The following SAP documents are:

1. ITN084AD_GRADUATE_SAP v1.0. This is the original signed version of the SAP
2. ITN084AD_Graduate_SAP_v1.0_Change_Log_2026-02-05. Signed post-database lock to correct an erroneous error in the Area Under the Curve (AUC) formula. This was a clerical error in the plan and did not impact any primary or secondary study results that utilized AUC calculations.

Both documents have been redacted for personal identifiable information.

STATISTICAL ANALYSIS PLAN

STUDY TITLE:

Grass Pollen Sublingual Tablet Immunotherapy plus Dupilumab for Induction of Tolerance in Adults with Moderate to Severe Seasonal Allergic Rhinitis

PROTOCOL NUMBER:

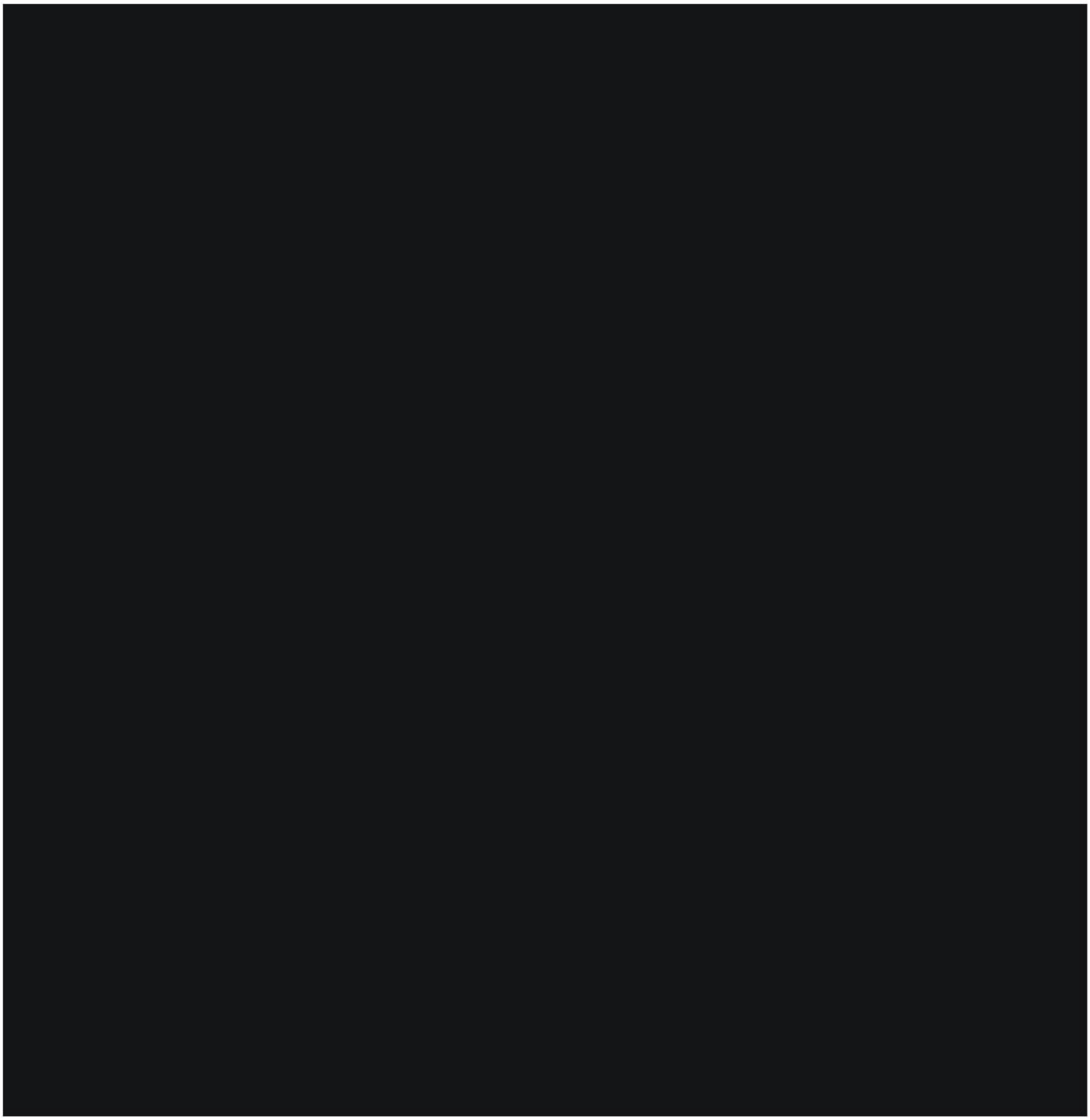
ITN084AD GRADUATE

SHORT TITLE:	**Gra**ss Pollen Immunotherapy plus **Du**pilumab for **T**olerance Induction
NCT#:	NCT04502966
COMPOUND #:	Not applicable
CLIENT:	The National Institute of Allergy and Infectious Diseases (NIAID) [redacted]
REGULATORY AGENCY IDENTIFIER NUMBER(S):	EudraCT: 2018-003456-20
PREPARED BY:	Rho [redacted]

[redacted]

[redacted]

ACKNOWLEDGEMENT AND SIGNATURE SHEET



VERSION HISTORY

SAP Version	Version Date	Change(s)	Rationale
1.0	13NOV2024	First approved version	

TABLE OF CONTENTS

[redacted]

[redacted]

LIST OF TABLES

[redacted]

1. LIST OF ABBREVIATIONS

Table 1: List of Abbreviations

Abbreviation	Term
AE	Adverse event
ARIA	Allergic Rhinitis and its Impact on Asthma
AUC	Area under the curve
BDR	Biomarker and Discovery Research
BiG	Bioinformatics Groups
CRF	Case report form
CSMS	Combined Symptom Medication Score
CSR	Clinical Study Report
DSMB	Data and Safety Monitoring Board
eCRF	Electronic case report form
FEV1	Forced expiratory volume in one second
ICH	International Conference on Harmonization
IgE	Immunoglobulin E
ITN	Immune Tolerance Network
ITT	Intent-to-treat
MAR	Missing at Random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MiniRQLQ	Mini Rhinoconjunctivitis Quality of Life Questionnaire
mITT	Modified intent-to-treat
MNAR	Missing Not at Random
MRSUI	Modified Rhinitis Symptom Utility Index
NAC	Nasal allergen challenge
NIAID	National institute of allergy and infectious disease
PNIF	Peak nasal inspiratory flow
PP	Per protocol
REML	Restricted Maximum Likelihood

SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SLIT	Sublingual Immunotherapy
SOC	System organ class
SQ	Standardized Quality
TNSS	Total Nasal Symptom Score (0-12)
VAS	Visual analogue scale
WAO	World Allergy Organization

2. PURPOSE OF THE ANALYSES

The purpose of this Statistical Analysis Plan (SAP) is to describe the planned efficacy and safety analyses as well as data displays to be submitted for publication and/or included in the Clinical Study Report (CSR) for Protocol ITN084AD. This document provides details on the analysis populations, derivation of variables, plans for the handling of missing data, and statistical methods to be used in these analyses.

The statistical analysis plan (SAP) is based on International Conference on Harmonization (ICH) guidelines E3 (Structure and Content of Clinical Study Reports) and E9 (Statistical Principles for Clinical Trials).

This statistical analysis plan (SAP) only includes analyses related to the clinical endpoints. Mechanistic analyses will be performed at the Immune Tolerance Network (ITN), and a separate analysis plan may be created to detail the planned analyses. Relevant clinical data from the study will be submitted to the ITN Biomarker and Discovery Research (BDR) and ITN Bioinformatics Groups (BiG) to augment the mechanistic analyses.

3. PROTOCOL SUMMARY

Title	Grass Pollen Sublingual Tablet Immunotherapy plus Dupilumab for Induction of Tolerance in Adults with Moderate to Severe Seasonal Allergic Rhinitis
Short Title	**Gra**ss Pollen Immunotherapy plus **Du**pilumab for **T**olerand**e** Induction.
Clinical Phase	Phase II
Number of Sites	1 site
CTA Sponsor/Eudra Number	DAIT NIAID, 2018-003456-20
Study Objectives	The primary objective is to compare the effect of the combination of SLIT and dupilumab versus double placebo on the nasal allergen challenge (NAC) response to grass pollen at year 3, one year after completion of study treatment.
Study Design	This is a randomized, double blinded trial in adults (n=108) with moderate to severe seasonal allergic rhinitis with allergic sensitization to grass pollen. Participants will be recruited from October 2020 through May 2021. Eligible participants will be randomized to one of the following 3 groups in a 1:1:1 ratio. • Combination of SLIT and dupilumab (n=36) • SLIT plus dupilumab placebo (n=36) • SLIT placebo plus dupilumab placebo (n=36)

	Grazax® is a sublingual grass immunotherapy product approved for clinical use in the United Kingdom and will be used as SLIT in this study. It (and its matching placebo) will be self-administered daily by participants through week 104. Dupixent® is the brand name for dupilumab and is a monoclonal antibody against the IL-4 receptor. Dupilumab and its placebo will be administered every two weeks by subcutaneous injection through week 104. With appropriate training by study personnel, alternate doses of dupilumab (and other occasional doses as required) may be self-administered by participants at home. Dupilumab is approved for self-administration by patients. The treatment phase of 104 weeks will be followed by an observation phase of 52 weeks.
Primary Endpoint	NAC (TNSS Area-under-Curve [$AUC_{0\text{-}1hr}$]) at year 3, one year after completion of treatment. The primary comparison will be between the combination of SLIT and dupilumab arm and the double-placebo arms (clinical tolerance endpoint).
Secondary Endpoint(s)	<u>Clinical tolerance endpoints at year 3, one year after completion of study treatment.</u> The comparison will be between the combination of SLIT and dupilumab and the double-placebo arms. 1. Peak nasal inspiratory flow (Delta PNIF $AUC_{0\text{-}1hr}$) 2. Peak NAC TNSS (0-1hr) 3. Size of early and late intradermal skin test response 4. Size of the skin prick test endpoint titration response 5. Seasonal outcomes 1. Weekly seasonal symptoms (Visual Analogue Scale [VAS] 0-10 cms) and combined symptom medication scores (CSMS) 2. Weekly rhinitis quality of life scores (Juniper mini-Rhinoconjunctivitis Quality of Life Questionnaire [miniRQLQ]) measured in-season 3. Modified Rhinitis Symptom Utility Index (MRSUI) measured in-season and out-of-season 4. Global Evaluations No. 1 and No. 2 after the season <u>Clinical desensitization endpoints at year 1 and 2 whilst on study treatment.</u> The comparison will be between the combination of SLIT and dupilumab and the double-placebo arms. 6. TNSS $AUC_{0\text{-}1hr}$ 7. Delta PNIF $AUC_{0\text{-}1hr}$ 8. Size of early and late intradermal skin test response 9. Size of skin prick test endpoint titration response

	10. Seasonal outcomes 1. Weekly seasonal symptoms (VAS 0-10 cms) and CSMS 2. Weekly rhinitis quality of life scores miniRQLQ measured in-season 3. MRSUI measured in-season and out-of-season 4. Global Evaluations No. 1 and No. 2 after the season <u>Safety Endpoints</u> The number, severity, and relatedness of local and systemic AEs and SAEs reported during screening, at baseline, during years 1 and 2 whilst on study treatment, and during year 3 (one year after completion of study treatment). Each treatment arm will be summarized separately.
Exploratory Endpoints	<u>Clinical tolerance endpoints at year 3, one year after completion of study treatment.</u> All pairwise comparisons not included in primary and secondary endpoints above will be reported. 1. TNSS $AUC_{0\text{-}1hr}$ 2. Peak TNSS 3. Total sneeze count (0-15 minutes) from NAC 4. Delta PNIF $AUC_{0\text{-}1hr}$ 5. Size of early and late intradermal skin test response 6. Size of skin prick test endpoint titration response 7. Seasonal outcomes 1. Weekly seasonal symptoms VAS 0-10 cms and CSMS 2. Weekly rhinitis quality of life scores miniRQLQ measured in-season 3. MRSUI measured in-season and out-of-season 4. Global Evaluations No. 1 and No. 2 after the season <u>Clinical desensitization endpoints at year 1 and 2 whilst on study treatment.</u> All pairwise comparisons not included in primary and secondary endpoints above will be reported. 8. TNSS $AUC_{0\text{-}1hr}$ 9. Delta PNIF AUC0-1hr 10. Size of early and late intradermal skin test response 11. Size of skin prick test endpoint titration response 12. Seasonal outcomes 1. Weekly seasonal symptoms VAS 0-10 cms and CSMS

	2. Weekly rhinitis quality of life scores miniRQLQ measured in-season 3. MRSUI measured in-season and out-of-season 4. Global Evaluations No. 1 and No. 2 after the season Mechanistic endpoints at year 3, one year after study treatment (i.e., immune tolerance) and at year 1 and 2 whilst on study treatment (i.e., immune desensitization). All pairwise comparisons will be reported. 13. Mechanistic assessments of local immune responses in the nasal mucosa before and after the NAC. 14. Mechanistic assessments of peripheral blood leukocytes (including eosinophils, mast cells, and basophils) and mononuclear cell subsets after the NAC. 15. Mechanistic assessments of local immune responses and peripheral blood leukocytes (including eosinophils, mast cells, and basophils) and mononuclear cell subsets during natural allergen exposure In versus Out of the pollen season.
Accrual and power calculation	90 of the 108 enrolled participants will complete the study assuming approximately a 16.6% dropout rate. Sample size for this trial was determined by calculating the total enrollment necessary to provide at least 80% power to detect an expected difference in the mean TNSS $AUC_{0\text{-}1hr}$ at year 3 comparing the combination of SLIT and dupilumab arm to the double-placebo arm.
Study Duration	4 years, Oct 2020 - Jan 2025. • Recruitment/screening: 11 months • Dosing: 28 months • Follow-up: 12 months
Treatment Description	Participants will receive daily sublingual Grazax® 75,000 SQ units (containing 15 mcg major allergen *Phl p* 5) or placebo tablets once daily and Dupixent® 300mg or placebo subcutaneous injection every two weeks for 2 years.
Inclusion Criteria	Individuals who meet all of the following criteria are eligible for enrollment as study participants: 1. Participant must be able to understand and provide informed consent. 2. Adults age 18 to 65 years.

	3. A clinical history of grass pollen-induced allergic rhinoconjunctivitis for at least 2 years with peak symptoms in May, June, or July. 4. A clinical history of moderate to severe rhinoconjunctivitis symptoms for at least 2 years interfering with usual daily activities or with sleep as defined according to the Allergic Rhinitis and its Impact on Asthma (ARIA) classification of rhinitis. 5. A clinical history of inadequately controlled rhinoconjunctivitis symptoms despite treatment with antihistamines and/or nasal corticosteroids during the grass pollen season for at least 2 years. 6. Positive skin prick test response at screening, defined as wheal diameter greater than or equal to 3 mm, to *Phleum pratense*. 7. Positive specific IgE at screening, defined as greater than or equal to IgE class 2 (0.7 kU/L) against *Phleum pratense*. 8. A positive response to NAC with Phleum pratense defined as a TNSS greater than or equal to 5 of 12 points. *Note this criterion was removed in protocol version 3.0*. 9. A woman of childbearing potential (WOCBP; for definition, see Appendix 6 of protocol), regardless of birth control history is required to consistently use one of the following highly effective methods of contraception throughout the study: hormonal (e.g. oral, transdermal, intravaginal, implant, or injection); intrauterine device (IUD) or system (IUS); vasectomized partner; bilateral tubal occlusion; or sexual abstinence (for definitions see contraceptive guidance in Appendix 6 of protocol).
Exclusion Criteria	Individuals who meet any of these criteria are not eligible for enrollment as study participants: 1. Inability or unwillingness of a participant to give written informed consent or comply with study protocol. 2. Prebronchodilator forced expiratory volume (FEV_1) less than 70% of predicted value at either screening or baseline visit. 3. A clinical history of asthma requiring regular inhaled corticosteroids for > 4 weeks per year outside of the grass pollen season. 4. A clinical history of moderate to severe allergic rhinitis, as defined according to the ARIA classification of rhinitis, caused by either: a. an allergen to which the participant is regularly exposed OR b. tree pollen during tree pollen season treated with regular antihistamine or intranasal corticosteroids.

	5. History of emergency visit or hospital admission for asthma in the previous 12 months. 6. History of chronic obstructive pulmonary disease. 7. History of recurrent acute sinusitis, defined as 2 episodes per year for the last 2 years, all of which required antibiotic treatment. 8. History of chronic sinusitis, defined as sinus symptoms lasting greater than 12 weeks that includes 2 or more major factors or 1 major factor and 2 minor factors. Major factors are defined as facial pain or pressure, nasal obstruction or blockage, nasal discharge or purulence or discolored postnasal discharge, purulence in nasal cavity, or impaired or loss of smell. Minor factors are defined as headache, fever, halitosis, fatigue, dental pain, cough, and ear pain, pressure, or fullness. 9. History of systemic disease affecting the immune system such as autoimmune diseases, immune complex disease or immunodeficiency. 10. Current symptoms of, or treatment for, upper respiratory tract infection, acute sinusitis, acute otitis media, or other relevant infectious process; serous otitis media is not an exclusion criterion. Participants may be re-evaluated for eligibility after symptoms resolve. 11. A past history of any malignant disease in the previous 5 years. 12. Any tobacco smoking within the last 6 months, or a history of greater than or equal to 10 pack years of cigarette use. Any vaping or electronic cigarette use within the last 6 months. 13. Previous immunotherapy with grass pollen allergen within the previous 5 years. 14. Previous treatment by dupilumab. 15. Previous Grade 4 anaphylaxis (WAO grading criteria) due to any cause. 16. History of anti-IgE, anti-IL-5, anti-IL-5 receptor, anti-IL-4/IL-13 receptor, or other monoclonal antibody treatment. 17. Current use of tricyclic antidepressants or monoamine oxidase inhibitors. 18. Ongoing systemic immunosuppressive treatment. 19. History of intolerance to the study therapy, rescue medications, or their excipients. 20. A positive pregnancy test. 21. Currently lactating/breast feeding.

	22. The use of any investigational drug within 30 days of the screening visit. 23. The presence of any medical condition that the investigator deems incompatible with participation in the trial. 24. Past or current medical problems or findings from physical examination or laboratory testing that are not listed above, which, in the opinion of the investigator, may pose additional risks from participation in the study, may interfere with the participant's ability to comply with study requirements or may impact the quality or interpretation of the data obtained from the study. 25. Eosinophilic esophagitis or a diagnosis of any hypereosinophilic syndrome. 26. Administration of live attenuated vaccines within four weeks of dupilumab or dupilumab placebo injections, before the first injection and throughout the treatment period.
Study Stopping Rules	• Any death that is possibly related or related to one of the investigational products (dupilumab/dupilumab placebo or SLIT/SLIT placebo) or a study procedure. • Any grade 4 AE that is possibly related or related to one of the investigational products (dupilumab/dupilumab placebo or SLIT/SLIT placebo) or a study procedure. • Any grade 3 AE that is possibly related or related to one of the investigational products (dupilumab/dupilumab placebo or SLIT/SLIT placebo) or a study procedure in three or more participants.

4. ANALYSIS SAMPLES

- Screened sample: All participants who undergo screening procedures for purposes of eligibility.
- Randomized sample: All participants who undergo random treatment assignment.
- Safety sample: All participants who undergo random treatment assignment and who receive at least one dose of a study drug.
- Modified intent-to-treat (mITT) sample: All participants who undergo random assignment, who receive at least one dose of both study treatments, and complete the baseline NAC assessment.
- Per protocol (PP) sample: All participants in the mITT sample for whom the primary endpoint is assessed and who are compliant with study treatment, defined as taking at least 50% of SLIT or SLIT placebo tablets and at least 75% of dupilumab or dupilumab placebo injections throughout the treatment period. See Section 7.3 for the formula to calculate treatment compliance for each study treatment; for SLIT or SLIT placebo, the primary method of calculating observed doses taken will be used for PP sample determination.

If a participant is administered the incorrect treatment, analyses based on the randomized, mITT, and PP samples will be based on the treatment arm to which they were randomized. Whereas analyses based on the safety sample will be based on the treatment arm associated with the medication they actually received, regardless of their random assignment.

For purposes of grouping participants into the treatment arm they actually received, the following rules apply:

- Participants randomized to double-placebo arm that received at least one dose of SLIT and at least one dose of dupilumab will be considered in the combination of SLIT and dupilumab arm.
    - Also, in the rare eventuality that a participant randomized to double-placebo arm receives zero doses of SLIT and at least one dose of dupilumab, then the actual treatment will be the combination of SLIT and dupilumab arm.
- Participants randomized to double-placebo arm that received at least one dose of SLIT and zero doses of dupilumab will be considered in the SLIT plus dupilumab placebo arm.
- Participants randomized to SLIT plus dupilumab placebo arm that received at least one dose of dupilumab will be considered in the combination of SLIT and dupilumab arm.
- Participants randomized to combination of SLIT and dupilumab arm will be considered in this arm, unless they received 100% placebo doses for both treatments, in which cases they will be considered in the double-placebo arm; or if they received all dupilumab placebo doses but some SLIT treatment, in which cases they will be considered in the SLIT plus dupilumab placebo arm.
- Otherwise, actual treatment arm will be set to the randomized treatment arm

5. ESTIMAND FOR PRIMARY ENDPOINT

The primary estimand is as follows:

- **Population:** The primary analysis population will be defined as all participants defined by the study inclusion/exclusion criteria who fall into the mITT analysis sample: All randomized participants who received at least one dose of both study treatments and completed the baseline NAC Total Nasal Symptom Score (TNSS) assessment.

- **Variable:** The primary endpoint will be the NAC TNSS area under the curve from 0 to 1 hours ($AUC_{0\text{-}1hr}$) at year 3.

- **Intercurrent Events:** No special handling will be employed for intercurrent events including missing data for any reason; data will be analyzed as observed using a treatment policy strategy. Any missing data prior to and including year 3 visit after treatment will be implicitly handled as Missing at Random (MAR) via the primary mixed-effect model described in Section 9.1.2, without explicit imputation.

- **Population-Level Summary:** The population-level summary will be the difference in least square means at year 3 visit between the combination of SLIT and dupilumab arm and the double-placebo arm (analyzed as randomized) from the primary mixed-effect model.

Sensitivity analyses for the primary estimand will modify the handling of intercurrent events as described in Section 9.1.3 below.

6. STUDY SUBJECTS

6.1. Disposition of Subjects

The disposition of all enrolled participants will be summarized in tables and listed.

The numbers and percentages of participants randomized and in each analysis sample will be displayed by randomized treatment group and overall. Reasons for early termination from the study and visit completion statistics will be presented. For participants discontinuing study treatment early, the reasons for discontinuing study treatment early will also be presented for each treatment. The number and percentage of participants randomized will be summarized.

The listing of disposition data will also include dates of the first dose, randomization, last visit, treatment discontinuation, and termination from study. The listing will be sorted by treatment group and subject identifier.

6.2. Demographic and Other Baseline Characteristics

Summary descriptive statistics for baseline and demographic characteristics will be reported for the randomized and mITT samples. Demographic data will include age, sex, race/ethnicity, body weight at screening, and height at screening. Other baseline characteristic variables will include baseline skin prick test wheal size (mm) to grass pollen and other allergens, and baseline grass pollen specific-IgE. Note that baseline skin prick test wheal size (mm) to grass pollen and other allergens will have the negative control subtracted out, for each participant. If this subtraction causes the wheal size to be <0mm, then a value of 0mm will be imputed for analysis purposes.

These data will be presented in the following manner:

- Continuous data (e.g., age and baseline skin prick test wheal size) will be summarized descriptively by mean, 95% confidence interval for the mean, SD, median, first quartile, third quartile, min, and max.
- Categorical data (e.g., sex and race/ethnicity) will be presented as frequencies and percentages.

Demographic and baseline characteristic data will also be listed.

6.3. Prior and Concomitant Medications

Medications will be coded according to the World Health Organization (WHO) Drug Dictionary (version September-2020, or later). Medications reported in this study include any non-study treatment medication taken, including antibiotics and vaccinations.

Medications reported on the CRF will be categorized for analysis as prior, concomitant, or after study treatment by comparing the medication start and stop dates with the first and last dose of study medication dates. Definitions for each type are as follows:

- Prior medications: both the medication start and stop dates prior to the first dose of study treatment date.

- After medications: both the medication start and stop dates after the last dose of study treatment date.
- Concomitant medications: all other medications not classified as prior or after. Concomitant indicates that the medication overlapped with study treatment by at least one day.

Partial or missing medication start and end dates will be imputed using the algorithm in Appendix 15.3.

The number and percentage of participants receiving prior, concomitant, and after medications will be presented. When reporting the number of participants receiving the medication, a participant will only be counted once if they ever received the medication within the drug class. Percentages will be based on the number of participants in the Safety sample.

Also, a separate summary of antibiotics taken concomitantly will be presented, with the number and percentage of participants, as well as the number of distinct courses/episodes of antibiotic use per treatment group. Antibiotics will be searched for using the WHO drug dictionary coding.

Separate data listings will be provided for prior, concomitant, and after medications.

6.4. Medical History

Medical history will be listed, including the verbatim investigator description of the relevant medical condition, the body system (as collected on eCRF), start date, end date, and whether or not the condition is ongoing.

7. STUDY OPERATIONS

7.1. Protocol Deviations

A protocol deviation is any noncompliance with the clinical trial protocol or International Council for Harmonisation (ICH) of Technical Requirements for Pharmaceuticals for Human Use Good Clinical Practice requirements. The noncompliance may be either on the part of the participant, the investigator, or the study site staff. Protocol deviations will be recorded on the source documents and with an explanation for the deviation.

Protocol deviations will be listed separately for randomized and non-randomized participants with information such as type of deviation, severity of the deviation (major or non-major), date of occurrence, the reason for the deviation, steps taken to address the deviation, and whether or not the deviation was IRB reportable. The number of major protocol deviations, the number of each category of violation, the number of consented participants with at least one deviation, the number of deviations leading to termination of a participant from study and the number resulting in an AE will be summarized.

7.2. Randomization

Participants who met eligibility criteria and chose to enroll underwent a baseline visit and then were randomized 1:1:1 to three parallel treatment arms as follows:

- Combination of SLIT and dupilumab
- SLIT plus dupilumab placebo
- SLIT placebo plus dupilumab placebo (i.e. double-placebo)

A listing of treatment assignments, including the randomized arm and actual treatment received will be provided following the final database lock.

7.3. Measurement of Treatment Compliance

Treatment compliance will be summarized by treatment arm through the total duration of study treatment (104 weeks). In each of the treatment arms, compliance rate will be calculated using the following formula:

$$Compliance\ (\%) = \frac{observed\ doses\ taken}{expected\ number\ of\ doses} \times 100$$

The observed doses taken for each study treatment will be calculated as follows:

- SLIT and SLIT placebo: participants were asked to return used SLIT and SLIT placebo blister packs and unused tablets at each study visit. The number of empty blisters was collected, along with a participant-reported quantity taken and a reliability grade for the participant-reported quantity recorded. The observed doses taken will be calculated in two ways:

  - Primary method: the total sum of a) the participant-reported quantity taken if the reliability grade = "reliable" plus b) the number of empty blisters returned if the reliability grade of the participant-reported quantity is not "reliable", across all dispensing/return visits
  - Supportive method: the total sum of number of empty blisters returned, across all dispensing/return visits
- Dupilumab and dupilumab placebo: the total sum of injection records where injection date is non-missing

The expected number of doses for each study treatment are as follows:

- SLIT and SLIT placebo: planned to be taken once per day for 104 weeks, which means:
  - $Expected\ number\ of\ doses\ =\ 104\ \text{weeks} \times 7\ \text{days per week} = 728\ \text{doses}$
- Dupilumab and dupilumab placebo: planned initial subcutaneous injection of 600mg (counted as two doses) at Week 0 followed by a single injection of 300mg every two weeks after for remaining 104 weeks, which means:
  - $Expected\ number\ of\ doses\ =\ 2 + \left(\frac{104}{2}\right) = 54\ \text{doses}$

8. GENERAL ANALYSIS AND REPORTING CONSIDERATIONS

The following is a list of general analysis and reporting conventions that will be used:

- Categorical variables will be summarized using counts (n) and percents (%) and will be presented in the form n (%). Percentages will be generally rounded to one decimal place.

- In general, first moment statistics, such as the mean, will be reported at 1 more significant digit than the precision of the data. Second moment statistics, such as standard deviation, will be reported at 2 more significant digits than the precision of the data. Other descriptive statistics related to the shape of the data, such as median and range (min/max), will be reported at the same level of precision as the data is recorded. The level of precision may be modified from this convention for specific displays based on clinical judgement.

- Following SAS® default rules, the median will be reported as the average of the two middle numbers if the dataset contains an even number of observations.

- Test statistics including *t* and *z* test statistics will be reported to two decimal places.

- Unless otherwise specified, all statistical tests will be 2-sided using type I error rate = 0.05. Estimates and confidence intervals will be reported to 1 more decimal than the original data. *P*-values will be reported to 3 decimal places. *P*-values less than 0.001 will be displayed as "<0.001"; *p*-values greater than 0.999 will be displayed as ">0.999". A p-value will be reported as 1.000 only if it is exactly 1.000 without rounding.

- No preliminary rounding should be performed; rounding should only occur after analysis. To round, the digit to right of last significant digit will be considered: if < 5 then it will be rounded down, if ≥ 5 then it will be rounded up.

- All analysis will be performed using the SAS System version 9.4.

If departures from these general conventions are present in the specific evaluations section of this SAP, then those conventions will take precedence over these general conventions.

8.1. Standard Calculations

The following standard calculations will be used:

- For purposes of longitudinal analyses of efficacy and key safety, the baseline value (unless otherwise noted) is defined as the last observed data value prior to receiving the first study treatment for treated participants. The baseline derivation can include unscheduled visits. In cases where there is not an actual time collected nor timepoint for the assessment of interest, an assessment date on the same date as the *first study treatment date* will be included for the baseline derivation.

- *First study treatment date* is defined as the earliest of: 1.) first SLIT or SLIT placebo treatment date and 2.) first dupilumab or dupilumab placebo injection date.

- Calculations using dates will be performed as noted below.
  - Study day for dates ≥ first study treatment date

    Study Day = Date – First study treatment date + 1

    Note: Therefore, the first day of study treatment will be considered Day 1.
  - Study day for dates prior to first study treatment date

    Study Day = Date – First study treatment date
  - Total duration of days

    Duration = End date – Start date + 1
- Values reported as greater than or less than some quantifiable limit (e.g., "< 2.0") will be summarized using the numeric equivalent, e.g. "<2.0" will be analyzed as 2.0 in summary tables and figures. When this occurs, a footnote will be added accordingly.

8.2. Multicenter Studies

This is a single site study.

8.3. Assessment Time Windows

All scheduled study visits will aim to occur within the time limits specified in Section 8.5 of the protocol. Unscheduled visits may also occur throughout the study.

For summaries, all data will be associated with the nominal study visit for which it was collected. Repeat assessments and unscheduled (including early withdrawal) visits will be included in study listings and may be separately summarized in tables, but will not be used in place of nominal timepoint data. In the case that an early withdrawal visit overlaps with the window for a missing scheduled visit, then it may be used in substitution of the scheduled visit data, if deemed appropriate to do so. Note that no efficacy endpoint data is planned to be collected at unscheduled (including early withdrawal) visits. For scheduled visits, it is not expected that there will be multiple measurements that occur within the same assessment time window. Should that occur, the closest measurement will be used in the analysis; an earlier measure will be used in case of a tie.

8.4. Timing of Analyses

A single final analysis for purposes of creating a clinical study report will be performed following final database lock using unblinded treatment data. Other intermediate analyses may occur but will be blinded and will not be used to assess the primary or secondary endpoints for any inference claims.

8.5. Multiple Comparisons/Multiplicity

To control the family-wise type I error rate for multiple tests of the primary and key secondary endpoints, the following statistical testing algorithm will be used. Note that all p-values for key

secondary endpoints may be displayed for descriptive purposes only, but only key secondary endpoints deemed significant from the below testing algorithm may be said to demonstrate statistically significant differences between treatments.

The testing algorithm will consist of 3 "sub-families" of endpoints each with their own testing algorithm. In general, the Hochberg testing algorithm will be employed for all sub-families. The Hochberg procedure is a multi-step, step-up testing procedure. The procedure starts with the largest p-value, which is compared to the largest endpoint-specific critical value (i.e. alpha). Then, if the first hypothesis test does not show statistical significance, testing proceeds to compare the second-largest p-value to alpha/2. Sequential testing continues in this manner until a p-value for an endpoint is statistically significant at alpha/m, whereupon that endpoint and all subsequent endpoints (with smaller p-values) are concluded to show statistical significance. In general, the Hochberg procedure is known to control the overall type I error rate when the tests are positively correlated among the endpoints being tested (FDA 2022). This is the case for the endpoints within each sub-family, given the similarity of the endpoints within each group.

1) Clinically-implicated sub-family

a. A modified Hochberg testing algorithm will be used for all endpoints within this sub-family such that the primary endpoint (NAC TNSS $AUC_{0\text{-}1hr}$ at year 3) will be tested first and foremost regardless of its p-value, prior to the testing of the key secondary endpoints within this sub-family.

b. Given there are 5 endpoints including the primary endpoint in this sub-family (see Table 2: Endpoints with Multiplicity Control below), the largest divisor (m) of alpha will be 5.

c. Therefore, the steps are as follows:

   i. Compare the p-value for the primary endpoint with alpha = 0.05. If the p-value is smaller than 0.05 then the treatment effect of the primary endpoint is considered statistically significant. Otherwise, it is considered *not* significant.

   ii. Regardless of the primary endpoint outcome, compare the largest p-value of the key secondary endpoints to alpha/2 = 0.025. If the p-value is smaller than 0.025 then the treatment effect of this key secondary endpoint (with the largest p-value) is considered statistically significant. Otherwise, continue to next step:

   iii. Compare the second largest p-value of the key secondary endpoints to alpha/3 = 0.0167. If the p-value is smaller than 0.0167 then the treatment effect of this key secondary endpoint is considered statistically significant. Otherwise, continue to next step:

   iv. Continue testing the remaining key secondary endpoints in decreasing p-value order at alpha/4 and alpha/5, until statistical significance is reached or all endpoint tests are exhausted.

   v. Note that at whatever step statistical significance is reached (could be step i, ii, iii, or after), the Hochberg testing algorithm will deem all untested

endpoints (which will have smaller p-values) to be statistically significant. Note that it is possible for all key secondary endpoints within this sub-family to be considered not significant, if testing proceeds to the smallest p-value and it is not less than alpha/5.

2) Biomarker sub-family

a. A traditional Hochberg testing algorithm will be used for all endpoints within this sub-family, starting with the largest p-value of the key secondary endpoints being compared to alpha, then second largest p-value being compared to alpha/2, and so on. An alpha of 0.05 will be used for this sub-family, independently of the other sub-families.

3) Survey questionnaire sub-family

a. A traditional Hochberg testing algorithm will be used for all endpoints within this sub-family, starting with the largest p-value of the key secondary endpoints being compared to alpha, then second largest p-value being compared to alpha/2, and so on. An alpha of 0.05 will be used for this sub-family, independently of the other sub-families.

Table 2: Endpoints with Multiplicity Control

1) Clinically-implicated sub-family	**2) Biomarker sub-family**	**3) Survey questionnaire sub-family**
Primary endpoint (tested first and foremost):		
a) NAC TNSS $AUC_{0\text{-}1hr}$ at year 3, for treatment effect of SLIT/dupilumab arm *versus* double-placebo arm	N/A	N/A
*Key secondary endpoints (in no particular order)**		
b) Peak nasal inspiratory flow (Delta PNIF $AUC_{0\text{-}1hr}$) at year 3 c) Peak TNSS at year 3 d) NAC TNSS $AUC_{0\text{-}1hr}$ at year 2 e) NAC TNSS $AUC_{0\text{-}1hr}$ at year 2, for treatment effect of SLIT/dupilumab arm *versus*	a) Size of early intradermal skin test response at year 3 b) Size of late intradermal skin test response at year 3 c) Skin prick test endpoint titration response, as defined by the Provocative Concentration at 5mm (PC5) at year 3	a) Weekly seasonal combined symptom medication scores (CSMS) at year 3 using "in-season" data b) Weekly rhinitis quality of life scores (miniRQLQ) measured in-season at year 3 using "in-season" data

SLIT plus dupilumab placebo arm		c) MRSUI measured in-season at year 3 d) Global Evaluation No. 1 at year 3 e) Global Evaluation No. 2 at year 3

*Note: For all key secondary endpoints, unless otherwise stated, the treatment effect being tested is SLIT/dupilumab arm *versus* double-placebo arm.

The following biological arguments support each sub-family being independent enough such that each sub-family will be tested at their own alpha of 0.05:

- The primary and key secondary endpoints within the clinically-implicated sub-family refer to responses after a grass pollen allergen challenge in the target organ. This target organ (i.e. the nose) shares a common lymphatic drainage to the sublingual mucosa, the site of application of grass pollen sublingual tablets.
- The biomarker sub-family refers to responses after a grass pollen allergen challenge in the skin that is remote from the nose and its lymphatic drainage and therefore is likely to have a distinct underlying mechanism.
- The survey questionnaire sub-family refers to participant-recorded outcomes of the severity of their hay fever during natural seasonal exposure to grass pollen, rather than a controlled experimental allergen challenge in either the nose or skin.
- For these reasons the 3 sub-families of endpoints may be regarded as distinct for the statistical analyses.

Secondary and exploratory endpoints detailed in this SAP which are not primary nor key-secondary, as well as time points and pairwise comparisons other than those listed above, may have p-values reported for descriptive purposes only, as a means for hypothesis generation.

9. ENDPOINT EVALUATION

9.1. Primary Endpoint

The primary endpoint is the NAC TNSS $AUC_{0\text{-}1hr}$ at year 3, one year after completion of treatment.

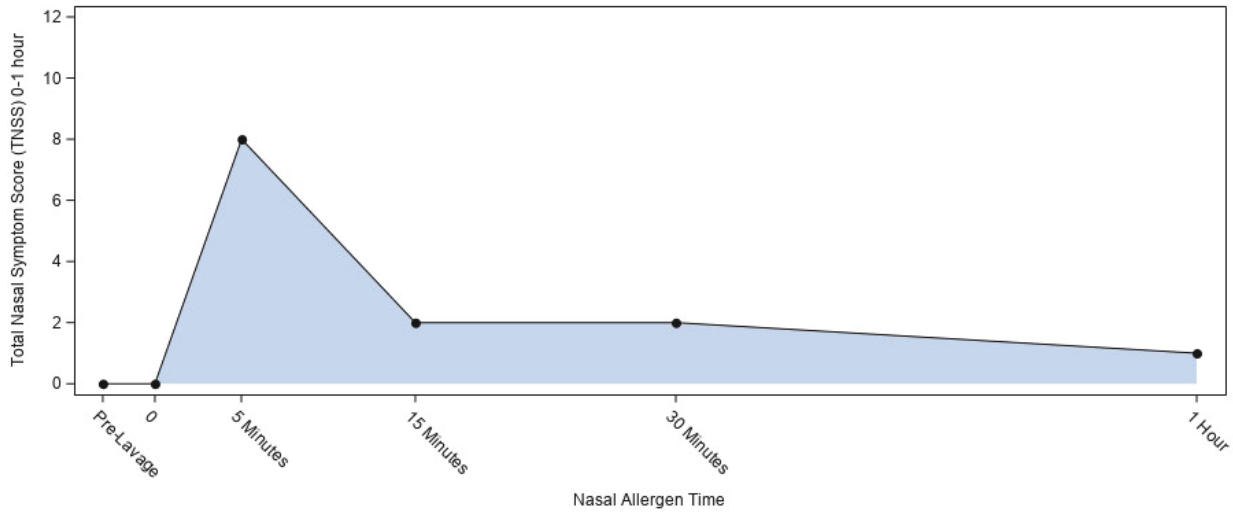
9.1.1. Computation of the Primary Endpoint

The NAC TNSS $AUC_{0\text{-}1hr}$ will be calculated as follows using the linear trapezoidal rule, for each visit the NAC is performed. The primary endpoint is the calculated AUC at the year 3 visit.

$$\text{AUC} = \sum_{i=1}^{n} \frac{\text{TNSS}_{i-1} + \text{TNSS}_i}{2(t_i - t_{i-1})}$$

where TNSS_i = TNSS at time i, n is the number of timepoints being summed through the 1 hour timepoint, and $(t_i - t_{i-1})$ is the nominal time difference in hours between timepoint i and timepoint $i - 1$ for each of the intervals in the time span.

For the first timepoint of the interval, the post-lavage result will used. This will only be used to establish a starting count to calculate first area under curve (the first element in the summation). An illustration of the AUC calculation (as shown by the shaded area) is presented below.



If the last timepoint for 1 hour is missing for a participant included in the primary endpoint analysis, then linear interpolation will be used to impute the missing value using a weighted average of TNSS for the time points on either side, weighted by their relative distance from the nominal time point over the distance between them. This is equivalent to connecting the prior and subsequent values to the missing value (at 30 minutes and 7 hours, respectively) and calculating where the 1 hour value would fall on that line.

9.1.2. Primary Analysis of the Primary Endpoint

The primary analysis will compare the mean TNSS $AUC_{0\text{-}1hr}$ between treatment arms using a longitudinal repeated measures model on data from the mITT sample at years 1, 2, and 3. The model will include fixed effects for treatment arm (all 3 levels), time (i.e. year), and treatment arm by time interaction and will include a covariate for baseline TNSS $AUC_{0\text{-}1hr}$. Since a nonlinear relationship between TNSS $AUC_{0\text{-}1hr}$ and time is expected, time will be treated as a categorical variable (with 3 levels = year 1, year 2, and year 3).

The covariance parameters for the model will be estimated using residual (restricted) maximum likelihood (REML) estimation, and the Kenward-Roger method will be used to compute the denominator degrees of freedom. An unstructured covariance matrix will be assumed to model the correlation among time points within a participant. If convergence issues are experienced with this matrix type, then Toeplitz, first-order auto-regressive, and compound symmetry types will be tried in order and the first matrix to result in successful convergence will be used. If all three alternative matrix types converge, then the best fitting and most parsimonious model may be investigated and chosen as the most suitable model.

The LS means with associated 95% confidence intervals (CIs) will be reported for each treatment arm and time point. Finally, the LS means treatment difference at each time point (year 1, 2, and 3) with associated two-sided p-value will be reported for each pairwise difference:

- Combination of SLIT and dupilumab arm *versus* double-placebo arm
- SLIT plus dupilumab placebo arm *versus* double-placebo arm
- Combination of SLIT and dupilumab arm *versus* SLIT plus dupilumab placebo arm

The primary objective will be addressed by the LS means treatment difference at year 3 between the combination of SLIT and dupilumab arm *versus* double-placebo arm.

Diagnostics for the validity of the primary endpoint analysis will be conducted. Specifically, to assess if the normality assumptions for the longitudinal repeated measures model are met, departures from normality will be investigated by inspecting the model residuals (including a quantile-quantile plot, histogram of residuals, etc.). If the normality assumption is violated, then another suitable statistical technique will be considered, for example including log-transforming the AUC values prior to modeling or performing a non-parametric test.

9.1.3. Sensitivity Analyses of the Primary Analysis

The following sensitivity analysis will be conducted on the primary estimand, by changing the handling of intercurrent events as follows:

1. Hypothetical strategy assuming missing at random

    a. For this sensitivity analysis, a hypothetical strategy will be used such that any missing post-baseline assessments will be imputed using multiple imputation (MI), assuming Missing at Random (MAR) for participants with monotone missing data due to early study discontinuation. Intermittent missing data will first

be imputed using the Markov Chain Monte Carlo (MCMC) method which is appropriate for non-monotonic missing data, with imputations done by treatment arm and adjusted for baseline covariates. Monotone missing data (i.e., data missing after participants discontinue the study early) will then be multiply-imputed using a regression method adjusting for treatment arm, baseline TNSS $AUC_{0\text{-}1hr}$ and other baseline covariates as appropriate. There will be 100 samples imputed for this MI analysis, and results will be summarized across the 100 MI samples. The overall estimate will be compared to the estimate derived from the primary estimand model to assess the potential magnitude and direction of bias.

2. Hypothetical strategy assuming missing not at random

    a. For this sensitivity analysis, a hypothetical strategy will be used such that any missing post-baseline assessments will be imputed using multiple imputation (MI), assuming Missing Not at Random (MNAR) for participants with monotone missing data due to early study discontinuation. The same first MCMC step will be carried out for non-monotonic missing data, as described above. However, for the second step for multiple imputation of the monotone missing data, the MI data will be modelled from data of similar participants from the double-placebo arm following the "copy reference" approach. This hypothetical strategy assumes that active treatment participant values follow that of the double-placebo group after early study discontinuation.

In addition to the above model-based methods of exploring the impact of missing data, demographic characteristics and baseline assessments among subgroups defined by study completion status (completers versus early discounters) will be summarized to identify possible differences of interest.

9.1.4. Supportive Analyses of the Primary Endpoint

The primary analysis of the primary endpoint will be repeated in the same manner as above, but using participants included in the PP sample instead.

For both mITT and PP samples, a plot of LS means (with associated 95% CI) for TNSS $AUC_{0\text{-}1hr}$ for each treatment arm will be plotted by year to illustrate the treatment effects over time. For the mITT sample, this same plot will additionally be repeated using the LS means and 95% CI from the first MI model above assuming MAR.

9.2. Secondary Endpoints

All secondary endpoints denoted as "key" per the multiplicity control testing algorithm (see Section 8.5) are alpha-controlled. The remaining secondary endpoints, including time points and pairwise comparisons that do not appear explicitly in Table 2: Endpoints with Multiplicity Control of this SAP, will be treated as supportive and are not alpha-controlled; therefore p-values are to be treated as descriptive only.

Unless otherwise noted, all secondary endpoint analyses will be performed on data from the mITT sample using a similar model methodology as per the primary endpoint analysis as described in Section 9.1.2 above. All the same model stipulations apply including the inclusion of fixed effects for treatment arm (all 3 levels), time (i.e. year), and treatment arm by time interaction. Any additional baseline covariates that apply for a given secondary endpoint are noted in each sub-section below.

All pairwise treatment differences will be presented for each year post-treatment, as estimated by the models. Unless otherwise specified, the secondary endpoints will focus on the comparison for:

- Combination of SLIT and dupilumab arm *versus* double-placebo arm

The exploratory endpoints will focus on the remaining pairwise treatment comparisons (see Section 9.3).

Clinical tolerance endpoints will come from the model estimates at year 3, one year after completion of study treatment; whereas clinical desensitization endpoints will come from the model estimates at year 1 and 2 whilst on study treatment.

If a large departure from normality is found from model assumption testing, then log-transforming the outcome variable of interest prior to modeling or switching to a non-parametric test at each time point, may be performed instead.

9.2.1. TNSS $AUC_{0\text{-}1hr}$ at years 1 and 2

The same model as for the primary analysis of TNSS $AUC_{0\text{-}1hr}$ above will be used to estimate the LS means treatment difference at years 1 and 2 for the pairwise treatment difference between the combination of SLIT and dupilumab arm *versus* double-placebo arm.

Note as per Section 8.5, only the year 2 treatment differences for a.) combination of SLIT and dupilumab arm *versus* double-placebo arm, and b.) SLIT/dupilumab arm *versus* SLIT plus dupilumab placebo arm, will be part of the multiplicity control. Therefore, the remaining time points and/or pairwise comparisons at years 1 and 2 will be considered supportive and may still have p-values reported for descriptive purposes only.

Also as an exploratory analysis, this same model will be used to estimate the LS means difference between year 3 TNSS $AUC_{0\text{-}1hr}$ and year 2 TNSS $AUC_{0\text{-}1hr}$, for the combination of SLIT and dupilumab arm.

9.2.2. Delta PNIF $AUC_{0\text{-}1hr}$

Peak nasal inspiratory flow, defined as the change in PNIF relative to the pre-allergen exposure PNIF measurement during the NAC (Delta PNIF $AUC_{0\text{-}1hr}$), will be analyzed using a longitudinal repeated measures model on data from years 1, 2, and 3 with a covariate for baseline Delta PNIF $AUC_{0\text{-}1hr}$. All model details per the primary endpoint analysis apply.

For each year's visit, the first step will be to calculate the change in PNIF by subtracting the post-lavage value at each "post" time point, as follows:

$$\text{Change in PNIF } \left({}^{\text{L}}/_{\text{min}}\right)_{\text{timepoint x}} = \left(PNIF_{\text{timepoint x}} - PNIF_{\text{post-lavage}}\right)$$

The change in PNIF will be calculated this way for all time points starting with "5 minutes post". Then, the $AUC_{0\text{-}1hr}$ will be calculated using all available change in PNIF values through the 1 hour time point, using the trapezoidal rule in the same manner as stated in Section 9.1.1. To establish a starting count to calculate first area under curve (the first element in the summation), a change value of 0 will be used. The same linear interpolation details apply if a participant is missing their change in PNIF for the 1 hour time point.

9.2.3. Peak NAC TNSS (0-1hr)

Peak TNSS (0-1hr) during the NAC will be analyzed using a longitudinal repeated measures model on data from years 1, 2, and 3 with a covariate for baseline Peak TNSS (0-1hr). All model details per the primary endpoint analysis apply.

For each year's visit, the peak TNSS will be calculated as the maximum TNSS across all time points during the NAC from 0 to 1 hour.

If any of the "post" timepoints where the peak is expected to occur for each participant (i.e. 5 minutes, 15 minutes, or 30 minutes post) are missing, then the peak TNSS will be set to missing for the main analysis. Sensitivity analyses will be done to include these participants where their missing timepoints will be filled in, as part of a similar multiple imputation (MI) analysis as described in Section 9.1.3 above (assuming MAR).

9.2.4. Size of early and late intradermal skin test response

The size of the early intradermal test response for *Phleum pratense* will be analyzed using a longitudinal repeated measures model on data from years 1, 2, and 3. A similar model will be used to analyze the late intradermal test response.

For analysis purposes, only the average of left and right arm sizes for *Phleum pratense* will be included in the model for each participant and visit (including baseline). In the rare case if only left (or only right) arm size is present, then the analysis value will be set to the single non-missing size. The sizes measured using Saline will not be analyzed.

For approximately 15 participants at the baseline visit, only a 1 BU dose of *Phleum pratense* test concentration was used instead of 10 BU. At the remaining visits for years 1, 2, and 3, the 10 BU concentration was used for all participants. The model-based analyses will be carried out in two ways:

1.) Main model for key secondary endpoint and multiplicity control: Using all participant results where both 1 BU and 10 BU test concentrations are considered. The following additional covariates will be included in the model, to adjust for the differing baseline values among participants:

   a) size of the test response at baseline

   b) categorical fixed effect for test concentration (1 BU versus 10 BU) at baseline

   c) two-way interaction effect between a) and b)

2.) Exploratory model: Including only participants where the intended 10 BU test concentration was used across all visits, i.e. excluding the participants who had a 1 BU test concentration at baseline. A covariate for the size of the test response at baseline will be included.

9.2.5. Size of skin prick test endpoint titration response

The wheal size of the skin prick test endpoint titration response will be assessed for increasing concentrations of *Phleum pratense*, at each year during the study. The following concentrations will be assessed: 10 SQ-U/mL, 100 SQ-U/mL, 1000 SQ-U/mL, 10000 SQ-U/mL, and 100000 SQ-U/mL.

For analysis purposes, the wheal size used for analyses will be calculated as follows:

- The average of left and right arm wheal sizes will be calculated at each year (including the baseline year) for each participant and concentration.
- After taking the average of the left and right arm wheal sizes, the negative control will be subtracted out.
- In the rare case if only left (or only right) arm wheal size is present, then the analysis value will be set to the single non-missing size, minus negative control.
- If the subtraction of the negative control causes the final average wheal size to be <0mm, then a value of 0mm will be imputed for analysis purposes

In addition to the average wheal size, the Provocative Concentration at 5mm (also known as PC5) will be calculated as follows:

- An individual's average wheal size will be plotted per year, for each increasing concentration in a dose-response fashion, on a base-10 log scale where 10 SQ-U/mL = 1, 100 SQ-U/mL = 2, 1000 SQ-U/mL = 3, and so on.
- Using linear interpolation, the PC5 will be derived as the *minimum* concentration (on the base-10 log scale) for which an average wheal size of 5mm was met. This is calculated by taking a weighted average of the concentrations for the two points on either side of 5mm, for the *first* time (i.e. minimum concentration) the average wheal size exceeds 5mm for a given participant. In this case, the PC5 will be fractional value; for example, a PC5 = 2.65 refers to a concentration (on the base-10 log scale) partway between 100 SQ-U/mL (i.e. 2 on log scale) and 1000 SQ-U/mL (i.e. 3 on log scale).
- In the case in which a participant's average wheal size is exactly 5mm and this is the earliest (i.e. *minimum* concentration) at which 5mm is crossed, then the PC5 will be derived as this integer concentration (on the base-10 log scale).
    - In the special case that a participant's average wheal size exceeds 5mm and then drops to exactly 5mm at a higher concentration, the PC5 will be still be derived using the linear interpolated calculated value from the *first* time 5mm is exceeded (as opposed to the later/higher concentration when 5mm was exactly met).

- In the case in which a participant's average wheal size does not exceed 5mm at any of the five concentrations, an arbitrary value of 1000000 SQ-U/mL (i.e. a factor 6 on the base-10 log scale) will be imputed for that participant/year.

The skin prick test endpoint titration response will be analyzed using two different modelling approaches:

1. Main model for key secondary endpoint and multiplicity control: The PC5 (on the base-10 log scale) will be analyzed using a longitudinal repeated measures model on data from years 1, 2, and 3 with a covariate for the PC5 at baseline (i.e. Visit 0).
2. Supportive model: A multivariate repeated measures model will be constructed using a direct Kronecker product covariance structure for the distinct repeated effects of year (3 levels) and concentration (5 levels); see Galecki 1994. This model allows each participant to have each of their concentrations and each of their years of post-baseline data considered in the same model. An unstructured covariance matrix product with a second unstructured covariance matrix (SAS option UN@UN) will be assumed to model the correlation among year and concentration within a participant. If there are convergence issues with this matrix type, then the best fitting and most parsimonious alternative matrix will be used. The LS means with associated 95% CIs will be plotted for each treatment arm by year and concentration to examine which concentrations of grass pollen allergen caused a clinically meaningful skin wheal size at different thresholds (including at 3mm and 5mm). The treatment effect will be estimated from the LS means treatment differences in wheal size for each combination of year and concentration.

9.2.6. Seasonal outcome 1: Weekly seasonal symptoms VAS 0-10, Medication scores, and CSMS

Weekly seasonal symptoms (Visual Analogue Scale [VAS] 0-10 cm), medication scores, and combined symptom medication score (CSMS) were collected over the course of 10 weeks from approximately the end of May through end of July, for each year of the study:

- 2021 (baseline season)
- 2022 (year 1)
- 2023 (year 2)
- 2024 (year 3)

Within a given year, the definition of a pollen season will be programmatically derived using Protocol Table 6 as shown below:

	Pollen Season	Peak pollen season	High pollen days
Start of season	1st day of 5 days (out of 7 consecutive days) each of these 5 days with ≥3 pollen/m^3 and with a sum of these 5 days of ≥30 pollen/m^3	1st day of 3 consecutive days, each with at least ≥50 pollen/m^3	The day(s) with at least 50 pollen/m^3
End of season	Last day of series of 5 days (out of 7 consecutive days) with ≥3 pollen/m^3 and with a sum of these 5 days of ≥30 pollen/m^3	Last day of at least 3 consecutive days, each with ≥50 pollen/m^3	

The following two timeframes will be used for analysis:

a. only the values collected during weeks which are determined to be "in-season" according to the above table's definition of "Pollen Season"

b. all data collected i.e. up to 10 values per participant per year

The main model for analyses (including for the key secondary endpoint of CSMS) will estimate treatment effects from only the "in-season" weeks, whereas supportive analysis models will use all data collected.

The following three outcomes will be analyzed separately:

1. VAS score
2. Medication score
3. CSMS (combination of VAS and medication score)

The score for each outcome will be analyzed for each of the 10 weeks at each year (including the baseline year). A multiple imputation (MI) approach, similar to the approach described in Section 9.1.3 above, will be used to fill in missing values prior to modelling the score of each outcome at each week per year. Missing at random (MAR) will be assumed given the sporadic missingness of the values at each week.

The score for each outcome at years 1, 2, and 3 will be analyzed using a multivariate repeated measures model constructed using a direct Kronecker product covariance structure for the distinct repeated effects of year (3 levels) and week (10 levels); see Section 9.2.5 above for more model details. For the baseline season, a mean score for each outcome will be calculated across the "in-season" weeks such that there is a single baseline mean score for each participant; this baseline mean score will be included in the model as an additional covariate. An unstructured covariance matrix product with a second unstructured covariance matrix (SAS option UN@UN) will be assumed to model the correlation among year and week within a participant. If there are convergence issues with this matrix type, then the best fitting and most parsimonious alternative matrix will be used (for example, SAS option UN@AR(1) which models the effect of week using a first-order autoregressive structure). The LS means with associated 95% CIs will be reported for each treatment arm, year, and week.

To estimate a "yearly average" treatment effect for the main analysis, the LS means for each treatment will be calculated from the model estimates using the average of only the "in-season" weeks, for each year. The LS means treatment difference with associated two-sided p-value will be reported for each pairwise difference, for each year (again using only the "in-season" week estimates).

There will be 100 samples imputed for each MI analysis, and results will be summarized across the 100 MI samples. Only participants with a baseline visit (for the score being modelled) will be included in the main analyses. Sensitivity analyses may be considered to include participants where their entirely missing baseline visit will be filled in as part of the MI imputations.

Note as per Section 8.5, only the mean CSMS model results at year 3 (using "in-season" data) will be part of the multiplicity control. Therefore, the supportive model for CSMS (using all data collected) and the models for the component scores of VAS and medication score will be considered supportive and may still have p-values reported for descriptive purposes only.

For the supportive analyses using all data collected, the LS means for each treatment (and their differences) will be calculated from the *same* main analysis model, using the model estimates from all 10 weeks per year.

It may be explored if adding a time-varying covariate for the weekly mean of grass pollen counts to the model significantly improves model fit, and if so, will be included as an extra covariate in the main analysis model.

9.2.7. Seasonal outcome 2: Weekly rhinitis quality of life scores miniRQLQ measured in-season

Weekly Juniper mini-Rhinoconjunctivitis Quality of Life Questionnaire (miniRQLQ) scores were collected over the course of 10 weeks from approximately the end of May through end of July, for each year of the study using the same collection period as CSMS.

At each week, the average Juniper mini-Rhinoconjunctivitis Quality of Life Questionnaire (miniRQLQ) score will be calculated as the average of the numeric responses (0-6) to all 14 questions in the miniRQLQ questionnaire. Higher scores indicate a worse outcome. If one of the components of the score is missing, then the average miniRQLQ score will also be set to missing at that week.

The miniRQLQ total score will be analyzed for each of the 10 weeks at each year (including the baseline year). A multiple imputation (MI) approach, similar to the approach described in Section 9.1.3 above, will be used to fill in missing values prior to modelling the total score at each week per year. Missing at random (MAR) will be assumed given the sporadic missingness of the values for each year.

The miniRQLQ total score will be analyzed using a multivariate repeated measures model constructed using a direct Kronecker product covariance structure in the same way as CSMS will be modelled (see full model details in Section 9.2.6 above). The "yearly average" treatment effects for the main analysis will be estimated from only the "in-season" weeks as done for CSMS, with a supportive analysis using estimates from all 10 weeks per year.

A covariate for baseline mean miniRQLQ total score will be included in the model.

There will be 100 samples imputed for each MI analysis, and results will be summarized across the 100 MI samples. Only participants with a baseline visit will be included in the main analysis. A sensitivity analysis will be done to include these participants where their entirely missing baseline visit will be filled in as part of the MI imputations.

9.2.8. Seasonal outcome 3: MRSUI measured in-season and out-of-season

The Modified Rhinitis Symptom Utility Index (MRSUI) questionnaire will be collected twice prior to treatment (at Visit -2 in-season and at Visit 0 out-of-season) and twice (in-season in June/July and out-of-season in November/December) in years 1, 2, and 3. At each visit, the total MRSUI score will be calculated as the sum of the numeric responses to all 10 questions. Higher scores indicate a worse outcome, where the maximum score per visit is 25.

The total MRSUI score will be analyzed using a longitudinal repeated measures model in two ways:

1. Main model for key secondary endpoint and multiplicity control: Modeling only in-season assessments (year 1, year 2, and year 3) with a covariate for baseline in-season total MRSUI score (from Visit -2).
2. Exploratory model: Modeling the in-season assessments minus the prior out-of-season assessment at each year. The time-varying impact of the out-of-season assessments will be accounted for by handling each prior out-of-season assessment as a new baseline value at each year. For example, the first in-season assessment at year 1 will subtract the baseline out-of-season total MRSUI score (from Visit 0).

9.2.9. Seasonal outcome 4: Global Evaluations No. 1 and No. 2 after the season

The Global Evaluation No. 1 questionnaire will be collected after the grass pollen season, annually in years 1, 2, and 3. At each visit, the total Global Evaluation No. 1 score will be calculated as the sum of the numeric responses (0-3) to all 6 questions. Higher scores indicate a worse outcome/severity, where the maximum score per visit is 18.

If 1 or 2 questions were not answered at a particular visit (including baseline), then the total Global Evaluation No. 1 score will be imputed using the following method. If 3 or more questions are missing then the total sum will be left missing.

- Calculate the average of the non-missing question numeric responses for the same participant/visit at which the missing question(s) exist
- Impute the numeric response for the missing question(s) with that average
- Calculate the total sum including the imputed numeric responses for the missing questions

The total Global Evaluation No. 1 score will be analyzed using a longitudinal repeated measures model on data from years 1, 2, and 3 with a covariate for baseline total Global Evaluation No. 1 score.

Global Evaluation No. 2 is assessed as a single answer to the question "How was your hay fever this year compared with years before you started immunotherapy treatment?". The single answer ranges from -3 to +3 as per the following table:

Much better (+3)	Better (+2)	A little better (+1)	The same (0)	A little worse (-1)	Worse (-2)	Much worse (-3)

The Global Evaluation No. 2 score will be analyzed using a longitudinal repeated measures model on data from years 1, 2, and 3 without any additional baseline covariates.

An exploratory responder analysis may be performed, where the Global Evaluation No. 2 score will be dichotomized as a yes/no responder status for each participant at each post-baseline year. A "responder" will be defined in two ways:

- A score of +2 or better (i.e. better or much better) compared with the years before
- A score of +1 or better (i.e. a little better, better, or much better) compared with the years before

The proportion of responders will be calculated for each treatment arm, separately at each post-baseline year. Each pairwise treatment difference in the proportion of responders will be calculated, along with the corresponding 95% confidence interval (calculated using the Newcombe method).

9.3. Other Endpoints

All exploratory endpoints will be treated as supportive and will mostly focus on point estimates and 95% confidence intervals. P-values computed for analyses of exploratory endpoints will not be adjusted for multiplicity and are to be treated as descriptive only. Unless otherwise noted, all analyses will be performed on data from the mITT sample and will be analyzed using the same models as described above for the secondary endpoints.

9.3.1. Clinical Tolerance and Clinical Desensitization Endpoints

These exploratory endpoints mirror all secondary endpoints, but focused on the other pairwise treatment comparisons at years 1, 2, and 3 for:

- SLIT plus dupilumab placebo arm *versus* double-placebo arm*
- Combination of SLIT and dupilumab arm *versus* SLIT plus dupilumab placebo arm

*Note that as per Section 9.2.1, the year 2 treatment difference for TNSS $AUC_{0\text{-}1hr}$ for the treatment comparison of SLIT/dupilumab arm *versus* SLIT plus dupilumab placebo arm, is an exception given this comparison is a key secondary endpoint and therefore part of the multiplicity control. All other endpoints for these two above pairwise treatment comparisons are considered exploratory. See Section 8.5 for details of which endpoints are part of the multiplicity control.

9.3.2. Total sneeze count (0-15 minutes) from NAC

The total number of sneezes from 0 to 15 minutes, as collected as part of the NAC, will be analyzed using a generalized Poisson linear mixed model (Stroup 2013, SAS Institute 2015). Often overdispersion occurs because of an incorrect variance function from count data. To adjust the count data models for the likely overdispersion, the data will be modeled using the generalized Poisson distribution, which introduces a scale parameter. The advantage of this distribution is that when no overdispersion is observed, the generalized Poisson reduces to a Poisson distribution.

The generalized Poisson linear mixed model will analyze the total sneeze count data from years 1, 2, and 3 and will include fixed effects for treatment arm (all 3 levels), time (i.e. year), and treatment arm by time interaction. Time will be treated as a categorical variable (with 3 levels = year 1, year 2, and year 3). The model will also include a random intercept to account for the longitudinal design. Note that the number of sneezes was not collected at the screening/baseline NAC so no adjustment for baseline sneeze count will be made.

Given the nature of modelling using a Poisson distribution, the LS means with associated 95% confidence intervals (CIs) for each treatment arm and time point (as well as LS means treatment differences for each pairwise comparison), will be obtained by exponentiating the model estimates.

9.3.3. Pulmonary Function Spirometry Tests

Descriptive statistics of spirometry results and change from baseline results will be summarized for each treatment group and overall. Data listings will also be provided.

9.4. Examination of Subgroups

All subgroup examinations will be exploratory in nature, with a purpose to explore the uniformity of any treatment effects found overall.

First, subgroup analyses of the primary endpoint will include repeating the primary analysis of the primary endpoint using a longitudinal repeated measures model adjusting for treatment arm, time, baseline TNSS $AUC_{0\text{-}1hr}$, and the baseline ratio of grass pollen specific-IgE to total IgE. An interaction term will also be included between the ratio of grass pollen specific-IgE to total IgE and treatment arm, to evaluate if the treatment effect on the primary endpoint is modified by differing levels of the ratio of grass pollen specific-IgE to total IgE. The interaction effect between the treatment group and the subgroup variable will be tested using a Type III fixed effects test.

The same modeling approach will be employed to evaluate if the treatment effect on the primary endpoint is modified by differing levels of:

- baseline grass pollen specific-IgE (i.e. the values alone, not as a ratio to total IgE)
- baseline grass pollen skin prick test wheal size (mm).

Second, subgroup analyses of the primary endpoint will repeat the primary analysis model for the primary endpoint using a *separate* model for each subgroup level. Repeat for the following subgroup levels:

- Age (years): < median and ≥ median
- Sex: male and female
- Race: white, black, Asian, mixed, and other (including unknown/not reported); race categories that have a low sample size may be combined

10. SAFETY EVALUATION

10.1. Overview of Safety Analysis Methods

All safety summaries will be created using the Safety sample, unless otherwise specified. Missing safety information will not be imputed.

Listings will be prepared for all safety measurements, as required for the clinical study report. All listings will be sorted in order of treatment, subject identifier (ID), and time of assessment (e.g., visit, date, and/or time).

10.2. Extent of Exposure

No other summaries for exposure will be created other than the treatment compliance summaries per Section 7.3 above.

10.3. Adverse Events

Adverse events for this study are defined as per the protocol section 12.2.1.

For purposes of analysis, treatment-emergent adverse events (TEAEs) will be defined as all reported AEs with a start date on or after the first date of study treatment.

In order to assign treatment-emergence status, partial or completely missing AE start dates will be imputed using the algorithm in Appendix 15.3.

All AEs will be classified by system organ class (SOC) and preferred term (PT), according to Medical Dictionary for Regulatory Activities (MedDRA), version 23.1 or later.

The severity of AEs will be classified using various grading criteria depending on the type of AE as follows:

Type of AE	**Name of Severity Grade Criteria**	**Grading Scale**
All local reactions to SLIT	Grading Table for Local Reactions to SLIT	Grade 1, 2, 3, or 4
All local reactions to allergen skin testing	Grading Table for Local Reactions to Allergen Skin Testing	Grade 1, 2, or 3
All local reactions to the NAC procedure	NAC procedure (local reactions)	Grade 1, 2, 3, or 4
Any immediate (0-1hr) systemic allergic reaction that occurred following administration of either SLIT or dupilumab All systemic reactions related to allergen skin testing or to the NAC procedure	WAO Subcutaneous Immunotherapy Systemic Reaction Grading System	Grade 1, 2, 3, 4, or 5 (where 5 = death)

All local reactions to dupilumab All other AEs	National Cancer Institute's (NCI's) Common Toxicity Criteria for Adverse Events (CTCAE) v5.0	Grade 1, 2, 3, 4, or 5 (where 5 = death)

Note: See protocol section 12.3.1 for the definitions of each grading criteria.

Each AE is entered on the case report form (CRF) once at the highest severity. As such, no additional data manipulation is needed to identify events.

AE summaries during screening (i.e., any time pre-treatment for treated participants, or during screening period for non-treated participants) will be performed using the Screened sample. All other AE summaries will be performed for only treatment-emergent events by treatment arm using the Safety sample.

An overall summary table will be created separately for 1.) during screening and 2.) only treatment-emergent events, to report the number of events and the number and percentage of participants having at least one event in the following categories. Categories with an asterisk (*) below are only relevant for the treatment-emergent summary (not during screening period).

- Any AE
- Any AE indicated as serious
- Any AE that led to study drug discontinuation*
- Any AE with an outcome of death
- Any AE that was reported as being related to any study drug*
    - Any AE that was reported as being related to SLIT
    - Any AE that was reported as being related to dupilumab
- Any AE that was reported as being related to any study procedure
    - Any AE that was reported as being related to skin prick test
    - Any AE that was reported as being related to NAC procedure
    - Any AE that was reported as being related to venipuncture
- AEs reported by maximum severity grade (all severity grade criteria combined)
- AEs reported by maximum severity grade (within each severity grade criteria separately)

For purposes of assigning relatedness to study drugs, "related" will be considered for analysis purposes if relationship entered on eCRF is "Possibly Related", "Related", or missing.

For purposes of assigning relatedness to study procedures, "related" will be considered for analysis purposes if relationship entered on eCRF is "Possibly Related" or "Related". A missing category will be separately considered.

In addition, AEs classified by MedDRA SOC and PT will be summarized for each treatment group and overall, separately for 1.) during screening and 2.) only treatment-emergent events. These SOC/PT summaries will be repeated for each of the following categories:

- Any AE
- AEs by maximum severity (all severity grade criteria combined)
- AEs by maximum severity (within each severity grade criteria separately)
- SAEs by maximum severity (all severity grade criteria combined)
- SAEs by maximum severity (within each severity grade criteria separately)
- AEs by relationship to study drug (only applicable to treatment-emergent summary)
- AEs by relationship to study procedure (both screening and treatment-emergent summaries)

Summary tables will present the total number of events, as well as the number and percentage of participants experiencing the events. If a participant experiences the same AE on multiple occasions, the event will be counted for each occurrence when reporting the number of events. When reporting the number of participants experiencing the events, a participant will only be counted once if they experience an event within the particular SOC or PT. Percentages will be based on the number of participants in the analysis sample of interest.

In addition to the above treatment-emergent summaries at any time post-treatment, AEs will be summarized during year 1, 2, and 3 separately using the following displays:

- Overall summary (repeated with same categories as above)
- Any AE by SOC/PT
- AEs by maximum severity (all severity grade criteria combined)
- SAEs by maximum severity (all severity grade criteria combined)
- AEs by relationship to study drug

For the above AE year intervals, the analysis windows are as follows for counting events:

Year Interval	**Lower Bound**	**Upper Bound**
Year 1 (Week 0 through <Week 54)	Study Day 1 (first treatment date)*	Study Day 377
Year 2 (Week 54 through < Week 108)	Study Day 378	Study Day 755
Year 3 (Week 108 onwards)	Study Day 756	No upper bound

*Study day is defined per Section 8 of this SAP.

The denominators for percentages will be for each time interval as follows:

Year Interval	Denominator Subset
Year 1 (Week 0 through <Week 54)	All participants in the Safety sample
Year 2 (Week 54 through < Week 108)	All participants in the Safety sample who remained in the study through Study Day 378
Year 3 (Week 108 onwards)	All participants in the Safety sample who remained in the study through Study Day 756

The study day for inclusion for denominators will be based on the maximum of 1.) end of study date, 2.) date of most-recent follow-up, and 3.) last AE start date.

10.4. Deaths, Serious Adverse Events, and Other Significant Adverse Events

Serious adverse events (SAEs) will be listed and summarized in the same manner described above. A separate listing of any death (including time to death and cause of death) as well as a listing of AEs that led to study drug discontinuation will also be created.

10.5. Clinical Laboratory Evaluation

Clinical laboratory measurements include serum chemistry, urinalysis, hematology, and IgE. The IgE collected includes grass pollen specific-IgE, birch pollen specific-IgE, and total IGE. Results will be converted to standardized units where possible.

For numeric laboratory results, descriptive statistics of laboratory values and the change from baseline of laboratory values will be presented for each treatment group. Included in the change from baseline summaries will be the worst post-baseline change from baseline value per parameter/participant. In cases, where the "worst" value can be defined in the low or high direction for a given parameter, both directions will be considered. Note that unscheduled visit results may be considered when deriving the worst value post-baseline.

For categorical laboratory results, the number and percentage of participants reporting each result will be presented for each treatment group.

Data listings will also be provided for clinical laboratory measurements, including the laboratory normal ranges and out-of-range flags to denote abnormal values as high (H) or low (L).

Per the protocol, any abnormal value or result from a clinical or laboratory evaluation may also be entered as an AE and will therefore appear in the AE summary tables.

10.6. Vital Signs, Physical Findings, and Other Observations Related to Safety

10.6.1. Vital Signs

Descriptive statistics of vital signs results and change from baseline of vital signs will be summarized for each treatment group. Data listings will also be provided for vital signs measurements.

10.6.2. Physical Examinations

Any abnormal (including clinically significant) physical examination findings post-randomization will be entered as an AE and will therefore appear in the AE summary tables.

10.6.3. Pregnancy Tests

Serum and urine pregnancy tests will be performed throughout the study, according to the protocol schedule of events. This data will be reported along with the clinical laboratory data.

11. OTHER ANALYSES

A separate analysis plan will be created to detail the planned mechanistic analyses.

12. INTERIM ANALYSES AND DATA MONITORING

The DSMB will receive at least annual safety reports on study participants. However, no formal interim analysis of safety data will be conducted. Also, no interim analysis of efficacy data is planned.

13. CHANGES TO THE ANALYSES PLANNED IN THE PROTOCOL

The following are changes in this analysis plan that differ from those stated in the protocol:

- New key secondary endpoint Peak NAC TNSS (0-1hr) was added, despite this not being part of the protocol list of secondary endpoints
- New exploratory endpoint for Total sneeze count (0-15 minutes) from NAC was added, despite this not being part of the protocol list of exploratory endpoints
- The protocol states that the p-values computed for analyses of secondary endpoints will not be adjusted for multiplicity. As per SAP Section 8.5 states, a multiplicity strategy will be used for key secondary endpoints.
- The protocol states that VAS 0-10 cm, CSMS, and miniRQLQ will be summarized using AUC calculated endpoints. Instead of AUC, the means of VAS, CSMS, and miniRQLQ will be calculated and analyzed at baseline, year 1, 2, and 3. Also, prior to calculation of the mean for each participant per year, multiple imputation will be performed to fill in sporadic missing data.
- The protocol states that the AUC of pollen counts collected weekly during the pollen season in year 1, 2, and 3 will be calculated. Instead, the AUC will be calculated from pollen counts collected *daily*.

14. REFERENCES

FDA (October 2022), "Multiple Endpoints in Clinical Trials: Guidance for Industry," Technical Report, U.S. Department of Health and Human Services, Food and Drug Administration.

Galecki, A. T. (1994), "General Class of Covariance Structures for Two or More Repeated Factors in Longitudinal Data Analysis," Communications in Statistics—Theory and Methods, 23, 3105–3109.

SAS Institute Inc. 2015. SAS/STAT® 14.1 User's Guide. Cary, NC: SAS Institute Inc. http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_glimmix_examples19.htm.

Stroup, Walter W. "Generalized Linear Mixed Models: Modern Concepts, Methods and Applications." CRC Press, 2013: 354-356.

15. APPENDICIES

15.1. Study Flow Chart

See Protocol Figure 2.

15.2. Schedule of Events

See Protocol Appendices 1-5.

15.3. Algorithm for Partial or Missing Start Dates

Partial or Missing Start Dates:

AE start dates that are partial or completely missing will be imputed using the following algorithm to assign treatment-emergence status. The same applies for partial/missing medical start dates for assigning concomitant status:

1. If the year and month are known:
    a. If the year and month are the same as the year and month of the treatment start date, impute to treatment start date.
    b. Otherwise, use the 1st of the month.
2. If only the year is known:
    a. If the year is the same as the year of treatment start date, impute to treatment start date.
    b. Otherwise, impute to January 1st of that year.
3. If the date is completely missing, impute to treatment start date.

Partial or Missing End Dates:

Partial AE (or medication) end dates will be imputed using the same algorithm as above for start dates, but with the following modifications:

1. Date of most-recent follow-up (including end of study date) replaces treatment start date
2. Use last day of month if year/month known, or December 31st if only year known
3. Do not impute completely missing end dates, but leave as missing (i.e. ongoing)

SAP Change Log

Client:	The National Institute of Allergy and Infectious Diseases (NIAID)
Protocol:	ITN084AD GRADUATE

After finalization of the Statistical Analysis Plan (SAP) and post-database lock, one change to the plan was identified, as summarized below:

Section 9.1.1 (Computation of the Primary Endpoint) of SAP v1.0 contained a typographical error (erroneous carriage return) in the area under the curve (AUC) formula. Using the linear trapezoidal rule as expressed in the text, the formula should have had the $(t_i - t_{i-1})$ term in the numerator rather than denominator. The corrected formula is as follows:

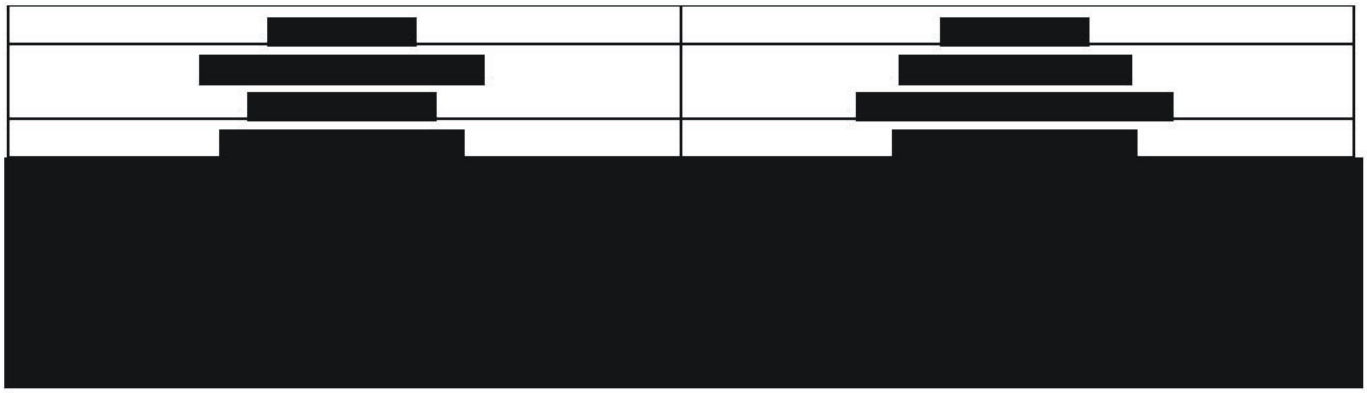
$$\text{AUC} = \sum_{i=1}^{n} \frac{\text{TNSS}_{i-1} + \text{TNSS}_i}{2} (t_i - t_{i-1})$$

where TNSS_i = TNSS at time i, n is the number of timepoints being summed through the 1 hour timepoint, and $(t_i - t_{i-1})$ is the nominal time difference in hours between timepoint i and timepoint $i - 1$ for each of the intervals in the time span.

The erroneous error in the formula was merely a clerical error in the plan and did not impact any primary or secondary study results that utilized AUC calculations; all calculations were conducted according to the correct formula for the linear trapezoidal rule.

Approvals

I am aware of the post-database lock changes made to the analysis plan for this study and approve of them as presented.



[redacted]