MT-12 STATISTICAL ANALYSIS PLAN COVER PAGE

Official trial title	A one-year placebo-controlled phase III trial evaluating the efficacy and safety of the house dust mite (HDM) SLIT-tablet in children (5-11 years of age) with HDM allergic rhinitis/rhinoconjunctivitis with or without asthma
NCT number	NCT04145219
Document date	24-May-2023

Statistical Analysis Plan

Trial ID: MT-12

A one-year placebo-controlled phase III trial evaluating the efficacy and safety of the house dust mite (HDM) SLIT-tablet in children (5-11 years of age) with HDM allergic rhinitis/rhinoconjunctivitis with or without asthma

Sponsor: ALK-Abelló A/S
Bøge Alle 6-8
DK-2970 Hørsholm

Investigational medicinal product: HDM SLIT-tablet

Phase: III

EudraCT No.: 2019-000560-22

Document status: Final

Date: 24-MAY-2023

Version number: 1.0

ALK approval of statistical analysis plan

Trial statistician:

[redacted]

Signature _____ Date _____

Trial manager:

[redacted]

Signature _____ Date _____

Medical writer:

[redacted]

Signature _____ Date _____

Head of biometrics:

[redacted]

Signature _____ Date _____

Head of GPCD:

[redacted]

Signature _____ Date _____

This document is approved by electronic signature, refer to signature page at the end of this document.

Table of contents

Table of figures

Page

Table of tables

Page

Version History

This statistical analysis plan for trial MT-12 is based on the protocol version 5.0 dated 19-Mar-2021.

SAP Version	*Date*	*Change*	*Rationale*
1	24-MAY-2023	Not applicable	Original version

1 List of abbreviations

AE	Adverse event
AR	Allergic rhinitis
ANOVA	Analysis of variance
CI	Confidence interval
CRF	Case report form
CSMS	Combined symptom and medication score (recommended by European academy of allergy and clinical immunology task force)
CSR	Clinical study report
DMS	Daily medication score
DPS	Data points set
DSS	Daily symptom score
ESI	Event of special interest
EAACI	European academy of allergy and clinical immunology
EudraCT	European Union drug regulating authorities clinical trials database
FAS	Full analysis set
FDA	Food and Drug Administration
FEV_1	Forced expiratory volume in 1 second
FVC	Forced vital capacity
GCP	Good clinical practice
GLMM	General linear mixed model
HDM	House dust mite
ICH	International council for harmonisation of technical requirements for pharmaceuticals for human use
ICS	Inhaled corticosteroid
ID	Identification
IgE	Immunoglobulin E

IMP	Investigational medicinal product
ITT	Intention-to-treat
LME	Linear mixed effects model
MAR	Missing at random
MedDRA	Medical dictionary for regulatory activities
MMRM	Mixed-effect model for repeated measurements
MNAR	Missing not at random
PRQLQ	Paediatric rhinoconjunctivitis quality of life questionnaire
PD	Protocol deviation
PT	Preferred term
QoL	Quality of life
SABA	Short-acting β_2-agonist
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SAS	Statistical analysis software
SD	Standard deviation
SE	Standard error
SOC	System organ class
SPT	Skin prick test
SLIT	Sublingual allergy immunotherapy
TC FU	Telephone follow-up
[redacted]	[redacted]
TCRS	Total combined rhinitis score
TCS	Total combined score (of rhinoconjunctivitis symptoms and medication)
TEAE	Treatment-emergent adverse event
WHO	World Health Organisation

2 List of definitions

AE	An AE is any untoward medical occurrence in a clinical trial subject and which does not necessarily have a causal relationship with the administered IMP
Completed	A randomised subject is considered as completed if he/she has not discontinued

subject	the trial
Concomitant medication	All medications being continued by a subject on entry into the trial and all medications given in addition to IMP and rescue medication
IMP	A pharmaceutical form of the active substance or placebo being tested
Solicited AE	A pre-specified AE recorded by the investigator based on the symptoms reported in the subject's eDiary during the first 28 days after randomisation
Source documents	Source documents are original documents, data, and records from which the subjects' eCRF data are obtained. These include, but are not limited to, hospital records (from which medical history and concomitant medication may be summarised into the eCRF), clinical and office charts, laboratory and pharmacy records, diaries, microfiches, radiographs, and correspondence
Trial completion	The trial is completed once the CSR is signed
Treatment-emergent AE	An AE that starts on or after the first IMP administration and no later than 7 days after the last IMP administration

3 Introduction

The statistical analysis plan is a supplement to the protocol and provides additional details regarding estimands (where appliable), analysis sets, endpoints and the statistical analyses.

Supporting documentation is provided in section 9. Section 9.1 'Derivations for the study' contains details about the derivations used in the study. In Section 9.2 'Further details pertaining to statistical analyses', a detailed description of how to code the statistical analyses is provided, and in section 9.3 'Subjects excluded from the sensitivity analysis' a list of subjects to be excluded from the sensitivity analysis is given.

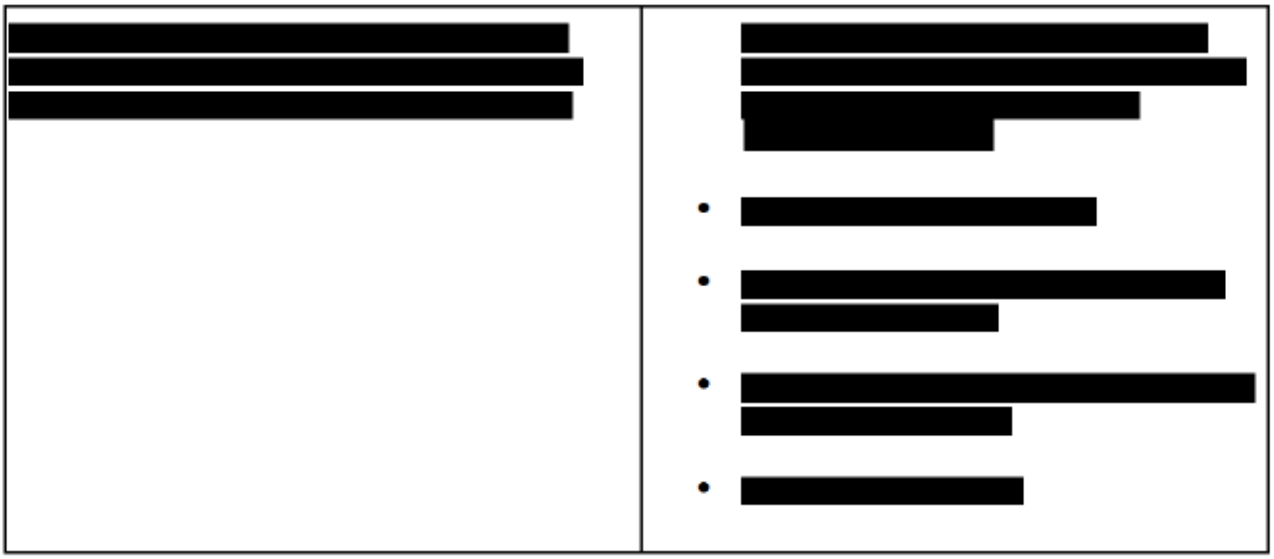
Changes to the analyses and changes that could impact the analyses described in the protocol are documented in section 6.10.

This statistical analysis plan has been written and approved before database lock and unblinding.

3.1 Objectives, endpoints, and estimands

Objectives	**Endpoints**
Primary objective	Primary endpoints
Demonstrate the efficacy of the HDM SLIT-tablet compared to placebo in the treatment of HDM AR in children (5-11 years of age) based	• The average daily TCRS during the primary efficacy assessment period

on total combined rhinitis symptoms and medication use during the primary efficacy assesment period.	
Key secondary objectives	Key secondary endpoints
Demonstrate the efficacy of the HDM SLIT-tablet compared to placebo during the primary efficacy assessment period based on rhinitis symptoms, rhinitis medication use, combined rhinoconjunctivitis symptoms and medication use, and rhinoconjunctivitis QoL.	• The average rhinitis DSS during the primary efficacy assessment period • The average rhinitis DMS during the primary efficacy assessment period • The average daily TCS during the primary efficacy assessment period
Secondary objectives	Secondary endpoints
Evaluate the HDM SLIT-tablet compared to placebo during the primary efficacy assessment period based on safety and tolerability, rhinoconjunctivitis symptoms, rhinoconjunctivitis medication use, rhinoconjunctivitis QoL, asthma symptoms and medication use, changes in immunological parameters, rhinitis mild days, rhinitis exacerbation days, rhinitis CSMS, rhinoconjunctivitis CSMS	• Safety and tolerability assessments • Average rhinoconjunctivitis DSS • Average rhinoconjunctivitis DMS • PRQLQ score • Average asthma DSS • SABA free days • Weekly number of puffs of as-needed SABA use • Changes in immunological parameters • Rhinitis mild days • Rhinitis exacerbation days • Average daily rhinitis CSMS (recommended by the EAACI) • Average daily rhinoconjunctivitis CSMS (recommended by the EACCI)
Explorative objectives	Explorative endpoints
Evaluate the following: [redacted]	• [redacted] • [redacted] • [redacted]



3.1.1 Estimands

The study was not designed using the estimand framework, but since the first version of the protocol the ICH E9 addendum (ICH 2019, 1998) has come into affect. Therefore the primary analysis using an observed case analysis is supplemented with analyses of the trial product estimand and the treatment policy estimand using the estimand framework.

For all estimands considered for the study, the population, the population-level summary, and the intercurrent events are given below.

3.1.1.1 Population

The population is defined by the inclusion and exclusion criteria of the trial to reflect the targeted patient population for approval.

3.1.1.2 Population level summary

The absolute difference in means between placebo and HDM SLIT-tablet.

3.1.1.3 Variable

The two estimands will be defined using the primary endpoint average TCRS, and the key secondary endpoints average rhinitis DSS, average rhinitis DMS, and average daily TCS, all during the primary efficacy assessment period.

3.1.1.4 Intercurrent events

In allergy immunotherapy clinical trials, use of rescue medication is allowed according to the protocol and is included in the treatment condition. Treatment interruptions are also considered part of the treatment condition. Discontinuation of IMP is considered an intercurrent event and will be handled by different strategies for the estimands considered for the study, either the hypothetical strategy, or the treatment policy strategy.

3.1.1.4.1 Hypothetical strategy

The intercurrent event of discontinuation of treatment will be handled by the **hypothetical strategy**, i.e. if a subject discontinues treatment, only diary data up until the time of treatment discontinuation will be included in the analysis, and any data recorded after discontinuation of treatment will be excluded from the analysis.

Thus, an estimand applying the hypothetical strategy is the absolute difference between placebo and HDM SLIT-tablet based on the endpoint, in the population defined by the trial inclusion and exclusion criteria, if all subjects adhered to the treatment regimen. In other words, the estimand is the expected treatment effect in subjects with HDM allergic rhinitis/rhinoconjunctivitis if all subjects adhered to the treatment regimen.

A such estimand is denoted a **trial product estimand**, as it assesses the anticipated effect of the treatment (trial product) if taken as instructed. This is considered to be of most relevance to the patient, as it describes the potential benefit they could obtain from the trial product if they adhere to treatment for the planned duration.

3.1.1.4.2 Treatment policy strategy

The intercurrent event of discontinuation of treatment, will be handled by the **treatment policy strategy**, i.e. all collected data also after the time of treatment discontinuation will be included in the analysis.

Thus, an estimand applying the treatment policy strategy is the absolute difference between placebo and HDM SLIT-tablet based on the endpoint, in the population defined by the trial inclusion and exclusion criteria, regardless of whether subjects complete treatment for the planned duration. In other words, the estimand is the effect of the treatment regimen in subjects with HDM allergic rhinitis/rhinoconjunctivitis regardless of whether the subject adhered to the treatment regimen.

A such estimand is denoted a **treatment policy estimand**, as it assesses the treatment effect regardless of adherence to treatment and provides a broad perspective of the treatment effect in clinical practice.

3.2 Study design

This trial is a randomised, parallel-group, double-blind, placebo-controlled, multinational phase III trial conducted in Europe and North America.

The trial was initiated on 10-OCT-2019 and three cohorts were included. Subjects received treatment for approximately 12 months.

It was planned that approximately 1370 subjects would be randomised (1:1) to receive treatment with the HDM SLIT-tablet or placebo.

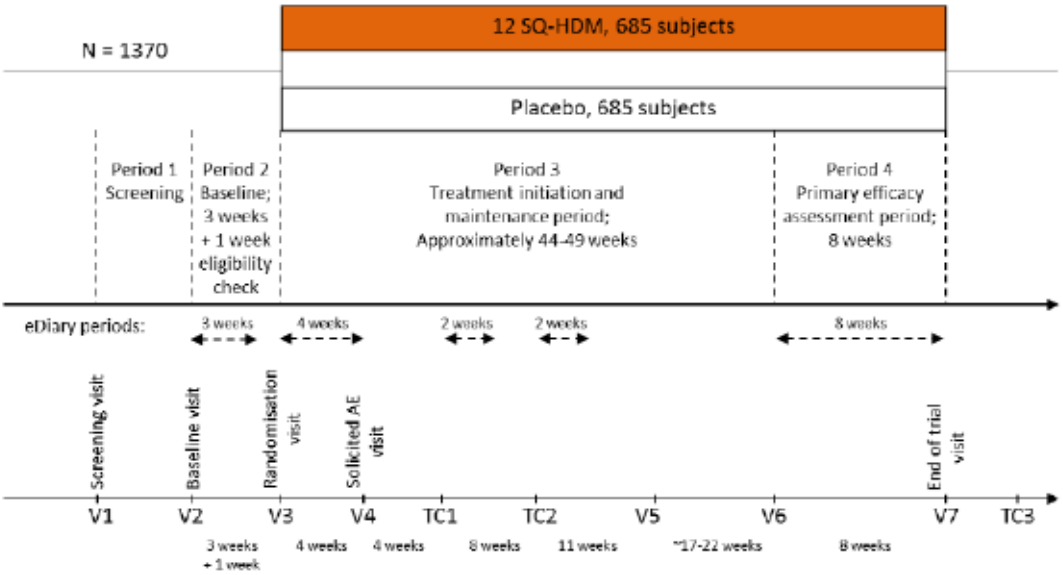


Figure 1 Study Design

The trial consisted of 4 periods: screening (Period 1), a baseline period (Period 2), a treatment initiation and maintenance period (Period 3), and a primary efficacy assessment period during the last 8 weeks (period 4) of approximately 12 months treatment.

Period 1 was the screening period. Subjects may have been screened up to 12 weeks before randomisation.

Period 2 was the baseline period and lasted 21 days plus 7 days for eligibility check. Subjects' baseline must have been recorded between 1 September and 1 April. In addition, pollen allergic subjects must have their baseline period outside the season of their pollen allergy. During the baseline period, subjects were allowed to take sponsor provided rhinoconjunctivitis and asthma rescue medication, and their rhinoconjunctivitis symptoms and use of rescue medication were scored on a daily basis in their eDiary. To be eligible for inclusion, subjects were required to have a rhinitis DSS of at least 6, or a score of at least 5 with one symptom being severe, on 8 of the last 14 days of the baseline period and were required to use rescue medication for the treatment of HDM allergic rhinitis during at least 8 of the last 14 days of the baseline period. The baseline period must have ended 1 week before randomisation to allow for eligibility check. Subjects with evidence of current, clinically significant, intercurrent illness (e.g., significant common cold or influenza) during the baseline period could be rescheduled for a repeat of the baseline on a resolution of their illness. The blood samples for safety testing were repeated if the results were more than 12 weeks old.

Period 3 was the treatment initiation and maintenance period. It began at V3 (randomisation visit) and lasted until V6 (44-49 weeks of treatment). During the first 28 days of period 3, 15 pre-specified symptoms/signs (based on recommendation by the WAO (Passalacqua et al. 2013) occurring after IMP intake were solicited and recorded by the parent/caregiver together with the subject in the eDiary. If pre-specified symptoms were reported in the eDiary, the investigator was to evaluate these and report them in the eCRF as solicited AEs at V4 (4 weeks of treatment). To evaluate efficacy, there were two 2 weeks efficacy assessment periods recorded in the eDiary at TC1 (after 8 weeks of treatment) and TC2 (after 16 weeks of treatment).

Period 4 was the 8-week primary efficacy assessment period. It began at V6 (44-49 weeks of treatment) and lasted until V7(end of trial). Subjects' primary efficacy assessment period had to be between 1 September and 1 April and the period was planned to include the dates that were included in the 3-week baseline period the previous year. In addition, pollen allergic subjects must have had their primary efficacy assessment period outside the season of their pollen allergy. During period 4, the parent/caregiver together with the subject rated the rhinoconjunctivitis symptoms and use of rescue medication on a daily basis in the eDiary.

The schedule of cohorts is presented in Table 1.

Table 1 Schedule of cohorts

	Cohort 1	**Cohort 2**	**Cohort 3**
First subject first visit (FSFV)	12-Oct-2019	07-Jul-2020	07-Jul-2021
Last subject randomised	20-Mar-2020	01-Apr-2021	01-Apr-2022
Last subject last visit	On-site (V7): 08-Jul-2021 Telephone (TC3): 22-Jul-2021	On-site (V7): 07-Apr-2022 Telephone (TC3): 25-Apr-2022	On-site (V7): 07-Apr-2023 Telephone (TC3): 21-Apr-2023

3.2.1 Flow chart

The flow chart of the study assessments by visits are shown below in

Table 2.

Table 2 Flow chart

Visit ID:	V1	V2	V3	V4	TC1	TC2	V5	V6	V7	TC 3	UV
Visit	Screening	Base-line	Randomi-sation	Solicited AEs					End of trial	Follow-up	Unschedul-ed visit
Time from randomisation (IMP initiation)	Max -12 weeks	-4 weeks -7 d		8 + 7 d	8 weeks ± 7 d	16 weeks ± 7 d	27 weeks ± 7 d	44-49 weeks ± 7 d	52-57 weeks ± 7 d[1]	+ 2 weeks from V7 + 7 d	
Informed consent	X										
Demography	X										
Medical history	X										

[1] V7 to be performed 8 weeks +14 days after V6 and no later than 1st April

Visit ID:	V1	V2	V3	V4	TC1	TC2	V5	V6	V7	TC 3	UV
Assess symptoms of eosinophilic oesophagitis	X	X	X	X	X	X	X	X	X	X	(X)
Record previous and concomitant medication	X	X	X	X	X	X	X	X	X	X	X
Physical examination	X								X		(X)
Oropharyngeal examination		X	X[2]	X			X	X			(X)
Height and weight[3]	X	X	X	X			X	X	X		X
Vital signs	X	X	X	X			X	X	X		(X)
Body temperature			X								
FEV_1			X						X		(X)
Urine pregnancy test, if applicable[4]	X	X	X	X			X	X	X		(X)
SPT	X	(X)[5]									
In-/exclusion criteria	X	X	X								
Blood and urine samples for safety laboratory assessments	X								X		(X)
Blood sample for specific IgE[6]	X[7]										(X)
Assess and record Aes in eCRF	X	X	X	X	X	X	X	X	X	X	X
Randomisation			X								
Issue and review local and systemic allergic reaction emergency plan			X								
PRQLQ			X				X	X	X		

[2] Oropharyngeal examinations was done before and 30 ±5 mins after IMP administration at Visit 3

[3] If applicable, adjust the local and systemic allergic reaction emergency plan

[4] For female subjects of childbearing potential

[5] SPT to be performed at visit 2, if not possible at visit 1, due to necessary wash-out of concomitant medication

[6] IgE against *D pteronyssinus and D farinae*

[7] Inform subjects that continued participation in trial depends on the result of the blood sample for specific IgE against *D. pteronyssinus* and *D. farinae*

Visit ID:	V1	V2	V3	V4	TC1	TC2	V5	V6	V7	TC 3	UV
[redacted]			X	X	X	X	X	X	X		(X)
[redacted]			X	X	X	X	X	X	X		(X)
Blood sample for pharmacogenetics biobank[8]									X		
Intake of IMP at clinic			X								(X)
Dispense IMP			X	X			X	X			(X)
Dispense rhinoconjunctivitis and asthma rescue medication and instruct in the use		X	X	X			X	X			(X)
Dispense adrenaline/ epinephrine auto-injectors[9]			X	X			X	X			(X)
Collect unused adrenaline/ epinephrine auto-injector									X		
Collect rescue medication as applicable and perform drug accountability			X	X			X	X	X		
Collect IMP, perform drug accountability and IMP compliance check				X			X	X	X		
Show and discuss trial video		X						X			
Issue and instruct parent/caregiver in use and activation of eDiary		X	X		X	X		X			X
Instruct in the recording of pre-specified symptoms in eDiary			X								

[8] For subjects where the subject/parent/guardian has given consent
[9] For countries where this is a regulatory requirement

Visit ID:	V1	V2	V3	V4	TC1	TC2	V5	V6	V7	TC 3	UV
Review eDiary and record solicited Aes in eCRF				X							
eDiary recording		3 weeks	4 weeks		2 weeks	2 weeks		8 weeks			
Check eDiary compliance			X	X					X		
Collect eDiary									X		

For Canada and Poland only:

Visit ID:	V1	V2	V3	V4	TC1	TC2	V5	V6	V7	TC 3	UV
Blood sample for immunological assessments	X								X		
Biobank blood sample[10]	X								X		

[10] For subjects where the subject/parent/guardian has given consent

3.2.2 Randomisation

The randomisation list was generated by a trial-independent statistician and was not accessible to trial personnel involved in the conduct of the trial until the database had been locked.

When a subject was randomised a unique number was assigned at the time of the first dispensing of IMP.

4 Statistical hypotheses

4.1.1 Multiplicity adjustment

The primary endpoint, key secondary endpoints, and the overall PRQLQ score at end of trial (using observed case analysis) will be controlled for multiplicity to ensure a maximum overall type I error rate of 5% in the hypothesis testing of these endpoints. The control for multiplicity is done by hierarchical testing, pre-specifying the order of the hypotheses to be tested. For all endpoints the null hypothesis to be tested is the hypothesis of no absolute difference in means between treatment groups. Let μ_1 denote the mean in the HDM SLIT-tablet group, and μ_2 denote the mean in the Placebo group. Then the null hypothesis (H_0) and the alternative hypothesis (H_A) are given as follows:

H_0: $\mu_1 = \mu_2$ and H_A: $\mu_1 \neq \mu_2$

The order of hypotheses to be tested is:

1. Superiority testing of the HDM SLIT-tablet over placebo with respect to the average daily TCRS during the primary efficacy assessment period.
2. Superiority testing of the HDM SLIT-tablet over placebo with respect to the average rhinitis DSS during the primary efficacy assessment period.
3. Superiority testing of the HDM SLIT-tablet over placebo with respect to the average rhinitis DMS during the primary efficacy assessment period.
4. Superiority testing of the HDM SLIT-tablet over placebo with respect to the average daily TCS during the primary efficacy assessment period.
5. Superiority testing of the HDM SLIT-tablet over placebo with respect to the overall PRQLQ score at end of trial.

All hypotheses in the hierarchy are tested on a 5% significance level. If the first hypothesis is statistically significant at the 5% level i.e., the null hypothesis of no difference is rejected, then the second hypothesis is tested, and so forth. A hypothesis is only tested if all previously tested hypothesis were statistically significant at the 5% level i.e., all previous null hypotheses were rejected.

5 Analysis sets

For the purposes of analysis, the following analysis sets are defined:

Analysis Set	Description
Total analysis set	The total analysis set comprises all subjects who signed informed consent and thus includes screening failures.
Full analysis set (FAS)	The FAS consists of all randomised subjects who received at least one dose of IMP.

	Subjects will be analysed as randomised i.e., according to their randomised assignment of treatment.
Safety analysis set (SAF)	The SAF consists of all randomised subjects, who received at least one dose of IMP. Subjects will be analysed as treated i.e., according to treatment they actually received.

The Total Analysis Set will be used for listing reasons for screening failure and for AEs before randomisation. The FAS will be used for baseline tables, efficacy and questionnaire endpoints, and the SAF will be used for safety reporting. FAS and all observed data for the endpoint is used for the observed case analyses (primary analysis for all endpoints).

The following data point sets (DPS) are defined:

DPS	Description
DPS1	For subjects who complete the study all data is included. For subjects who discontinue IMP, post-discontinuation data will not be included. Includes data until discontinuation of treatment.
DPS2	For subjects who complete the study all data is included. For subjects who discontinue IMP, all observed data will be included.

FAS and DPS1 are used to estimate the trial product estimand for the primary and key secondary objectives, whereas FAS and DPS2 are used to estimate the treatment policy estimand for the primary and key secondary objectives.

6 Statistical analyses

Statistical analyses will be carried out by ALK-Abelló A/S (hereafter ALK) Biometrics, Hørsholm, Denmark. All computation will be performed using the statistical analysis system SAS® version 9.4 or later.

The analyses described in this section along with the supporting information provided in section 9.2 specify the statistical analyses.

6.1 General considerations

In addition to the observed case analysis specified in the protocol, the primary endpoint and the key secondary endpoints are analysed using the estimand framework. For all other efficacy

endpoints, only the observed case analysis is performed. All efficacy endpoints are summarised by treatment and visit, and the primary and key secondary endpoints are listed.

All the statistical tests described in this section use a significance level of 5% and all tests and confidence intervals are two-sided. The null hypothesis is the hypothesis of no difference between treatment groups and the alternative to the null hypothesis is the hypothesis of difference. For each specific analysis the exact test is described in more detail.

Variation between countries/regions in the primary endpoint is expected and may result from variation in geography and thus HDM exposure, variation in standard treatment procedures, and possible differences in the conduct of the trial protocol. Variation between sites is also expected but is assumed to be small compared to the variation between countries and regions. The trial is not powered to detect differences or interactions in treatment effect within the country/region. Due to low number of subjects, France, Germany, and Spain will be pooled into the region Western Europe. The rest of the countries will not be pooled. Thus "country/region" will denote the effect of the individual countries as well as Western Europe.

6.1.1 Potential data issues

During study conduct, it was discovered that 11 subjects for various reasons had been randomised in error. Out of these, 10 subjects violated 1-2 inclusion criteria, and the last subject violated 1 exclusion criteria. As the inclusion criteria were systematically checked for all subjects in the trial, and because the violated inclusion criteria indicate that subjects potentially have a less severe HDM allergy, these subjects will be excluded from a sensitivitiy analysis.

However, as only one of these 10 subjects completed the trial, and the rest of the subjects discontinued prior to the start of the primary efficacy assessment period, only 1 of these subjects actually contributed to the observed case analyses. A list of the subject ID of the 10 subjects who violated at least one of the systematically checked inclusion criteria can be found in section 9.3.

The study includes several siblings. When siblings are included in the same cohort, there is a potential risk of mixing up their treatments at home, as well as mixing up/discussion of their diary recordings. Therefore siblings in the same cohort (56 in total) will be excluded from a sensitivity analysis. A list of the subject ID of the siblings can be found in section 9.3.

Several (8) subjects in Ukraine were temporarily without IMP due to the UKR crisis. These subjects will also be excluded from a sensitivity analysis. A list of the subject ID of the subjects in Ukraine without IMP can be found in section 9.3.

6.2 Primary endpoint analyses

6.2.1 Definition of endpoint

The primary endpoint is the average daily total combined rhinitis score (TCRS) of all observed daily values during the primary efficacy assessment period.

6.2.2 Primary analysis

The primary endpoint will be analysed using all subjects in FAS with at least 1 eDiary record during the primary efficacy assessment period. The presence of at least 1 eDiary record means that the primary endpoint can be calculated, so it is an observed case analysis. The analysis

will be performed using an LME. The model includes the square root of the average daily TCRS during the primary efficacy assessment period as response variable, treatment group and cohort as fixed factors, the square root of the baseline average daily TCRS as a covariate, country/region within cohort as a random effect, and with different residual errors specified for each treatment group. For further details see section 9.2.

Potential data issue: The primary observed case analysis will be repeated, but the subjects with potential data issues (randomised in error, siblings in same cohort, subjects in Ukraine without IMP as described in section 6.1.1) will be excluded from the analysis. The patients excluded from the analysis will be presented in a listing.

Sensitivity analysis: The primary analysis will be repeated using a modified DMS scoring to investigate the influence of subjects using more rescue medication than the recommended daily dose. The modified DMS scoring is defined in section 9.1.2.

Sensitivity analysis using bootstrapping: An additional sensitivitiy analysis will be performed using bootstrapping to calculate the 95% CI for the relative difference. For further details see section 9.2.

The primary analysis will also be repeated adding the cohort by treatment interaction, to investigate potential changes in treatment effect over time.

6.2.3 Applying the estimand framework

The study was not designed using the estimand framework, but since the first version of the protocol the ICH E9 addendum (ICH 2019, 1998) has come into affect. Therefore the primary analysis using an observed case analysis is supplemented with analyses of the trial product estimand and the treatment policy estimand using the estimand framework. However, as the study was not designed with these analyses in mind, data after IMP discontinuation was not collected for most subjects who discontinued the trial.

6.2.3.1 Trial product estimand

The primary endpoint will be compared between treatment groups using the **trial product estimand** for the FAS population and DPS1, following a hypothetical strategy.

6.2.3.1.1 Main analytical approach

Subjects for whom the primary endpoint is missing or unobserved (because of either missing diary data or the exclusion of diary data due to IMP discontinuation) will be included in the analysis through multiple imputation under the hypothetical situation where subjects continued to take study treatment as planned.

Multiple imputation for subjects missing the primary endpoint will be conducted defined as M1 in the following:

- For subjects who discontinue IMP due to lack of efficacy, multiple imputation of the missing endpoint will be from the placebo group. This assumes that had the subject continued to take study treatment, they would have experienced similar efficacy to subjects in the placebo group. Data is assumed to be MNAR.
- For subjects who do not discontinue IMP, and for subjects who discontinue IMP due to other reasons, multiple imputation of the missing endpoint will be from their own treatment group. This assumes that the level of efficacy that would have been

experienced by the subject if they had continued to take study treatment is unrelated to the reason for discontinuation and, if they had continued to take study treatment, they would have experienced similar efficacy to subjects in their own treatment group. Data is assumed to be MAR.

Analysis of the imputed datasets will be performed using the same model as the primary analysis using observed case analysis. The resulting treatment effect from the model is the main estimator. For more details see section 9.2.

6.2.3.1.2 Sensitivity analytical approach

Three sensitivity estimators will investigate/relax the assumptions about data missing at random.

Sensitivity estimator 1:

A sensitivity estimator for the trial product estimand, will relax the assumption about data missing at random for subjects discontinuing treatment due to IMP-related adverse events by imputing defined as M2 in the following:

- For subjects who discontinue IMP due to lack of efficacy or due to IMP-related adverse events, multiple imputation of the missing endpoint will be from the placebo group. This assumes that had the subject continued to take study treatment, they would have experienced similar efficacy to subjects in the placebo group. Data is assumed to be MNAR.
- For subjects who do not discontinue IMP, and for subjects who discontinue IMP due to other reasons, multiple imputation of the missing endpoint will be from their own treatment group. This assumes that the level of efficacy that would have been experienced by the subject if they had continued to take study treatment is unrelated to the reason for discontinuation and, if they had continued to take study treatment, they would have experienced similar efficacy to subjects in their own treatment group. Data is assumed to be MAR.

Sensitivity estimator 2:

Another sensitivity estimator will further relax the assumption about data missing at random by imputing defined as M3 in the following:

- For subjects who do not discontinue IMP, and for subjects who discontinue IMP due to lost to follow-up or withdrawal of consent, multiple imputation of the missing endpoint will be from their own treatment group. This assumes that the level of efficacy that would have been experienced by the subject if they had continued to take study treatment is unrelated to the reason for discontinuation and, if they had continued to take study treatment, they would have experienced similar efficacy to subjects in their own treatment group. Data is assumed to be MAR.
- For subjects who discontinue IMP due to other reasons, multiple imputation of the missing endpoint will be from the placebo group. This assumes that had the subject continued to take study treatment, they would have experienced similar efficacy to subjects in the placebo group. Data is assumed to be MNAR.

Sensitivity estimator 3:

In addition, in order to investigate the MAR assumption for the imputation in sensitivity estimator 2 above, a tipping point analysis will be performed as follows: A penalty (a number) is added to all imputed values in the active treatment group, and the analysis is repeated. The penalty is gradually increased until the point, the tipping point, where the null hypothesis is no longer rejected. If the tipping point is considered a clinical plausible difference, the tipping-point analysis does not support the sensitivity estimator.

6.2.3.2 Treatment policy estimand

In addition to the trial product estimand, the primary endpoint will be compared between treatment groups using the **treatment policy estimand** for the FAS population and DPS2, following a treatment policy strategy.

6.2.3.2.1 Main analytical approach

Subjects for whom the primary endpoint is missing (because of missing diary data) will be included in the analysis through multiple imputation as follows (M2).

- For subjects discontinuing IMP due to lack of efficacy or IMP-related adverse events, multiple imputation of the missing endpoint will be from the placebo group. Data is assumed to be MNAR.
- For subjects who do not discontinue IMP, and for subjects discontinuing IMP due to other reasons, multiple imputation of the missing endpoint will be from their own treatment group. This assumes that the missing data are MAR.

Analysis of the imputed datasets will be performed using the same model as the primary analysis using observed case analysis. The resulting treatment effect from the model is the main estimator. For more details see section 9.2.

Note that the above analysis is almost identical to the first sensitivity estimator for the trial product estimand, as data is collected after discontinuation of IMP for very few subjects only.

6.2.3.2.2 Sensitivity analytical approach

Sensitivity estimator 1: A sensitivity estimator will impute all missing values of the endpoint from the placebo group (M4).

6.3 Key secondary endpoint analyses

6.3.1 Definition of endpoints

The key secondary efficacy endpoints are the average of all observed daily values during the primary efficacy assessment period for rhinitis DSS, rhinitis DMS, and TCS.

6.3.2 Primary analysis

The key secondary endpoints “average rhinitis DSS”, “average rhinitis DMS”, and “average daily TCS”, will be analysed using an observed case analysis similar to the primary analysis of the primary endpoint. All subjects in the FAS with at least 1 eDiary record during the primary efficacy assessment period will be included. The analysis will be performed using an LME. The model includes the square root of the endpoint as response variable, treatment group and

cohort as fixed factors, the square root of the baseline endpoint as a covariate, country/region as a random effect, and with different residual errors specified for each treatment group.

Sensitivity analysis: For "average rhinitis DMS" and "average daily TCS" the primary analysis will be repeated using a modified DMS scoring to investigate the influence of subjects using more rescue medication than the recommended daily dose. The modified DMS scoring is defined in section 9.1.2.

6.3.3 Applying the estimand framework

In analogy with the supplemental analyses for the primary endpoint, the analysis of the key secondary endpoints will also be supplemented with analyses of the trial product estimand and the treatment policy estimand. The same main analytical approach as for the primary endpoint will be used for both estimands . No sensitivity estimators will be performed.

6.4 Summary of primary and key secondary endpoint analyses

Table 3 Overview of analyses for the primary and key secondary endpoints

Endpoint	Descriptions	Population	DPS	Model	Imputation Method	Analysis name
Average daily TCRS during the primary efficacy assessment period	Observed case	FAS	All	LME [b]	None	**Primary** [a]
						Sensitivity
						Potential data issue
	Trial product estimand		DPS1	LME with MI [b]	M1	Main analytical approach
					M2	Sensitivity 1
					M3	Sensitivity 2
					M3 tipping point	Tipping point
	Treatment policy estimand		DPS2	LME with MI [b]	M2	Main analytical approach
					M4	Sensitivity 1
Average rhinitis DSS during the primary efficacy assessment period	Observed case	FAS	All	LME [b]	None	**Primary** [a]
	Trial product estimand		DPS1	LME with MI [b]	M1	Main analytical approach
	Treatment policy estimand		DPS2	LME with MI [b]	M2	Main analytical approach

Average rhinitis DMS during the primary efficacy assessment period	Observed case	FAS	All	LME [b]	None	**Primary** [a]
						Sensitivity
	Trial product estimand		DPS1	LME with MI [b]	M1	Main analytical approach
	Treatment policy estimand		DPS2	LME with MI [b]	M2	Main analytical approach
Average daily TCS during the primary efficacy assessment period	Observed case	FAS	All	LME [b]	None	**Primary** [a]
						Sensitivity
	Trial product estimand		DPS1	LME with MI [b]	M1	Main analytical approach
	Treatment policy estimand		DPS2	LME with MI [b]	M2	Main analytical approach

[a] **Primary** refers to the analyses included in the test hierarchy

[b] Endpoints are square-root transformed before applying the model

6.5 Additional secondary efficacy endpoint analyses

The additional secondary efficacy endpoints, the average of all observed daily values during the primary efficacy assessment period for **rhinoconjunctivitis DSS**, for **rhinoconjunctivitis DMS**, for **rhinitis CSMS**, and for **rhinoconjunctivitis CSMS**, will be analysed using the same observed case analysis as specified for the primary endpoint. All subjects in the FAS with at least one eDiary record during the primary efficacy assessment period will be included. The analysis will be performed using an LME. The model includes the square root of the endpoint as response variable, treatment group and cohort as fixed factors, the square root of the baseline endpoint as a covariate, country/region as a random effect, and with different residual errors specified for each treatment group.

The **overall PRQLQ score at end of trial** will be analysed similar to the primary endpoint, but without the square root transformation, i.e. the LME will include the overall PRQLQ score as response variable, treatment group and cohort as fixed factors, the baseline overall PRQLQ score as a covariate, country/region as a random effect, and with different residual errors specified for each treatment group.

The overall **PRQLQ** score at V5 and V6 will be analysed separetely, similarly to the overall PRQLQ score at end of trial, and the model adjusted means for each treatment group will be plotted as a function of visit.

The endpoint **rhinitis mild days** during the primary efficacy assessment period consists of the daily binary responses “rhinitis mild day yes/no”, and will be analysed using a parametric GLMM with a logit link function. The model includes “rhinitis mild day yes/no” as the response variable, treatment group and cohort as fixed factors, the baseline average rhinitis TCRS as a covariate, and country/region within cohort, and subject as random effects. From the GLMM model, the

odds ratio for having a rhinitis mild day for active treatment relative to placebo will be presented together with the coherent p-values and 95% confidence limits.

Rhinitis exacerbation days will be analysed using the same model as for the analysis of rhinitis mild days, apart from that baseline average rhinitis DSS is included instead of baseline average rhinitis TCRS.

In the subgroup of subjects with asthma at baseline the following endpoints will be analysed:

- The average **asthma DSS** of all observed daily values during the primary efficacy assessment period, will be analysed using the same observed case analysis as specified for the key secondary endpoints.
- **SABA free days** will be analysed similar to "rhinitis mild days", but with baseline average asthma DSS included instead of baseline average rhinitis TCRS.
- The average of all observed values during the primary efficacy assessment period for **weekly number of puffs of as-needed SABA use**, will be analysed using the same observed case analysis as specified for the primary endpoint, however with the difference that both the endpoint and the baseline value will not be square root transformed.

6.5.1 Immunology

For Canada and Poland only, blood samples for immunological assessments were taken at screening and at end of trial. Data for the two IgE and IgG4 species, and total IgE, will be log10-transformed, whereas IgE-blocking factor will not be transformed. The assessments will be summarised by visit and treatment including change from baseline.

[redacted]

[redacted]

6.7 Safety analyses

All safety summary tables and listings will be based on the SAF.

6.7.1 Extent of exposure

6.7.1.1 Treatment duration

The duration of treatment for each subject is calculated from the date of first dose up until (and including) the date of last IMP administration. For subjects who complete the trial this information is not available, instead the date of the last visit will be used.

Treatment duration will be summarised by treatment group.

6.7.1.2 Exposure

Exposure is a priori calculated as the difference between the number of tablets dispensed and the number of tablets returned or lost. But for a substantial number of subjects the information about tablets lost is missing, and is therefore assumed to be 0. To account for this, exposure will be calculated as the minimum of the difference between the number of tablets dispensed and the number of tablets returned or lost, and the treatment duration.

The number of tablets taken will be summarised by treatment group.

6.7.1.3 IMP compliance

Percentage compliance will be calculated for each subject as treatment taken divided by the treatment duration. The average percentage compliance will be presented by treatment group.

6.7.1.4 eDiary compliance for TCRS

eDiary compliance for TCRS is calculated for each subject as the number of daily records of TCRS in the period, divided by the expected number of days in the period. Note that daily TCRS is calculated based on rhinitis symptoms and rhinitis medication score, so a daily record of both is needed in order to have a daily record of TCRS.

The expected number of days in the period is 14 for baseline, TC1, and TC2, and 56 for the primary efficacy assessment period. The average percentage eDiary compliance will be presented by visit and treatment group.

6.7.2 Adverse events

6.7.2.1 Treatment emergent adverse events

An AE is considered a TEAE if the AE starts on or after the first IMP administration and no later than 7 days after the last IMP administration.

6.7.2.2 Solicited AEs

During the first 28 days after randomisation (V3), the subjects (together with a parent/caregiver) record in the eDiary whether they experience any of the 15 pre-specified symptoms/signs after IMP intake. These symptoms/signs (solicited term) and their corresponding PTs are listed in Table 4.

At Visit 4 the symptoms/signs reported in the eDiary are discussed with the subject. If they are assessed to be AEs, they are either reported as solicited AEs (by one of the solicited terms) or as regular AEs. Thus, there is not a one-to-one correspondence between the reporting of a symptom/sign in the eDiary and a solicited AE.

For solicited AEs, the PTs are assigned as shown in Table 4.

Table 4 Correspondance between solicited symptom/sign and PT

Solicited symptom/sign	PT
Food tastes different	Dysgeusia
Mouth ulcer	Mouth ulceration
Swelling in the back of the mouth	Mouth swelling
Itching in the mouth	Oral pruritus
Itching in the ear	Ear pruritus
Swelling of the lips	Lip swelling
Swelling of the tongue	Swollen tongue
Tongue pain	Glossodynia
Tongue ulcer	Tongue ulceration
Throat irritation/tickle	Throat irritation
Throat swelling	Pharyngeal swelling
Stomach pain	Abdominal pain upper
Nausea (feel like throwing up)	Nausea

Vomiting	Vomiting
Diarrhoea	Diarrhoea

6.7.2.3 Events of special interest

Events of special interest for this trial are:

- Anaphylactic reactions, anaphylaxis and/or systemic allergic reactions
- Events treated with adrenaline/epinephrine
- Severe local swelling of the mouth and/or throat
- Eosinophilic oesophagitis

6.7.2.4 Reporting of AEs

All TEAEs will be summarised by treatment group according to causality, severity (mild, moderate, severe), seriousness, action taken, outcome, and whether the event led to IMP discontinuation. The number of subjects in treatment group, the frequency and proportion of subjects having the event, as well as the number of events and proportion of events will be displayed. A similar summary table will be produced for all IMP-related TEAEs.

All TEAEs and all IMP-related TEAEs will be summarised (in separate tables) by treatment group, system organ class, and preferred term, displaying number of subjects in treatment group, and number and frequency of subjects having the event. The tables will be sorted according to most frequent SOC and PT in the active treatment group, and the most common TEAEs (preferred term from MedDRA version 23.0) defined as those that are present in at least 2% of subjects in the active treatment group will be marked.

In addition IMP-related solicited TEAEs will be summarised by treatment group, and presented by PT.

Severe TEAEs will be listed only, if there are fewer than 10 events in total. Otherwise severe TEAEs will also be summarised and presentented by SOC and PT similar to how TEAEs are presented. The same goes for SAEs, TEAEs leading to IMP discontinuation, and IMP-related TEAEs leading to IMP disruption.

For most frequent IMP-related TEAEs (present in at least 2% of subject in the active treatment group) onset and duration will be presented as follows: Average time to onset and average duration both in days, will be presented by PT and treatment. In addition, for day 1 of treatment average time to onset in minutes will be presented by PT and treatment.

For the most frequent IMP-related TEAEs which are recurrent, the average daily duration and the average duration in days from first to last occurrence will be presented by PT and treatment.

A table with the overall summary of the safety profile will be produced showing subjects with TEAEs, IMP-related TEAEs, IMP-related severe TEAEs, IMP-related treatment-emergent SAEs, IMP-related TEAEs leading to IMP discontinuation, IMP-related solicited TEAEs, and each type of IMP-related treatment-emergent ESI by treatment group.

Serious TEAEs, severe TEAEs, TEAEs leading to IMP discontinuation, IMP-related TEAEs leading to IMP interruption, Individual ESIs, TEAEs, and all non-treatment-emergent AEs for the total analysis set will be listed. Medication errors will also be listed.

6.7.3 Additional safety assessments

Shift tables will be produced for physical examination and oropharyngeal examination including the categories normal, abnormal not clinically significant, abnormal clinically significant.

Shift tables will be produced for laboratory assessments including the categories low, normal, high, and missing as relevant, with the exception is urinalysis, where the shift table will include the categories normal, abnormal not clinically significant, and abnormal clinically significant.

Vital signs and the pulmonary function test parameters (FVC, FVC percent predicted, FEV_1, and FEV_1 percent predicted) will be summarised by visit and treatment group.

6.8 Subgroup analyses

No formal statistical subgroup analyses were planned in the protocol. But as the trial includes subjects with HDM allergic rhinitis/rhinoconjunctivitis with or without asthma, the subgroup of subjects with asthma at baseline will be examined. The observed case analysis for the primary and key secondary endpoints will be repeated adding asthma subgroup and asthma subgroup by treatment interaction.

6.9 Interim analysis

No interim analyses will be performed.

6.10 Changes and/or deviations to protocol-planned analyses

The sensitivitiy analyses of the primary endpoint are not performed as specified in the protocol, and neither is the specified MMRM analysis. Instead the planned observed case analyses are supplemented with analyses using the estimand framework and several corresponding sensitivity analyses. A few other sensitivity analyses are also added for the primary endpoint. For the primary endpoint and the key secondary endpoints using medication score, a sensitivity analysis is added to investigate the influence of subjects using more medication than the recommended daily dose.

The "overall PRQLQ score at end of trial" was included as the 5th and last test in the testing hierarcy.

Additional secondary and explorative endpoints are added:

- "[redacted]"
- "[redacted]"

Some endpoints are redefined to make them more clear:

- The endpoint "average daily use of SABA" is changed into two endpoints: "SABA free days" and "average weekly number of puffs of as-needed SABA use"
- The "endpoint CSMS" is specified as "average daily rhinitis CSMS" and "average daily rhinoconjunctivitis CSMS"

The distribution of IMP discontinuations will be visualised by a band plot, instead of by a Kaplan-meier plot and a cumulative incidence plot, as the band plot also shows the different reasons for discontinuation. The band plot is also used to visualize time to discontinuation due to TEAEs.

Subgroup analyses of the primary and the key secondary endpoints according to asthma status have been added.

7 Sample size determination

The primary efficacy endpoint is the "average daily TCRS during the 8-week primary efficacy assessment period".

7.1 Assumptions

The primary efficacy analysis is based on treatment comparison performed using a two-sided t-test on a 5% level of significance in a linear mixed effect (LME) model assuming unequal variances. The TCRS is square root transformed in the LME model in order to obtain a better approximation to a normal distribution. Least squares means are then back-transformed to the original scale by taking the square. In addition, the relative difference of the back-transformed least squares means will be calculated together with 95% confidence limits. The latter will be calculated based on Fieller's theorem.

Superiority of active treatment compared to placebo is confirmed when the p-value for the primary efficacy analysis is <0.05 in the t-test. An additional FDA acceptance criteria for the trial is a point estimate of the relative difference to placebo of no more than -15% with an associated upper bound of the 95% CI being at most -10%.

Data from previous adult phase III trials of the HDM tablet conducted in Europe (MT-06) and North America (P001) and data from a previous paediatric trial (5-17 years of age) of the HDM tablet conducted in Japan (TO-203-3-3) is used for power calculation. Table 5 shows the mean and standard deviation of the square root transformed TCRS by treatment group for the MT-06 trial. The power calculation is based on the mean for placebo and the standard deviation for the placebo and the 12 SQ-HDM treatment group observed in the MT-06 trial due to the similar nature of the trial design for the MT-12 and MT-06 trial. Table 6 shows the estimated effect sizes in MT-06, P001 and TO-203-3-3 for different age groups. Power calculation is performed based on the assumption that the effect size is similar to the effect size seen in the paediatric trial TO-203-3-3 for the age group 5-11 years (~21%). In addition, power calculation is based on the more conservative assumption that the effect size is similar to the effect sizes observed in the previous adult trials MT-06 and P001 conducted in Europe and North America (~17-18%).

Table 5 Mean and standard deviation of TCRS by treatment group in MT-06

Square root transformed TCRS	MT-06	
	Placebo	12 SQ-HDM
Raw mean	2.65	2.40
Standard deviation	0.89	0.95

Table 6 Effect size of TCRS in MT-06, P001, and TO-203-3-3

Trial	MT-06	P001		TO-203-3-3		
Region	Europe	North America		Japan		
Age group	18-65	18-65	12-17[A]	5-17	5-11[A]	12-17[A]
N	582	1186	160	427	247	180
Effect size	-18.1%[B]	-17.2%[C]	-22.4%[C]	-22.5%[B]	-20.8%[B]	-25.9%[B]

[A] Subgroup analysis; [B] Relative difference (Active-Placebo)·100%/Placebo in adjusted means; [C] Relative difference (Active-Placebo)·100%/Placebo in medians

7.2 Results of sample size calculation

Table 7 shows the results for the power calculation. Power calculation for fulfilling the acceptance criteria of a p-value <0.05 is based on Satterthwaite unpooled two-sided t-test on a 5% level of significance. Power calculation for fulfilling the acceptance criteria of an upper bound of the confidence limit being at most -10% (95% UCL ≤ -10%) is conducted by simulations using Fieller's theorem (Fieller 1954).

Table 7 Power calculations based on different assumptions of the effect size using observed case approach

Effect size	-18%		-21%	
Acceptance criterion	p-value <0.05	95%UCL ≤-10	p-value <0.05	95%UCL ≤-10
Sample size per arm	Power			
300 subjects	91%	35%	97%	58%
500 subjects	99%	51%	100%	80%
580 subjects	100%	58%	100%	85%

95% UCL: upper bound of 95% confidence interval.

Based on Table 7 a sample size of 580 per treatment arm is chosen. With an effective sample size of 580 subjects per treatment arm, the MT-12 trial have:

- 85% power (2-sided, α=0.05) to obtain an upper bound of the 95% CI for a relative difference of no more than -10% if assuming an effect size of -21%
- >90% power (2-sided, α=0.05) to detect a statistically significant difference (p-value<0.05) if assuming an effect size of -18%.

When further adjusting for a drop-out that may be approximately 15% based on experience from MT-06 and P001, the proposed sample size per treatment arm for the MT-12 trial is 682 subjects (Ntotal=1364 ~ 1370 subjects).

8 Demographics and baseline characteristics

Summaries for numerical variables will display the descriptive statistics mean, SD, median, minimum and maximum.

Summaries for categorical variables will be frequency tables displaying numbers and percentages.

8.1 Screening failures

Reasons for screening failure will be summarised for the total analysis set. Reasons for screening failure will also be listed.

8.2 Protocol deviations

Important PDs and important subject-level PDs will be summarised. In addition important trial-level, country-level, site-level, and subject-level protocol deviations will be listed.

8.3 Subject disposition

Subject disposition will be summarised by number and percentage of subjects screened, randomised, included in the analysis sets, discontinued and the primary reason for discontinuation by treatment group will be presented.

IMP discontinuations over time by different reasons will be visualised by a band plot showing proportions on IMP, having discontinued IMP due to AE, withdrawal by subject, and all other reasons (pooled). IMP and trial discontinuations will be listed.

8.4 Baseline characteristics

Demographic variables (age, sex, race, ethnic origin, and country/region) will be summarised by treatment group. Demographic data will also be listed.

Baseline characteristics (mono/poly sensitised, duration of HDM AR/C, skin prick test results, skin pric test wheal sizes (for HDM), specific HDM IgE (both continuous and categorised), asthma status and severity, and use of rescue medication for rhinitis at baseline) will be summarised by treatment group.

8.5 Medical history

Medical history (including history of rhinitis, conjunctivitis, asthma, atopic dermatitis and food allergy) will be summarised by treatment group. Separate tables will be created for allergy, asthma, and atopic dermatitis medical history, and other medical history.

8.6 Prior and concomitant therapy

Prior and concomitant medication will be summarised separately by treatment group.

9 Supporting documentation

9.1 Derivations for the study

9.1.1 Derivation of daily symptom and medication scores

Symptoms and medication use are recorded daily by the subject in the eDiary. For each day it is only possible for the subject to either record all or nothing at all.

9.1.1.1 Rhinitis DSS

The rhinitis DSS consists of the 4 rhinitis symptoms: runny nose, blocked nose, sneezing, and itchy nose. Each of the symptoms is scored from 0 to 3 by the subject daily. No symptoms=0, mild symptoms=1, moderate symptoms=2, and severe symptoms=3. The rhinitis DSS is the sum of the 4 individual rhinitis symptom scores and ranges from 0 to 12.

[redacted]

9.1.1.3 Rhinoconjunctivitis DSS

The rhinoconjunctivitis DSS is a sum of rhinitis DSS and conjunctivitis DSS, and ranges from 0 to 18.

9.1.1.4 Rhinitis DMS

The rhinitis DMS is the sum of the total daily scores for each type of rescue medication (Table 8).

Note that the recommended dose of Desloratadine is scored 4, and that doses below are scored proportionally, i.e. if a 12 years old takes 5 ml once daily, this is scored as 2. Similarly lower than recommended use of Loratadine tablets and Mometasone furoate nasal spray is also scored proportionally based on the recommended daily dose. Some subjects took Loratadine tablets, although they should have taken Desloratadine oral solution. Their dosing is scored the same as if they should have taken tablets, i.e. a score of 4 for 1 tablet, and a score of 2 for half a tablet. The rhinitis DMS ranges from 0 to 12.

Table 8 Rhinitis DMS scoring

Rescue medication	Subject dosing	Score/Dose unit	Maximum daily score
Desloratadine oral solution, 0.5 mg/ml	5 years old: 2.5 ml (1.25 mg) once daily	4	4
	6-11 years old: 5 ml (2.5 mg) once daily		

	≥12 years old: 10 ml (5 mg) once daily		
Loratadine tablets, 10 mg	6-12 years old and > 30 kg: 1 tablet (10 mg) once daily		
	>12 years old: 1 tablet (10 mg) once daily		
Mometasone furoate nasal spray, 50 micrograms/dose	<12 years: 1 puffs in each nostril once daily	4	8
	≥12 years: 2 puffs in each nostril once daily	2	
Max. daily rhinitis medication score			12

For Desloratadine and Loratadine, the dose unit corresponds to the recommended dose. For Mometasone, the dose unit is 1 puff in 1 nostril.

[redacted]

[redacted]

[redacted]

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

[redacted]

[redacted]

9.1.1.6 Rhinoconjunctivitis DMS

The rhinoconjunctivitis DMS is the sum of the rhinitis DMS and [redacted] and ranges from 0 to 20.

9.1.1.7 TCRS

The TCRS is the sum of the rhinitis DSS and the rhinitis DMS, and ranges from 0 to 24.

[redacted]

9.1.1.9 TCS

The TCS is the sum of the rhinoconjunctivitis DSS and the rhinoconjunctivitis DMS, and ranges from 0 to 38.

9.1.1.10 Asthma DSS

The 4 symptoms of asthma: cough, wheezing, chest tightness, and shortness of breath, are scored from 0 to 3 by the subject daily. No symptoms =0, mild symptoms =1, moderate symptoms =2, and severe symptoms =3. The asthma DSS is the sum of the 4 individual asthma symptom scores, and ranges from 0 to 12.

9.1.1.11 Rhinitis CSMS

The rhinitis CSMS recommended by the EAACI consists of symptom and medication scoring.

The 4 EAACI rhinitis symptoms: runny nose, blocked nose, sneezing, and itchy nose are scored. Each symptom is scored from 0 to 3 by the subject daily. No symptoms =0, mild symptoms =1, moderate symptoms=2, and severe symptoms 3. The rhinitis CSMS symptom score is constructed as the sum of the individual symptoms score divided by the number of symptoms (4) and thus range from 0 to 3.

The rhinitis CSMS medication score is a stepwise scoring method, which was not implemented in this trial. However, the rhinitis CSMS medication scoring is still implemented where the highest medication step score is used as the medication score. The summary of the steps is presented in Table 10.

Table 10 Rhinitis Medication Scoring for CSMS by EAACI task force

Step	Medication	Rhinitis Medication Score

1	Desloratadine oral solution or Loratadine tablets	1
2	Mometasone furoate nasal spray	2

If the medication is not used, then the medication score is 0. The range of medication score is from 0 to 2. Note that in Pfaar et al (Pfaar et al. 2014) a step 3 of oral corticosteroids was also included, but in this trial oral corticosteroids were not provided as rescue medication, hence the step 3 is not possible.

The rhinitis CSMS by EAACI is the sum of rhinitis symptoms and the rhinitis medication scores, and ranges from 0 to 5.

9.1.1.12 Rhinoconjunctivitis CSMS

The rhinoconjunctivitis CSMS recommended by the EAACI consists of symptom and medication scoring.

The 6 EAACI rhinoconjunctivitis symptoms are scored: runny nose, blocked nose, sneezing, itchy nose, gritty feeling/red/itchy eyes, and watery eyes. Each symptom is scored from 0 to 3 by the subject daily. No symptoms=0, mild symptoms=1, moderate symptoms=2, and severe symptoms=3. The rhinoconjunctivitis symptoms score is constructed as the sum of the individual symptoms score divided by the number of symptoms (6) and thus range from 0 to 3.

The rhinoconjunctivitis CSMS medication score is a stepwise scoring method, which was not implemented in this trial. However, the rhinitis CSMS medication scoring is still implemented where the highest medication step score is used as a medication score. The summary of the steps is presented in Table 11.

Table 11 Rhinoconjunctivitis medication scoring for CSMS by EAACI task force

Step	Medication	Rhinitis Medication Score
1	Desloratadine oral solution or Loratadine tablets or/and Olopatadine eye drops	1
2	Mometasone furoate nasal spray	2

If the medication is not used, then the medication score is 0. The range of medication score is from 0 to 2. Note that in Pfaar et al (Pfaar et al. 2014) a step 3 of oral corticosteroids was also included, but in this trial oral corticosteroids were not provided as rescue medication, hence the step 3 is not possible.

The rhinoconjunctivitis CSMS by EAACI is the sum of rhinoconjunctivitis symptoms and the rhinoconjunctivitis medication scores, and ranges from 0 to 5.

9.1.2 Medication scores used for sensitivity analyses

DMS_sens scores will be used for sensitivity analyses for the primary endpoint average daily TCRS, and the key secondary endpoints average rhinitis DMS and average daily TCS, all during the primary efficacy assessment period.

9.1.2.1 Rhinitis DMS_sens

The rhinitis DMS_sens is the sum of the total daily scores for each type of rescue medication (Table 12).

Note that the recommended dose of Desloratadine is scored 4, and that doses above or below are scored proportionally, i.e. if a 12 years old takes 5 ml once daily, this is scored as 2. And if the same subject takes 20 ml once daily, this is scored as 8.

Similarly doses of Loratadine tablets and Mometasone furoate nasal spray are also scored proportionally based on the recommended daily dose. Some subjects took Loratadine tablets, although they should have taken Desloratadine oral solution. Their dosing is scored the same as if they should have taken tablets, i.e. a score of 4 for 1 tablet, and a score of 2 for half a tablet. If a subject took 2 tablets (exceeded the recommended dose) this is scored as 8.

The rhinitis DMS_sens ranges from 0 to 12, unless the recommend dose is exceeded, in which case the max of 12 is also exceeded.

Table 12 Rhinitis DMS_sens scoring

<table>
<tr><th>Rescue medication</th><th>Subject dosing</th><th>Score/Dose unit</th><th>Score for recommended dose</th></tr>
<tr><td rowspan="3">Desloratadine oral solution, 0.5 mg/ml</td><td>5 years old:
2.5 ml (1.25 mg) once daily</td><td rowspan="5">4</td><td rowspan="5">4</td></tr>
<tr><td>6-11 years old:
5 ml (2.5 mg) once daily</td></tr>
<tr><td>≥12 years old:
10 ml (5 mg) once daily</td></tr>
<tr><td rowspan="2">Loratadine tablets, 10 mg</td><td>6-12 years old and > 30 kg:
1 tablet (10 mg) once daily</td></tr>
<tr><td>>12 years old: 1 tablet (10 mg) once daily</td></tr>
<tr><td rowspan="2">Mometasone furoate nasal spray, 50 micrograms/dose</td><td><12 years: 1 puff in each nostril once daily</td><td>4</td><td rowspan="2">8</td></tr>
<tr><td>≥12 years: 2 puffs in each nostril once daily</td><td>2</td></tr>
<tr><td>Max. daily rhinitis medication score for recommended dose</td><td colspan="3">12</td></tr>
</table>

For Desloratadine and Loratadine, the dose unit corresponds to the recommended dose. For Mometasone, the dose unit is 1 puff in 1 nostril.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

9.1.2.3 TCRS_sens

The TCRS_sens is the sum of the rhinitis DSS and the rhinitis DMS_sens, and ranges from 0 to 24, unless the recommended daily dose is exceeded.

9.1.2.4 TCS_sens

The TCS_sens is the sum of the rhinoconjunctivitis DSS and the rhinoconjunctivitis DMS_sens (sum of rhinitis DMS_sens and [redacted]), and ranges from 0 to 38, unless the recommended daily dose is exceeded.

9.1.3 Derivation of average symptom and medication scores

For each observation period the average score for that period is calculated as the mean of all observed (i.e., non-missing) daily values in that period.

The observation periods considered are:

- Baseline period: last 2 weeks of the 3 weeks recorded after V2 (baseline visit)
- Maintenance period 1: starts at TC1 (8 weeks after IMP initiation) and lasts for 2 weeks
- Maintenance period 2: starts at TC2 (16 weeks after IMP initation) and lasts for 2 weeks
- Primary efficacy period: starts at V6 (44-49 weeks after IMPinitation) and lasts for 8 weeks

The averages are calculated for the following scores:

- TCRS
- Rhinitis DSS
- Rhinitis DMS
- TCS
- Rhinoconjuctivitis DSS
- Rhinoconjuctivitis DMS
- Asthma DSS
- Rhinitis CSMS
- Rhinoconjunctivitis CSMS
- TCSS
- [redacted]
- [redacted]
- TCRS_sens
- Rhinitis DMS_sens
- TCS_sens

9.1.4 Derivation of daily binary variable

9.1.4.1 Rhinitis mild day

Rhinitis mild day is a day with no symptoms or mild symptoms (i.e. rhinitis DSS= 0 or rhinitis DSS=1) and no use of symptoms medication (i.e. rhinitis DMS = 0). Rhinitis mild day is a binary endpoint where 1 is a rhinitis mild day, and 0 otherwise.

9.1.4.2 Rhinitis exacerbation day

A rhinitis exacerbation days is a day with a rhinitis DSS of at least 6 or a day with a rhinitis DSS of 5 and 1 individual symptom scored 3. Rhinitis exacerbation day is a binary endpoint where 1 is a day with rhinitis exacerbation, and 0 otherwise.

9.1.4.3 SABA free day

A SABA free days is a day where the subject did not use SABA. A SABA free day is a binary endpoint where 1 is a day where the subject did not use SABA, and 0 otherwise.

[redacted]

[redacted]

9.1.5 Other derivations of endpoints

9.1.5.1 Average weekly number of puffs of as-needed SABA use

For each observation period the average daily number of puffs of as-needed SABA use for that period is calculated as the mean of all observed (i.e., non-missing) daily values in that period.

The observation periods considered are:

- Baseline period: starts at V2 (baseline visit) and lasts for 2 weeks
- Primary efficacy period: starts at V6 (44-49 weeks of treatment) and lasts for 8 weeks

Then the average weekly number of puffs of as-needed SABA use is calculated as 7 times the average of the daily number of puffs of as-needed SABA use.

9.1.5.2 PRQLQ

The PRQLQ has 23 questions in 5 domains (nasal symptoms, eye symptoms, practical problems, activity limitation, and other symptoms (Juniper et al. 1998).

The PRQLQ is completed by an interviewer. Subjects are asked to think about how they have been during the previous week and to respond to each of the 23 questions on a 7-point scale (from 0 = not bothered/none of the time to 6 = extremely bothered/all of the time). The overall PRQLQ score is the mean of all 23 responses and the individual domain scores are the means of the items in those domains.

[redacted]

[redacted]

9.1.6 Imputation of dates

9.1.6.1 Partial dates in AE reporting

Partial dates and time for **start** of AEs are imputed as follows:

- Missing day is imputed as maximum of first day of month and day of first IMP
- Missing month is imputed as maximum of first day in January and date of first IMP

Partial dates and time for **end** of AEs are imputed as follows:

- Missing day is imputed as minimum of last day in month and day of last contact
- Missing month is imputed as minimum last day of December and date of last contact

With these imputation rules AEs are considered treatment-emergent when dates are partial and the existing information does not exclude the possibility of the event being treatment-emergent.

9.1.6.2 Incomplete date for last IMP

Any partial date of last dose is imputed as follows:

- Missing day is imputed as minimum of last day in month and day of last visit.
- Missing month is imputed as minimum of last day of December and date of last visit.

9.1.6.3 Imputation for date of birth

When only month and year of birth is known, the date is imputed as the last day of the month. When only year is known, the date is imputed as the last day of the year.

9.1.7 Other derivations

9.1.7.1 Subjects with reported asthma

Subjects with reported asthma are defined as the subjects with allergic asthma at baseline according to the eCRF, along with subjects with asthma medical history as registered by MHLLTCD=10003553 or MHLLTCD=10003555.

9.1.7.2 eDiary periods

For some subjects the eDiary period has by mistake been longer than planned. In these cases it is decided to define the period to be the last part of the eDiary period. Thus for baseline the last 14 days up until one day before first dose will be used, for TC1 and TC2 the last 14 days will be used, and for the primary efficacy assessment period, the last 8 weeks of the eDiary period will be used.

9.1.7.3 Age for DMS

For each subject and each eDiary period, age is calculated at the first day of the eDiary period, and this age is used throughout the period.

9.2 Further details pertaining to statistical analyses

9.2.1 LME analyses

Adjusted means for each treatment group, the absolute treatment difference (Placebo – Active) with 95% confidence interval and p-value, and the relative treatment difference (Placebo – Active/ Placebo) with 95% confidence interval will be presented.

Each endpoint is calculated as the simple average of all the observed daily values during the relevant period as defined in section 9.1.7.2.

9.2.1.1 Without multiple imputation

When the endpoint is not transformed

The endpoint is analysed as the response variable in an LME which includes treatment group and cohort as fixed factors, the baseline value as a covariate, country/region as a random effect, and with different residual errors specified for each treatment group. Denominator degrees of freedom will be calculated using the Kenward and Roger's approximation (Kenward and Roger 1997). The p-value for the absolute difference is reported as the test result. The 95% confidence interval for the relative difference will be calculated using Fieller's theorem (Fieller 1954).

When the endpoint is square root transformed

The endpoint is square root transformed and analysed as the response variable in an LME which includes treatment group and cohort as fixed factors, the square root of the baseline value as a covariate, country/region within cohort as a random effect, and with different residual errors specified for each treatment group. Denominator degrees of freedom will be calculated using the Kenward and Roger's approximation (Kenward and Roger 1997). The p-value for the absolute difference is reported as the test result.

The results are back-transformed as follows; from the LME, estimated least square means on the square root transformed scale are output along with associated covariance matrix. For the absolute difference the SE is approximated by using the first order Delta method (first order Taylor approximation), and from this the 95% CI is calculated. For the relative difference, Fieller's theorem (Fieller 1954) is used to calculate the 95% CI for X/Y, and then the 95% confidence bounds for $(Y^2-X^2)/Y^2 = 1 - (X/Y)^2$ are found by applying the monotone transformation $f(r)= 1 - r^2$ to the confidence interval.

Sensitivity analysis using bootstrapping

For the primary endpoint, a sensitivity analysis will be performed where the 95% CI for the relative difference on the original scale, instead of using Fieller's theorem and a monotone bacl-transformation, will be calculated using bootstrapping. The bootstrapping which will be used is the "non-parametric residual bootstrap" by Carpenter et al (Carpenter et al. 2003).

Model checking:

The assumption of normally distributed residuals underlying the LME will be evaluated by visual inspection of quantile-quantile plots for the primary and key secondary endpoints. If the square root transformation does not result in a good approximation to the normal distribution, no transformation will be applied if this improves the approximation to a normal distribution

Convergence issues:

If the model does not converge, country/region will be pooled into larger regions. If the model still does not converge, the random effect of region within cohort will be removed from the model.

9.2.1.2 With multiple imputation

Imputation will be done using the method of unrestricted random sampling with replacement (seed=686) and 1000 multiple imputed datasets will be created.

When the endpoint is not transformed

Each of the 1000 multiple imputed datasets is analysed using the same model as without imputation. The endpoint is analysed as the response variable in an LME which includes treatment group and cohort as fixed factors, the baseline endpoint as a covariate, country/region as a random effect, with different residual errors specified for each treatment group, and with denominator degrees of freedom calculated using the Kenward and Roger's approximation (Kenward and Roger 1997).

Rubin's rule is used to combine the 1000 analysis results across the multiple imputed datasets. The p-value for the absolute difference is reported as the test result. The 95% confidence interval for the relative difference is calculated using Fieller's theorem (Fieller 1954).

When the endpoint is square root transformed

The endpoint is square root transformed. Each of the 1000 multiple imputed datasets is analysed using the same model as without imputation. The square root transformed endpoint is analysed as the response variable in an LME which includes treatment group and cohort as fixed factors, the square root of the baseline endpoint as a covariate, country/region as a random effect, with different residual errors specified for each treatment group, and with

denominator degrees of freedom calculated using Kenward and Roger's approximation (Kenward and Roger 1997).

Rubin's rule is used to combine the 1000 analysis results across the multiple imputed datasets. The p-value for the absolute difference is reported as the test result.

The results are back-transformed as follows; from the LME, estimated least square means on the square root transformed scale are output along with associated covariance matrix. For the absolute difference the SE is approximated by using the first order Delta method (first order Taylor approximation), and from this the 95% CI is calculated. For the relative difference, Fieller's theorem (Fieller 1954) is used to calculate the 95% CI for X/Y, and then the 95% confidence bounds for $(Y^2-X^2)/Y^2 = 1 - (X/Y)^2$ are found by applying the monotone transformation $f(r)= 1 - r^2$ to the confidence interval.

9.2.1.3 Square root transformation

The LME relies on the assumption of normally distributed residuals. Therefore the square root transformation is generally applied, as this usually results in a good approximation to the normal distribution. In addition, a lot of subjects do not use rescue meddication, so the average DMS score is often 0, and this is also accounted for by using the square root transformation, and a reason the logarithm transformation is not applied.

9.2.1.4 Mean and SE

Let μ_{Active}, μ_{Active}, $\sigma_{Placebo}$, and σ_{Active}be the resulting means and variances for placebo and active treatment and cov be the covariance. Then the back-transformed means are given by:

- $\hat{\mu}_{Placebo} = (\mu_{Placebo})^2$
- $\hat{\mu}_{Active} = (\mu_{Active})^2$
- $\hat{\mu}_{Difference} = (\mu_{Placebo})^2 - (\mu_{Active})^2$

and the back-transformed standard errors are given by

- $SE_{Active} = \sqrt{(2\mu_{Active})^2\sigma_{Active}}$
- $SE_{Placebo} = \sqrt{(2\mu_{Placebo})^2\sigma_{Placebo}}$
- $SE_{Difference} = \sqrt{(2\mu_{Placebo})^2\sigma_{Placebo} + (-2\mu_{Active})^2\sigma_{Active} + 2(2\mu_{Placebo})(-2\mu_{Active})cov}$

Table 14 SAS code for primary efficacy analysis

Data set	SAS code

All subjects in FAS with at least one diary record in the primary efficacy assessment	proc mixed method=reml plots=studentpanel(marginal conditional);
	class treatment cohort creg;
	model $TCRS_{sqrt}$= treatment cohort base_$TCRS_{sqrt}$ / ddfm = KenwardRogers residual;
	random creg (cohort) ;
	repeated/group=treatment;

period	lsmeans treatment / cl diff om; run;

$TCRS_{sqrt}$ is the square root of the average TCRS, $base\_TCRS_{sqrt}$ is the square root of the average TCRS at baseline, and creg is country/region, and the option om is used for the analysed population at baseline.

9.2.2 GLMM analyses

The daily binary responses will be analysed using a GLMM with a logit link function. The model includes the daily binary response as the response variable, treatment group and cohort as fixed factors, the relevant baseline value (from the model specification) as a covariate, and country/region within cohort, and subject as random effects.

The odds ratio for the response for active treatment relative to placebo will be presented together with the coherent p-values and 95% confidence limits.

Convergence issues:

If the model does not converge, country/region will be pooled into larger regions. If the model still does not converge, the random effect of region within cohort will be removed from the model.

Table 15 SAS code for additional efficacy analysis (binary endpoint)

Data set	SAS code
All subjects in FAS with at least one eDiary record in the primary efficacy assessment period	proc glimmix; class treatment cohort creg subject; model binary_variable (event='1') = treatment cohort baseline/ ddfm=kr solution dist=binary link=logit oddsratios; random creg (cohort) subject; lsmeans treatment /diff=control('Placebo') oddsratio cl ilink om; run;

creg is country/region, and baseline is the relevant baseline value from the model specification.

9.3 Subjects excluded from the sensitivity analysis

The subject ID for the 10 subjects who violated at least one inclusion criteria are:

[redacted]

The subject ID for the siblings participating in the same cohort are:

[redacted]

[redacted]

The subject ID for the subjects in [redacted] temporarily without IMP are:

[redacted]

10 References

Carpenter, James R; Goldstein, Harvey; Jon, Rasbash (2003): A Novel Bootstrap Procedure for Assessing the Relationship between Class Size and Achievement 52 (4), pp. 431–443. DOI: 10.1111/1467-9876.00415.

Fieller, E. C. (1954): Some problems in interval estimation. In *J. R. Stat. Soc. Series B (Methodological)* 16 (2), pp. 175–185.

ICH (1998): ICH Harmonised Tripartite Guideline Topic E9: Statistical Principles for Clinical Trials.

ICH (2019): ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials (step 4).

Juniper, E. F.; Howland, W. C.; Roberts, N. B.; Thompson, A. K.; King, D. R. (1998): Measuring quality of life in children with rhinoconjunctivitis. In *J. Allergy Clin. Immunol.* 101 (2 Pt 1), pp. 163–170. DOI: 10.1016/s0091-6749(98)70380-x.

Kenward, M. G.; Roger, J. H. (1997): Small sample inference for fixed effects from restricted maximum likelihood. In *Biometrics* 53 (3), pp. 983–997.

Passalacqua, Giovanni; Baena-Cagnani, Carlos E.; Bousquet, J.; Canonica, Giorgio W.; Casale, T. B.; Cox, Linda S. et al. (2013): Grading local side effects of sublingual immunotherapy for respiratory allergy: speaking the same language. In *J. Allergy Clin. Immunol.* 132 (1), pp. 93–98. DOI: 10.1016/j.jaci.2013.03.039.

Pfaar, O.; Demoly, P.; van Gerth Wijk, R.; Bonini, S.; Bousquet, J.; Canonica, Giorgio W. et al. (2014): Recommendations for the standardization of clinical outcomes used in allergen immunotherapy trials for allergic rhinoconjunctivitis: an EAACI Position Paper. In *Allergy Eur. J. Allergy Clin. Immunol.* 69 (7), pp. 854–867. DOI: 10.1111/all.12383.

Signature Page for VV-CLIN-004538 v1.0

Approval Task	[redacted] 24-May-2023 07:42:34 GMT+0000
Approval Task	[redacted] 24-May-2023 07:45:14 GMT+0000
Approval Task	[redacted] 24-May-2023 07:46:05 GMT+0000
Approval Task	[redacted] 24-May-2023 08:14:03 GMT+0000
Approval Task	[redacted] 24-May-2023 12:20:52 GMT+0000

Signature Page for VV-CLIN-004538 v1.0