo in adult patients with chronic rhinosinusitis with nasal polyps (CRSwNP) who have had an inadequate response to standard-of-care treatments.

This SAP is based on Section 6 (Statistical Considerations and Analysis Plan) of the protocol for Study GA39855. However, the analyses specified in this document supersede the high-level analysis plan described in the protocols. Note that Studies GA39688 and GA39855 are independent studies with identical study designs as described in the Protocol Synopsis for Study GA39855 in Appendix 1.

Prior to the initiation of Studies GA39688 and GA39855, the Sponsor requested advice from both the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA) on the nasal polyps program for omalizumab, including the statistical considerations for the CRSwNP program. The feedback from these agencies have been incorporated into a draft SAP. After initiation of the studies, the Sponsor requested advice from the FDA on a draft SAP. Additional feedback regarding the draft SAP was obtained from the FDA and this feedback has also been incorporated into this document.

# 2.	STUDY DESIGN

Studies GA39688 and GA39855 are Phase III, randomized, multi-center, double-blind, placebo-controlled, clinical trials that are run in parallel. In each study, a total of approximately 120 patients with CRSwNP whose disease remains inadequately controlled despite daily treatment with intranasal corticosteroid (CS) therapy will be enrolled.

Written informed consent for participation in the study must be obtained before performing any study-specific tests or evaluations. The 5-week screening/run-in period will include two visits ("1st screening visit" [V1] and "2nd screening visit" [V2]) (see Figure 1). During the screening/run-in period, at both the V1 and V2, patients will undergo video endoscopy to quantify the size of the polyps and be assigned a nasal polyp score (NPS) prior to randomization. The site physician performing the nasal endoscopy does not score the nasal polyps. The bilateral NPS at each assessment will be determined by central readers, according to a modified '2+1 paradigm' (Ahmad et al. 2015) consisting of two central readers chosen from a pool of five and up to one central adjudicator from a pool of three based on the criteria presented below in Table 1. Specifically, if the two central readers assigned disagree on NPS by 1 or more points, a central adjudicator will determine which of the two central reader scores is utilized. The NPS ranges from 0−8 (sum of 0−4 on the left nostril and 0−4 on the right nostril).

Table 1 Nasal Polyps Scoring System

Polyp Score	Polyp Size
0	No polyps
1	Small polyps in the middle meatus not reaching below the inferior border of the middle turbinate
2	Polyps reaching below the lower border of the middle turbinate [a]
3	Large polyps reaching the lower border of the inferior turbinate or polyps medial to the middle turbinate
4	Large polyps causing complete obstruction of the inferior nasal cavity

Note: Scoring system is used to evaluate polyp size in each nasal passage by means of video nasal endoscopy. Nasal polyp score is the sum of unilateral polyp scores for each nasal passage.

[a] The scoring is modified to accommodate patients who have had a middle turbinectomy, such that the polyp must reach the top of the inferior turbinate to be graded as Score 2.

Source: Gevaert et al. 2013.

During the screening/run-in period, beginning at V1, patients will be asked to standardize their intranasal CS to a regimen of mometasone, 200 μg twice a day (BID) (400 μg total daily dosage). Patients deemed by the investigator to be intolerant to a BID regimen of mometasone may remain on a stable dosage of mometasone once a day (QD) during the run-in period and throughout the duration of the treatment period (two sprays/nostril, both nostrils, 50 μg/spray QD for a total daily dosage of 200 μg). Any patient transitioned to QD mometasone should remain on this regimen for the remainder of study. At V1, patients should receive appropriate training on proper self-administration of mometasone nasal spray.

Other assessments and activities occurring during the screening/run-in period are presented in the Schedule of Assessments (SOA) (see Appendix 2).

If patients do not meet the inclusion criteria (e.g., development of an acute infection at the time of screening necessitating treatment with antibiotics), they may be eligible for re-screening up to two times, after discussion with a Medical Monitor. Additionally, if video endoscopy at V2 is determined to be of insufficient quality to allow for scoring of NPS, patients may repeat nasal endoscopy if sufficient time remains to allow for central reading of that endoscopy prior to randomization; in such cases, before proceeding with repeat endoscopy, investigators should confirm that sufficient time does indeed remain for central reading of this endoscopy prior to randomization.

After screening/run-in has been completed, approximately 120 eligible patients per study will be randomly allocated in a 1:1 ratio to receive double-blind treatment with omalizumab or placebo. Randomization will be stratified on comorbid asthma and aspirin sensitivity status at baseline (3 levels: asthmatic and aspirin sensitive, asthmatic

not aspirin sensitive, all other) and geographic region (North America [including Canada, Mexico, or USA] and ex-North America).

A modified version of the approved E.U. Summary of Product Characteristics (SmPC) omalizumab dosing table (see Appendix 3) will be used; the dosing table from the SmPC has been modified to treat only the higher body weights reflective of the adult population being studied. The first dose of study drug will be administered on the same day as randomization Week 0 (Day 1). Dosing will be repeated every 2 or 4 weeks during a 24-week placebo-controlled treatment period.

Safety and efficacy will be assessed throughout the placebo-controlled treatment period, as detailed in the SOA (see Appendix 2). The co-primary efficacy endpoints are changes from baseline in NPS and nasal congestion score (NCS) at Week 24.

Video nasal endoscopy will be performed at V1 and V2, and at Weeks 4, 8, 16, and 24 (for a total of 6 endoscopies). NCS will be assessed in the mornings daily via eDiary throughout the study.

After the treatment period ends (at Week 24), patients will be followed for four additional weeks as part of safety follow-up (SFU). However, if patients enroll at Week 24 into the open-label extension (OLE) Study WA40169 of omalizumab in nasal polyps, then the SFU will be deferred until the end of Study WA40169 and not considered a part of this study.

All patients who discontinue study drug early during the treatment period will be asked to continue the planned study assessments through Week 24 and then to complete a 4-week SFU period. Patients who are unwilling or unable to continue with the planned assessments in the treatment period will instead complete a dosing-termination visit and then enter a 4–week SFU period. Patients who require nasal polypectomy or require two or more courses of treatment with systemic CS for ≥3 consecutive days will discontinue study drug but continue in the study with assessments and treatments.

An SOA is provided in Appendix 2.

Figure 1 Study Schema
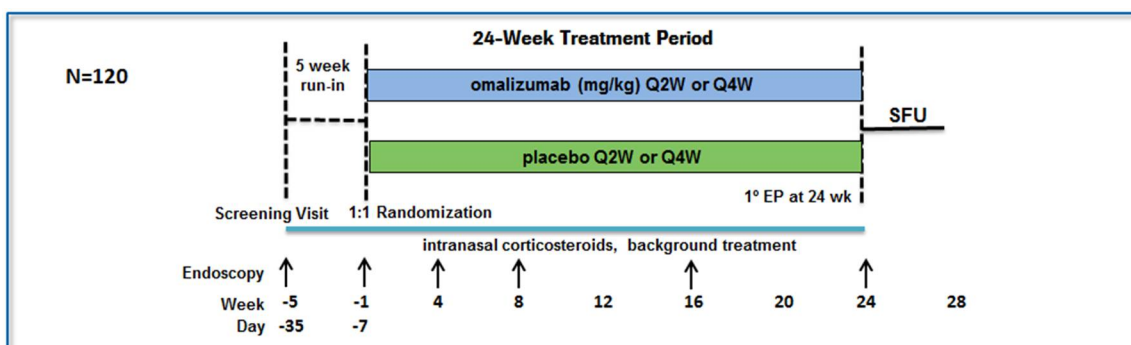
1 EP=co-primary end point; N=number of patients; Q2W=every 2 weeks; Q4W=every 4 weeks; SFU=safety follow-up; wk=week.

Note: All patients will be treated during the entire study with intranasal corticosteroids (mometasone nasal spray) as background therapy.

2.1   PROTOCOL SYNOPSIS

Studies GA39688 and GA39855 are independent studies with identical study designs. The Protocol Synopsis for Study GA39855 is in Appendix 1. For additional details on planned study assessments, see the SOA in Appendix 2.

2.2   ENDPOINTS

Baseline is defined differently for different endpoints depending on the SOA (see Appendix 2). Definitions of baseline for each assessment are provided in Section 4.2.3.

2.2.1   Co-Primary Efficacy Endpoints

The co-primary efficacy endpoints are as follows:

- Change from baseline at Week 24 in average daily NCS
- Change from baseline at Week 24 in NPS

2.2.2   Secondary Efficacy Endpoints

The secondary efficacy endpoints are as follows:

- Change from baseline at Week 24 in the average daily total nasal symptom score (TNSS)
- Change from baseline at Week 24 in the average daily sense of smell score (SSS)
- Change from baseline at Week 24 in the average daily posterior rhinorrhea score (PRS)
- Change from baseline at Week 24 in the average daily anterior rhinorrhea score (ARS)
- Change from baseline at Week 24 in patient-reported health-related quality of life (HRQoL) as assessed by the total Sino-Nasal Outcome Test-22 (SNOT-22)
- Change from baseline at Week 16 in the average daily NCS

- Change from baseline at Week 16 in NPS

- Change from baseline at Week 24 in sense of smell, as assessed by the University of Pennsylvania Smell Identification Test (UPSIT)

- Reduction in the need for nasal polypectomy by Week 24, as defined by an NPS of ≤ 4 (unilateral score of ≤ 2 on each side) and improvement in SNOT-22 score of ≥ 8.9

- Requirement of rescue treatment (systemic CS for ≥ 3 consecutive days or having had nasal polypectomy) through Week 24

- Requirement of rescue medication (systemic CS for ≥ 3 consecutive days) through Week 24

- Having had nasal polypectomy through Week 24

- Change from baseline at Week 24 in Asthma Quality of Life Questionnaire (AQLQ) of ≥ 0.5 (in patients with comorbid asthma only)

2.2.3 Exploratory Efficacy Endpoints

The exploratory efficacy endpoints are as follows:

- Change from baseline at Week 24 in SNOT-22 of at least the minimal important difference (MID) (8.9 points)

- Change from baseline at Week 24 as assessed by the EuroQol 5-Dimension 5-level Questionnaire (EQ-5D-5L)

2.2.4 Safety Endpoints

The safety endpoints are as follows:

- Incidence of adverse events (AEs)

- Incidence of serious adverse events (SAEs)

- Incidence of AEs leading to omalizumab/placebo discontinuation

- Clinically significant change in laboratory values

2.2.5 Pharmacokinetic/Pharmacodynamic Endpoints

The pharmacokinetic and pharmacodynamic endpoints are as follows:

- Serum concentration of omalizumab at specified timepoints outlined in the SOA

- Serum levels of total and free IgE at specified timepoints outlined in the SOA

2.3 DETERMINATION OF SAMPLE SIZE

A total of approximately 120 patients will be enrolled in this study. Patients will be randomly allocated in a 1:1 ratio to receive treatment with omalizumab or placebo, in addition to intranasal steroids.

The sample size of 120 patients (102 patients divided by 0.85 assuming a 15% early withdrawal rate) will provide at least 85% power to detect both a 0.56-point difference between treatment groups in change from baseline at Week 24 in the average daily NCS

and a 1.50-point difference between treatment groups in change from baseline at Week 24 in NPS.

Without any early patient withdrawals from the study, the sample size of n=102 patients will provide approximately 92% power to detect a 0.56-point treatment group difference in change from baseline at Week 24 in NCS (SD=0.83) and approximately 93% power to detect a 1.50-point treatment group difference in change from baseline at Week 24 in NPS (SD=2.2), for an overall power of at least 85% ($0.93 \times 0.92 > 0.85$).
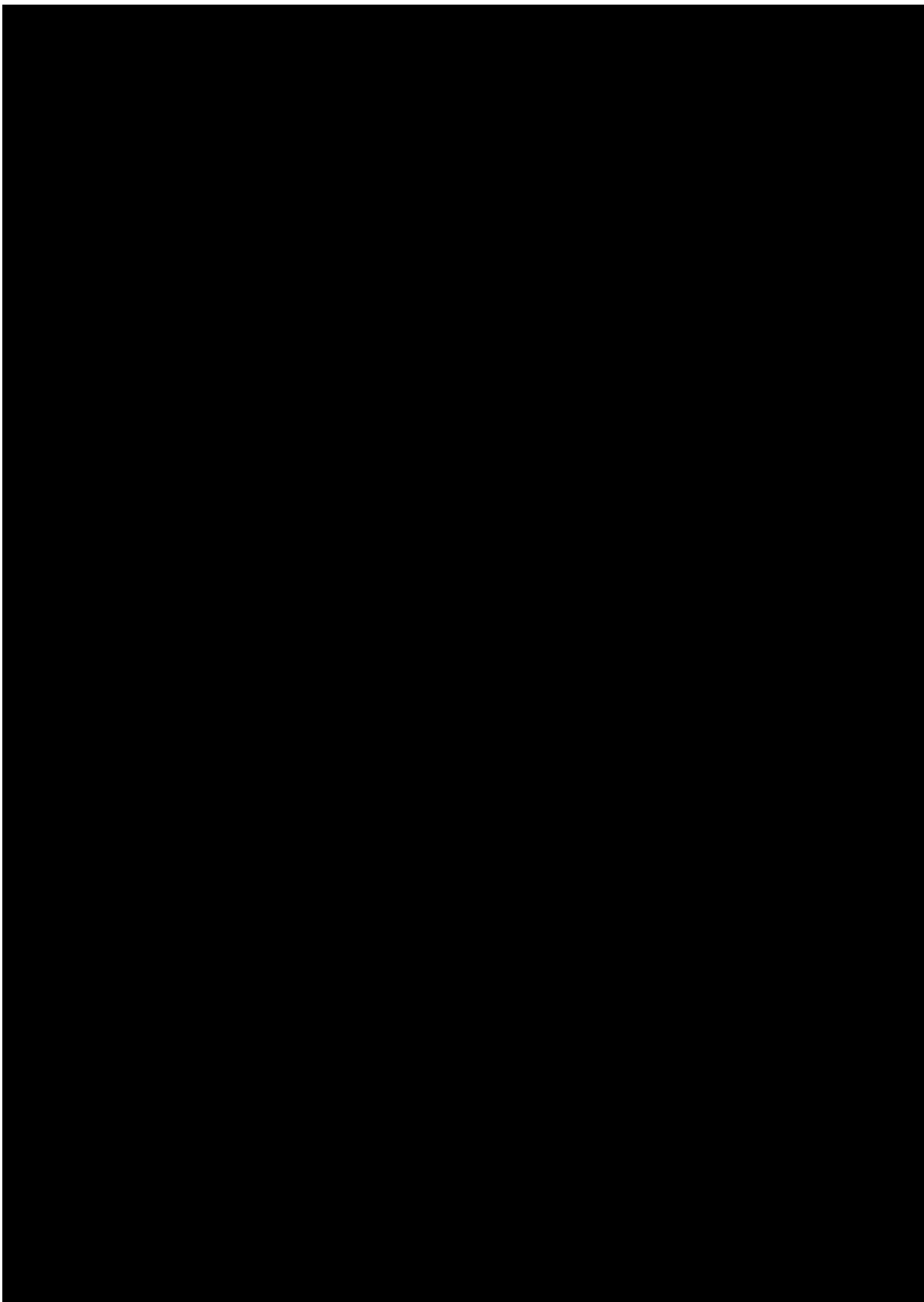
2.4   ANALYSIS TIMING

The analysis of complete data from the study, including data from the SFU period for patients not entering into the OLE Study WA40169, will be performed when all patients have met one of the following three criteria: 1) discontinued the study early, 2) completed the SFU period, or 3) enrolled in the OLE Study WA40169, and additionally all data from the study are in the database and the database is cleaned and locked.

3.   STUDY CONDUCT

3.1   RANDOMIZATION

At the Day 1 visit, after verification of inclusion/exclusion criteria, patients are randomized to receive either omalizumab or placebo in approximately a 1:1 ratio using an interactive web-based response system (IWRS) provided by ███████. Randomization will be stratified by comorbid asthma and aspirin sensitivity status at baseline (three levels: asthmatic and aspirin sensitive, asthmatic not aspirin sensitive, all other) and geographic region (two levels: North America [including Canada, Mexico, or USA] and ex-North America) which are crossed for a total of six strata.

3.2   ANAPHYLAXIS ADJUDICATION COMMITTEE

All potential anaphylaxis cases reported by investigators to the Sponsor will be subsequently submitted for adjudication to a blinded three-member anaphylaxis adjudication committee (AAC) composed of external experts in allergic diseases. The committee will assess whether the reported event meets criteria for an anaphylaxis event (based on Sampson's criteria) and whether the reported anaphylaxis event is causally related to blinded study drug. Further details will be provided in the Anaphylaxis Adjudication Charter.

3.3   EARLY STUDY DRUG DISCONTINUATION AND EARLY STUDY WITHDRAWAL

The primary reason for study drug (placebo or omalizumab) discontinuation will be documented on the appropriate electronic Case Report Form (eCRF).
When discontinuing study drug, the patients are encouraged to continue with all planned assessments through Week 28. When patients complete the blinded treatment period (24 weeks) and are eligible for enrollment into the OLE Study WA40169, if they choose, they will skip the 4-week SFU period and enter directly into the OLE Study and will be considered to have completed Study GA39855. Patients who discontinue study drug early do not qualify for enrollment into the OLE Study WA40169. When patients discontinue drug early or withdraw from the study early, the sites are instructed to enter the most informative reason available for early study drug discontinuation or early study withdrawal according to the following hierarchy from most informative to least informative: pregnancy, death, progressive disease (PD), AE, lack of efficacy (LOE), lost to follow-up, non-compliance with study drug, protocol deviation, withdrawal by subject, physician decision, study terminated by sponsor, other. Free text is collected if the reason is due

to an AE, protocol deviation, withdrawal by subject, physician decision, or other, and every effort will be made to accurately record the reason for discontinuation. These steps were taken to help ensure that missing data are handled appropriately in the efficacy analysis.

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATIONS

4.1.1 Full-Analysis Set

The full-analysis set (FAS) will include all randomized patients in Study GA39855 grouped according to the treatment assigned at randomization.

4.1.2 Pooled Full-Analysis Set

The pooled full-analysis set (PFAS) will include all randomized patients in both Studies GA39688 and GA39855 grouped according to the treatment assigned at randomization.

4.1.3 Safety Analysis Set

The safety analysis set will consist of all patients in Study GA39855 who received at least one dose of study drug, with patients grouped according to treatment received defined as:

- Placebo: Patients who received only (and at least one) placebo injections (i.e., no active treatment) during the treatment period.
- Omalizumab: Patients who received at least one omalizumab injection during the treatment period.

4.2 DATA HANDLING CONVENTIONS

4.2.1 Treatment Failure

Treatment failure is defined as a patient requiring rescue treatment. Patients receiving nasal polypectomy at any time during the interval between the date of randomization and the planned date of their Week 24 visit inclusive will be considered to have received rescue treatment. In addition, patients requiring ██████████████ systemic CS for nasal polyps symptoms for ≥3 consecutive days ████████████████████ ██████████████████ beginning after randomization through the planned date of their Week 24 visit, will be considered to have received rescue treatment. ████████ ██ ████████████████████████████

Only CS administered ██, at least in part, to alleviate symptoms of nasal polyposis as recorded in the eCRF will be considered when determining treatment failure status. ████████████████████ ██

Treatment failure status for each patient will be determined prior to unblinding. The treatment failure status, having had nasal polypectomy, concomitant systemic CS medication use after randomization, and indication for systemic CS use for all patients within the FAS will be provided in a single listing.

4.2.2 Definition of Study Day

For purpose of this document, study Day 1 is defined as the day of randomization. The day immediately following randomization is study Day 2 and so on. Study Day -1 is defined as the day that is one day prior to randomization. There is no study Day 0 by this definition of study day.

4.2.3 Definition of Baseline

Baseline for safety and efficacy endpoints will be defined as below in Table 3.

Table 3   Definition of Baseline

Assessment	Baseline Definition
Nasal Polyp Score	The last assessment on or before the date of randomization (i.e., V2 or V1 if V2 value is missing)
Nasal Congestion, Smell, Anterior Rhinorrhea, Posterior Rhinorrhea	Average of the daily values recorded during the 7-day interval ending on the latest day prior to randomization such that the prior 7-day interval includes a recorded value on at least 4 of the 7 days of that interval
SNOT-22	The last assessment on or before the date of randomization (i.e., Day 1 or V1 if Day 1 is missing)
UPSIT, EQ-5D-5L, AQLQ	The last assessment on or before the date of randomization (i.e., Day 1 is the only planned assessment that qualifies)
Total IgE	The first assessment before but not on the day of randomization. Cases where this assessment of IgE and body weight does not meet protocol inclusion criteria will be excluded from calculation of baseline value.
All other non-efficacy (i.e., free IgE, vital signs, lab assessments)	The last assessment on or before the date of first dose of study drug

AQLQ=asthma quality of life questionnaire; EQ-5D-5L=EuroQol 5-Dimension 5-level Questionnaire; SNOT-22=sino-nasal outcome test-22; UPSIT=University of Pennsylvania Smell Identification Test.

4.2.4 Daily eDiary Assessments

The following assessments are recorded via eDiary each morning: NCS, SSS, ARS, and PRS.

Table 4   Analysis Timepoints and Windows

Analysis Timepoints	First Study Day Included (-13 days)	Planned Timepoint Study Day	Last Study Day Included (+14 days)
Week 4	15	28	42
Week 8	43	56	70
Week 12	71	84	98
Week 16	99	112	126
Week 20	127	140	154
Week 24	155	168	186*

*the last day of the Week 24 interval is 4 days longer to minimize missing data at the Week 24 timepoint.

The following calculation will be performed for each of the following: NCS, ARS, PRS, SSS, and TNSS (defined as the sum of the four individual daily scores). For each study day (Day 1 through Day 186), a score will be calculated using an average of the prior 7 days (among the available days within the specified window, excluding the study day itself) if a value has been recorded by the patient on at least 4 of the prior 7 days, otherwise the 7-day prior average for that study day will be considered missing. For example the 7-day prior average NCS at Day 56 can be interpreted as "the average NCS in the 7 days prior to study Day 56. The 7-day prior average NCS will be calculated for study Days 1 through 186 with non-missing data in at least 4 of 7 days prior. (See Table 5 for examples). Each calculated (non-missing) 7-day prior average for each study day will be assigned to an analysis timepoint at Weeks 4, 8, 12, 16, 20, and 24 according to the intervals specified in Table 4. For each analysis timepoint, one calculated (non-missing) 7-day prior average will be selected for analysis according to the study day with nearest proximity to the planned timepoint study day, with the earlier selected in the case of a tie. For further details on handling of missing data and imputation rules for analysis of NCS, SSS, PRS, ARS, and TNNS (see Section 4.5).

Table 5   Calculation Examples for 7-Day Prior Average of NCS at Week 8

Example Scenario	Calculation
A daily NCS value was collected for Days 48, 49, 50, 51, 52, 53, 54, and 55	The 7-day prior average for study Day 56 is the average of the NCS on Days 49, 50, 51, 52, 53, 54, and 55
A daily NCS value was collected for Days 48, 49, 50, 51, 52 but missing for Days 53, 54, and 55.	With 4 or more of the 7 days collected the 7-day prior average for study Day 56 is the average of the NCS on Days 49, 50, 51, and 52. This average is used for analysis timepoint at Week 8.
A daily NCS value was collected for Days 48, 53, 54, 55, and 56 but missing for Days 49, 50, 51, and 52.	The 7-day prior average for study Day 56 is missing because 3 or fewer days were collected. The 7-day prior average is checked for Days 55 and found invalid because 4 days among Days 48-54 are missing. The 7-day prior average is checked for Days 57 and used and calculated as the average of Days 53, 54, 55, and 56 and used for the analysis timepoint at Week 8.
A daily NCS value was collected for Days 48, 49, 50, and 51 but missing for Days 52, 53, 54, and 55.	The 7-day prior average for study Day 56 is missing because 3 or fewer days were collected from Days 49 through 55. The 7-day prior average for study Day 55 is valid because at least 4 days are non-missing. The 7-day prior average for study Day 55 is calculated from Days 48, 49, 50, and 51 and used for analysis timepoint at Week 8.

4.2.5   Assessment Intervals and Assigning Timepoints

Data collected outside the visit intervals specified in the protocol will not necessarily be excluded from the analysis. For a given safety, efficacy, pharmacodynamic, or pharmacokinetic assessment, all data collected within the specified window of a planned timepoint study day (see Table 4) will be considered for an assigned timepoint for analysis. Within this interval for a given planned visit day, the assessment with the date nearest to the planned timepoint study day (for that assessment) will be selected with assessments occurring prior to the planned timepoint study day taking priority in case of a tie. With respect to the last timepoint (Week 24), for patients enrolled in the OLE Study WA40169, assessments recorded after open-label omalizumab dosing will not be considered part of data collected under this study protocol and therefore will not be used in the analysis for this study.

4.3   ANALYSIS OF STUDY CONDUCT

Descriptive statistics will be used to evaluate the conduct of the study. The number of patients randomized will be tabulated by country, study site, and randomized treatment arm in the FAS. Patient disposition (the number and percentage of patients randomized, treated, completing the treatment, and completing the study) will be tabulated by

randomized treatment arm in the FAS. The number and percentage of patients who undergo premature study drug discontinuation and study discontinuation, as well as reasons for discontinuations, will be summarized. The number and percentage of patients with eligibility criteria deviations, dosing errors and other major protocol deviations will be summarized.

4.3.1 Intra-Reader Reliability of Nasal Polyp Score Assessments

For details on the NPS assessment paradigm including adjudication, see Section 4.5.1.2. In order to assess the consistency of reader performance during the study, intra-reader reliability of NPS assessments will be evaluated using combined data from a random sample of 20 videos (10 from each of Studies GA39688 and GA39855). The following mixed-effect model (Shrout and Fleiss 1979; Gwet 2014) will be used for the NPS reading y_{ijk} associated with video i, reader j, and replicate k:

$$y_{ijk} = \mu + s_i + r_j + (\text{sr})_{ij} + e_{ijk}$$

Where

- μ is the expected score,
- s_i the random video effect,
- r_j the fixed reader effect due to the reader j,
- $(\text{sr})_{ij}$ the random video-reader interaction,
- and e_{ijk} the random error term.
- Each of $s_i, (\text{sr})_{ij},$ and e_{ijk}

    are assumed to follow Normal distributions with variances $\sigma_s^2, \sigma_{sr}^2$ and σ_e^2 respectively.

The intra-reader reliability will be assessed by the intraclass correlation coefficient (ICC):

$$\text{ICC}_a(3,1) = \frac{\text{MSS} + r\text{MSI} - (r+1)\text{MSE}}{\text{MSS} + r\text{MSI} + (rm - r - 1)\text{MSE}},$$

Where $r=5$ the number of readers and $m=2$ the number of replications.

MSS is the mean of squares for subjects and is calculated by summing the squared differences $(\bar{y}_{i..} - \bar{y})^2$ over all $n=20$ videos and multiplying the summation by rm (n-1). Note that $\bar{y}$ is the overall average NPS while $\bar{y}_{i..}$ is the average of all NPS associated with video i.

MSI is the mean of squares for rater=subject interaction and is calculated by summing the squared differences $(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y})^2$ over all n=20 videos and r=5 readers and multiplying the summation by $m/[(r\text{-}1)(n\text{-}1)]$. The term $\bar{y}_{ij\cdot}$ represents the average of all NPSs associated with reader j and video i. The term $\bar{y}_{\cdot j\cdot}$ is the average of all measurements associated with reader j.

MSE is the mean of squares for errors and calculated by summing the squared differences $(y_{ijk} - \bar{y}_{ij\cdot})^2$ over all rnm=200 readings, and by dividing the summation by $rn(m\text{-}1)$.

ICC will range from 0 to 1 where 0 represents no reliability and 1 perfect reliability.

To provide the data for the model, 10 videos will be randomly selected from each study from the set of videos captured in the study at the time of sampling to be re-read by the original readers assigned, and also read twice by the three other readers at least 4 weeks after the initial read. There are five readers total and all five will read videos in each study in approximately equal proportion. The selected original reads, re-reads, and reads/re-reads from non-assigned readers for these 20 videos will be used for purpose of assessment of intra-rater reliability. Scores which are the result of re-reads for purpose of this assessment will not be used in the primary efficacy analysis. If a reader assigns a NPS for a video in one assessment, but deems this video unreadable in another assessment, a NPS of either 0 or 8 will be imputed for that non-evaluable assessment, whichever leads to the poorest intra-rater agreement for that reader and video. This will provide approximately 200 reads (5 readers per video, 10 videos per study, 2 reads per video) for calculation of an empirical estimate of ICC.

4.3.2   Inter-Reader Reliability of Nasal Polyp Score Assessments

To further assess the reader performance in the study, the inter-reader reliability of NPS assessment of endoscopy videos will be assessed at each planned timepoint with point estimates of Fleiss' Kappa (Fleiss 1971) with linear weights (see Table 6) using all available data from all randomized patients from Studies GA39688 and GA39855 pooled for each planned timepoint (V1, V2, Weeks 4, 8, 16, and 24). Fleiss' Kappa may be interpreted as <0='poor reliability', 0 to 0.20='slight', 0.21 to 0.40='fair', 0.41 to 0.60='moderate', 0.61 to 0.80='substantial', >0.81='almost perfect' reliability (Landis and Koch 1977).

Table 6 Linear Weights

NPS Reader 2	NPS Reader 1								
	0	1	2	3	4	5	6	7	8
0	1.00	0.875	0.75	0.625	0.5	0.375	0.25	0.125	0
1	0.875	1.00	0.875	0.75	0.625	0.5	0.375	0.25	0.125
2	0.75	0.875	1.00	0.875	0.75	0.625	0.5	0.375	0.25
3	0.625	0.75	0.875	1.00	0.875	0.75	0.625	0.5	0.375
4	0.5	0.625	0.75	0.875	1.00	0.875	0.75	0.625	0.5
5	0.375	0.5	0.625	0.75	0.875	1.00	0.875	0.75	0.625
6	0.25	0.375	0.5	0.625	0.75	0.875	1.00	0.875	0.75
7	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1.00	0.875
8	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1.00

NPS=nasal polyp score.

4.4 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographic and baseline characteristics, such as age, sex, race/ethnicity, NPS, NCS, SSS, ARS, PRS, TNSS, UPSIT, and SNOT-22 will be summarized with descriptive statistics by treatment arm in the FAS. Baseline characteristics will be summarized by treatment group using mean, standard deviation, median, and range for continuous variables and number and percentages of patients for categorical variables, as appropriate.

The number and percentage of patients who report prior medical conditions will be reported by coded condition and by coded condition class within the FAS. The following targeted medical conditions will be summarized separate from the general medical conditions: allergic rhinitis, allergic sinusitis, asthma (from eCRF), otitis media, aspirin exacerbated respiratory disease (from eCRF), atopic dermatitis, and obstructive sleep apnea.

4.5 EFFICACY ANALYSIS

All analyses of efficacy outcomes will be performed within the FAS. In addition, following an assessment of the appropriateness of a combined data analysis by comparing trends in treatment benefit, baseline characteristics, and treatment characteristics between the two studies, the analysis of the following secondary endpoints will be performed based on the PFAS (pooled data) from both studies due to the small number of expected events or limited subgroup sample size for some endpoints:

- Requirement of rescue treatment (systemic CS for ≥ 3 consecutive days or having had nasal polypectomy) through Week 24

- Requirement of rescue medication (systemic CS for ≥ 3 consecutive days) through Week 24

- Having had nasal polypectomy through Week 24
- Change from baseline at Week 24 in Asthma Quality of Life Questionnaire (AQLQ) of ≥0.5 (in patients with comorbid asthma only)

While two-sided unadjusted p-values will be reported, the overall study-level type 1 error will be controlled at a two-sided α=0.05 level according to the type 1 error control plan detailed in Section 4.5.2.

All efficacy models will be adjusted by the stratification variables used in the randomization procedure: a two-level categorical covariate for geographic region (two levels: North America [including Canada, Mexico, or USA] and ex-North America), a three-level categorical covariate for baseline asthma comorbidity/aspirin sensitivity status (three levels: asthma with aspirin sensitivity, asthma without aspirin sensitivity, and non-asthma regardless of aspirin sensitivity). The baseline asthma/aspirin sensitivity status will be derived from that collected on the eCRF, which unlike the data collected in the IWRS, is subject to data monitoring, query, and revision in cases when an error is made, and is therefore most accurate.

4.5.1 Co-Primary Efficacy Endpoints

As NCS and NPS are co-primary endpoints, both null hypotheses for NCS and NPS must be rejected, with parameter estimates indicating a benefit of omalizumab over placebo, for the study to be deemed positive. The primary analysis will separately test the null hypotheses that no treatment group difference exists in terms of each of the co-primary endpoints:

$$H_0^{(1)}: \Delta_A^{(1)} = \Delta_P^{(1)} \text{ and } H_0^{(2)}: \Delta_A^{(2)} = \Delta_P^{(2)}$$

- $\Delta_A^{(1)}$ and $\Delta_P^{(1)}$ are the change from baseline at Week 24 in the average daily NCS in the FAS in the omalizumab arm and in the placebo arm respectively.
- $\Delta_A^{(2)}$ and $\Delta_P^{(2)}$ are the change from baseline at Week 24 in NPS in the FAS in the omalizumab arm and in the placebo arm respectively.

The two-sided alternative hypotheses are:

$$H_a^{(1)}: \Delta_A^{(1)} \neq \Delta_P^{(1)} \text{ and } H_a^{(2)}: \Delta_A^{(2)} \neq \Delta_P^{(2)}$$

Because both NCS and NPS hypotheses need to be rejected to demonstrate efficacy of omalizumab in patients with CRSwNP, there is no adjustment for multiplicity for these co-primary endpoints and each will be tested at the two-sided α=0.05 level.

4.5.1.1 Nasal Congestion Score
4.5.1.1.1 Estimand

The first of the co-primary estimands is defined as follows:

- Population: Adult patients with a physician diagnosed CRSwNP for ≥12 months, with and without comorbid asthma, who have large nasal polyps with moderate to severe symptoms, and an inadequate response to standard of care therapy (which includes at least 8 weeks of intranasal CS and may include systemic CS and/or nasal polypectomy)

- Variable: Change from baseline in average daily NCS at Week 24.

- Intercurrent events: (1) Had rescue treatment not been made available prior to Week 24; (2) regardless of study drug discontinuation due to AE/PD/LOE. All values after the intercurrent event would be coded to worst observed event value carried forward (worst observed value post-randomization and up to and including the day of the intercurrent event for that patient).

- Population-level summary: Difference in least squares variable means between omalizumab and placebo treatment groups

4.5.1.1.2 Variable Definition

NCS will be scored by patients as not at all, mild, moderate, or severe (a score of 0, 1, 2, or 3, respectively; see Fairley et al. 1993) every morning from V1 through Week 28 via eDiary. Baseline average NCS will be calculated according to the method previously described in Table 3 while post-baseline (Weeks 4, 8, 12, 16, 20, and 24) will be calculated according to the method described in Section 4.2.4 and Table 4. The absolute change from baseline at Week 24 in the average daily NCS values over the 7 days prior to the planned study visit day will be defined for each patient as the average NCS assigned to the Week 24 timepoint minus the average daily NCS assigned to baseline (and similarly for Weeks 4, 8, 12, 16, and 20).

4.5.1.1.3 Primary Estimator

Treatment group comparisons of absolute change from baseline at Week 24 in average daily NCS between treatment groups will be assessed using a mixed-effect model repeated measurement (MMRM) model with unstructured covariance.
The variance-covariance matrices for each treatment group will be assumed equal.
The Kenward-Rogers approximation (Kenward and Roger 1997) will be used to calculate the denominator degrees of freedom. The dependent variable will be absolute change from baseline in average daily NCS. In addition to adjustment for geographic region and asthma/aspirin sensitivity comorbidity status, the model will include terms for treatment group, timepoint (Weeks 4, 8, 12, 16, 20, and 24), baseline value of the dependent variable (baseline average daily NCS in this case), treatment by timepoint interaction, and baseline value of dependent variable by timepoint interaction. Point estimates, 95% confidence intervals, and p-values for the treatment effect (omalizumab vs. placebo) on absolute change from baseline at a timepoint in average daily NCS will be calculated on the basis of the model for all post-baseline analysis timepoints:

Weeks 4, 8, 12, 16, 20, and 24, using appropriate difference in least square (LS)-means. Patients with baseline and no post-baseline data will be included in the model for the calculation of the LS means. The LS-means will be calculated using the coefficients of the independent variables weighted according to that observed in the FAS, regardless of treatment assignment. This marginal weighting of coefficients for LS-means may impact the estimates of the means in each treatment arm but would not impact the difference between means in arms because the same covariate values are applied to calculate the means in both arms. The null hypothesis will be tested by the t-test arriving from the treatment group difference in LS-means of the change from baseline at Week 24 in average daily NCS. If issues with convergence arise, the unstructured covariance assumption will be replaced with a heterogeneous compound symmetry covariance assumption and later ANCOVA at each timepoint if issues persist.

Data collected after intercurrent events will be handled as follows before calculation of the 7-day prior averages:

1. <u>Receive rescue treatment (nasal polypectomy or systemic CS) prior to Week 24</u>: For patients who received rescue treatment as defined in Section 4.2.1, all data collected after treatment failure will be set to missing and the worst observed daily NCS (observed post-randomization on or before the day of rescue treatment) will be used to impute on each day after the date of treatment failure through the date of the planned Week 24 visit. The date of treatment failure is defined as the date of nasal polypectomy or the start date of the earliest course of systemic CS that qualifies as treatment failure per the criteria in Section 4.2.1. It would be unethical to deny the patient rescue treatment. Yet, assuming that rescue treatment is effective in lowering the NCS, incorporation of data collected post rescue treatment risks biasing the comparison in favor of the experimental treatment arm with higher rescue treatment rates (and lower efficacy). Therefore, data collected after rescue treatment will not be used in the analysis model. At the same time simply excluding the data would assume missing at random (MAR) which is unreasonable so a missing not at random (MNAR) imputation rule is preferred to excluding the data. Although rescue treatment is temporarily effective, it is associated with clinically significant complications and recurrence and therefore suboptimal. It is also reasonable to assume that for patients who require rescue treatment, the NCS would have been poor if rescue treatment was not available. For these reasons the worst observed post-baseline score is chosen for imputation after rescue treatment.

2. <u>Discontinuation of study drug early due to PD, AE, or LOE</u>: For patients who discontinue study drug early due to PD, AE, or LOE all data collected after the last dose of study drug will be set to missing and the worst observed daily NCS (observed post-randomization on or before the last dose of study drug) will be used to impute on each day after the last dose of study drug through the date of the planned Week 24 visit. The rationale is that, given the severity of nasal polyposis, patients who discontinue drug due to PD or LOE are likely to require rescue treatment after dropping out of the study. The rationale for imputing the worst observed post-baseline on or before last dose of study drug value of NCS for drop

outs after discontinuing drug due to AE is to avoid potential bias in favor of an experimental drug associated with AEs.

Missing data not explicitly covered above will be assumed MAR (i.e., given the observed outcomes and other variables in the statistical model missingness is independent of the unobserved outcomes). This type of missing data will not be explicitly imputed.

Patients who have a pre-treatment value but lack at least one post-randomization value assigned to planned analysis timepoint to be used in the MMRM model for NCS assessment will only contribute to the analysis in the calculation of LS-means in each arm in change from baseline via their contribution to the observed marginal (within the FAS) coefficients for independent variables (baseline asthma/aspirin sensitivity comorbidity, geographic region, and baseline NCS). This marginal weighting of coefficients for LS-means may impact the estimates of the means in each treatment arm but would not impact the difference between LS-means in arms because the same covariate values are applied to calculate the means in both arms.

The order of operations is as follows:

1. Assess the occurrence of intercurrent events and impute the daily NCS after the intercurrent event as per above at the level of study day.

2. Using all observed and imputed daily NCS data, calculate the 7-day prior average daily NCS for all study days up to and including Week 24 (up to and including Day 186) for those with at least 4 days of the 7-day prior interval are non-missing as per Section 4.2.4.

3. Using the calculated 7-day prior average daily NCS for the study days, assign one to the analysis timepoint according to proximity to the planned study day per the windowing rules in Section 4.2.5.

4.5.1.1.4   Sensitivity Analyses of Estimand Assumptions
4.5.1.1.4.1   Handling of Missing Data

In order to test the sensitivity of the co-primary NCS estimate with respect to the MAR assumption for non-intermittent missing data not following an intercurrent event due to early study discontinuation, a tipping point sensitivity analysis will be performed if the percentage of patients with missing data at Week 24 not following an intercurrent event exceeds 5% in either treatment group. Additional sensitivity analyses of the MAR assumption may be performed as deemed appropriate whether or not the 5% condition is met.

In order to evaluate the sensitivity of the treatment group difference in NCS estimate with respect to the variability associated with single imputation following an intercurrent event in the primary analysis, a second tipping point analysis will be performed via multiple imputation. Non-intermittent missing data not preceded by an intercurrent event will be

assumed MAR and not imputed (same as in the primary analysis). Data preceded by an intercurrent event will be omitted and imputed using multiple imputation with delta adjustment at various MNAR assumptions and with MAR as a reference. The data post intercurrent event will be multiply imputed in separate simple linear regression imputation models in each arm using the same baseline covariates as used in the primary analysis (baseline NCS value, asthma/aspirin status, geographic region) and NCS at prior timepoints. Each of the imputed and MNAR adjusted (tipped) datasets will be analyzed with the same MMRM model as used in the primary analysis. Estimates grouped according to distinct tipping point assumptions will be combined separately according to Rubins rules and the estimate and nominal p-value will be reported for each distinct tipping point assumption.

4.5.1.1.4.2   Sensitivity Analyses on Baseline Covariates

Baseline mometasone treatment and anti-leukotriene treatment may explain some of the variability in the change from baseline in NCS, but the number of patients with prescribed anti-leukotriene treatment at baseline or baseline mometasone dose <400 μg per day is expected to be small. Therefore, pending sufficient data for model convergence, the same model as specified for the primary estimator will be fitted with two additional baseline categorical covariates defined as the mometasone dose prescribed at baseline (2 levels: ≥400 , <400 μg per day) and prescribed anti-leukotriene treatment at baseline (2 levels: Yes, No).

4.5.1.1.5   Supplementary Analysis
4.5.1.1.5.1   Cumulative Proportion of Responders Analysis Graph

No minimal important difference in the change from baseline in the average daily NCS has been identified; therefore there is no generally accepted definition of a responder. In order to aid in the interpretation of the results, cumulative proportion of responders analysis graphs (Farrar et al. 2006) will be provided with responder defined separately as having achieved a change from baseline at Weeks 16 and 24 in 7-day prior average daily NCS $\leq X$ without having been classified as a treatment failure, using the same augmented dataset (same imputation rules) as those specified for the primary estimator. Curves will be displayed for both treatment groups with proportion of responders on the y-axis and X on the x-axis ranging from the maximum to the minimum change from baseline at Weeks 16 and 24 (in separate figures) in 7-day prior daily average NCS. Patients with missing Week 16 or 24 values not due to an intercurrent event will be excluded from these graphs.

4.5.1.2   Nasal Polyp Score

The estimand and primary estimator method as well as all sensitivity and supplementary analyses for NPS will follow those used for NCS. The variable for NPS is defined below.

Polyp size in each nostril will be graded through an assessment of the video nasal endoscopy (0–4 integer scale per nostril, see Table 1; Gevaert et al. 2013) for both pivotal studies (GA39688 and GA39855) by a central panel of five independent sinus

surgeons who are blinded to treatment assignment and source study. The NPS used in the primary analysis is determined according to a modified '2+1 paradigm' (Ahmad et al. 2015). It is modified from the example in the referenced publication because the NPS is determined initially by two central readers rather than one central reader and a site physician. The site physician performing the endoscopy does not score the nasal polyp and thus does not contribute to the score used for efficacy analysis. Each subject is assigned two independent central readers from a pool of five central readers who read that subject's videos throughout the study including those obtained during screening (V1 and V2) and post-randomization (Weeks 4, 8, 16, and 24) when possible. NPS is the sum of the polyp scores in both nostrils (maximum score of 8). Once two central readers independently assign an NPS (0–8) their scores are then compared and if there is any difference the video and both reader's assessments (by nostril) are sent to one central adjudicator among a pool of three central adjudicators distinct from and blinded to the identity of the central readers. The adjudicator is required to choose between the two reader's assessments and the chosen score (sum of individual scores in each nostril) is used for efficacy analyses. The absolute change from baseline at each timepoint in NPS for each patient will be defined as the NPS assessment assigned to the analysis timepoint as per Section 4.2.5 minus the NPS at baseline as per Section 4.2.3. The same intercurrent events and imputation rules that apply to NCS will apply to NPS, but with a NPS worst post-baseline value observed prior to or on the day of rescue treatment/last does of study drug will be imputed for post-intercurrent event analysis timepoints.

4.5.2 Secondary Efficacy Endpoints

4.5.2.1 Type 1 Error Control for Co-Primary and Secondary Endpoints

Along with the co-primary endpoints (change from baseline at Week 24 in NPS and NCS), all secondary endpoints will be included in the type 1 error control.

The following hypotheses will be tested in the FAS of Study GA39855. The corresponding null hypotheses are denoted as

- H1: No difference between the treatment groups for change from baseline at Week 24 in NPS

- H2: No difference between the treatment groups for change from baseline at Week 24 in average daily NCS

- H3: No difference between the treatment groups for change from baseline at Week 24 in the average daily SSS

- H4: No difference between the treatment groups for change from baseline at Week 24 in average daily PRS

- H5: No difference between the treatment groups for change from baseline at Week 16 in NPS

- H6: No difference between the treatment groups for change from baseline at Week 16 in average daily NCS

- H7: No difference between the treatment groups for HRQoL as assessed by the change from baseline at Week 24 in total SNOT-22

- H8 : No difference between the treatment groups for change from baseline at Week 24 in average daily ARS

- H13 : No difference between treatment groups for reduction in the need for nasal polypectomy by Week 24, as defined by an NPS of ≤ 4 (unilateral score of ≤ 2 on each side) and improvement in SNOT-22 score of ≥ 8.9

- H14 : No difference between treatment groups for change from baseline at Week 24 in the average daily TNSS

- H15: No difference between the treatment for sense of smell, as assessed by the change from baseline at Week 24 in UPSIT

Studies GA39688 and GA39855 are identical in design. The following endpoints will be tested using the PFAS (combined data) from Studies GA39688 and GA39855 if the analysis findings from the two studies are consistent. The corresponding hypotheses are denoted as

- H9: No difference between treatment groups for requirement of rescue medication (systemic CS for ≥ 3 consecutive days) through Week 24

- H10: No difference between treatment groups for having had nasal polypectomy through Week 24

- H11: No difference between treatment groups for change from baseline at Week 24 in Asthma Quality of Life Questionnaire (AQLQ) of ≥ 0.5 (in patients with comorbid asthma only)

- H12: No difference between treatment groups for requirement of rescue treatment (systemic CS or nasal polypectomy).

The alternative hypothesis is that there is a difference between treatment groups for the respective endpoints.

Within Study GA39855, the co-primary hypotheses (H1 and H2) are tested simultaneously at first. If both of the co-primary hypotheses are rejected at a two-sided significance level of 0.05, then the secondary hypotheses are tested in a sequential order of H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14 and H15 switching between FAS and PFAS as specified above. A secondary hypothesis can be tested for statistical significance at a two-sided significance level of 0.05 if all of the previous hypotheses are rejected at a two-sided significance level of 0.05. Otherwise, testing for statistical significance will stop within the study. For example, H8 can be tested at a two sided significance level of 0.05 if all of H1, H2, H3, H4, H5, H6, and H7 are rejected. H9, H10, H11, and H12 are based on PFAS (combined data from both studies) and must meet the additional condition that in separate analyses in FAS of both Studies GA39688 and GA39855 for a given endpoint, the findings must be consistent prior to formal testing based on PFAS (pooled data).

4.5.2.2 Type 1 Error Control for Co-Primary and Secondary Endpoints for EMA

The same type 1 error control procedure described in the Section 4.5.2.1 will be applied for EMA but with a different set (and order) of secondary endpoints as agreed with EMA.

Along with the co-primary endpoints (NPS at Week 24 and NCS at Week 24), the following secondary endpoints will be included in the type 1 error control:

- Requirement of rescue treatment (systemic CS for ≥3 consecutive days) or having had nasal polypectomy through Week 24

- Change from baseline at Week 24 in the average daily TNSS

- Change from baseline at Week 24 in patient-reported HRQoL as assessed by the total SNOT-22

- Change from baseline at Week 24 in sense of smell, as assessed by the UPSIT

- Change from baseline at Week 16 in NPS

- Change from baseline at Week 16 in the average daily NCS

The following hypotheses will be tested in the FAS of Study GA39855. The corresponding null hypotheses are denoted as

- H1: No difference between the treatment groups for change from baseline at Week 24 in NPS

- H2: No difference between the treatment groups for change from baseline at Week 24 in average daily NCS

- H3: No difference between the treatment groups for change from baseline at Week 24 in average daily TNSS

- H4: No difference between the treatment groups for change from baseline at Week 24 HRQoL as assessed by the total SNOT-22

- H5: No difference between the treatment groups for change from baseline at Week 24 in UPSIT

- H6 : No difference between the treatment groups for change from baseline at Week 16 in NPS

- H7 : No difference between the treatment groups for change from baseline at Week 16 in average daily NCS

Studies GA39688 and GA39855 are identical in design. The following endpoints will be tested using the PFAS (combined data) from Studies GA39688 and GA39855 studies if the analysis findings from the two studies are consistent. The corresponding hypotheses are denoted as

- H8: No difference between treatment groups for requirement of rescue treatment (systemic CS or nasal polypectomy).

The alternative hypothesis for each of the hypotheses is that there is a difference between treatment groups for the respective endpoint.

Within Study GA39855 the co-primary hypotheses (H1 and H2) are tested simultaneously first. If both of the co-primary hypotheses are rejected at two-sided significance level of 0.05, then the secondary hypotheses are tested in a sequential order of H3, H4, H5, H6, H7, and H8.

4.5.2.3 Change from Baseline at Week 24 in Sino-Nasal Outcome Test-22

The SNOT-22, a disease specific HRQoL measure, has a range 0–110 (lower score indicating less disease and better HRQoL). The same estimand and primary estimator method for NPS will be used for SNOT-22.

Supplementary Analysis for EMA

The relative (percent) change from baseline at Week 24 in SNOT-22 will be analyzed using the same method as the absolute change from baseline at Week 24 in SNOT-22 except that the covariates baseline SNOT-22 and baseline SNOT-22 by treatment interaction will be removed from the model.

4.5.2.4 Change from Baseline at Week 24 in Nasal Symptoms Other than NCS

In addition to daily NCS, the following nasal symptoms will be measured daily in the morning:

- Loss of smell

- Posterior rhinorrhea

- Anterior rhinorrhea

As with the NCS, these symptoms are each scored on the integer scale ranging from 0 (not at all) to 3 (severe). For each of these symptoms, the change from baseline in average daily symptom score over the 7 days prior to planned study visit day at each timepoint will be calculated and analyzed according to the assumptions and procedures outlined for NCS in Section 4.5.1.1.

4.5.2.5 Change from Baseline at Week 24 in Total Nasal Symptom Score

Patient-reported TNSS will be the sum of the four equally weighted individual daily symptom scores, as measured every morning: NCS, SSS, PRS, and ARS. Each daily subscore ranges on the integer scale from 0 [not at all] to 3 [severe], and the TNSS ranges from 0 to 12. For the TNSS, the change from baseline for average daily TNSS over the 7 days prior to planned study visit day will be calculated and analyzed according to the estimand and primary estimator method for NCS. The imputation for observations after the intercurrent events will be performed at the day level and

according to the worst observed daily TNSS post-randomization and on or prior to intercurrent event as per Section 4.5.1.1.

4.5.2.6 Change from Baseline at Week 16 in the Average Daily NCS

The same MMRM model used in Section 4.5.1.1 for the change from baseline at Week 24 in the average daily NCS will be used for the change from baseline at Week 16 in average daily NCS. The null hypothesis will be tested by the t-test associated with the treatment group difference in LS-means (derived from the MMRM model for the co-primary endpoint) of the change from baseline in average daily NCS at Week 16.

4.5.2.7 Change from Baseline at Week 16 in the NPS

The same MMRM model used in Section 4.5.1.2 for the change from baseline at Week 24 in the NPS will be used for the change from baseline at Week 16 in the NPS. The null hypothesis will be tested by the t-test associated with the treatment group difference in LS-means (derived from the MMRM model for the co-primary endpoint) of the change from baseline in the NPS at Week 16.

4.5.2.8 Change from Baseline at Week 24 in UPSIT

The UPSIT is a 40-question instrument that measures an individual's ability to detect odors and ranges from 0−40. The same estimand and analysis methods as will be used for NPS will be used for UPSIT.

4.5.2.9 Reduction in the Need for Nasal Polypectomy by Week 24, as Defined by an NPS of ≤4 (Unilateral Score of ≤2 on Each Side) and Improvement in SNOT-22 Score of ≥8.9

The following estimand of interest will be estimated:

- Population: same as for NCS estimand

- Variable: binary response variable indicating reduction in the need for nasal polypectomy by Week 24, defined as if an NPS of ≤4 (unilateral score of ≤2 on each side) and improvement in SNOT-22 score of ≥8.9 then will be coded "yes", if both NPS and SNOT-22 are non-missing and either NPS >4 at Week 24 or SNOT-22 is not improved ≥8.9 at Week 24 will be coded "no"

- Intercurrent event: (1) Had rescue treatment not been made available prior to Week 24; (2) regardless of study drug discontinuation due to AE/PD/LOE. All patients having an intercurrent event will be imputed as "no", not having had the event

- Population-level summary: Treatment group odds-ratio.

A reduced need for nasal polypectomy by Week 24 is defined by a reduction in NPS to ≤4 (unilateral score of ≤2 on each side) and with improvement from baseline at Week 24 in SNOT-22 score of ≥8.9 without rescue treatment (systemic CS or surgery). Patients who have not had an intercurrent event and without valid Week 24 assessments of both NPS and SNOT-22 at Week 24 will have a missing outcome.

The treatment group odds-ratio at Week 24 with a reduced need for nasal polypectomy will be estimated using a logistic regression model analysis adjusted for geographic region, asthma/aspirin sensitivity comorbidity status, as well as baseline value of the measures used to define the endpoint (in this case, baseline NPS and SNOT-22). The null hypothesis for treatment group difference in the reduction in the need for nasal polypectomy by Week 24 will be assessed by the Wald Chi-square test of the treatment term in the logistic regression model. If model convergence is an issue, then Fisher's Exact test will be used, unadjusted for baseline covariates.

4.5.2.10　Requirement of Rescue Treatment (Systemic Corticosteroids for ≥3 Consecutive Days or Having Had Nasal Polypectomy) through Week 24

Due to the low frequency of this clinically meaningful endpoint, the analysis will be conducted for Study GA39855 and also conducted on pooled data from Studies GA39688 and GA39855 in a separate analysis if the analysis findings are consistent for the two studies.

The following estimand of interest will be estimated:

- Population: same as for NCS estimand

- Variable: binary event variable defined as if requirement of rescue treatment through Week 24 defined by having taken systemic CS for ≥3 consecutive days or having had nasal polypectomy at any point between randomization and Week 24 inclusive then coded yes; if follow-up is ≥155 days and no rescue treatment is required then coded "no".

- Intercurrent event: regardless of study drug discontinuation due to AE/PD/LOE. All patients having an intercurrent event will be imputed as "yes"

- Population-level summary: Treatment group odds-ratio.

In this analysis, requiring rescue treatment is defined as having taken systemic CS for ≥3 consecutive days or having had sinus nasal polypectomy at any point between randomization and Week 24, inclusive. The treatment group odds-ratio of patients requiring rescue treatment will be estimated using a logistic regression model analysis adjusted for geographic region, asthma/aspirin sensitivity comorbidity status, and study (Study GA39688, Study GA39855 for pooled analysis). Patients whose last visit falls prior to Day 155 (Week 24 minus 13-day interval is the first day of the Week 24 analysis visit window) and did not already meet the criteria for the event and intercurrent event will have a missing outcome.

The null hypothesis for treatment group difference in the requirement of rescue treatment by Week 24 will be assessed by the Wald Chi-square test of the treatment term in the logistic regression model. If model convergence is an issue, then Fisher's Exact test will be used.

4.5.2.11 Requirement of Rescue Medication (Systemic Corticosteroids for ≥3 Consecutive Days) through Week 24

Due to the low frequency of this clinically meaningful endpoint the analysis will be conducted for Study GA39855 and also conducted on pooled data from Studies GA39688 and GA39855 in a separate analysis if the analysis findings are consistent for the two studies. The same estimand, procedures, and imputation rules outlined in Section 4.5.2.10 will be performed with a change in definition of the event variable to having taken systemic CS for ≥3 consecutive days at any point between randomization and Week 24 (regardless of surgical status). A two-level study covariate will be added to the model for analyses of pooled data.

4.5.2.12 Having Had Nasal Polypectomy through Week 24

Due to the low frequency of this clinically meaningful endpoint the analysis will be conducted for Study GA39855 and also conducted on pooled data from Studies GA39688 and GA39855 in a separate analysis if the analysis findings are consistent for the two studies. The same estimand, procedures, and imputation rules outlined in Section 4.5.2.10 will be performed with a change in the definition of the event variable to having had nasal polypectomy at any point between randomization and Week 24 (regardless of systemic CS usage). A two-level study covariate will be added to the model for analyses of pooled data.

4.5.2.13 Change from Baseline at Week 24 in AQLQ of ≥0.5 (in Patients with Comorbid Asthma Only)

This analysis will be conducted in the subgroup of patients with comorbid asthma at baseline in Study GA39855 and will also be conducted on pooled data from Studies GA39688 and GA39855 in a separate analysis if the analysis findings are consistent for the two studies.

The following estimand of interest will be estimated:

- Population: same as for NCS estimand

- Variable: binary response variable defined as: if a change from baseline at Week 24 in AQLQ of ≥0.5 coded as "yes; if non-missing change from baseline at Week 24 in AQLQ < 0.5 codes as "no".

- Intercurrent event: (1) Had rescue treatment not been made available prior to Week 24; (2) regardless of study drug discontinuation due to AE/PD/LOE. All patients having an intercurrent event will be imputed as "no", not having had the event

- Population-level summary: Treatment group odds-ratio.

This analysis will be conducted only in the comorbid asthma subgroup of the FAS as defined in Section 4.1.1. The AQLQ is a 32-item measure of asthma-related quality of life (QoL) with a total score (the mean of all 32 responses) ranging from 1 to 7; a higher score indicates a better QoL. Patients whose last visit falls prior to Day 155 (Week 24

minutes 13 day interval) without assessments of AQLQ at Week 24 and have not had an intercurrent event will have a missing outcome.

The treatment group odds-ratio of patients with a change from baseline at Week 24 in AQLQ of ≥ 0.5 will be estimated using the same methods in Section 4.5.2.9 with an additional two-level covariate of the study (Study GA39688 or Study GA39855). The null hypothesis for treatment group odds-ratio of patients with a change from baseline at Week 24 in AQLQ of ≥ 0.5 will be assessed by the Wald Chi-square test of the treatment term in the logistic regression model. If model convergence is an issue, then Fisher's Exact test will be used, unadjusted for baseline covariates.

4.5.3 <u>Exploratory Efficacy Endpoints</u>

There will be no type 1 error control for exploratory endpoints. Unadjusted two-sided p-values will be reported.

Change from baseline at Week 24 in SNOT-22 of at least the MID (8.9 points)

The estimand and primary estimator method for change from baseline at Week 24 in SNOT-22 of at least the MID (8.9 points improvement) will follow those used in Section 4.5.2.9.

Change from baseline at Week 24 in EQ-5D-5L

The absolute value and change from baseline to Week 16 and Week 24 of the visual analogue scale (VAS) for the EQ-5D-5L will be summarized by treatment group using mean, standard deviation, median, and range. The percentage of response to each of the five questions will be summarized at baseline, Week 16 and Week 24.

Change from baseline at Week 24 in average daily NCS of at least 0.5 and 1 point

The estimand and primary estimator method for the proportions of patients with ≥ 0.5 and ≥ 1 point improvement in change from baseline at Week 24 of daily NCS averaged over the prior 7 days will follow those used in Section 4.5.2.9. Patients with baseline NCS that did not meet inclusion criteria of NCS >1 will be excluded from the analysis. If model convergence is an issue, then Fisher's Exact test will be used.

Change from baseline at Week 24 in NPS of at least 1 and 2 points

The estimand and primary estimator method for the proportions of patients with ≥ 1 point and ≥ 2 point improvement in change from baseline at Week 24 of NPS will follow those used in Section 4.5.2.9.

4.6 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

The analyses of all pharmacodynamic endpoints will be based on the safety analysis set (see Section 4.1.3). Serum total and free IgE concentrations will be measured at V1

(total IgE only), Day 1, Week 16, and Week 24 during the treatment period, and at Week 28 during the SFU period according to the SOA (see Appendix 2). Descriptive summaries (mean, SD, coefficient of variation, median, minimum, and maximum) of these pharmacodynamic concentrations will be produced at each timepoint by treatment group. Only pharmacodynamic assessments made on or prior to date of last dose for patients who discontinue study drug early will be considered for assignment to a planned timepoint for analysis per Section 4.2.5.

Serum omalizumab concentrations are measured at the following timepoints: Day 1, Week 16, Week 24, and Week 28 in patients randomized to omalizumab only. Descriptive summaries (mean, SD, coefficient of variation, median, minimum, and maximum) of these omalizumab concentrations will be produced at each for omalizumab patients only based on the safety analysis set (see Section 4.1.3). While treatment groups will be defined according to the treatment received, a review of dosing information including dosing errors, if any, and the pharmacokinetic data for patients with unusual dosing patterns will be conducted to assess for any potential impact on results. Only pharmacokinetic assessments made on or prior to date of last dose for patients who discontinue study drug early will be considered for assignment to a planned timepoint for analysis per Section 4.2.5.

Additional pharmacokinetic analyses may be conducted as appropriate. Population pharmacokinetic/pharmacodynamic modeling may be performed to characterize inter-individual variability, which may be reported separately from the clinical study report.

4.7 SAFETY ANALYSES

Safety will be assessed through the summary of AEs, laboratory test results (hematology and serum chemistry), and vital signs. Safety summaries may include data collected at unscheduled visits, dosing termination visits, early termination visits, SFU visits.

Safety outcomes will be summarized based on the safety analysis set with patients grouped according to the actual treatment received (see Section 4.1.3). Safety summaries will be presented by treatment group for all treated patients.

4.7.1 Exposure of Study Medication

Patients were dosed with study drug according to their screening IgE and body weight based on the dosing table listed in Appendix 5 of the protocol. In cases where initial screening IgE, in combination with body weight, did not meet the protocol inclusion criteria, the IgE could be re-assessed at the investigator's discretion, with this re-assessed level used to determine the study drug dose. The number and percentage of patients receiving each dose of study drug will be summarized by treatment group. The overall duration of exposure (weeks), and number of doses received will be summarized by descriptive statistics (mean, SD, median, range). The overall duration of exposure will be summarized by number of patients and percentage categorically (i.e.,

$>4-8$ weeks, $>8-12$ weeks, etc.). Duration of treatment will be defined based on the difference (in days) between the dates of the first and last dose of study drug plus 1 day.

4.7.2　　Adverse Events

Verbatim descriptions of treatment-emergent adverse events (TEAEs) will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA). A TEAE is defined as any new AE or any worsening of an existing condition with an onset after the first study drug administration. Summaries of treatment-emergent events will be provided for each of the following categories:

- All AEs
- All AEs by severity
- AEs assessed as related to study drug by the investigator
- SAEs
- AEs leading to discontinuation of study drug
- Adverse events of special interest (AESI) (i.e., anaphylaxis/anaphylactoid reactions, drug induced liver injury, and suspected transmission of an infectious agent by the study drug)
- Deaths

The AESIs will be identified as defined below. Summaries by treatment arm will include tabulations by MedDRA preferred terms and/or listings, as appropriate.

- <u>Anaphylactic, anaphylactoid, and hypersensitivity reactions</u>: Potential cases of anaphylactic, anaphylactoid, and hypersensitivity reactions will be identified and sent for adjudication by an independent AAC (see Section 3.2). Members of the AAC review blinded data to adjudicate cases as anaphylaxis per Sampson's Criteria and for relatedness to study drug. A detailed description of the process for identification of potential events and data flow is provided in the AAC Charter. Events adjudicated by the AAC as meeting Sampson's Criteria for anaphylaxis and relatedness to study drug will be summarized. Additionally, a listing of all possible cases of anaphylaxis will be produced.

- <u>Suspected transmission of an infectious agent by the study drug</u>: Events will be identified using the Preferred Term (PT) "Suspected transmission of an infectious agent via product".

- <u>Drug-induced liver injury</u>: Events will be identified and summarized if any one of the following three criteria are met: 1) treatment-emergent AST or $ALT>3\times ULN$ in combination with total bilirubin $>2 \times ULN$; 2) treatment-emergent AST or $ALT>3\times ULN$ in combination with clinical jaundice as represented by any one of the following preferred terms: ocular icterus, jaundice acholuric, jaundice hepatocellular; or 3) if the patient is reported to have the preferred term drug-induced liver injury.

The following AEs identified as risks associated with omalizumab will be listed and summarized separately:

- Serum Sickness Syndrome/Serum Sickness Like Disease (SSLD) (PTs "Serum sickness" and "Serum sickness-like reaction")

- Antibody formation to omalizumab (PTs "Drug specific antibody present", "Human anti-human antibody test", and "Drug specific antibody")

- The following AEs identified as risks associated with omalizumab will be first identified according to the broad search criteria below. Events will be evaluated by a Sponsor scientist and only confirmed events will be summarized. Events identified by the broad search will be listed with the events confirmed by the sponsor scientist flagged:

  – Churg Strauss Syndrome (CSS)/Hyper Eosinophilic Syndrome (HES)/Eosinophilic Granulomatosis with Polyangiitis (EGPA) (broad search criteria: HLGT "Vascular inflammations" and HLT "Eosinophilic disorders")

  – Thrombocytopenia (SMQ [Standardized MedDRA Query] "Haematopoietic thrombocytopenia" [Broad] and PT "Immune thrombocytopenic purpura" [ITP]).

  – Arterial Thrombotic Events (ATEs) (SMQs [broad]: "Myocardial Infarction" and "Other ischaemic heart disease", SMQs [narrow]: "Ischaemic central nervous system vascular conditions", "Haemorrhagic central nervous system vascular conditions" and PTs "Hemiparesis", "Hemiplegia", "Sudden cardiac death", "Sudden death", and "Cardiac death").

  – Malignant neoplasms (SMQ "Malignancies" [broad])

  – Parasitic infections (broad search criteria: the MedDRA High Level Group Terms (HGLT) of "Helminthic disorders", "Mycobacterial infectious disorders", and "Protozoal infectious disorders", MedDRA High Level Term (HLT) of "Listeria infections")

4.7.3 Laboratory Data

Serum chemistry and hematology values and changes from baseline in clinical laboratory values will be summarized by descriptive statistics (mean, SD, median, minimum, maximum) at baseline and throughout the study by treatment group. Worst post-baseline World Health Organization (WHO) grade for clinical laboratory values of platelets, hemoglobin, neutrophils, creatinine, aspartate transaminase (ALT), alanine transaminase (AST), total bilirubin, alkaline phosphatase, sodium, and potassium will be presented by treatment group. Worst post-baseline FDA Healthy Volunteers grade for clinical laboratory values of eosinophils and white blood cells will be presented by treatment group. Shift tables of highest WHO grade post-baseline by baseline grade for platelets, AST, ALT, and total bilirubin will be presented by treatment group.

The baseline value of any variable will be defined as the last available value prior to the first administration of study drug for the treatment period (Day 1), unless otherwise specified.

4.7.4   Vital Signs

Vital signs (pulse rate, systolic and diastolic blood pressure while the patient is in a seated position) and changes from baseline will be summarized over the course of the study by descriptive statistics (mean, SD, median, minimum, and maximum) by treatment group.

4.8   MISSING DATA

If the baseline asthma/aspirin status based on the eCRF data is missing or unknown, then the value entered into IWRS at randomization will be used (i.e., asthmatic and aspirin sensitive, asthmatic not aspirin sensitive, all other).

In all efficacy analyses, missing data handling is specific to the efficacy endpoint and details are given within Section 4.5 for all efficacy endpoints. For analysis of change from baseline endpoints if a patient is randomized but no baseline value exists for that endpoint, then the patient will be excluded from the analysis of that endpoint.

4.9   INTERIM ANALYSES

No efficacy interim analyses are planned.

5. <u>REFERENCES</u>

Ahmad H, Gottlieb K, Hussain F. The 2 + 1 paradigm; an efficient algorithm for central reading of Mayo endoscopic subscores in global multicenter phase 3 ulcerative colitis clinical trials. Gastroenterology Report 2015; 4(1): 35-38.

Bachert C, Mannent L, Naclerio RM, et al. Effect of subcutaneous dupilumab on nasal polyp burden in patients with chronic sinusitis and nasal polyposis: a randomized clinical Trial. JAMA 2016;315:469-79.

Fairley JW, Yardley MPJ, Durham LH, et al. Reliability and validity of a nasal symptom questionnaire for use as an outcome measure in clinical research and audit of functional endoscopic sinus surgery. Clin Otolaryngol 1993;18:436–42.

Farrar JT, Dworkin RH, Max MB. Use of the cumulative proportion of responders analysis graph to present pain data over a range of cut-off points: Making clinical trial data more understandable. Journal of Pain and Symptom Management. 2006;31:369–77.

Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin 1971; 76:378.

Gevaert P, Van Bruaene N, Cattaert T, et al. Mepolizumab, a humanized anti-IL-5 mAb, as a treatment option for severe nasal polyposis. J Allergy Clin Immunol 2011;128(5):989-95.

Gevaert P, Calus L, Van Zele T, et al. Omalizumab is effective in allergic and nonallergic patients with nasal polyps and asthma. J Allergy Clin Immunol 2013;131:110-16.

Gwet KL. Handbook of Inter-Rater Reliability. 4th ed. Gaithersburg, MD USA: Advanced Analytics, LLC, 2014.

Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics, 1997, 53:983–97.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–74.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979; 86: 420-28.

Appendix 1
Protocol Synopsis

TITLE:	A PHASE III, RANDOMIZED, MULTICENTER, DOUBLE-BLIND, PLACEBO-CONTROLLED CLINICAL TRIAL OF OMALIZUMAB IN PATIENTS WITH CHRONIC RHINOSINUSITIS WITH NASAL POLYPS
PROTOCOL NUMBER:	GA39855
VERSION NUMBER:	2
EUDRACT NUMBER:	2017-001718-28
IND NUMBER:	5369
TEST PRODUCT:	Omalizumab (IGE025)
PHASE:	Phase III
INDICATION:	Chronic rhinosinusitis with nasal polyps
SPONSOR:	F. Hoffmann-La Roche Ltd

Objectives and Endpoints

The purpose of this study is to determine the efficacy and safety of omalizumab compared with placebo in adult patients with CRSwNP who have had an inadequate response to standard-of-care treatments. Specific objectives and corresponding endpoints for the study are outlined below.

Primary Efficacy Objective	Corresponding Co-Primary Endpoints
• To evaluate the efficacy of omalizumab compared with placebo	Co-primary endpoints: • Change from baseline at Week 24 in average daily nasal congestion score (NCS) • Change from baseline at Week 24 in nasal polyps score (NPS)
Secondary Efficacy Objective [a]	**Corresponding Endpoints**
• To evaluate the efficacy of omalizumab compared with placebo	• Change from baseline at Week 24 in the average daily total nasal symptom score (TNSS) • Change from baseline at Week 24 in the average daily sense of smell score • Change from baseline at Week 24 in the average daily posterior rhinorrhea score • Change from baseline at Week 24 in the average daily anterior rhinorrhea score • Change from baseline at Week 24 in patient-reported health-related quality of life as assessed by the total SNOT-22 • Change from baseline at Week 16 in the average daily NCS

	• Change from baseline at Week 16 in NPS
	• Change from baseline at Week 24 in sense of smell, as assessed by the UPSIT
	• Reduction in the need for surgery by Week 24, as defined by an NPS of ≤ 4 (unilateral score of ≤ 2 on each side) and improvement in SNOT-22 score of ≥ 8.9
	• Requirement of rescue treatment (systemic CS for ≥ 3 consecutive days) or having had surgery for nasal polyps through Week 24
	• Requirement of rescue treatment (systemic CS for ≥ 3 consecutive days) through Week 24.
	• Having had surgery for nasal polyps through Week 24
	• Change from baseline at Week 24 in AQLQ of ≥ 0.5 (in patients with comorbid asthma only)
Exploratory Efficacy Objective	**Corresponding Endpoint**
• To evaluate the efficacy of omalizumab compared with placebo	• Change from baseline at Week 24 in SNOT-22 of at least the MID (8.9 points).
• To evaluate health status utility scores compared with placebo	• Change from baseline at Week 24 as assessed by EQ-5D
Safety Objective	**Corresponding Endpoints**
• To evaluate the safety of omalizumab compared with placebo	• Incidence of adverse events
	• Incidence of serious adverse events
	• Incidence of adverse events leading to omalizumab/placebo discontinuation
	• Clinically significant change in laboratory values
Pharmacokinetic and Pharmacodynamic Objectives	**Corresponding Endpoint**
• To evaluate the pharmacokinetics and pharmacodynamics of omalizumab	• Serum concentration of omalizumab at specified timepoints outlined in the Schedule of Activities
	• Serum levels of total and free IgE at specific timepoints outlined in the Schedule of Activities

AQLQ = Asthma Quality of Life Questionnaire; CS = corticosteroid; MID = minimal important difference; NCS = nasal blockage/congestion score; NPS = nasal polyp score; SNOT-22 = Sino-Nasal Outcome Test-22; TNSS = total nasal symptom score; UPSIT = University of Pennsylvania Smell Identification Test.

a Order of endpoints presented here does not determine order in a sequential type I error control procedure. Details on type I error control procedures will be given in the Statistical Analysis Plan at a later date.

Study Design

Description of Study

This study is a Phase III, randomized, multicenter, double-blind, placebo-controlled, clinical trial that will be run in parallel with a replicate study, under Protocol GA39688.

The study consists of a 5-week screening/run-in period, a 24-week treatment period, and a 4-week study follow-up period. The 4-week follow-up period will be for all patients unless they enroll at Week 24 into another available sponsor-permitted study of omalizumab in nasal polyps.

The screening/run-in period will include two visits ("1st screening visit" at Day -35 and "2nd screening visit" at Day -7), during which patients will undergo video endoscopy to quantify the size of the polyps and assign a nasal polyps score (NPS) prior to baseline.

Beginning at Day -35 (1st screening visit), patients will be asked to standardize their nasal CS to a regimen of mometasone, 200 μg twice a day (400 μg total daily dosage). Patients deemed by the investigator to be intolerant to a BID regimen of mometasone may remain on a stable dosage of mometasone QD during the run-in period and throughout the duration of the treatment period (two sprays/nostril, both nostrils, 50 μg/spray QD for a total daily dosage of 200 μg). Any patient transitioned to daily mometasone should remain on this regimen for the remainder of study.

After screening/run-in has been completed, eligible patients will be randomly allocated in a 1:1 ratio to receive double-blind treatment with omalizumab or placebo. Randomization will be stratified on comorbid asthma and aspirin sensitivity status at baseline (3 levels: asthmatic and aspirin sensitive, asthmatic not aspirin sensitive, all other) and geographic region (North America, ex-North America).

A modified version of the approved E.U. Summary of Product Characteristics (SmPC) omalizumab dosing table will be used; the SmPC has been modified to treat only the higher body weights reflective of the adult population being studied. The first dose of study drug will be administered on the same day as randomization ([Day 1, Week 0]). Dosing will be repeated every 2 or 4 weeks during a 24-week placebo–controlled treatment period.

Video nasal endoscopy will be performed at Day -35 (1st screening visit) and at baseline (Day -7), and at Weeks 4, 8, 16, and 24 (for a total of 6 endoscopies). Change from baseline in NPS at Week 24 will be used as a co-primary endpoint to evaluate the benefit of omalizumab, where the baseline measurement is performed at Day -7 (2nd screening visit) prior to randomization.

After the treatment period ends (at Week 24), patients will be followed for 4 additional weeks as part of safety follow-up, unless they enroll at Week 24 into another available sponsor-permitted study of omalizumab in nasal polyps.

Number of Patients

A total of approximately 120 patients with CRSwNP whose disease remains inadequately controlled despite daily treatment with intranasal corticosteroid (CS) therapy will be enrolled.

Target Population

Inclusion Criteria

Patients must meet the following criteria for study entry:

- Signed Informed Consent Form
- Age 18–75 years, inclusive, at time of signing Informed Consent Form
- Ability to comply with the study protocol, in the investigator's judgment
- NPS ≥ 5, with a unilateral score of ≥ 2 for each nostril, at screening (Day -35), and on Day -7 (as assessed by a central panel of independent central readers)
- SNOT-22 score ≥ 20 at screening (Day -35) and at randomization (Day 1)
- Treatment with nasal mometasone at least 200 μg per day, or equivalent daily dosing of another nasal CS, for at least 4 weeks before screening (Day -35)
- Treatment with nasal mometasone 200 μg twice a day (or once a day if intolerant to twice daily) during the run-in period with an adherence rate of at least 70%.
- Presence of nasal blockage/congestion with NCS ≥ 2 (1-week recall) at Day -35 and a weekly average at randomization of NCS > 1 with at least one of the following symptoms prior to screening: nasal discharge (anterior/posterior nasal drip) and/or reduction or loss of smell
- Eligibility per the study drug–dosing table (serum IgE level ≥ 30 to ≤ 1500 IU/mL and body weight ≥ 30 to ≤ 150 kg) and ability to be dosed per the dosing table
- Willingness to maintain all background medications stable for the duration of the treatment and follow-up periods

- Willingness and ability to use electronic device to enter study-related information in electronic devices (electronic diary [eDiary]/electronic tablet [eTablet])
- Demonstration of at least 70% adherence to eDiary daily symptom assessment during run-in period, with fully completed entries on at least 4 days in the week prior to randomization
- For women of childbearing potential: agreement to remain abstinent (refrain from heterosexual intercourse) or use acceptable contraceptive methods during the treatment period and for 60 days after the last dose of study drug.
  - Abstinence is only acceptable if it is in line with the preferred and usual lifestyle of the patient. Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not acceptable methods of contraception.
  - *Acceptable methods of contraception include surgical sterilization (e.g., bilateral tubal ligation, vasectomized partner), hormonal contraception (e.g., implantable, injectable, patch, oral), and intrauterine device (IUD).*
  - Women of childbearing potential must have a negative serum pregnancy test result during the screening period.

Exclusion Criteria

Patients who meet any of the following criteria will be excluded from study entry:

- Known history of anaphylaxis/hypersensitivity to omalizumab
- Treatment with investigational drugs within 12 weeks or 5 half-lives (whichever is longer) prior to screening (Day -35)
- Treatment with monoclonal antibodies (e.g., omalizumab, mepolizumab) for 6 months prior to screening (Day -35)
- Current treatment with leukotriene antagonists/modifiers, unless patient has been on stable dosing of such medication for at least 1 month prior to screening (Day -35)
- Treatment with non-steroid immunosuppressants (e.g., cyclosporine, methotrexate, azathioprine, mycophenolate, siroliumus, tacrolimus) within 2 months or 5 half-lives, whichever is longer, prior to screening (Day -35)
- Treatment with systemic corticosteroids (CS), except when used as treatment for nasal polyposis, within 2 months prior to screening (Day -35)
- Usage of systemic CS during the run-in period. Patients requiring systemic CS during run-in may be rescreened after completing systemic CS.
- Treatment with intranasal CS drops or CS−administering devices (e.g., OptiNose® device or stents) within 1 month prior to screening (Day -35) or during the run-in period
- History of nasal surgery (including polypectomy) within 6 months prior to screening
- History of sinus or nasal surgery modifying the structure of the nose such that assessment of NPS is not possible
- Uncontrolled epistaxis requiring surgical or procedural intervention, including nasal packing, within 2 months prior to screening
- Known or suspected diagnosis of cystic fibrosis, primary ciliary dyskinesia (e.g., Kartagener syndrome) or other dyskinetic ciliary syndromes, hypogammaglobulinemia or other immune deficiency syndrome, chronic granulomatous disease and granulomatous vasculitis, granulomatosis with polyangiitis (e.g., Wegener's Granulomatosis), or eosinophilic granulomatous with polyangiitis (EGPA) (e.g., Churg-Strauss syndrome)
- Presence of antrochoanal polyps
- Concomitant conditions that interfere with evaluation of primary endpoint:
  - Nasal septal deviation occluding one or both nostrils
  - Ongoing rhinitis medicamentosa

- – Acute sinusitis, nasal infection, or upper respiratory infection during the run-in period
- – Known or suspected invasive or expansive fungal rhinosinusitis
- Known HIV infection at screening
- Known acute and chronic infections with hepatitis C virus (HCV) and hepatitis B virus (HBV) at screening
- *History of myocardial infarction, unstable angina, cerebrovascular accident, or transient ischemic attack or a known history of a hypercoagulable disorder*
- Infection that meets any of the following criteria:
  - – Resulted in hospital admission within 4 weeks prior to screening
  - – Required treatment with intravenous or intramuscular antibiotics within 4 weeks prior to screening
  - – Any active infection that required treatment with oral antibiotics within 2 weeks prior to screening
  - – Active parasitic infection, including nematodes (e.g., Ascaris, Ancylostoma), platyhelminths (e.g., Schistosoma), or Listeria monocytogenes infection within 6 months prior to screening

  Note: Antibiotics are considered to include any antimicrobial therapy used to treat bacterial, fungal, parasitic, viral, or other infections.
- Active tuberculosis requiring treatment within 12 months prior to screening (Day -35)

  Patients who have completed treatment for tuberculosis at least 12 months prior to screening (Day -35) and have no evidence of recurrent disease are permitted.
- Initiation of or change in allergen immunotherapy within 3 months prior to screening (Day -35) or during the run-in period
- Initiation of or change in aspirin desensitization within 4 months prior to screening (Day -35) or during the run-in period
- Pregnant or breastfeeding, or intending to become pregnant during the study or within 60 days after the last dose of omalizumab
- Current malignancy or history of malignancy within 5 years prior to screening, except for appropriately treated carcinoma in situ of the cervix or non-melanoma skin carcinoma that has been treated or excised and is considered resolved
- Any serious medical condition (including but not limited to significant arrhythmia, uncontrolled hypertension, significant pulmonary disease other than asthma) or abnormality in clinical laboratory tests that precludes the patient's safe participation in and completion of the study
- History of alcohol, drug, or chemical abuse within 6 months of screening

End of Study

The end of the study is defined as the date of the last patient's last visit (LPLV). The LPLV is expected to occur at a maximum of 28 weeks after the last patient is randomized.

Length of Study

Each patient will be followed for up to 33 weeks (5-week screening/run-in period, 24-week treatment period, 4-week off-drug safety follow-up period). The 4-week follow-up period will be for all patients unless they enroll into another available sponsor-permitted study of omalizumab in nasal polyps.

Investigational Medicinal Products

The investigational medicinal product (IMP) for this study is omalizumab.

Test Product (Investigational Drug)

Study drug will be administered subcutaneously to patients by qualified personnel who are not involved with conducting safety or efficacy evaluations using a disposable 25-gauge needle in

the deltoid region of the right or left arm. Alternatively, the injections can be administered in the thigh, if medically significant reasons preclude administration in the deltoid region.

Study drug will be administered subcutaneously every 2 or 4 weeks. Because the solution is slightly viscous, the injection may take 5−10 seconds to administer. The dose (mg) and dosing frequency will be determined by serum total IgE level (IU/mL) (measured before the start of treatment) and body weight (kg) at Day -35. Assignment of the study drug dose will be determined by using the study drug−dosing table. Doses of > 150 mg are divided among more than one injection site to limit injections to no more than 150 mg per site. The reconstituted vial is to be used for single-dose administration only.

Comparator

The placebo will be administered according to the same dose, route, and dosing regimen as omalizumab, shown above.

Non-Investigational Medicinal Products

In this study, mometasone furoate monohydrate nasal spray is considered a non-IMP and is used as background therapy only.

Statistical Methods

Primary Analysis

The analysis of data from the 24-week treatment period may be performed after all patients have either completed the Week 24 visit or discontinued from the treatment period prematurely, and all data from the treatment period are in the database and have been cleaned and verified. Patients who discontinue early will not be replaced.

The analysis of complete data from the study, including data from the safety follow-up period will be performed when all patients have either discontinued the study early or completed the safety follow-up period, all data from the study are in the database, and the database is cleaned and locked.

Determination of Sample Size

A total of approximately 120 patients will be enrolled in this study. Patients will be randomly allocated in a 1:1 ratio to receive treatment with omalizumab or placebo, in addition to intranasal steroids.

The sample size of 120 patients will provide at least 85% power to independently detect both a 0.56-point difference between treatment groups in change from baseline at Week 24 in the average daily NCS and a 1.50-point difference between treatment groups in change from baseline at Week 24 in NPS.

Appendix 2 Schedule of Assessments

	Screening/ Run-In Period		Treatment [a]													SFU [b]	UV	Dosing Term./ Early Term.
Day (Window)	-35	-7 (± 5)	1 (-5/+7) [c]	14 (± 3)	28 (± 3)	42 (± 3)	56 (± 3)	70 (± 3)	84 (± 3)	98 (± 3)	112 (± 3)	126 (± 3)	140 (± 3)	154 (± 3)	168 (± 3)	196 (± 3)		
Visit/Week	V1	V2	W0	W2	W4	W6	W8	W10	W12	W14	W16	W18	W20	W22	W24	W28		
Patient-reported nasal CS usage (eDiary) [e]	<<——>>																	
Nasal symptoms (eDiary) [f]	<<———————————————————————————————>>																	
Adverse events			<<———————————————————————————>>															
Concomitant medications			<<———————————————————————————>>															
Video endoscopy, NPS [g]	x	x			x [h]		x [h]			x [h]				x [h]				
UPSIT			x				x				x				x			
Patient-Reported Outcomes																		
EQ-5D-5L			x								x				x			
SNOT-22	x		x		x		x				x				x			
AQLQ [i]			x								x				x			

Appendix 2   Schedule of Assessments (cont.)

	Screening /Run-In Period		Treatment [a]													SFU [b]	UV	Dosing Term./ Early Term.
Day (Window)	-35	-7 (± 5)	1 (-5/+7) c	14 (± 3)	28 (± 3)	42 (± 3)	56 (± 3)	70 (± 3)	84 (± 3)	98 (± 3)	112 (± 3)	126 (± 3)	140 (± 3)	154 (± 3)	168 (± 3)	196 (± 3)		
Visit/Week	V1	V2	W0	W2	W4	W6	W8	W10	W12	W14	W16	W18	W20	W22	W24	W28		
Laboratory Tests																		
Urine pregnancy test [j]		x	x		x		x		x		x		x		x	x		x
Serum pregnancy test [j]	x																	
PK [k]			x								x				x	x		x
Total IgE [l, m]	x [l]		x								x				x	x		x
Free IgE [m]			x								x				x	x		x
Hematology [n]	x		x								x				x	x		x
Chemistry [o]	x		x												x	x		x
Coagulation [p]	x																	
Viral serology [q]	*x*																	
Urinalysis [r]	x																	
Stool ova and parasite [s]		x																
Optional DNA [t]			x															

Appendix 2   Schedule of Assessments (cont.)

eDiary=electronic diary; EQ-5D-5L=EuroQol 5-Dimension 5-Level Questionnaire; HBcAb=hepatitis B core antibody; HBsAg=hepatitis B surface antigen; HCV=hepatitis C virus; IWRS=interactive web-based response system; NCS=nasal blockage/congestion score; NPS=nasal polyp score; PK=pharmacokinetic; Q2W=every 2 weeks; RBR=Roche Biosample Repository; SNOT-22=Sino-Nasal Outcome Test-22; SFU=safety follow-up; Term.=termination; UPSIT=University of Pennsylvania Smell Identification Test; UV=unplanned visit; W=week.

Notes: Cells shaded in gray pertain to patients on Q2W dosing schedule only.

[a]   Patients requiring omalizumab dosing every 2 weeks, based on IgE levels and weight as in dosing table (see Appendix 5), will return to clinic every 2 weeks for study drug administration between baseline and Week 22. Vital signs (blood pressure and pulse rate) should be assessed prior to all study drug administration, including administration at Weeks 2, 6, 10, 14, 18, and 22 in patients receiving study drug every 2 weeks.

[b]   4-week safety follow-up period for patients unless they enroll at Week 24 into another available sponsor-permitted study of omalizumab in nasal polyps.

[c]   Baseline/Day 1 visit should be targeted for 7 days after the Day -7 visit but may occur as late as 14 days after the Day -7 visit.

[d]   The 1st screening visit at Day -35 is defined as Visit 1 (V1) and the 2nd screening visit at Day -7 is defined as Visit 2 (V2). Thereafter, because some patients will require visits every 2 weeks and other patients every 4 weeks (depending on frequency of study drug administration), all remaining visits are referenced by their timing in relationship to baseline (e.g., Week 0 [W0], Week 2 [W2], Week 4 [W4]).

[e]   Patients will remain on mometasone intranasally throughout the study as specified in Section 4.3.3. At each visit the investigator must ensure that the patient has the necessary doses up to the next visit.

[f]   Patients will be instructed to complete the questions in their eDiary in the morning, within approximately 1 hour of awakening. NCS (1-week recall) at Day -35 used for inclusion criteria will be collected via the EDC system.

[g]   Video endoscopy will be read centrally.

[h]   If initial video endoscopy done during visit is of insufficient quality to allow for assessment of nasal polyps score, patient should return to clinic within 10 working days to repeat video endoscopy.

[i]   AQLQ assessed only in patients with asthma

[j]   For women of childbearing potential

[k]   PK samples should be obtained prior to study drug administration. Residual PK samples will be stored and may be used for further PK analysis and/or ADA analysis. See Section 4.5.8 for further details.

[l]   Dosing of omalizumab should be based on IgE and weight levels from Day -35, and patients must meet criteria for dosing based on dosing table in Appendix 5 using the IgE value from Day -35. Exceptions to the specific timing of IgE measurement may be made for cases in which there are problems with sample processing (e.g., sample destroyed in shipment requiring repeat phlebotomy).

[m]   Total IgE and free IgE samples need to be drawn before study drug administration. Residual samples will be stored and may be used for specific IgE testing. See Section 4.5.8 for further details.

Appendix 2 Schedule of Assessments (cont.)

[n] Hematology, including: WBC count, RBC count, hemoglobin, hematocrit, platelet count, differential count (neutrophils, eosinophils, basophils, monocytes, lymphocytes, other cells)

[o] Chemistry panel (serum or plasma), including: sodium, potassium, chloride, bicarbonate, glucose, BUN or urea, creatinine, total protein, albumin, phosphorus, calcium, total and direct bilirubin, alkaline phosphatase, ALT, AST, uric acid, LDH

[p] Coagulation: INR, aPTT, PT

[q] *Viral serology: HIV, HBsAg, total HBcAb, HCV antibody. Viral serology will not be assessed if prohibited by local regulations or ethics committees.*

[r] Urinalysis, including dipstick (pH, specific gravity, glucose, protein, ketones, blood) and microscopic examination (sediment, RBCs, WBCs, casts, crystals, epithelial cells, bacteria)

[s] Stool ova and parasite examination should be performed on Day -7 in patients with an eosinophil count > 2 times the upper limit of normal on Day -35 and risk factors for parasitic disease, as per Section 4.5.8. Stool ova and parasite examination will be performed by a local laboratory.

[t] Optional DNA collection for donation to RBR. These samples will only be collected from patients who give specific consent for his or her samples to be stored for optional exploratory research.

Appendix 3
Omalizumab Dosing Table for Nasal Polyps

Baseline IgE (IU/mL) (Day -35)	Body Weight (kg) (Day -35)							
	>30-40	> 40–50	> 50–60	> 60–70	> 70–80	> 80–90	> 90–125	>125–150
≥ 30–100	75 mg Q4wk	150 mg Q4wk	150 mg Q4wk	150 mg Q4wk	150 mg Q4wk	150 mg Q4wk	300 mg Q4wk	300 mg Q4wk
> 100–200	150 mg Q4wk	300 mg Q4wk	300 mg Q4wk	300 mg Q4wk	300 mg Q4wk	300 mg Q4wk	450 mg Q4wk	600 mg Q4wk
> 200–300	225 mg Q4wk	300 mg Q4wk	300 mg Q4wk	450 mg Q4wk	450 mg Q4wk	450 mg Q4wk	600 mg Q4wk	375 mg Q2wk
> 300–400	300 mg Q4wk	450 mg Q4wk	450 mg Q4wk	450 mg Q4wk	600 mg Q4wk	600 mg Q4wk	450 mg Q2wk	525 mg Q2wk
> 400–500	450 mg Q4wk	450 mg Q4wk	600 mg Q4wk	600 mg Q4wk	375 mg Q2wk	375 mg Q2wk	525 mg Q2wk	600 mg Q2wk
> 500–600	450 mg Q4wk	600 mg Q4wk	600 mg Q4wk	375 mg Q2wk	450 mg Q2wk	450 mg Q2wk	600 mg Q2wk	
> 600–700	450 mg Q4wk	600 mg Q4wk	375 mg Q2wk	450 mg Q2wk	450 mg Q2wk	525 mg Q2wk		
> 700–800	300 mg Q2wk	375 mg Q2wk	450 mg Q2wk	450 mg Q2wk	525 mg Q2wk	600 mg Q2wk		
> 800–900	300 mg Q2wk	375 mg Q2wk	450 mg Q2wk	525 mg Q2wk	600 mg Q2wk			
> 900–1000	375 mg Q2wk	450 mg Q2wk	525 mg Q2wk	600 mg Q2wk				
> 1000–1100	375 mg Q2wk	450 mg Q2wk	600 mg Q2wk	DO NOT ADMINISTER				
> 1100–1200	450 mg Q2wk	525 mg Q2wk	600 mg Q2wk					
> 1200–1300	450 mg Q2wk	525 mg Q2wk						
> 1300–1500	525 mg Q2wk	600 mg Q2wk						

Q2wk = once every 2 weeks; Q4wk = once every 4 weeks.

Lighter gray shading with **black text** indicates doses to be administered by subcutaneous injection **every 4 weeks**.

Darker gray shading with **white text** indicates doses to be administered by subcutaneous injection **every two weeks**.

Source: Xolair® E.U. Summary of Product Characteristics, October 2016.