Official Protocol Title:	A Phase III, Randomized, Placebo-Controlled Clinical Trial to Study the Efficacy and Safety of MK-3641, a Ragweed (Ambrosia artemisiifolia) Sublingual Immunotherapy Tablet, in Children With a History of Ragweed-Induced Rhinoconjunctivitis With or Without Asthma
NCT number:	NCT02478398
Document Date:	30-Nov-2018

Supplemental Statistical Analysis Plan (sSAP)

1 INTRODUCTION

This supplemental SAP (sSAP) is a companion document to the protocol. In addition to the information presented in the protocol SAP which provides the principal features of confirmatory analyses for this trial, this supplemental SAP provides additional statistical analysis details/data derivations and documents modifications or additions to the analysis plan that are not "principal" in nature and result from information that was not available at the time of protocol finalization.

2 SUMMARY OF CHANGES

Date	Page	Changes
Oct-21-2015	9	Added the sensitivity analysis for rescue medications not supplied by the Sponsor
Dec-1-2016	10	Added the sensitivity analysis for rescue medications not supplied by the Sponsor, for the two exploratory endpoints related to asthma medication.
Nov-30-2018	2	Added further definition of pollen regions.
Nov-30-2018	9	Added sensitivity analysis to include non-study provided allergy rescue medications for the DMS.
Nov-30-2018	10	Deleted sensitivity analysis for the nocturnal awakenings exploratory endpoints as related data was not collected. These analysis were originally added inadvertently.
Nov-30-2018	15	Added the adverse events identified by the WAO as local side effects of SLIT, the WAO method of grading AE intensity, and definition of the modified WAO intensity grading.
Nov-30-2018	18	Added the data analysis conventions for cross-treated subjects.

3 ANALYTICAL AND METHODOLOGICAL DETAILS

3.1 STATISTICAL ANALYSIS PLAN SUMMARY

This section contains a brief summary of the statistical analyses for this trial. Full detail is in the Statistical Analysis Plan (SAP) Section 3.2.

3.1.1 Efficacy Analyses

Primary Efficacy Analysis

The primary analysis will be conducted on the Full Analysis Set (FAS) population. The FAS population includes all subjects who receive at least one dose of study treatment. Subjects will be analyzed according to the treatment group to which they are randomized.

The primary efficacy endpoint for the current trial is the average TCS during the peak ragweed season. It will be analyzed using the analysis of variance (ANOVA) model, which includes fixed effects of treatment, baseline asthma status (yes, no), age group (5 to 11 years, 12 to 17 years), pollen season, and pollen region nested within pollen season. Pollen region will be defined based on pollen station and may include several sites within an acceptable distance between the pollen counters and corresponding sites. To avoid failure in model convergence, observed in model testing on the blinded data, pollen region will be defined by participating countries, except for the United States, which will be split into 2 pollen regions.

Two-sided 95% confidence interval (CI) of the treatment difference in adjusted means will be presented. Also, the treatment difference in adjusted means relative to the adjusted mean of the placebo group will be presented as a percentage with corresponding 2-sided 95% CI derived using the bootstrap method. No missing data will be imputed.

The normality assumption of the ANOVA model will be checked using the Shapiro-Wilk test and inspection of the Q-Q plot. If a severe violation is observed, the primary analysis will be based on Wilcoxon Rank Sum test; the p-value will be reported together with the associated Hodges-Lehmann estimate of the treatment difference and its 2-sided 95% CI. Also, the difference in the treatment group medians relative to the median of the placebo group will be presented as a percentage with the corresponding 2-sided 95% CI derived using the bootstrap method.

A Per Protocol (PP) analysis will be conducted for the primary efficacy endpoint, where the same ANOVA model and missing data approach will be used. See Section 3.2.5.1 for further sensitivity analyses of the primary efficacy endpoint.

Key Secondary Efficacy Analyses

The key secondary efficacy endpoints for the current trial include:

- Average TCS during the entire RS;
- Average rhinoconjunctivitis DSS during the peak RS;
- Average rhinoconjunctivitis DMS during the peak RS.

These endpoints will be analyzed based on the FAS population using the same ANOVA model as the primary efficacy endpoint. A fixed sequence procedure will be applied to control multiplicity, where the primary efficacy endpoint will be tested first, and then the key secondary efficacy endpoints will be tested in the order stated above.

In a situation when there are more than 30% of the daily rhinoconjunctivitis DMS equal to zero, the zero-inflated lognormal model will be used as the primary analysis method for rhinoconjunctivitis DMS as appropriate; this model takes the average rhinoconjunctivitis DMS during the peak RS as response and adjusts for the same terms as in the ANOVA model.

[Table 1] summarizes the key analysis strategy for the primary and key secondary efficacy endpoints.

Table 1 Analysis Strategy for Key Efficacy Endpoints

Endpoint	Statistical Method[a]	Analysis Population	Missing Data Approach
Primary Efficacy Endpoint			
Average TCS during the peak RS	ANOVA	FAS	Observed data only
			LOCF
			Multiple Imputation[b]
		PP	Observed data only
	Wilcoxon Rank Sum Test;	FAS	Observed data only
	Hodges-Lehmann estimate of treatment difference[c]	FAS	Model based
	Longitudinal Data Analysis		
Key Secondary Efficacy Endpoints			
Average TCS during the entire RS	ANOVA	FAS	Observed data only
Average rhinoconjunctivitis DSS during the peak RS	ANOVA	FAS	
Average rhinoconjunctivitis DMS during the peak RS	ANOVA	FAS	
	Zero-inflated log-normal[c]		

TCS = Total Combined Score; DSS = Daily Symptom Score; DMS = Daily Medication Score; RS = Ragweed Season.
ANOVA = Analysis of Variance; LOCF = Last non-missing observation carried forward.
FAS = Full Analysis set; PP = Per Protocol.
[a] Details of the statistical models are described in Section 3.2.5.
[b] Missing data from both treatment groups will be imputed using the sample distribution of TCS observed from the placebo group.
[c] This method will be applied if excessive zeros are observed in the data.

3.1.2 Safety Analyses

An All-Subjects-as-Treated (ASaT) population will be used for safety analyses. The ASaT population includes all subjects who receive at least one dose of study treatment. Subjects will be included in the treatment group corresponding to the study treatment they actually receive during the trial.

The analysis of safety endpoints will follow a tiered approach. For this study, Tier 1 safety endpoints include:

- Proportion of subjects reporting pre-specified local application site reactions (including adverse events related to lip swelling/edema, mouth swelling/edema, palatal swelling/edema, swollen tongue/edema, oropharyngeal swelling/edema, pharyngeal edema/throat tightness, oral pruritus, throat irritation, tongue pruritus, and ear pruritus);

- Proportion of subjects reporting anaphylaxis and/or systemic allergic reactions;
- Proportion of subjects treated with epinephrine.

Tier 1 safety endpoints will be subject to inferential testing for statistical significance with p-values and 95% confidence intervals provided for between-group comparisons. They will be analyzed using the stratified Miettinen and Nurminen method [1985] with baseline asthma status and age group as stratification factors.

3.1.3 Power and Sample Size

A total of approximately 1000 subjects will be randomized in a 1:1 ratio to either MK-3641 or placebo. Assuming a 15% dropout rate, this gives approximately 425 evaluable subjects per treatment group.

With 425 subjects per arm, the study will have:

- approximately 90% power (2-sided, $\alpha = 0.05$) to have the upper bound of the 95% CI for relative difference below -10%, and
- more than 90% power (2-sided, $\alpha = 0.05$) to have an estimated relative difference below -15%.

See Section 3.2.7 for further details of sample size calculation.

3.1.4 Interim Analyses

No efficacy interim analysis is planned for this trial. Safety data will be reviewed by an external Data Monitoring Committee (eDMC) (see Section 7.3.3 of the protocol).

3.2 STATISTICAL ANALYSIS PLAN

This section outlines the statistical analysis strategy and procedures for the study. If, after the study has begun, but prior to any unblinding, changes are made to primary and/or key secondary hypotheses, or the statistical methods related to those hypotheses, then the protocol will be amended (consistent with ICH Guideline E-9). Changes to exploratory or other non- confirmatory analyses made after the protocol has been finalized, but prior to final database lock, will be documented in a supplemental SAP (sSAP) and referenced in the Clinical Study Report (CSR) for the study. Post hoc exploratory analyses will be clearly identified in the CSR.

3.2.1 Responsibility for Analyses/In-House Blinding

The statistical analysis of the data obtained from this study will be the responsibility of the designee of the sponsor.

This study will be conducted as a double-blind study under in-house blinding procedures. The official, final database will not be unblinded until medical/scientific review has been

performed, protocol violators have been identified, and data have been declared final and complete.

The randomized allocation schedule will be generated by the sponsor and implemented by the vendor of the study interactive voice response system (IVRS).

Safety will be monitored by the external Data Monitoring Committee (eDMC) on an ongoing basis and the eDMC will make recommendations to the sponsor as appropriate.

3.2.2 Hypotheses

Objectives and hypotheses of the study are stated in Section 3 of the protocol.

3.2.3 Analyses Endpoints

Efficacy and safety endpoints that will be evaluated for within- and between-treatment differences are listed below, followed by the descriptions of the derivations of selected endpoints.

3.2.3.1 Efficacy Endpoints

- Average TCS during the peak RS.

Key Secondary Efficacy Endpoints

- Average TCS during the entire RS;
- Average rhinoconjunctivitis DSS during the peak RS;
- Average rhinoconjunctivitis DMS during the peak RS.

Tertiary Efficacy/Immunologic Endpoints

- Average rhinoconjunctivitis DSS during the entire RS
- Change from baseline in IgE level against *Ambrosia artemisiifolia* at Visit 6 and at Visit 8;
- Change from baseline in IgG_4 level against *Ambrosia artemisiifolia* at Visit 6 and at Visit 8.

Exploratory Efficacy Endpoints

- Average Asthma DSS during the peak RS, the entire RS, and the whole treatment period;
- Average daily number of puffs of as-needed SABA used during the peak RS and the entire RS;

- Average weekly number of nights with nocturnal awakening due to asthma symptoms requiring SABA use during the peak RS and the entire RS.

3.2.3.2 Safety Endpoints

Tier 1 Safety Endpoints

- Proportion of subjects reporting pre-specified local application site reactions (including adverse events related to lip swelling/edema, mouth swelling/edema, palatal swelling/edema, swollen tongue/edema, oropharyngeal swelling/edema, pharyngeal edema/throat tightness, oral pruritus, throat irritation, tongue pruritus, and ear pruritus);
- Proportion of subjects reporting anaphylaxis and/or systemic allergic reactions;
- Proportion of subjects treated with epinephrine.

Tier 2 Safety Endpoints

- Proportion of subjects with any AE;
- Proportion of subjects with any serious AE;
- Proportion of subjects with any drug-related AE;
- Proportion of subjects with any serious and drug-related AE;
- Proportion of subjects who discontinue due to an AE;
- Proportion of subjects with specific AEs or SOCs (incidence ≥ 1% subjects in one or more of the treatment groups).

Tier 3 Safety Endpoints

- Change from baseline in laboratory test parameters at each post-baseline visit;
- Change from baseline in vital sign parameters at each post-baseline visit;
- Summaries by system organ class (SOC) for general AEs, serious AEs, drug-related AEs, and AEs resulting in discontinuation;
- Average duration of AEs in minutes on Day 1;
- Average duration of AEs in days over Day 2 to Day 14;
- Average duration of AEs in days over Day 15 to Day 28.

3.2.3.3 Derivations of Efficacy/Immunologic Endpoints

Ragweed Season

Entire Ragweed Season: the start of the entire ragweed season is the first day of 3 consecutive recorded days with a pollen count of ≥10 grains/m^3; the end of the entire ragweed season is the last day of 3 consecutive recorded days with a pollen count of ≥10 grains/m^3. This is specific for each study site.

Peak ragweed season: the 15 consecutive recorded days within the entire ragweed season with the highest 15-day moving average pollen count. This is specific for each study site.

TCS, Rhinoconjunctivitis DSS, Rhinoconjunctivitis DMS, Asthma DSS

These are defined in Section 4.2.3.1 of the protocol.

3.2.3.4 Derivations of Safety Endpoints

For laboratory tests and vital sign parameters, the baseline value is defined as the last measurement taken prior to randomization. Change from baseline in laboratory and vital signs parameters is calculated by on-treatment value minus the baseline value.

3.2.4 Analysis Populations

3.2.4.1 Efficacy Analysis Populations

The Full Analysis Set (FAS) population will serve as the primary population for the analysis of efficacy data in this study. The FAS population consists of all randomized subjects who receive at least one dose of study treatment. Subjects will be analyzed according to the treatment group to which they are randomized.

A supportive analysis using the Per-Protocol (PP) population will be performed for the primary efficacy endpoint. The PP population excludes subjects due to important deviations from the protocol that may substantially affect the results of the primary efficacy endpoints. The major protocol violations that would lead to exclusion of a subject or a data point are described below:

- Subjects with asthma requiring prescribed high daily doses of inhaled corticosteroids (as defined in the protocol);
- Subjects with chronic sinusitis during the previous 2 years;
- Subjects treated with immunotherapy within 5 years prior to screening;
- Subjects unable to meet medication washout requirements;
- Subjects randomized in the study more than once;

- Subjects who participated in the same study at another site;
- Subjects with pre-seasonal duration of treatment exposure < 56 Days;
- Subjects with a negative skin prick test response to ragweed;
- Subjects with specific IgE to ragweed < 0.7 kU/L;
- Subjects with overall treatment compliance < 75% between Visit 2 and Visit 8;
- Subjects who took prohibited medications as defined in the protocol, with the exception of antihistamines. Subjects who have taken antihistamines (other than sponsor provided rescue medications) will be considered protocol violators if they have taken the medication for 2 or more consecutive days between Visit 6 and Visit 8;
- Subjects who had their blinded treatment randomization code broken.

Analyses related to SABA use and nocturnal awakening will be based on subjects with asthma in the FAS population.

3.2.4.2 Safety Analysis Populations

The All-Subjects-as-Treated (ASaT) population will be used for the analysis of safety data in this study. The ASaT population consists of all randomized subjects who received at least one dose of study treatment. Subjects will be grouped by the study treatment actually received.

3.2.5 Statistical Methods

Statistical testing and inference for efficacy and safety analyses are described in Sections 3.2.5.1 and 3.2.5.2. Controlling of family-wise Type I error rate is described in Section 3.2.6, Multiplicity. Nominal p-values will be computed for tertiary and exploratory efficacy analyses as a measure of strength of association between the endpoint and the treatment effect rather than formal tests of hypotheses. Unless otherwise stated, all statistical tests will be conducted at α=0.05 (2-sided) level.

3.2.5.1 Statistical Methods for Efficacy/Immunologic Analyses

Primary Efficacy Analysis

The primary efficacy endpoint for the current trial is the average TCS during the peak ragweed season. It will be analyzed using the ANOVA model, which includes fixed effects of treatment, baseline asthma status (yes, no), age group (5 to 11 years, 12 to 17 years), pollen season, and pollen region nested within pollen season. Pollen region will be defined based on pollen station and may include several sites within an acceptable distance between the pollen counters and corresponding sites.

Two-sided 95% CI of the treatment difference in adjusted means will be presented. Also, the treatment difference in adjusted means relative to the adjusted mean of the placebo group will be presented as a percentage with corresponding 2-sided 95% CI derived using the bootstrap method. No missing data will be imputed.

The normality assumption of the ANOVA model will be checked using the Shapiro-Wilk test and inspection of the Q-Q plot. If a severe violation is observed, the primary analysis will be based on Wilcoxon Rank Sum test; the p-value will be reported together with the associated Hodges-Lehmann estimate of the treatment difference and its 2-sided 95% CI. Also, the difference in the treatment group medians relative to the median of the placebo group will be presented as a percentage with the corresponding 2-sided 95% CI derived using the bootstrap method.

Sensitivity Analysis

The PP analysis will be conducted for the primary efficacy endpoint, where the same ANOVA model and missing data approach will be used. In addition, the following sensitivity analyses will be performed for the primary efficacy endpoint.

- The same ANOVA model based on the FAS population will be used. Missing data will be imputed by the LOCF approach. For each subject, only the observations within the pollen season are eligible to be carried forward.

- The same ANOVA model based on the FAS population will be used. Missing data will be handled by the multiple imputation method. Missing data from both treatment groups will be imputed using the sample distribution of TCS observed from the placebo group, which is considered to be a conservative approach for imputations.

- Longitudinal Data Analysis model based on the FAS population will be used. Daily TCS during the peak season will be the response variable and covariates will include treatment, time, the interaction of treatment by time, baseline asthma status, age group, pollen season, and pollen region nested within pollen season. The Toeplitz structure will be used to model the covariance of the repeated TCS measurements over time within subjects. Additional covariance structures, such as unstructured covariance, may be considered if convergence issues are encountered with the analysis model.

Furthermore, if there are more than 2% of subjects taking prednisone during any one ragweed season in the course of the study, a sensitivity analysis will be conducted for the primary efficacy endpoint, where the usage of prednisone is included in the definition of the rhinoconjunctivitis DMS. The medication scores for prednisone usage are defined in [Table 2].

Table 2 Medication Score for Prednisone Usage

Rescue Medication	Subject Dosing Instructions	Score/Dose Unit	Maximum Daily Score
Prednisone tablet 5 mg	Day 1: 1 mg/kg/day, Max 50 mg/day	1.6 (per tablet)	16
	Day 2+: 0.5 mg/kg/day, Max 25 mg/day	1.6 x 2 (per tablet)	16

In addition, in order to minimize missing data, subjects will be allowed to complete the daily e-diary up to 9 am the following morning (in case the subject cannot complete it at their regular time that day). A sensitivity analysis will be conducted for the FAS population, where only the data entered on the same day (instead of next morning) are included. The same ANOVA model and missing data approach will be used.

Moreover, the primary analysis will only consider the rescue medications provided by the study (i.e., reported in the e-diary). Some subjects may take rescue medications that are not provided by the study. A sensitivity analysis will be conducted for the DMS by including these rescue medications if a medication score was pre-specified in the protocol.

Key Secondary Efficacy Analyses

The key secondary efficacy endpoints for the current trial include:

- Average TCS during the entire RS;
- Average rhinoconjunctivitis DSS during the peak RS;
- Average rhinoconjunctivitis DMS during the peak RS.

These endpoints will be analyzed based on the FAS population using the same ANOVA model as the primary efficacy endpoint.

Based on the rescue medication usage from previous AIT trials, many subjects used little rescue medication during the trial, which resulted in large amounts of rhinoconjunctivitis DMS records equal to zero. Data with excessive zeroes may not conform to the normality assumption for the ANOVA model and may also prevent the proper application of non-parametric method. In a situation when there are more than 30% of the daily rhinoconjunctivitis DMS equal to zero, the zero-inflated lognormal model [16] will be used as the primary analysis method for rhinoconjunctivitis DMS as appropriate; this model takes the average rhinoconjunctivitis DMS during the peak RS as response and adjusts for the same terms as in the ANOVA model.

Tertiary Efficacy/Immunologic Analyses

The tertiary efficacy/immunologic endpoints for the current trial include:

- Average rhinoconjunctivitis DSS during the entire RS;
- Change from baseline in IgE level against *Ambrosia artemisiifolia* at Visit 6;
- Change from baseline in IgE level against *Ambrosia artemisiifolia* at Visit 8;
- Change from baseline in IgG_4 level against *Ambrosia artemisiifolia* at Visit 6;
- Change from baseline in IgG_4 level against *Ambrosia artemisiifolia* at Visit 8.

Average rhinoconjunctivitis DSS during the entire RS will be analyzed based on the FAS population using the same ANOVA model as the primary efficacy endpoint.

IgE and IgG_4 related endpoints will be analyzed separately using the constrained LDA (cLDA) model. The cLDA model assumes a common mean across treatment groups at baseline and different means for different treatments at each post-baseline time point. In this model, the response vector consists of the baseline value and the values observed at each post-baseline time point. The model will adjust for time, the interaction of treatment by time, baseline asthma status, age group, pollen season, and pollen region nested within pollen season. Time is treated as a categorical variable so that no restriction is imposed on the trajectory of the means over time. A Toeplitz covariance matrix will be used to model the correlation among repeated measurements. Further details of the model specification, assumptions, and SAS implementation codes are provided in the Appendix.

Exploratory Efficacy Analyses

The exploratory efficacy endpoints for the current trial include:

- Average Asthma DSS during the peak RS;
- Average Asthma DSS during the entire RS;
- Average Asthma DSS during the whole treatment period;
- Average daily number of puffs of as-needed SABA used during the peak RS;
- Average daily number of puffs of as-needed SABA used during the entire RS;
- Average weekly number of nights with nocturnal awakening due to asthma symptoms requiring SABA use during the peak RS;
- Average weekly number of nights with nocturnal awakening due to asthma symptoms requiring SABA use during the entire RS.

These endpoints will be analyzed using the same ANOVA model as the primary efficacy endpoint. Asthma DSS related endpoints will be analyzed based on the FAS population, while SABA use and nocturnal awakening related endpoints will be analyzed based on subjects with asthma in the FAS population.

In addition, components of the rhinoconjunctivitis DSS (ocular symptoms and nasal symptoms) will be summarized for the peak RS and the entire RS.

Moreover, the exploratory efficacy analyses will only consider the rescue medications provided by the study (i.e., reported in the e-diary). Some subjects may take rescue medications that are not provided by the study. A sensitivity analysis will be conducted for the following endpoints by including these rescue medications:

- Average daily number of puffs of as-needed SABA used during the peak RS;
- Average daily number of puffs of as-needed SABA used during the entire RS.

[Table 3] summarizes the key analysis strategy for the efficacy/immunologic endpoints.

Table 3 Efficacy/Immunologic Endpoints

Endpoint	Statistical Method[a]	Analysis Population	Missing Data
Primary Efficacy Endpoint			
Average TCS during the peak RS	ANOVA	FAS	Observed data only
			LOCF
			Multiple Imputation[b]
		PP	Observed data only
	Wilcoxon Rank Sum Test; Hodges-Lehmann estimate of treatment difference[c]	FAS	Observed data only
	Longitudinal Data Analysis	FAS	Model based
Key Secondary Efficacy Endpoints			
Average TCS during the entire RS	ANOVA	FAS	Observed data only
Average rhinoconjunctivitis DSS during the peak RS	ANOVA	FAS	
Average rhinoconjunctivitis DMS during the peak RS	ANOVA	FAS	
	Zero-inflated log-normal[c]		
Tertiary Efficacy/Immunologic Analyses			
Average rhinoconjunctivitis DSS during the entire RS	ANOVA	FAS	Observed data only
Change from baseline in IgE level against *Ambrosia artemisiifolia* at Visit 6 and Visit	constrained Longitudinal	FAS	Model based
Change from baseline in IgG_4 level against *Ambrosia artemisiifolia* at Visit 6 and Visit	constrained Longitudinal	FAS	Model based
Exploratory Efficacy Analyses			
Average Asthma DSS during the peak RS, the entire RS, and the whole treatment	ANOVA	FAS	Observed data only
Average daily number of puffs of as-needed SABA used during the peak RS and the entire RS	ANOVA	FAS (asthma only)	Observed data only
Average weekly number of nights with nocturnal awakening due to asthma symptoms requiring SABA use during the peak RS and the entire RS	ANOVA	FAS (asthma only)	Observed data only

ANOVA = Analysis of Variance; DSS = Daily Symptom Score; DMS = Daily Medication Score; FAS = Full Analysis set; IgE = immunoglobulin E; IgG_4 = immunoglobulin G_4; LOCF = Last non-missing observation carried forward; PP = Per Protocol; RS = Ragweed Season; SABA = short- acting $beta_2$-agonist; TCS = Total Combined Score.
[a] Details of the statistical models are described in Section 3.2.5.
[b] Missing data from both treatment groups will be imputed using the sample distribution of TCS observed from the placebo group.
[c] This method will be applied if excessive zeros are observed in the data.

Handling of Missing Data

The missing data approaches are specified for the primary efficacy endpoint sensitivity analyses (including multiple imputation, LOCF, and model based) and immunologic endpoints (model based). All other analyses will be conducted based on the observed data only. Proportion of subjects with missing data will be summarized for each efficacy/immunologic endpoint.

3.2.5.2 Statistical Methods for Safety Analyses

Safety and tolerability will be assessed by clinical review of all relevant parameters including adverse experiences (AEs), laboratory tests, and vital sign measurements.

The analysis of safety results will follow a tiered approach [Table 4]. The tiers differ with respect to the analyses that will be performed. Tier 1 safety endpoints will be subject to inferential testing for statistical significance with p-values and 95% CIs provided for between- group comparisons. Tier 2 safety endpoints will be evaluated via point estimates and 95% CIs for between-group comparisons. Tier 3 safety endpoints will be evaluated via point estimates only.

For this trial, Tier 1 safety endpoints include:

- Proportion of subjects reporting pre-specified local application site reactions (including adverse events related to lip swelling/edema, mouth swelling/edema, palatal swelling/edema, swollen tongue/edema, oropharyngeal swelling/edema, pharyngeal edema/throat tightness, oral pruritus, throat irritation, tongue pruritus, and ear pruritus);
- Proportion of subjects reporting anaphylaxis and/or systemic allergic reactions;
- Proportion of subjects treated with epinephrine.

For the analysis of "Proportion of subjects treated with epinephrine", the denominator will include all subjects in the ASaT population, regardless of whether self-injectable epinephrine was provided or not. Meanwhile, a sensitivity analysis will be conducted where the denominator will include subjects in the ASaT population who were provided with self-injectable epinephrine, and the numerator will be subjects from the denominator subjects who were treated with epinephrine.

Tier 2 safety endpoints for this trial include:

- Proportion of subjects with any AE;
- Proportion of subjects with any serious AE;
- Proportion of subjects with any drug-related AE;
- Proportion of subjects with any serious and drug-related AE;
- Proportion of subjects who discontinue due to an AE;
- Proportion of subjects with specific AEs or SOCs (incidence ≥ 1% subjects in one or more of the treatment groups).

The Tier 1 and Tier 2 safety endpoints will be analyzed using the stratified Miettinen and Nurminen method [1985] with baseline asthma status and age group as stratification

factors.

Tier 3 safety endpoints for this trial include:

- Change from baseline in laboratory test parameters at each post-baseline visit;
- Change from baseline in vital sign parameters at each post-baseline visit;
- Change from baseline in pulmonary function test parameters at each post-baseline visit;
- Summaries by SOC for general AEs, serious AEs, drug-related AEs, serious and drug- related AEs, and AEs resulting in discontinuation;
- Average duration of AEs in minutes on Day 1;
- Average duration of AEs in days over Day 2 to Day 14;
- Average duration of AEs in days over Day 15 to Day 28.

Point estimates will be provided for the Tier 3 safety endpoints. In addition, AE summaries (general AEs, serious AEs, drug-related AEs, serious and drug-related AEs, and AEs resulting in discontinuation) by asthma status will also be provided. For pulmonary function test parameters, a cLDA model-based analysis will also be conducted (see Section 3.2.5.1 for more details about the cLDA model).

[Table 4] summarizes the safety tier and level of analysis for the safety endpoints.

Table 4 Analysis Strategy for Safety Parameters

Safety Tier	Safety Endpoint[a]	p-Value	95% CI for Treatment Comparison	Descriptive Statistics
Tier 1	Proportion of subjects reporting pre-specified local application site reactions (including adverse events related to lip swelling/edema, mouth swelling/edema, palatal swelling/edema, swollen tongue/edema, oropharyngeal swelling/edema, pharyngeal edema/throat tightness, oral pruritus, throat irritation, tongue pruritus, and ear pruritus); Proportion of subjects reporting anaphylaxis and/or systemic allergic reactions; Proportion of subjects treated with epinephrine.	X	X	X
Tier 2	Any AE; Any Serious AE; Any Drug-Related AE; Any Serious and Drug-Related AE; Discontinuation due to AE; Specific AEs or SOCs[b] (incidence ≥1% of subjects in		X	X
Tier 3	Change from baseline in laboratory test parameters at each post-baseline visit; Change from baseline in vital sign parameters at each post-baseline visit; Change from baseline in pulmonary function test parameters at each post-baseline visit; Summaries by SOC for general AEs, serious AEs, drug- related AEs, serious and drug-related AEs, and AEs resulting in discontinuation; AE summaries by asthma status; Average duration of AEs in minutes on Day 1; Average duration of AEs in days over Day 2 to Day 14; Average duration of AEs in days over Day 15 to			X

AE = adverse event; CI = confidence interval; SOC = System Organ Class; X = results will be provided.
[a] Adverse Experience references refer to both Clinical and Laboratory AEs.
[b] Includes only those endpoints not pre-specified as Tier 1 or Tier 2 endpoints.

In addition, the SLIT Report Card (Section 4.2.3.2 of the protocol) will be completed by the subject/parent/guardian, daily for the first ~28 days of dosing, to collect information on local side effects of SLIT that occur within the first 60 minutes after each daily study drug intake. Adverse events identified by the WAO as local side effects of SLIT include the following: dysgeusia, oral pruritus, swelling of lips, mucosal edema, ear pruritus, swollen tongue, glossodynia, mouth ulceration, tongue ulceration, throat irritation, pharyngeal edema, nausea, abdominal pain upper/abdominal pain, vomiting, and diarrhea.

Identified local AEs will be graded via a programmatic approach and clinical review as mild, moderate, or severe. An exploratory review of the intensity grading via 2 different methods, the programmatic approach using the WAO grading system and grading as determined by the investigator's review of the AEs with their subjects/parents/guardians, will be conducted.

The WAO grading system for SLIT local adverse events defines intensity as follows:

Mild (Grade 1)	Not troublesome and no symptomatic treatment required and no discontinuation of SLIT because of local side effects
Moderate (Grade 2)	Troublesome or requires symptomatic treatment and no discontinuation of SLIT because of local side effects
Severe (Grade 3)	Grade 2 and SLIT discontinued because of local side effects

The WAO local AE intensity grading was modified in that subjects collect the information on the presence of the AEs on the SLIT Report Card but do not indicate if they consider the AE "troublesome" or "not troublesome." Subjects also indicate if a medication is utilized to treat the AE and information regarding discontinuation due to the AE is also collected. The WAO intensity grading for local AEs will be determined programmatically. Essentially, mild intensity (Grade 1) will be assigned to events that do not result in symptomatic use of a medication. If an AE results in use of a symptomatic medication to treat the AE, it is assigned an intensity of moderate (Grade 2) and considered troublesome; if an AE results in the use of a symptomatic medication and in discontinuation from treatment, it is assigned an intensity of severe (Grade 3).

3.2.6 Summaries of Baseline Characteristics, Demographics, and Other Analyses

Demographic and Baseline Characteristics

The number and percentage of subjects screened, randomized, the primary reasons for screening failure, and the primary reason for discontinuation will be displayed.

Demographic variables (including age, gender, race, weight, height, and body mass index), baseline characteristics (including baseline asthma status, inhaled corticosteroid use, duration of allergic rhinitis, sensitization type, ragweed specific IgE, wheal size from skin prick test, and number of puffs of bronchodilator used for FEV_1 reversibility), primary and secondary diagnoses, and prior and concomitant therapies will be summarized by treatment either by descriptive statistics or categorical tables. The comparability of the treatment groups for each relevant characteristic will be assessed by the summary tables. No statistical hypothesis tests will be performed on these characteristics.

Skin Prick Test (SPT) Analysis

Descriptive statistics (mean, standard deviation, etc.) of the wheal size determined from the SPT will be provided for all randomized subjects.

3.2.7 Multiplicity

A fixed sequence procedure will be applied to control multiplicity. The primary and key secondary efficacy endpoints will be tested in the following order:

- Average TCS during the peak RS;
- Average TCS during the entire RS;
- Average rhinoconjunctivitis DSS during the peak RS;
- Average rhinoconjunctivitis DMS during the peak RS.

A lower order endpoint will be tested only if all higher order endpoints have been tested and claimed statistically significant.

3.2.8 Sample Size and Power Calculations

A total of approximately 1000 subjects will be randomized in a 1:1 ratio to either MK-3641 or placebo. Assuming a 15% dropout rate, this gives approximately 425 evaluable subjects per treatment group.

With 425 subjects per arm, the study will have:

- approximately 90% power (2-sided, $\alpha = 0.05$) to have the upper bound of the 95% CI for relative difference below -10%, and
- more than 90% power (2-sided, $\alpha = 0.05$) to have an estimated relative difference below -15%.

The calculations are based on the assumptions that the true difference between treatment arms in average TCS during peak RS is -2.12, that the average TCS during peak RS from the placebo group is 8.9, and that the standard deviation is 5.60. These assumptions, as well as the assumption on dropout rate, are based on the two MK-3641 (AIT ragweed) adult studies, P05233 and P05234, and the pediatric trial P05239 for MK-7243 (AIT grass).

3.2.9 Subgroup Analyses and Effect of Baseline Factors

Analysis for the primary efficacy endpoint will be provided for the following subgroups of baseline factors:

- Baseline asthma status (yes, no);
- Age group (5 to 11 years, 12 to 17 years);
- Gender (male, female);
- Race (Caucasians, non-Caucasians);

- ICS use for subjects with asthma (yes, no);
- Allergen sensitization type (ragweed only, ragweed + others);
- Geographic region (e.g., US, Canada, Europe). Region classification for each study site will be determined before database lock.

In addition, subgroup analysis on the primary efficacy endpoint will be provided for:

- Pollen counts (low, high): The subgroups will be defined based on the median of the subject level accumulated pollen counts during the first 21 days of the pollen season;
- Local application site reaction (had local application site reactions, did not have local application site reactions).

The same ANOVA model as in the primary efficacy analysis will be applied for the subgroup analyses.

3.2.10 Interim Analyses

No efficacy interim analysis is planned for this trial. Safety data will be reviewed by an eDMC.

3.2.11 Compliance (Medication Adherence)

For each subject, percent compliance will be calculated using the following formula:

$$\text{Percent Compliance} = \frac{\text{Number of Days on Therapy}}{\text{Number of Days Should Be on Therapy}} \times 100\%$$

A day within the study will be considered an "On-Therapy" day if the subject takes all required medication as instructed in Section 5.2 of the protocol. When a subject takes less than or more than the required medication on a day, that day is not considered an On-Therapy day.

For subjects who are followed for the entire study period, the "Number of Days Should be on Therapy" is the number of days from the first scheduled treatment day to the last scheduled treatment day. For subjects who discontinue from the study, the "Number of Days Should Be on Therapy" is the number of days from the first scheduled treatment day to the last dose day.

Summary statistics will be provided on percent compliance for the ASaT population.

3.2.12 Extent of Exposure

The duration of treatment for each subject will be evaluated by calculating the number of days on therapy. The range and mean for days on therapy will be calculated for the ASaT population.

3.2.13 Cross Treated Subjects

The following data analysis conventions will be used for cross treated subjects:

Safety Analysis:

- Scenario 1: if the incorrect kit corresponded to the assigned treatment arm then the subject will be analyzed to the group randomized (no issue).
- Scenario 2: If the subject incorrectly **switched from active to placebo** then the subject will be analyzed in the active treatment arm (conservative approach).
- Scenario 3: If the subject incorrectly **switched from placebo to active** then the subject will be analyzed in the active treatment arm (conservative approach).

For subjects that incorrectly switched from placebo to active, they will be excluded from the by-time-adverse event analyses.

For any adverse experiences (AE) occurring during the period that the incorrect study medications were taken, the cross-treatment information will be provided in the by-subject-listing of AEs and the narratives if the subject had a serious AE or discontinued from the study due to an AE.

Subject Compliance / Extent of Exposure:

The cross-treated information will be included in the clinical study report (CSR). The actual study drug exposure information within the clinical database for these subjects will be used to calculate the subject compliance and the duration of treatment presented in the relevant CSR tables (eg, Extent of Exposure), with the days exposed to the incorrect study drug being treated as days not exposed / non-compliant to the randomized treatment.

Efficacy Analysis:

No impact for the FAS analysis. As pre-specified in the protocol, subjects will be included in the Full Analysis Set (FAS) population and will be analyzed according to the treatment group to which he/she was randomized.

APPENDIX: Technical Details for supplemental SAP

APPENDIX A

CONSTRAINED LONGITUDINAL DATA ANALYSIS (CLDA) METHOD (WITH ADJUSTMENT FOR BASELINE VALUES) – TECHNICAL DETAILS FOR MODEL SPECIFICATION, ASSUMPTIONS, AND SAS IMPLEMENTATION CODES

Model

Let Y_{ijt} be the response for subject *i,* with treatment assignment *j*, at time *t*. The marginal mean responses of the cLDA model can be formulated as

$$E(Y_{ij0}) = y_0, \qquad j = 0, 1.$$

and

$$E(Y_{ijt}) = y_0 + y_{jt}, \qquad j = 0, 1;\ t = 1, 2, 3, 4.$$

The mean response γ_0 at $t = 0$ is constrained to be the same for all treatment groups due to randomization. The effect γ_{jt} denotes the change from baseline for treatment *j* at time *t*. The cLDA model assumes that baseline and post-baseline values have a joint multivariate normal distribution. A Toeplitz covariance matrix can be specified in the mixed model to account for within subject correlation at times $t \geq 0$ (including baseline).

At each time point *t*, *t = 1, 2, 3, 4*, the mean change from baseline (LSMEANS) for j^{th} dose of test drug and control are γ_{jt} and γ_{0t}, respectively, as defined in the cLDA model above.

The treatment difference for j^{th} treatment group vs. control for the mean change from baseline at time point *t*, $t = 1, 2, 3, 4$ is defined as:

$$\eta_{jt} = y_{jt} - y_{0t}.$$

This longitudinal model provides valid statistical inference in the presence of possible missing data if the missing data mechanism is ignorable (or more specifically, missing at random [MAR] or missing completely at random [MCAR]). This missing data mechanism requires that the probability of a data point being missing does not depend on the missing data after adjusting for the observed data.

Reasons for discontinuation from the trial may include lack of efficacy, clinical or laboratory adverse experiences, relocation, withdrawal of consent, protocol violations, and/or data processing issues. Missing data caused by relocation and data processing issues, are likely to be MCAR. On the other hand, missing data caused by discontinuation due to lack of efficacy may belong to MAR because the discontinuation may depend on the observed efficacy outcomes. The

MAR or MNAR mechanisms might each underlie the other reasons to some extent. If treatment in large part determines the loss of data for these other reasons (such as clinical or laboratory adverse experiences), the mechanism may be close to MAR since treatment assignment is an observed variable and included in the analysis model. Based on the prior trial results, missing data due to other reasons is relatively infrequent.

Model Convergence

If the Toeplitz covariance model fails to converge with the default algorithm, then the AR(1) structure can be used to provide initial values of the covariance parameters.

Example SAS Codes

```
**********************************************************************;
** data step necessary prior to running SAS PROC MIXED;
**********************************************************************; DATA long; SET long;
ARRAY T{5} t0-t4; * time indicator variables; ARRAY TT1{5} tt10-tt14;

** define week times treatment 1 indicator variables; DO i = 1 TO 5;
t{i} = (time=(i-1));
tt1{i} = t{i}*(trt=1);
tt2{i} = t{i}*(trt=2);
END; DROP i; RUN;
```

```
**********************************************************************;
** fitting the cLDA model using SAS PROC MIXED;
**********************************************************************; PROC MIXED DATA=long;

CLASS subj time stratum;  ** subj is the subject id number **; MODEL y=time tt11 tt12 tt13 tt14 tt21 tt22 tt23 tt24 stratum;
REPEATED time / SUBJECT=subj TYPE=TOEP;

ESTIMATE 'T1 Diff (MK - Placebo)' tt11 1; ESTIMATE 'T2 Diff (MK - Placebo)' tt12 1; ESTIMATE 'T3 Diff (MK - Placebo)' tt13 1;
ESTIMATE 'T4 Diff (MK - Placebo)' tt14 1;

ESTIMATE 'T1 Placebo LSM' time -1 1 0 0 0;
ESTIMATE 'T2 Placebo LSM' time -1 0 1 0 0;
ESTIMATE 'T3 Placebo LSM' time -1 0 0 1 0;

ESTIMATE 'T4 Placebo LSM' time -1 0 0 0 1;

ESTIMATE 'T1 MK LSM' time -1 1 0 0 0 tt11 1;
ESTIMATE 'T2 MK LSM' time -1 0 1 0 0 tt12 1;
ESTIMATE 'T3 MK LSM' time -1 0 0 1 0 tt13 1;
ESTIMATE 'T4 MK LSM' time -1 0 0 0 1 tt14 1;

ODS OUTPUT Estimates=outm1; RUN;
```