

Application Type	BLA
STN	125592/0
CBER Received Date	February 9, 2016
Goal Date	February 8, 2017
Division / Office	OVR
Committee Chair	Colleen Sweeney, R.N., M.S.
Clinical Reviewer(s)	Kathleen Hise , M.D.
Project Manager	Matthew Steele, Ph.D. Taruna Khurana, Ph.D.
Priority Review	No
Reviewer Name(s)	Zhong Gao, Ph.D. Mathematical Statistician
Review Completion Date / Stamped Date	
Supervisory Concurrence	Lihan Yan, Ph.D. Team Leader, Bacterial and Allergenic Team
	Amelia Dale Horne, Dr. P.H. Chief, Vaccine Evaluation Branch, DB, OBE
Applicant	Merck Sharp & Dohme Corp.
Established Name	House Dust Mites Allergen Extract
(Proposed) Trade Name	
Pharmacologic Class	Allergenic Extract
Formulation(s), including Adjuvants, etc	Tablet
Dosage Form(s) and Route(s) of Administration	Sublingual use only (Place the tablet under the tongue until dissolved. Do not swallow for at least 1 minute)
Dosing Regimen	12 SQ-HDM; one tablet daily
Indication(s) and Intended Population(s)	Immunotherapy for the treatment of house dust mite (HDM)-induced allergic rhinitis, with or without conjunctivitis, confirmed by positive skin test or <i>in vitro</i> testing for <i>Dermatophagoides farinae</i> or <i>Dermatophagoides pteronyssinus</i> IgE antibodies, in adults 18 through 65 years of age.

Table of Contents

1. Executive Summary 4

2. Clinical and Regulatory Background..... 5

 2.1 Disease or Health-Related Condition(s) Studied 5

 2.2 Currently Available, Pharmacologically Unrelated Treatment(s)/Intervention(s) for the Proposed Indication(s) 5

 2.4 Previous Human Experience with the Product (Including Foreign Experience)..... 5

 2.5 Summary of Pre- and Post-submission Regulatory Activity Related to the Submission 5

 2.6 Other Relevant Background Information 6

3. Submission Quality and Good Clinical Practices 6

 3.1 Submission Quality and Completeness..... 6

 3.2 Compliance with Good Clinical Practices and Data Integrity..... 6

4. Significant Efficacy/Safety Issues Related to Other Review Disciplines..... 6

5. Sources of Clinical Data and Other Information Considered in the Review 6

 5.1 Review Strategy 6

 5.2 BLA/IND Documents That Serve as the Basis for the Statistical Review..... 6

 5.3 Table of Studies/Clinical Trials 7

 5.4 Consultations 7

 5.5 Literature Reviewed (if applicable) 7

6. Discussion of Individual Studies/Clinical Trials 7

6.1 Study P001: A One-year Placebo-Controlled Study Evaluating the Efficacy and Safety of the House Dust Mite Sublingual Allergen Immunotherapy Tablet (SCH 900237/MK-8237) in Children and Adult Subjects With House Dust Mite-Induced Allergic Rhinitis/Rhinoconjunctivitis With or Without Asthma 7

 6.1.1 Objectives..... 7

 6.1.2 Design Overview..... 8

 6.1.3 Population 9

 6.1.4 Study Treatments or Agents Mandated by the Protocol..... 9

 6.1.6 Sites and Centers 9

 6.1.7 Surveillance/Monitoring.....10

 6.1.8 Endpoints and Criteria for Study Success10

 6.1.9 Statistical Considerations & Statistical Analysis Plan10

 6.1.10 Study Population and Disposition13

 6.1.11 Efficacy Analyses.....15

 6.1.12 Safety Analyses21

6.2 Study P015 (MT-06): A One-year Trial Evaluating the Efficacy and Safety of the ^{(b) (4)} House Dust Mite Allergy Immunotherapy Tablet in Adult Subjects with House Dust Mite Allergic Rhinitis22

 6.2.1 Objectives.....23

 6.2.2 Design Overview.....23

 6.2.3 Population23

 6.2.4 Study Treatments or Agents Mandated by the Protocol.....24

 6.2.6 Sites and Centers24

 6.2.7 Surveillance/Monitoring.....24

 6.2.8 Endpoints and Criteria for Study Success24

 6.2.9 Statistical Considerations & Statistical Analysis Plan24

 6.2.10 Study Population and Disposition27

 6.2.11 Efficacy Analyses.....28

6.2.12 Safety Analyses	31
6.3 Study P003: A Phase IIb, Randomized, Placebo-Controlled, Dose-Finding Clinical Trial to Study the Safety and Efficacy of MK-8237 using an Environmental Exposure Chamber in Subject with House Dust Induced Allergic Rhinitis/Rhinoconjunctivitis.....	32
6.3.1 Objectives.....	32
6.3.2 Design Overview.....	32
6.3.3 Population	33
6.3.4 Study Treatments or Agents Mandated by the Protocol.....	33
6.3.6 Sites and Centers	33
6.3.7 Surveillance/Monitoring.....	33
6.3.8 Endpoints and Criteria for Study Success	33
6.3.9 Statistical Considerations & Statistical Analysis Plan	34
6.3.10 Study Population and Disposition	35
6.3.11 Efficacy Analyses.....	36
6.3.12 Safety Analyses.....	38
7. Integrated Overview of Efficacy.....	39
7.1 Indication #1 Immunotherapy for the treatment of house dust mite (HDM)-induced allergic rhinitis, with or without conjunctivitis in adults 18 through 65 years of age.....	39
7.1.1 Methods of Integration	39
7.1.2 Demographics and Baseline Characteristics	40
7.1.4 Analysis of Primary Endpoint(s).....	40
7.1.5 Analysis of Secondary Endpoint(s)	41
7.1.6 Other Endpoints	42
7.1.7 Subpopulations	42
7.1.10 Additional Efficacy Issues/Analyses.....	42
7.1.11 Efficacy Conclusions.....	42
8. Integrated Overview of Safety	43
8.1 Safety Assessment Methods	43
8.2 Safety Database	43
8.2.1 Studies/Clinical Trials Used to Evaluate Safety.....	43
8.2.2 Overall Exposure, Demographics of Pooled Safety Populations	43
8.3 Caveats Introduced by Pooling of Data Across Studies/Clinical Trials.....	44
8.4 Safety Results.....	44
8.4.1 Deaths.....	44
8.4.2 Nonfatal Serious Adverse Events.....	44
8.4.3 Study Dropouts/Discontinuations.....	44
8.4.4 Common Adverse Events.....	44
8.4.5 Clinical Test Results.....	47
8.4.6 Systemic Adverse Events	47
8.4.7 Local Reactogenicity.....	47
8.4.8 Adverse Events of Special Interest.....	47
9. Additional Statistical Issues	47
10. Conclusions.....	47
10.1 Statistical Issues and Collective Evidence	47
10.2 Conclusions and Recommendations.....	48

1. EXECUTIVE SUMMARY

Merck Sharp & Dohme Corp. (Merck) submitted a Biologics License Application (BLA) for MK-8237 (House Dust Mites Allergen Extract) sublingual immunotherapy tablets. The product is indicated as immunotherapy for the treatment of house dust mite (HDM)-induced allergic rhinitis, with or without conjunctivitis, confirmed by positive skin test or in vitro testing for *Dermatophagoides farinae* or *Dermatophagoides pteronyssinus* immunoglobulin E (IgE) antibodies, in adults 18 through 65 years of age.

Efficacy: The Center for Biologics Evaluation and Research (CBER) and Merck agreed during the End of Phase II meeting that the pivotal North American study (P001) should be designed to demonstrate that the upper bound of the 95% confidence interval (CI) for treatment difference of MK-8237 12DU relative to placebo in average Total Combined Rhinitis Score (TCRS) during the last 8 weeks of treatment is lower than or equal to -10%. The study showed that the difference between MK-8237 12DU and placebo in average TCRS during the last 8 weeks of treatment was statistically significantly different from zero ($p < 0.001$); the treatment difference of MK-8237 12DU relative to placebo was -17.2% (95% CI, -25.0%, -9.7%). The upper bound of the 95% CI of relative treatment difference, calculated from the Full Analysis Set (FAS) using the bootstrap method, was slightly higher than the predefined threshold of -10%. As additional evidence, the applicant also provided results from the Phase II environmental exposure chamber (EEC) study P003. The study showed that the difference between MK-8237 12DU and the placebo in the average total nasal symptom score (TNSS) determined during the chamber session at Week 24 was statistically significantly different from zero ($p < 0.001$); the percent treatment difference of MK-8237 12DU relative to placebo was -48.6% (95% CI, -60.2%, -35.3%).

Safety: In the pivotal North American Study P001, there were more subjects with adverse events (AEs), drug-related AEs, discontinuations due to AEs, and discontinuations due to drug-related AEs in the MK-8237 12 DU group than in the placebo group. Specifically, the percentage of subjects with drug-related AEs was substantially higher in the MK-8237 group (84.0%) than in the placebo group (40.8%). The percentage of subjects who discontinued due to one or more adverse events was higher in the MK-8237 group (9.8%) than in the placebo group (2.6%); the percentage of subjects who discontinued due to one or more drug-related adverse events was higher in the MK-8237 group (8.3%) than in the placebo group (0.8%). The percentages of subjects with serious adverse events (SAEs) were generally similar between the two treatment groups (1.5% for MK-8237 vs. 0.9% for placebo) in this study. The safety findings were similar in the integrated safety analysis based on combined Phase II and III AR/C Trials (P003, MT-06, and P001) and the integrated safety analysis with all combined Phase II and III Trials (MT-02, MT-04, P003, MT-06, and P001).

I defer to the medical officer for further consideration of the totality of the evidence for efficacy and safety.

2. CLINICAL AND REGULATORY BACKGROUND

Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc. (Merck), submitted a Biologics License Application (BLA) for MK-8237 (House Dust Mites Allergen Extract) sublingual immunotherapy tablets. MK-8237 is a house dust mite (HDM) sublingual immunotherapy tablet (SLIT-tablet) that is a fast-dissolving pharmaceutical formulation (oral lyophilisate) containing standardized HDM allergen extract of *Dermatophagoides pteronyssinus* (*D. pteronyssinus*) and *Dermatophagoides farinae* (*D. farinae*). The applicant proposed that MK-8237 HDM SLIT-tablet be indicated as immunotherapy for the treatment of house dust mite (HDM)-induced allergic rhinitis, with or without conjunctivitis, confirmed by positive skin test or in vitro testing for *Dermatophagoides farinae* or *Dermatophagoides pteronyssinus* immunoglobulin E (IgE) antibodies, in adults 18 through 65 years of age.

2.1 Disease or Health-Related Condition(s) Studied

House dust mite allergy

2.2 Currently Available, Pharmacologically Unrelated Treatment(s)/Intervention(s) for the Proposed Indication(s)

Currently, the treatments for allergic diseases are based on allergen avoidance, allergy pharmacotherapy, and allergy immunotherapy (AIT). HDM allergen (*D. farinae* and *D. pteronyssinus* allergens) products for subcutaneous use (SCIT) are commercially available in North America.

2.4 Previous Human Experience with the Product (Including Foreign Experience)

On 30 Aug 2015, a decentralized procedure (DCP) comprising 11 EU countries (Austria, Czech Republic, Denmark, Finland, France, Germany, Italy, Norway, Poland, Slovakia, and Sweden) was completed for the HDM SLIT-tablet. The first national Marketing Authorization was issued to ALK-Abelló A/S (ALK) by Denmark on 23 September 2015, under the tradename of ACARIZAX. ACARIZAX is indicated in adult patients (18-65 years) diagnosed by clinical history and a positive test of house dust mite sensitization (skin prick test and/or specific IgE) with at least one of the following conditions: persistent moderate to severe house dust mite allergic rhinitis despite use of symptom-relieving medication; house dust mite allergic asthma not well controlled by inhaled corticosteroids and associated with mild to severe house dust mite allergic rhinitis. In September 2015, the HDM SLIT-tablet, licensed by ALK to Torii Pharmaceutical Co., Ltd. under the trade name of MITICURE, was approved for use in Japan. MITICURE is indicated for adults and adolescents (12 through 64 years) as hypo-sensitization therapy (allergy immunotherapy) for the treatment of allergic rhinitis caused by house dust mites. Merck and ALK are collaborating on the development of MK-8237 for North America.

2.5 Summary of Pre- and Post-submission Regulatory Activity Related to the Submission

Based on discussion at the end-of-Phase-II meeting (28-Mar-2014), to support the US Biologics License Application (BLA), the applicant conducted one clinical in-field Phase

III trial (study P001) in North America to evaluate efficacy and safety of MK-8237 in the HDM allergic rhinitis/rhinoconjunctivitis (AR/C) population. This study was designed to demonstrate a statistically significant and clinically meaningful treatment effect of MK-8237 compared to placebo. In addition to demonstrating statistical significance, P001 was also designed to meet the clinical relevance criterion required by the FDA of achieving the upper bound of the 95% confidence interval for the between-treatment difference in the TCRS relative to placebo $\leq -10\%$. Following completion of the Phase III North American efficacy trial, P001, the applicant and CBER had a Pre-BLA meeting on September 8, 2015. CBER concurred that the efficacy and safety data can be submitted for review for the proposed indication.

2.6 Other Relevant Background Information

N/A

3. SUBMISSION QUALITY AND GOOD CLINICAL PRACTICES

3.1 Submission Quality and Completeness

The submission is adequately organized for conducting a complete statistical review.

3.2 Compliance with Good Clinical Practices and Data Integrity

The submission generally complied with good data integrity.

4. SIGNIFICANT EFFICACY/SAFETY ISSUES RELATED TO OTHER REVIEW DISCIPLINES

N/A

5. SOURCES OF CLINICAL DATA AND OTHER INFORMATION CONSIDERED IN THE REVIEW

5.1 Review Strategy

This review focuses on the Phase III North American pivotal study P001, Phase III study MT-06 (conducted in EU, also known as study P015), and Phase II environmental exposure chamber (EEC) challenge study P003.

5.2 BLA/IND Documents That Serve as the Basis for the Statistical Review

- STN 125592/0 Module 2.5. Clinical Overview
- STN 125592/0 Module 2.7.3. Summary of Clinical Efficacy
- STN 125592/0 Module 2.7.4. Summary of Clinical Safety
- STN 125592/0 Module 2.7.6. Synopses of Individual Studies
- STN 125592/0 Module 5.3.5.1. Study P001 Clinical Study Report
- STN 125592/0 Module 5.3.5.1. Study P015 (MT-06) Clinical Study Report
- STN 125592/0 Module 5.3.5.1. Study P003 Clinical Study Report

- STN 125592/0 Module 5.3.5.3. Integrated Summary of Safety
- STN 125592/6, 125592/7, 125592/9, 125592/10, 125592/13, 125592/24: Merck’s response to CBER information request

5.3 Table of Studies/Clinical Trials

The clinical development program includes eight studies (Table 1).

Table 1 Overview of Completed Trials in the MK-8237 Clinical Development Program

Trial	Phase	Region	Age / years	Main Objective	Inclusion Criterion	MK-8237 Doses (DU) once daily
MT-01 (P011)	I	EU	18-63	Tolerability and safety	AA ± AR	1, 2, 4, 8, 16, 32
MT-03 (P013)	I	EU	5-14	Tolerability and safety	AA ± AR	0.5, 1, 3, 6, 9, 12
P008	I	US	12-17	Tolerability and safety	AR ± asthma	6, 12
MT-02 (P012)	II	EU	14-74	Efficacy in AA	AA + AR	1, 3, 6
P003	II	EU	18-65	Efficacy in AR (EEC)	AR ± asthma	6, 12
MT-04 (P014)	III	EU	17-83	Efficacy in AA	AA + AR	6, 12
MT-06 (P015)	III	EU	18-66	Efficacy in AR	AR ± AA	6, 12
P001	III	NA	12-85	Efficacy in AR	AR ± asthma	12

Note: AR: allergic rhinitis; AA: allergic asthma; DU: development unit; EEC: environmental exposure chamber; EU: European Union; NA: North America; US: United States.

Source: adapted from Table 2.5: 1 in Clinical Overview

5.4 Consultations

N/A

5.5 Literature Reviewed (if applicable)

N/A

6. DISCUSSION OF INDIVIDUAL STUDIES/CLINICAL TRIALS

6.1 Study P001: A One-year Placebo-Controlled Study Evaluating the Efficacy and Safety of the House Dust Mite Sublingual Allergen Immunotherapy Tablet (SCH 900237/MK-8237) in Children and Adult Subjects With House Dust Mite-Induced Allergic Rhinitis/Rhinoconjunctivitis With or Without Asthma

6.1.1 Objectives

Primary objective:

- To evaluate the efficacy of MK-8237 compared to placebo in the treatment of HDM-induced allergic rhinitis/rhinoconjunctivitis based on the average total combined rhinitis score (TCRS) during the last 8 weeks of treatment.
- To evaluate the safety and tolerability of MK-8237 administered sublingually once daily to subjects with a history of HDM-induced allergic rhinitis/rhinoconjunctivitis.

Secondary objectives:

To compare the following between the MK-8237 and placebo groups:

- Average rhinitis daily symptom score (DSS) during the last 8 weeks of treatment;
- Average rhinitis daily medication score (DMS) during the last 8 weeks of treatment;
- Average total combined score (TCS) during the last 8 weeks of treatment;
- Average allergic rhinitis/rhinoconjunctivitis symptoms assessed by Visual Analogue Scale (VAS) during the last 8 weeks of treatment.

6.1.2 Design Overview

This was a double-blind, randomized, placebo-controlled, multicenter trial to evaluate the efficacy and safety of MK-8237 12 DU sublingual tablets in subjects 12 years of age and older with a clinical history of house dust mite (HDM)-induced allergic rhinitis/rhinoconjunctivitis of 1 year duration or more, with or without asthma. The trial was designed to randomize a total of approximately 1500 subjects to receive (in a 1:1 ratio) either MK-8237 12 DU or placebo once daily for up to 52 weeks. Randomization was stratified by asthma status (yes/no) and age (<18/≥18 years). Figure 1 illustrates the overall study design.

Subjects who met all trial entry criteria and experienced the appropriate symptom threshold (during the Run-in period (maximum of 6 weeks) or historically with positive allergen provocation challenge to HDM) were randomized to study drug. As shown in Figure 1, subjects completed at least 12 visits: Screening; Run-in; 9 visits during the treatment phase (including randomization and 2 telephone contacts); and a final follow-up telephone contact 2 weeks following the last visit. Subjects were required to record rhinitis, conjunctivitis, and asthma symptom scores in an e-diary in the morning during the Run-in period and again from Visits 9 to 11. Rescue medications were provided to the trial sites to be given to the subjects as pre-defined, open-label medications to be taken in a step-wise fashion depending on the persistence, severity, and type of symptoms for allergic rhinoconjunctivitis during the last 12 weeks of the treatment period starting from Visit 9. Requirements for the start of rescue medication use included total symptom score of ≥4 and/or persistent eye symptoms. Efficacy was measured during approximately the last 8 weeks of the treatment period.

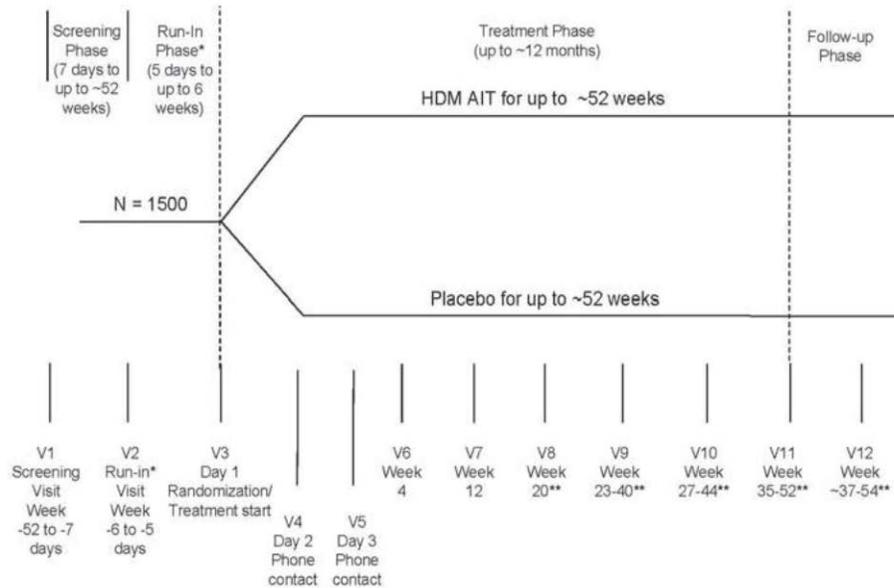


Figure 1 Trial Design Schematic

Note: AIT - allergy immunotherapy tablet; HDM - house dust mite

* Selected pre-approved sites did not perform the Run-in and combined Visits 1 and 2. This visit occurred -6 weeks to -5 days before randomization.

** Subjects randomized after Aug 10, 2014 followed a modified schedule per the trial flowchart (Visit 8 = Week 18, Visit 9 = Week 21, Visit 10 = Week 25, Visit 11 = Week 33, Visit 12 = Week 35.)

Source: Figure 9-1 in the P001 Clinical Study Report

6.1.3 Population

Adolescent and adult (12 years of age and older) males and females with a history of HDM-induced allergic rhinitis/rhinoconjunctivitis with or without asthma were eligible for the trial. Subjects were also required to have a history of prior treatment for HDM-induced allergic rhinitis/rhinoconjunctivitis in the previous year before the Screening Visit. Subjects were required to be sensitized to HDM with both a positive skin test of ≥ 5 mm compared with saline control and serum specific IgE of ≥ 0.7 kU/L to either *D. farinae* and/or *D. pteronyssinus*. Furthermore, subjects were required to demonstrate an allergic rhinitis symptom threshold before randomization at a time of year when subjects were expected to have natural exposure to HDM and when allergy-relieving medications were removed. Alternatively, at selected sites preapproved by the applicant, allergic rhinitis symptoms confirmed by a positive HDM chamber challenge could be used for eligibility in lieu of the Run-in period.

6.1.4 Study Treatments or Agents Mandated by the Protocol

- MK-8237 12 DU sublingual tablet, once daily up to approximately 52 weeks;
- Placebo sublingual tablet, once daily up to approximately 52 weeks.

6.1.6 Sites and Centers

This trial was conducted at 182 trial sites: 157 in the United States and 25 in Canada.

6.1.7 Surveillance/Monitoring

N/A

6.1.8 Endpoints and Criteria for Study Success

- Primary Efficacy Endpoint: Average Total Combined Rhinitis Score (TCRS) during the last 8 weeks of treatment. The TCRS was the sum of rhinitis Daily Symptom Score (DSS) and rhinitis Daily Medication Score (DMS) with a range of 0 to 24 points.
- Secondary Efficacy Endpoints
 - Average rhinitis DSS during the last 8 weeks of treatment. Rhinitis DSS was the sum of 4 nasal symptoms (runny nose, blocked nose, sneezing, itchy nose), each rated on a scale from 0 (no symptoms) to 3 (severe symptoms), with a range of 0 to 12 points.
 - Average rhinitis DMS during the last 8 weeks of treatment. Rhinitis DMS consists of rhinitis symptomatic medication scores with a range of 0 to 12.
 - Average Total Combined Score (TCS) during the last 8 weeks of treatment. TCS was the sum of rhinoconjunctivitis DSS and the rhinoconjunctivitis DMS with a range of 0 to 38 points.
 - Average allergic rhinitis/rhinoconjunctivitis Visual Analogue Scale (VAS) score during the last 8 weeks of treatment. VAS was a weekly self-reported assessment on symptom severity with a range of 0 (“no symptoms”) to 100 (“severe symptoms”).
- Safety Endpoints
 - Tier 1 Safety Endpoints:
 - Proportion of subjects reporting pre-specified local application site reactions, including lip swelling/edema; mouth edema; palatal edema; swollen tongue/edema; oropharyngeal swelling/edema; pharyngeal edema/throat tightness; oral pruritus; throat irritation; tongue pruritus; ear pruritus;
 - Proportion of subjects reporting anaphylaxis and/or systemic allergic reactions;
 - Proportion of subjects with events treated with epinephrine.
 - Tier 2 Safety Endpoints:
 - Proportion of subjects with any AE, any SAE, any drug-related AE, serious and drug-related AE, and specific AEs or system organ classes (SOCs) (incidence $\geq 1\%$ subjects in 1 of the treatment groups);
 - Proportion of subjects who discontinued due to an AE;
 - Proportion of subjects with upper respiratory viral infections.
 - Tier 3 Safety Endpoints: Tier 3 safety endpoints included laboratory tests, vital signs, and AEs that were not classified as Tier 1 or Tier 2 safety endpoints.

6.1.9 Statistical Considerations & Statistical Analysis Plan

- Hypotheses and Statistical Methods for Efficacy Endpoints

Primary Efficacy Hypothesis: Administration of MK-8237 sublingual tablet (12 DU), compared with placebo, results in significant reduction in the average TCRS during the last 8 weeks of treatment.

Secondary Efficacy Hypothesis: Administration of MK-8237 12 DU, compared with placebo, results in significant score reduction on the following endpoints:

- Average rhinitis DSS during the last 8 weeks of treatment;
- Average rhinitis DMS during the last 8 weeks of treatment;
- Average TCS during the last 8 weeks of treatment;
- Average allergic rhinitis/rhinoconjunctivitis symptoms assessed by VAS during the last 8 weeks of treatment.

The primary efficacy endpoint of the average TCRS during the last 8 weeks of treatment was analyzed based on the full analysis set (FAS) population. Wilcoxon Rank Sum test was used to test the null hypothesis that there is no treatment difference in the efficacy endpoints between the active treatment group and placebo group, with the p-value reported. The Hodges-Lehmann estimate of treatment difference and the corresponding 2-sided 95% CI were also presented. Also, the difference in medians between the treatment groups relative to the median of the placebo group was presented as a percentage relative treatment difference, with the corresponding 95% CI derived using the bootstrap method. A sensitivity analysis based on the PP population was also conducted.

Four sensitivity analyses using the square root transformed TCRS as response variable were performed for the primary endpoint based on the FAS population, including:

- Analysis of covariance (ANCOVA) model with treatment, baseline asthma status (yes, no), age group (<18, ≥18), and region (Northern US, Southern US, and Canada) as fixed effects and baseline endpoint value as a covariate.
- Analysis using the same ANCOVA model with multiple imputations. Missing average TCRS during the last 8 weeks of treatment from both groups were imputed using the sample distribution of TCRS observed from the placebo group.
- Analysis using the same ANCOVA model with last observation carried forward (LOCF) as the imputation method. The LOCF approach was implemented with the following restriction: only the TCRS scores recorded during the last 8 weeks of treatment were used to impute the missing scores for this period.
- Analysis using the Longitudinal Data Analysis (LDA) model. Because only a small proportion of subjects had efficacy assessment longer than 8 weeks, the response vector for the LDA model consisted only of daily TCRS scores during the first 56 days of the efficacy assessment (starting from the day of Visit 10) to avoid model convergence issues and to improve precision of the estimates. The analysis model included treatment, day, treatment-by-day interaction, baseline asthma status, age group, and region as fixed effects, subject as random effect, and included baseline endpoint value as a covariate. The Toeplitz covariance matrix was used to model the correlation among repeated measurements.

For the key secondary efficacy analyses, the rhinitis DSS, TCS, and VAS were analyzed based on the FAS population in a similar fashion as described for the primary efficacy endpoint using the non-parametric approach. Because the majority of subjects did not use rescue medications, 337 (59.5%) and 336 (54.2%) subjects in the MK-8237 12 DU and placebo treatment groups, respectively, had a rhinitis DMS equal to zero. Hence, the zero-inflated log-normal model was used, as pre-specified in the protocol, to analyze the average rhinitis DMS during the last 8 weeks of treatment.

- **Sample size determination**

Subjects were randomized in a 1:1 ratio to either MK-8237 or placebo. Assuming a 17% drop out rate, a total of 750 subjects per treatment group were planned to be randomized, so approximately 620 subjects per treatment group would be eligible for the evaluation period, i.e., the last 8 weeks of treatment. The applicant indicated that, with approximately 620 subjects per group, the study would be able to detect a difference of -1.45 in the primary endpoint between MK-8237 and placebo with > 99% power at the significance level of 2-sided $\alpha=0.05$, assuming a common standard deviation of 4.38. In addition, the applicant indicated that there would be approximately 90% power to detect a treatment difference (MK-8237 – placebo) of at least 10% of placebo effect for the primary endpoint at significance level of 2-sided $\alpha=0.05$, i.e., the upper bound of the 95% CI for $100*(MK-8237 - placebo)/placebo$ being less than -10%, assuming a mean score of 6.40 for subjects in the placebo group. The assumptions for treatment difference, placebo mean, standard deviation, and the dropout rate were based on the two North American grass studies.

- **Multiplicity adjustment**

The primary and key secondary endpoints were tested in a stepwise procedure, where statistical conclusions were made on the key secondary efficacy endpoints only if statistical significance was demonstrated in the primary efficacy endpoint. Multiplicity within the 4 key secondary endpoints was also controlled by the sequential testing procedure in the following order: rhinitis DSS, rhinitis DMS, TCS, and VAS. Statistical conclusion of any key secondary endpoint could be made only if statistical significance was demonstrated in all preceding key secondary endpoints.

- **Blinding**

A double-blind/masking technique was used: MK-8237 12 DU and matching image placebo were packaged identically so that treatment blind/masking was maintained. The subject, Investigator, and Sponsor personnel or delegate(s) who were involved in the treatment or clinical evaluation of the subjects were unaware of the treatment group assignments.

- **Randomization**

Randomization was performed centrally using an Interactive Voice Response System (IVRS)/Interactive Web Response System (IWRS). Subjects were assigned randomized treatment in a 1:1 ratio to MK-8237 12 DU or placebo. Randomization was stratified according to asthma status (yes/no) and age (<18 years [i.e., 12-17 years] and ≥ 18 years).

- Definitions of analysis populations
 - The Full Analysis Set (FAS) population served as the primary population for the analysis of efficacy data in this trial. The FAS population considered all randomized subjects who received at least 1 dose of study drug. Subjects were analyzed according to the treatment group to which they were randomized.
 - The Per Protocol (PP) population excluded subjects due to important deviations from the protocol that may have substantially affected the results of the primary and key secondary efficacy endpoints.
 - Safety Analysis Population: The All-Subjects-as-Treated (ASaT) population was used for the analysis of safety data in this trial. The ASaT population consisted of all randomized subjects who received at least 1 dose of study drug. Subjects were included in the treatment group corresponding to the study drug they actually received during the trial.

- Missing data handling

The primary analysis methods were based on observed data only; subjects with no data on a given endpoint during the efficacy assessment period were not evaluable for that specific endpoint under this approach. Sensitivity analyses were implemented to address different aspects of the missing data issues for the primary efficacy endpoint, including multiple imputation, last observation carried forward (LOCF), and the LDA model. The multiple imputation approach focused on the missing data due to early dropout before the efficacy assessment period. All subjects in the FAS population were evaluable in the multiple imputation analysis if the subject had non-missing baseline value. Within each imputation, missing endpoint values for subjects from both treatment groups were imputed by random samples drawn from the distribution of average TCRS from placebo treated subjects. A total of 50 imputations were performed and Rubin's strategy was used to combine the multiple analysis results generated from the same ANCOVA model. The LOCF and LDA approach focused on the missing data during the efficacy assessment period by using direct and model-based imputation, respectively.

6.1.10 Study Population and Disposition

6.1.10.1 Populations Enrolled/Analyzed

Of the 4497 subjects screened for inclusion, 3015 subjects were excluded and 1482 were randomized. A total of 1482 subjects were randomized 1:1 to receive either MK-8237 (741 subjects) or placebo (741 subjects). One subject randomized to the MK-8237 group never received study drug. Therefore, 1481 subjects were randomized and treated (MK-8237, 740 subjects; placebo, 741 subjects).

There were 1481 subjects in the FAS (MK-8237, 740 subjects; placebo, 741 subjects). Excluding subjects from the FAS population with protocol violations, there were 1296 subjects included in the PP population (MK-8237, 651 subjects; placebo, 645 subjects). Of the 1482 randomized subjects, 1481 received at least 1 dose of study drug (1 subject randomized to receive MK-8237 was not treated). Three subjects randomized to receive placebo received the incorrect treatment during the trial; these 3 subjects were analyzed as MK-8237-treated subjects in the ASaT population. Consequently, 1481 subjects were included in the ASaT population (MK-8237, 743 subjects; placebo, 738 subjects).

6.1.10.1.1 Demographics

Overall, the treatment groups were similar with respect to baseline demographics (Table 2).

Table 2 Subject demographics (all subjects randomized)

	MK-8237 12 DU n (%)	Placebo n (%)	Total n (%)
Subjects in population	741	741	1482
Gender			
Male	296 (39.9)	311(42.0)	607 (41.0)
Female	445 (60.1)	430 (58.0)	875 (59.0)
Age (Years)			
12 to <18	94 (12.7)	95 (12.8)	189 (12.8)
18 to <50	527 (71.1)	527 (71.1)	1054 (71.1)
50 to <65	109 (14.7)	105 (14.2)	214 (14.4)
>=65	11 (1.5)	14 (1.9)	25 (1.7)
Mean	34.9	35.2	35.1
Range	12 to 77	12 to 85	12 to 85
Race			
White	567 (76.5)	564 (76.1)	1131 (76.3)
Asian	48 (6.5)	51 (6.9)	99 (6.7)
Black or African American	80 (10.8)	75 (10.1)	155 (10.5)
Multi-Racial	39 (5.3)	46 (6.2)	85 (5.7)
American Indian or Alaska Natives	6 (0.8)	4 (0.5)	10 (0.7)
Unknown	1 (0.1)	1 (0.1)	2 (0.1)
Ethnicity			
Hispanic/Latino	68 (9.2)	63 (8.5)	131 (8.8)
Non-Hispanic/Latino	662 (89.3)	657 (88.7)	1319 (89.0)
Unknown	11 (1.5)	21 (2.8)	32 (2.2)

Source: adapted from Table 10-5 in the P001 Clinical Study Report

6.1.10.1.2 Medical/Behavioral Characterization of the Enrolled Population

N/A

6.1.10.1.3 Subject Disposition

The majority of randomized subjects (79.2%) completed the double-blind treatment period. The primary reasons for trial discontinuation were withdrawal by the subject, discontinuation due to AE, and loss to follow-up. The percentage of subjects who discontinued from the trial was higher in the MK-8237 group (24.2%) than in the placebo group (17.3%). The differential drop out between the MK-8237 and the placebo groups was primarily driven by the higher rate of discontinuations due to AEs in the MK-8237 group (9.9% vs. 2.4% in the placebo group).

6.1.11 Efficacy Analyses

All primary and secondary efficacy analyses were performed based on available data during approximately the last 8 weeks of treatment for the FAS population. For subjects eligible for analysis within the FAS population, a total of 1186 subjects had diary-based data between Visit 10 (start of the efficacy assessment) and Visit 11 (end of the efficacy assessment). Within the PP population used for sensitivity analysis, a total of 1034 subjects had data between Visit 10 and Visit 11 (MK-8237, 498 subjects; placebo, 536 subjects).

6.1.11.1 Analyses of Primary Endpoint(s)

The average TCRS during the last 8 weeks of treatment was lower in the MK-8237 group than in the placebo group (Table 3). For the null hypothesis that there is no difference in TCRS between the active treatment and placebo group, the Wilcoxon Rank Sum test showed that between-treatment difference was statistically significant ($p < 0.001$). The treatment difference relative to placebo (%) between the groups was -17.2% (95% CI, -25.0%, -9.7%). The pre-defined clinical relevance criterion is that the upper bound of the 95% CI for treatment difference relative to placebo should be less than -10%. Here the upper bound of the 95% CI of the treatment difference relative to placebo was -9.7%, slightly missing the clinical relevance threshold of -10%.

Table 3 Analysis of average Total Combined Rhinitis Score (TCRS) during the last 8 weeks of treatment using nonparametric analysis (Full Analysis Set)

Treatment	N	Baseline Mean (SD)	Average score during the last 8 weeks of treatment Mean (SD)	Average score during the last 8 weeks of treatment Median (lower quartile, upper quartile)	Treatment difference MK-8237 12 DU vs. Placebo (95% CI) ¹ p-value	Treatment difference relative to placebo (%) MK-8237 12 DU vs. Placebo (95% CI) ²
MK-8237 12 DU	566	7.94 (1.72)	4.67 (3.55)	4.10 (2.00, 6.40)	-0.80 (-1.20, -0.40) p-value < 0.001	-17.2 (-25.0, -9.7)
Placebo	620	7.94 (1.76)	5.49 (3.82)	4.95 (2.70, 7.55)		

Note: N- number of subjects included in the analysis; SD- standard deviation; CI- confidence interval.

1. The 95% CI for median difference was based on the Hodges-Lehmann estimator.
 2. Treatment difference relative to Placebo based on median was calculated by $(MK8237 - Placebo)/Placebo * 100\%$; CI was calculated by the bootstrap method using 10,000 iterations
- Source: adapted from Table 11-2 in the P001 Clinical Study Report

The applicant performed sensitivity analysis based on the Per-Protocol (PP) population. In the PP population, the average TCRS during the last 8 weeks of treatment was lower in the MK-8237 group than in the placebo group (Table 4). The treatment difference relative to placebo with 95% CI was -16.7% (-25.5%, -8.5%).

Table 4 Analysis of average Total Combined Rhinitis Score (TCRS) during the last 8 weeks of treatment using nonparametric analysis (Per-Protocol Set)

Treatment	N	Baseline Mean (SD)	Average score during the last 8 weeks of treatment Mean (SD)	Average score during the last 8 weeks of treatment Median (lower quartile, upper quartile)	Treatment difference MK-8237 12 DU vs. Placebo (95% CI) ¹ p-value	Treatment difference relative to placebo (%) MK-8237 12 DU vs. Placebo (95% CI) ²
MK-8237 12 DU	498	7.94 (1.74)	4.56 (3.53)	4.00 (2.00, 6.30)	-0.70 (-1.10, -0.30) p-value < 0.001	-16.7 (-25.5, -8.5)
Placebo	536	7.95 (1.74)	5.29 (3.68)	4.80 (2.50, 7.30)		

Note: N- number of subjects included in the analysis; SD- standard deviation; CI- confidence interval.

1. The 95% CI for median difference was based on the Hodges-Lehmann estimator.
2. Treatment difference relative to Placebo based on median was calculated by (MK8237 – Placebo)/Placebo*100%; CI was calculated by the bootstrap method using 10,000 iterations

Source: adapted from Table 11-3 in the P001 Clinical Study Report

The applicant also conducted four additional sensitivity analyses based on the FAS population: the ANCOVA model, the ANCOVA model with multiple imputation method and with LOCF, and the longitudinal data analysis (LDA) model. The ANCOVA model (with observed data only) provided an alternative approach to the primary non-parametric approach to analyze the data, and the other 3 sensitivity analyses, ANCOVA model with multiple imputation and with LOCF imputation as well as the LDA method, assessed the impact of missing data on the primary analysis result (Table 5). The sensitivity analyses showed statistical significance (p-values ranging from < 0.001 to 0.013) for testing the null hypothesis that there is no difference between the active treatment and placebo group. The upper bounds of the 95% CIs of % treatment difference relative to placebo ranged from -6.5% to -8.8%.

Table 5 Sensitivity Analysis of Average Total Combined Rhinitis Score (Full Analysis Set)

Sensitivity analysis	Treatment	N	Baseline Mean (SD)	Average TCRS during the last 8 weeks of treatment Mean (SD)	Treatment difference (95% CI) p-Value	Treatment difference relative to placebo (%) (95% CI) ¹
ANCOVA Model	MK-8237 12 DU	566	7.94 (1.72)	4.67 (3.55)	-0.75 (-1.18, -0.32) p-Value < 0.001	-17.5 (-25.2, -8.8)
	Placebo	620	7.94 (1.76)	5.49 (3.82)		
LDA Model	MK-8237 12 DU	566	7.94 (1.72)	4.66 (4.19)	-0.71 (-1.22, -0.20) p-Value < 0.001	-18.4 (-31.0, -6.5)
	Placebo	620	7.94 (1.76)	5.42 (4.43)		
ANCOVA Multiple Imputation	MK-8237 12 DU	740	7.95 (1.75)	4.67 (3.55)	-0.54 (-0.96, -0.11) p-Value = 0.013	-12.3 (-17.8, -6.9)

Sensitivity analysis	Treatment	N	Baseline Mean (SD)	Average TCRS during the last 8 weeks of treatment Mean (SD)	Treatment difference (95% CI) p-Value	Treatment difference relative to placebo (%) (95% CI) ¹
	Placebo	741	7.92 (1.75)	5.49 (3.82)		
ANCOVA LOCF	MK-8237 12 DU	566	7.94 (1.72)	4.67 (3.56)	-0.74 (-1.17, -0.31) p-Value < 0.001	-17.3 (-25.2, -8.5)
	Placebo	620	7.94 (1.76)	5.48 (3.81)		

Note: ANCOVA - Analysis of covariance; CI - confidence interval; LDA = Longitudinal Data Analysis; LOCF = Last observation carried forward; N = number of subjects included in the analysis; SD = standard deviation; TCRS = Total Combined Rhinitis Score.

1. Treatment difference relative to placebo based on LS means was calculated by (MK-8237 – Placebo)/placebo*100%; confidence interval was calculated by the bootstrap method using 10,000 iterations for ANCOVA model and ANCOVA model with LOCF or using Fieller’s Theorem for LDA and ANCOVA model with multiple imputation.

Reviewer Comments:

During the End of Phase II meeting, the applicant and CBER agreed that the pivotal North American study (Study P001) would be designed to demonstrate the upper bound of 95% CI of relative treatment effect being ≤ -10%. The study results showed that the upper bound of the 95% CI of treatment effect relative to placebo, calculated using the bootstrap method, was -9.7% which slightly missed the predefined threshold of -10%. The reviewer performed 500 runs of bootstrap analysis on FAS to assess the potential impact of fluctuation of bootstrapping process on the estimation of the upper bound of the 95% CI of relative treatment effect (Figure 2). The analysis showed that the estimate of the upper bound ranged from -8.70% to -10.20%. The median was -9.68%, which is very close to the applicant’s estimate.

The applicant’s sensitivity analysis based on the PP population showed that treatment difference relative to placebo was -16.7% (95% CI, -25.5%, -8.5%), using the bootstrap method. Other sensitivity analyses on the FAS reported that the upper bound of the 95% CI of treatment difference relative to placebo ranged from -6.5% to -8.8%. The reviewer defers to the medical officer for further consideration of totality of evidence.

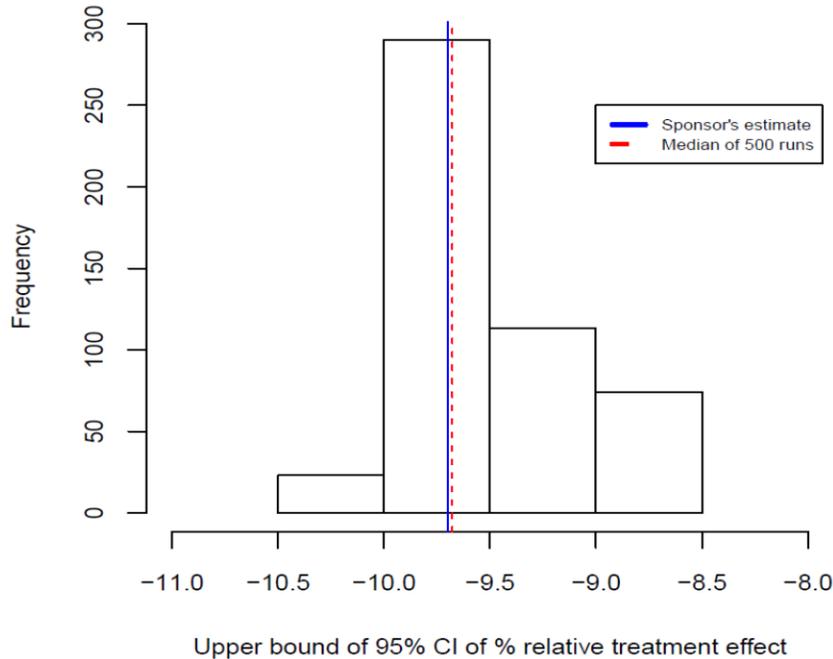


Figure 2 Histogram of the upper bound of the 95% CI of the percent treatment effect relative to placebo, generated from 500 runs of analysis on FAS using the bootstrap method. The dotted line represents the median of the upper bound of the 95% CI of percent relative treatment effect; the solid line represents the applicant's estimate of the upper bound.

6.1.11.2 Analyses of Secondary Endpoints

The analysis of average rhinitis DSS for the FAS population during the last 8 weeks of treatment, using the nonparametric approach, showed a statistically significant and lower average rhinitis DSS in the MK-8237 group as compared with the placebo group during the last 8 weeks of treatment (Table 6).

Because rescue medications were not used by the majority of subjects, 337 (59.5%) and 336 (54.2%) subjects in the MK-8237 12 DU and placebo treatment groups, respectively, had a rhinitis DMS equal to zero. The applicant used the zero-inflated log-normal model, as pre-specified in the protocol, to analyze the average rhinitis DMS for the FAS population during the last 8 weeks of treatment. The average rhinitis DMS was numerically lower in the MK-8237 group than in the placebo group. However, the treatment difference was not statistically significant compared to placebo (Table 6).

The analysis of average Total Combined Rhinoconjunctivitis Score (TCS) for the FAS population during the last 8 weeks of treatment, using the nonparametric approach, showed that the average TCS was lower in the MK-8237 group than in the placebo group (Table 6). The applicant indicated that, although the nominal p-value for this difference was <0.001, the result cannot be considered as confirmatory due to the pre-specified multiplicity control strategy.

The nonparametric analysis of the average VAS for the FAS population during the last 8 weeks of treatment showed subjects from the MK-8237 group reported fewer symptoms on the VAS compared to the placebo group (Table 6). Again, the applicant indicated that, despite the nominal p-value for this difference being <0.001, the result cannot be considered confirmatory due to the pre-specified multiplicity control strategy for this trial.

Table 6 Rhinitis Daily Symptom Score (DSS), Rhinitis Daily Medication Score (DMS), Total Combined Rhinoconjunctivitis Score (TCS), and Average Visual Analogue Scale (VAS) during the last 8 weeks of treatment

	Treatment	N	Mean (SD)	Median	Treatment Difference (%) Relative to Placebo (95% CI)	Treatment Difference (95% CI) p-value
RDSS	MK-8237 12 DU	566	3.83 (2.64)	3.55	-15.5 (-24.4,-7.3)	-0.60 (-1.00, -0.30) p<0.001
	Placebo	620	4.46 (2.80)	4.20		
RDMS	MK-8237 12 DU	566	0.84 (1.817)	-	-18.4 (-41, 4.3)	-0.15 (-0.35, 0.05) p=0.154
	Placebo	620	1.03 (2.074)	-		
TCS	MK-8237 12 DU	566	6.40 (5.16)	5.50	-16.7 (-24.6,-4.0)	-1.10 (-1.70, -0.60) p<0.001
	Placebo	620	7.62 (5.48)	6.60		
Average VAS	MK-8237 12 DU	540	42.29 (23.57)	41.40	-16.0 (-22.7, -8.3)	-6.10 (-9.10, -3.10) p<0.001
	Placebo	585	47.96 (23.66)	49.30		

Source: adapted from Table 2-2 in the P001 Clinical Study Report

6.1.11.3 Subpopulation Analyses

The applicant performed subgroup analysis on the average TCRS during the last 8 weeks of treatment, including age, gender, race, asthma status, inhaled corticosteroids (ICS) use, allergen sensitivity, and geographic location. Some differences were observed between the subgroups of age, gender, and baseline asthma status (Figure 3). Since the trial was not powered to show efficacy in the specific subgroups, the observed differences might not be interpretable or conclusive.

Per CBER’s request, the applicant performed a post-hoc analysis on the average TCRS during the last 8 weeks of treatment for age subgroups (12 to < 18 years old, 18 to 65 years old, and > 65 years old). As shown in Table 7, the percent relative treatment difference of subjects 18 to 65 years old was -16.0% with 95% CI (-23.2%, -5.3%). Since this is a post-hoc analysis with limited statistical rigor, the results are descriptive in nature. The point estimate of the relative treatment difference of subjects 12 to < 18 years old was -22.4%, as compared with -16.0% for subjects 18 to 65 years old. Because of

limited number of subjects in the age subgroups, the observed differences might not be interpretable or conclusive.

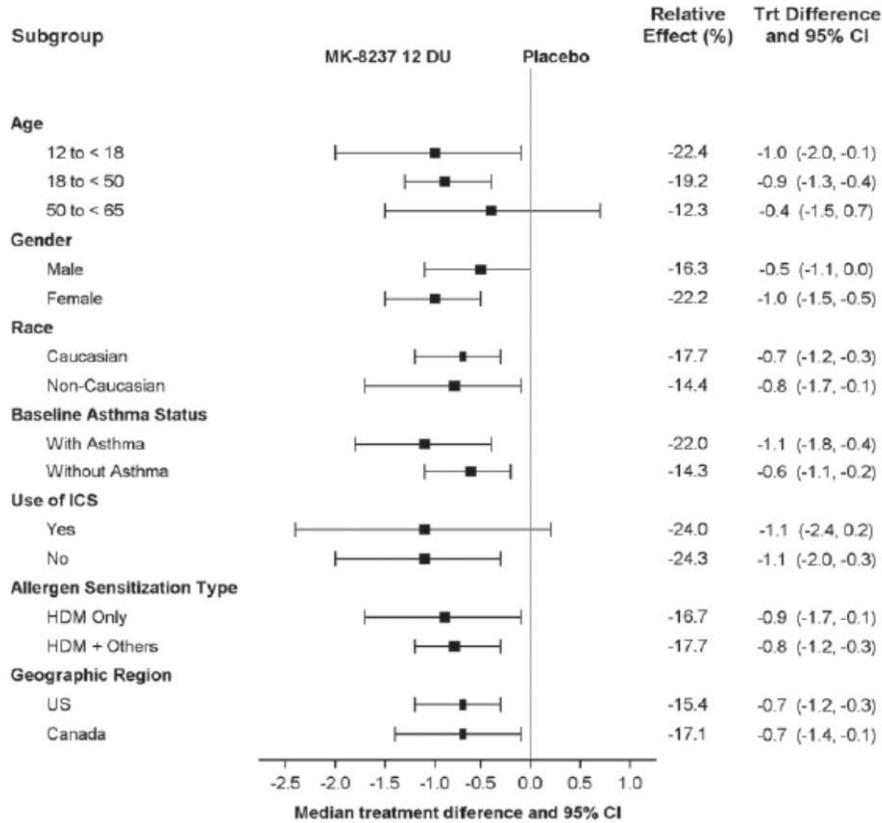


Figure 3 Average TCRS during the last 8 weeks of treatment by demographic Subgroups (Full Analysis Set)

Source: adapted from Figure 11-3 in the P001 Clinical Study Report

Table 7 Post-hoc analysis of average Total Combined Rhinitis Score (Nonparametric Analysis) during the last 8 weeks of treatment by age subgroup (Full Analysis Set)

Age	Treatment	N	Baseline Mean (SD)	Last 8 weeks Mean (SD)	Last 8 weeks Median (LQ, UQ)	Last 8 weeks Hodges-Lehmann Estimate (95% CI)	Last 8 weeks Relative treatment difference (%) (95% CI)
12 to <18	MK-8237 12 DU	76	8.06 (1.60)	3.64 (2.69)	3.3 (1.35, 5.10)	-1.00 (-2.00, -0.10)	-22.4 (-42.6, 8.1)
	Placebo	84	7.82 (1.64)	4.83 (3.42)	4.3 (2.50, 6.55)		
18 to 65	MK-8237 12 DU	482	7.93 (1.74)	4.77 (3.61)	4.20 (2.10, 6.60)	-0.80 (-1.20, -0.30)	-16.0 (-23.2, -5.3)
	Placebo	526	7.96 (1.78)	5.60 (3.90)	5.00 (2.70, 7.70)		
>65	MK-8237 12 DU	8	7.09 (1.74)	8.40 (3.92)	6.50 (5.50, 11.85)	1.80 (-0.40, 7.30)	28.7*
	Placebo	10	7.86 (1.69)	5.53 (1.89)	5.05 (4.40, 6.70)		

Note: N- number of subjects, SD- standard deviation, LQ- Lower Quartile, UQ- Upper Quartile

* No confidence interval on treatment difference relative to Placebo was produced due to insufficient number of subjects within the category.

Source: Table 1 in the applicant’s response to FDA (Amendment 10)

6.1.11.4 Dropouts and/or Discontinuations

Please refer to section 6.1.10.1.3.

6.1.11.5 Exploratory and Post Hoc Analyses

N/A

6.1.12 Safety Analyses

Safety analyses were performed based on data collected during the treatment period for the ASaT population. A total of 1481 treated subjects (MK-8237, 743 subjects; placebo, 738 subjects) were included in the safety analyses. There were 3 cross-treated subjects originally randomized to placebo but who received MK-8237 for a period of time during the trial.

The majority of subjects (82.0%) reported at least one adverse event (AE). The majority (62.5%) of subjects had AEs determined by the Investigator to be related to study drug. The most frequently reported AEs were throat irritation, oral pruritus, and ear pruritus. There were more subjects with AEs, drug-related AEs, discontinuations due to AEs, and discontinuations due to drug-related AEs in the MK-8237 group than in the placebo group (Table 8). The greatest difference between the groups was observed in the percentage of subjects with drug-related AEs. The percentages of subjects with serious adverse events (SAEs) were generally similar between the two treatment groups.

Table 8 Adverse Event Summary (All-Subjects-as-Treated)

	MK-8237 12 DU n (%)	Placebo n (%)	Total n (%)
Subjects in population	743	738	1481
With one or more adverse events	676 (91.0%)	539 (73.0%)	1215 (82.0%)
With drug-related adverse events	624 (84.0%)	301 (40.8%)	925 (62.5%)
With serious adverse events	11 (1.5%)	7 (0.9%)	18 (1.2%)
With serious drug-related adverse events	2 (0.3%)	0	2 (0.1%)
Discontinued due to an adverse event	73 (9.8%)	19 (2.6%)	92 (6.2%)
Discontinued due to a drug-related adverse event	62 (8.3%)	6 (0.8%)	68 (4.6%)
Discontinued due to a serious adverse event	3 (0.4%)	1 (0.1%)	4 (0.3%)
Discontinued due to a serious drug-related adverse event	0	0	0

Source: adapted from Table 2-3 in the P001 Clinical Study Report

6.1.12.1 Methods

The applicant conducted the safety analysis based on a tiered approach. Safety parameters or adverse experiences of special interest that were identified a priori constitute Tier 1 safety endpoints that were subject to inferential testing. Tier 2 parameters were assessed via point estimates with 95% CIs for between-group differences; only point estimates by treatment group were provided for Tier 3 safety parameters.

6.1.12.3 Deaths

No deaths were reported during the trial.

6.1.12.4 Nonfatal Serious Adverse Events

SAEs were infrequent and reported in a similar percentage of subjects in the MK-8237 and placebo groups. There were no drug-related SAEs in the MK-8237 group reported as life-threatening, resulting in persistent or significant disability/incapacity, resulting in or prolonging an existing inpatient hospitalization, a congenital anomaly/birth defect, or considered by the Investigator to be another important medical event. However, two subjects had drug-related AEs (oral pruritus and throat irritation, and oral pain, respectively) associated with an overdose and were reported as SAEs per protocol definition, but did not meet the ICH criteria for seriousness.

6.1.12.5 Adverse Events of Special Interest (AESI)

The most frequently reported local application site reactions were throat irritation, oral pruritus, and ear pruritus. All local application site reactions were more frequently reported in the MK-8237 group than in the placebo group. Severe pre-specified local application site reactions in the ASaT population occurred in 8 (0.5%) subjects; 7 of the 8 subjects had drug-related local application site reactions. All 8 subjects received MK-8237.

A total of 823 (55.6%) subjects had a pre-specified local application site reaction considered to be drug-related by the Investigator. Drug-related local application site reactions were more frequently reported for subjects in the MK-8237 group (80.1%) than for subjects in the placebo group (30.9%). The most frequently reported drug-related pre-specified local application site reactions included throat irritation, oral pruritus, and ear pruritus.

Anaphylactic and/or systemic allergic reactions were reported by a total of 6 subjects (3 subjects in the MK-8237 group and 3 subjects in the placebo group). One of the 3 subjects in the MK-8237 treatment group had an MK-8237-related systemic allergic reaction, which occurred on Day 1. Also, a total of 7 subjects experienced adverse events that were treated with epinephrine (4 in the MK-8237 group and 3 in the placebo group).

6.1.12.6 Clinical Test Results

N/A

6.1.12.7 Dropouts and/or Discontinuations

Please refer to section 6.1.10.1.3.

6.2 Study P015 (MT-06): A One-year Trial Evaluating the Efficacy and Safety of the (b) (4) House Dust Mite Allergy Immunotherapy Tablet in Adult Subjects with House Dust Mite Allergic Rhinitis

This Phase III study was conducted by the applicant's alliance partner ALK from 2011 to 2013, as part of its European clinical development program. ALK communicated with the

European Union (EU) national regulatory agencies on the study design and conduct. The U.S. FDA was not consulted regarding this study.

6.2.1 Objectives

Primary objective:

To evaluate the efficacy of the HDM allergy immunotherapy tablet given once daily compared to placebo in the treatment of HDM allergic rhinitis. The primary endpoint was the average TCRS during the last 8 weeks of treatment.

Secondary objectives:

- Key secondary objectives: To determine the effect of the HDM allergy immunotherapy tablet on average allergic rhinitis DSS, average allergic rhinitis DMS, average overall RQLQ score, and average total combined rhinoconjunctivitis score in the efficacy evaluation period.
- Additional secondary objectives:
 - To evaluate the safety and tolerability of the HDM allergy immunotherapy tablet.
 - To determine the effect of the HDM allergy immunotherapy tablet on individual rhinitis and conjunctivitis symptoms, medication use, onset of action, treatment satisfaction, and generic and disease-specific quality of life. Immunological parameters were investigated for a limited subset of the subjects.

6.2.2 Design Overview

This study was a randomized, parallel-group, double-blind, placebo-controlled, multinational, multisite trial in Europe. Approximately 900 subjects were planned to be randomized 1:1:1 to receive treatment with the HDM allergy immunotherapy tablet in doses of 6 DU or 12 DU or with placebo. Subjects received treatment for approximately 12 months.

Between screening and randomization, the subjects were asked to fill in an electronic diary daily for 15 days, to capture information on rhinitis symptoms, use of symptomatic medications, and impact of rhinitis on daily life. At randomization and throughout the rest of the trial, eligible subjects received investigational medicinal product (IMP) for daily administration and were provided with nasal steroid, oral antihistamine, and antihistamine eye drops to be used as needed. The subjects were asked about rhinitis symptoms, health and quality of life throughout the trial. They filled in an electronic diary during 1 week following each visit and daily during the last 8 weeks of treatment.

6.2.3 Population

Subjects 18-65 years of age, with a clinical history consistent with moderate to severe persistent HDM allergic rhinitis (with or without asthma) for at least one year prior to trial entry, with allergic rhinitis symptoms despite having received symptomatic treatment, and meeting all of the inclusion criteria and none of the exclusion criteria.

6.2.4 Study Treatments or Agents Mandated by the Protocol

Each subject was randomly assigned to receive active treatment (HDM allergy immunotherapy tablet) or placebo. The HDM allergy immunotherapy tablet is a 1:1 mixture of allergen extracts derived from 2 species of cultivated HDM (*D. pteronyssinus* and *D. farinae*).

6.2.6 Sites and Centers

A total of 100 trial sites in 12 countries, i.e., Austria, Bosnia and Herzegovina, Croatia, Czech, Denmark, France, Germany, Latvia, Poland, Romania, Serbia, and Ukraine.

6.2.7 Surveillance/Monitoring

N/A

6.2.8 Endpoints and Criteria for Study Success

Primary endpoints:

The average TCRS during the efficacy evaluation period

Key Secondary Endpoints:

- The average total allergic rhinitis DSS during the efficacy evaluation period.
- The average total allergic rhinitis DMS during the efficacy evaluation period.
- The average overall Rhinitis quality of life questionnaire RQLQ(S) score during the efficacy evaluation period.
- The average total combined allergic rhinoconjunctivitis score during the efficacy evaluation period.

Safety Endpoints:

Safety assessments included adverse events (AEs), AE discontinuations, serious adverse events (SAEs), vital signs, safety laboratory assessments, FEV1, and physical examinations.

6.2.9 Statistical Considerations & Statistical Analysis Plan

• Hypotheses and Statistical Methods for Efficacy Endpoints

The primary efficacy analysis was based on a linear mixed effect (LME) model and performed on the FAS using a multiple imputation strategy of Rubin for missing data. The response variable in the LME was the square root of the TCRS, and covariates included the average rhinitis DSS at baseline and country. The primary outcome was the pairwise comparison among all 3 treatment groups using a t-test in the LME model. The resulting p-values were reported together with the associated difference in (back-transformed) adjusted means with 95% confidence intervals.

Fisher's least significant difference (LSD) procedure was used to control for multiplicity in the primary efficacy analysis. Using an F-test in the LME model, the first hypothesis to be tested was the global hypothesis of no difference in means among the 3 groups: placebo, 6 DU, and 12 DU. If, and only if, this global hypothesis was rejected ($p < 0.05$),

all pairwise comparisons between treatment groups were performed (12 DU versus placebo, 6DU versus placebo, and 12 DU versus 6 DU).

Additional analyses of the primary endpoint included analyses using the same LME model on all non-missing observations of the FAS, on the PP, and on the FAS with imputation of missing data using the method of last observation carried forward (LOCF).

The applicant indicated that the minimal clinically relevant difference between active and placebo in the TCRS was predefined to be 20%, which corresponds to an absolute difference of 1 in the TCRS.

***Reviewer Comments:** This clinical relevance criterion was defined by ALK, the applicant's partner, for its European clinical development program. For U.S. registration, the applicant and CBER agreed that the clinical relevance criterion should be the upper bound of the 95% confidence interval for the between-treatment difference in the TCRS relative to placebo being $\leq -10\%$.*

Multiplicity for the key secondary endpoints was controlled for by hierarchical testing in the following order:

1. The average total allergic rhinitis DSS during the efficacy evaluation period.
2. The average total allergic rhinitis DMS during the efficacy evaluation period.
3. The average overall RQLQ(S) score during the efficacy evaluation period.
4. The average total combined allergic rhinoconjunctivitis score during the efficacy evaluation period.

The 4 key secondary hypotheses were first to be tested for the 12 DU group and then, if all were statistically significant, for the 6 DU group. The key secondary efficacy analyses were based on LME models and performed on the FAS, using a multiple imputation strategy for missing data for the key secondary endpoints 1 and 2 and on non-missing observations of FAS for all of the key secondary endpoints.

- **Sample size determination**

The primary endpoint was the average TCRS during the efficacy evaluation period. Multiplicity due to multiple comparisons was controlled by the following pre-specified order of hypotheses to be tested:

- The first hypothesis tested was the global hypothesis that all 3 groups (placebo, 6 DU, 12 DU) are equal.
- If, and only if, this global hypothesis could be rejected at the 5% level, then each and all 3 pairwise comparisons could be tested at the 5% level.

Therefore, the sample size was based on the power calculation for this 2-step procedure known as Fisher's least significant difference. First, the power to reject the null hypothesis, that all 3 groups were equal, was considered. Second, the power to additionally reject each of the 2 null hypotheses of main interest, that active treatment equals placebo, was considered.

The power calculation was based on the following assumptions:

- Analysis was performed based on multiple imputations, and subjects who did not contribute any diary data during the last 8 weeks of treatment would be imputed as sampled from the observed placebo distribution of the TCRS.
- Equal proportions of 10% were imputed in each treatment group.
- The mean TCRS value for placebo was assumed to be 4.10. A relative difference was assumed to be 20% and 25% for 6DU and 12DU, respectively, which corresponded to an absolute difference of 0.82 and 1.03, respectively.
- A pooled SD corresponding to a CV of 82% based on information from ALK MT-02 trial.
- The global hypothesis was tested with an F-test on 2 degrees of freedom at 5% level of significance.
- The pairwise hypotheses were tested with a 2-sided t-test at 5% level of significance.

Based on these assumptions, the applicant estimated that 300 randomized subjects per treatment group (i.e., a total of 900 subjects) would provide about 90% power to reject the global hypothesis of no difference between any of the treatment groups with an F-test at 5% level of significance. The power to reject the global null hypothesis as well as the null hypothesis regarding 12 DU vs. placebo was estimated to be 88%.

- **Multiplicity adjustment**

Multiplicity was controlled for the multiple comparisons of the treatment groups on the primary endpoint and on the key secondary endpoints, and for multiple endpoints to be tested which included the primary and 4 key secondary endpoints.

Multiplicity for the primary endpoint was controlled using Fisher's LSD procedure. In the first step of the procedure, a global test was performed for the null hypothesis that the means of each of the 3 treatment groups were equal. If this global null hypothesis could be rejected at the pre-specified level of significance of 5%, then in the second step of the procedure, one was permitted to perform all pairwise comparisons at the same level of significance. Fisher's LSD procedure preserves the experiment-wise type I error rate at the nominal level of significance, if the number of treatment groups is 3.

To further test key secondary hypotheses, the hierarchical testing procedure was conducted. If, and only if, the hypothesis of no difference between 12 DU and placebo on the primary endpoint was rejected ($p < 0.05$), the procedure continued in the hierarchical order:

1. 12 DU versus placebo in average total allergic rhinitis DSS.
2. 12 DU versus placebo in average total allergic rhinitis DMS.
3. 12 DU versus placebo in average overall RQLQ.
4. 12 DU versus placebo in average total combined allergic rhinoconjunctivitis score.

The procedure continued if a hypothesis was rejected ($p < 0.05$), otherwise it stopped with no further statistical conclusions allowed. If all 4 key secondary hypotheses for the comparison of 12 DU to placebo could be rejected ($p < 0.05$), and if the hypothesis of no

difference between 6 DU and placebo on the primary endpoint was rejected ($p < 0.05$), the procedure continued in the following hierarchical order:

1. 6 DU versus placebo in average total allergic rhinitis DSS.
 2. 6 DU versus placebo in average total allergic rhinitis DMS.
 3. 6 DU versus placebo in average overall RQLQ.
 4. 6 DU versus placebo in average total combined allergic rhinoconjunctivitis score.
- Definitions of analysis populations
 - Total-analysis set: All subjects who entered the trial (i.e., signed an informed consent). This analysis set included screening failures. The total population would be used for listing reasons for screening failures and AEs before randomization.
 - Full Analysis Set (FAS): All randomized subjects in accordance with the ICH intent-to-treat principle. This analysis set was the primary set for all efficacy analyses.
 - Per-Protocol Set (PP): All subjects who did not have major protocol deviations that would affect the primary endpoint.
 - Safety Set (SS): All randomized subjects, i.e., the SS was identical to the FAS. The SS was used for safety tables and listings.

Reviewer Comments: *ALK defined the safety set to be identical to the FAS, i.e., all subjects randomized. In this case, safety analysis was performed on treatment assigned in randomization instead of actual treatment received. This is different from the All-Subjects-as-Treated (ASaT) approach which was also used by Merck in the US pivotal study P001. The reviewer defers to the medical officers for further evaluation of the potential impact on safety analysis.*

- Missing data handling

For the primary analysis of the primary endpoint and for the two key secondary endpoints “average total allergic rhinitis DSS during the efficacy evaluation period” and “average total allergic rhinitis DMS during the efficacy evaluation period,” a multiple imputation strategy of Rubin was applied for missing data. Missing data in all treatment groups were sampled from the observed data in the placebo group using the method of unrestricted random sampling with replacement. Thus, all subjects with missing data were included in the analysis as if no treatment effect was present. In an additional analysis of the primary endpoint, missing data were replaced using the method of LOCF if the average TCRS was available from previous visits in the treatment period.

6.2.10 Study Population and Disposition

6.2.10.1 Populations Enrolled/Analyzed

A total of 1425 subjects were screened for the trial. Of the screened subjects, 433 (30%) were screening failures, resulting in 992 subjects eligible for randomization. The FAS comprised a total of 992 subjects: 338 subjects in the placebo group, 336 subjects in the 6 DU group, and 318 subjects in the 12 DU group. The PP comprised a total of 805 subjects (81% of the FAS); 272 subjects in the placebo group, 269 subjects in the 6 DU group, and 264 subjects in the 12 DU group.

6.2.10.1.1 Demographics

The 3 treatment groups were generally similar with regard to the distribution of males and females, different ethnic origins, nationality, and smoking history. The trial population consisted of similar proportions of males and females (approximately 50% of each). The majority of the subjects were Caucasian (98%). The countries recruiting most subjects for the trial were Poland, Germany, Romania, and Czechoslovakia with 25%, 14%, 12%, and 11% of the trial population, respectively.

6.2.10.1.2 Medical/Behavioral Characterization of the Enrolled Population

N/A

6.2.10.1.3 Subject Disposition

A total of 1425 subjects were screened for the trial. Of the screened subjects, 433 (30%) were screening failures, resulting in 992 subjects eligible for randomization. A total of 877 (88%) of the randomized subjects completed the trial, with no overall difference between the 3 treatment groups. One hundred fifteen (12%) subjects discontinued the trial, with no overall difference between treatment groups, except that a slightly higher proportion of subjects discontinued due to AEs in the active groups compared to placebo.

6.2.11 Efficacy Analyses

6.2.11.1 Analyses of Primary Endpoint(s)

The primary efficacy analysis on the FAS with multiple imputation (FAS-MI) showed that the observed difference in the mean TCRS between the 3 treatment groups was statistically significant (global analysis, $p = 0.003$). As shown in Table 9, statistical significance was demonstrated for the 6 DU and the 12 DU groups, respectively, for testing the null hypothesis that there is no difference between the respective active treatment group and the placebo group ($p = 0.004$ for both groups). Additionally, statistical significance for testing the null hypothesis was also demonstrated in the analysis of the FAS with observation (Table 9), PP population, and the FAS with imputation by LOCF (data not shown).

Table 9 Result of the analyses of the TCRS during the efficacy evaluation period

Analysis set	Treatment group	N	Adjusted mean TCRS (95% CI)	Absolute difference (95% CI)*	Relative difference (%) (95% CI)*	p-value
FAS-MI (N=992)	Global	992	-	-	-	0.003
	Placebo	338	6.81 (6.48, 7.13)	-	-	-
	6 DU	336	5.74 (5.42, 6.05)	-1.07 (-1.80, -0.34)	-	0.004
	12 DU	318	5.71 (5.40, 6.02)	-1.09 (-1.84, -0.35)	-16.1 (-25.8, -5.7)	0.004
FAS with observations (N=879)	Placebo	298	6.76 (5.94, 7.63)	-	-	-

Analysis set	Treatment group	N	Adjusted mean TCRS (95% CI)	Absolute difference (95% CI)*	Relative difference (%) (95% CI)*	p-value
	6 DU	297	5.58 (4.81, 6.40)	-1.18 (-1.91, -0.45)		0.002
	12 DU	284	5.53 (4.77, 6.35)	-1.22 (-1.96, -0.49)	-18.1 (-27.6, -7.7)	0.001

Note: N = number of subjects included in the analysis; FAS = full analysis set; FAS-MI = FAS with imputation; CI = confidence interval; Global refers to the global null hypothesis of no difference between the mean TCRS between the 3 treatment groups

* In the MT-06 Clinical Study Report, absolute treatment difference was calculated by Placebo – Treatment; percent treatment difference relative to placebo was calculated by (Placebo – Treatment)/Placebo × 100%. For consistency with presentation of results in study P001 and P003, in this table, absolute treatment difference was calculated by Treatment - Placebo; percent treatment difference relative to placebo was calculated by (Treatment - Placebo)/Placebo × 100%.

Source: adapted from panel 9-2 in the MT-06 Clinical Study Report and Table 2 in Amendment 30.

Reviewer Comments: *The reviewer’s analysis obtained similar numerical results. The primary efficacy analyses showed statistical significance for testing the null hypothesis that there was no difference between the active treatment group (12 DU) and the placebo group. However, both analyses using FAS-MI and FAS with observation showed that the primary efficacy endpoint did not meet the clinical relevance criterion of the upper bound of the 95% confidence interval for the between-treatment difference in the TCRS relative to placebo being ≤ -10%.*

6.2.11.2 Analyses of Secondary Endpoints

Table 10 Results of the key secondary efficacy analyses

Key secondary endpoint	Analysis set	Treatment group	N	Adjusted mean estimate (95% CI)	Absolute difference (95% CI) ¹	Relative difference (%) (95% CI) ¹	p-value
Rhinitis DSS	FAS-MI (N = 992)	Placebo	338	3.31 (3.20;3.43)	-	-	-
		6 DU	336	2.94 (2.81;3.06)	-0.38 (-0.74;-0.01)	-	0.042
		12 DU	318	2.84 (2.73;2.96)	-0.47 (-0.82;-0.11)	-14.1% (-23.8%;-3.9%)	0.01
	FAS with observations (N = 879)	Placebo	298	3.30 (2.84;3.80)	-	-	-
		6 DU	297	2.90 (2.46;3.38)	-0.40 (-0.76;-0.03)	-12.1% (-22.0%;-1.1%)	0.032
		12 DU	284	2.76 (2.34;3.22)	-0.54 (-0.89;-0.18)	-16.2% (-25.7%;-5.8%)	0.003
Rhinitis DMS	FAS-MI (N = 992)	Placebo	338	2.86 (2.68;3.05)	-	-	-
		6 DU	336	2.23 (2.06;2.39)	-0.63 (-1.15;-0.11)	-	0.017
		12 DU	318	2.32 (2.17;2.48)	-0.54 (-1.07;-0.01)	-18.9% (-34.7%;-1.3%)	0.045
	FAS with observations (N = 879)	Placebo	298	2.83 (2.27;3.44)	-	-	-
		6 DU	297	2.13 (1.66;2.67)	-0.69 (-1.20;-0.18)	-24.5% (-38.9%;-7.2%)	0.008
		12 DU	284	2.22 (1.73;2.78)	-0.60 (-1.13;-0.08)	-21.4% (-36.6%;-3.2%)	0.024

Key secondary endpoint	Analysis set	Treatment group	N	Adjusted mean estimate (95% CI)	Absolute difference (95% CI) ¹	Relative difference (%) (95% CI) ¹	p-value
Overall RQLQ	FAS (excl. RS, HR, and BA) ² (N = 711)	Placebo	240	1.58 (1.39;1.76)	-	-	-
		6 DU	242	1.45 (1.26;1.64)	-0.13 (-0.31;0.05)	-8.1% (-18.5%; 3.4%)	0.162
		12 DU	229	1.38 (1.20;1.57)	-0.19 (-0.37;-0.02)	-12.3% (-22.3%;-1.2%)	0.031
Combined rhinoconjunctivitis score	FAS (excl. RS and HR) (N = 754) ³	Placebo	257	9.12 (7.87;10.47)	-	-	-
		6 DU	256	7.74 (6.56;9.02)	-1.38 (-2.45;-0.32)	-15.2% (-25.5%;-3.7%)	0.011
		12 DU	241	7.91 (6.72;9.21)	-1.21 (-2.28;-0.13)	-13.2% (-23.7%;-1.5%)	0.029

N = number of subjects included in the analysis; FAS = full analysis set; FAS-MI = FAS with multiple imputation; DSS = daily symptom score; DMS = daily medication score; RQLQ = rhinoconjunctivitis quality of life questionnaire; CI = confidence interval.

Note: 1. In the MT-06 Clinical Study Report, absolute treatment difference was calculated by Placebo – Treatment; percent treatment difference relative to placebo was calculated by (Placebo – Treatment)/Placebo × 100%. For consistency with presentation of results in study P001 and P003, in this table, absolute treatment difference was calculated by Treatment - Placebo; percent treatment difference relative to placebo was calculated by (Treatment - Placebo)/Placebo × 100%.

2. For the RQLQ analysis, subjects from Serbia (RS), Croatia (HR), and Bosnia and Herzegovina (BA) were excluded from the FAS because no validated RQLQ questionnaire was available in local language in these countries.

3. For the analysis of the combined rhinoconjunctivitis score, subjects from RS and HR were excluded from the FAS.

Source: adapted from panel 9-3 in the MT-06 Clinical Study Report and Table 1 in Amendment 32.

For analysis of average total allergic rhinitis DSS and average total allergic rhinitis DMS using the FAS-MI and the FAS with observations, statistical significance was demonstrated for the mean rhinitis DSS between the respective treatment group (6 DU or 12 DU) and placebo group for testing the null hypothesis that there was no difference between the active treatment group and the placebo group.

For overall RQLQ score, the difference between 6 DU and placebo was not statistically significant (p = 0.162). In the 12 DU group, the overall RQLQ score was statistically significantly reduced by 0.19 in absolute numbers compared to placebo (p = 0.031). The total combined allergic rhinoconjunctivitis score was statistically significantly more reduced in the 6 DU group and the 12 DU group compared to placebo. According to the hierarchical testing strategy, the statistical analysis of the combined rhinoconjunctivitis score was not controlled for multiplicity for the 6 DU group, since the preceding statistical analysis of the RQLQ score in this group resulted in a non-significant treatment effect.

6.2.11.3 Subpopulation Analyses

The treatment difference to placebo in the TCRS was numerically larger for subjects ≥30 years (point estimate -1.85 for 12 DU) compared to subjects <30 years (point estimate

-0.69 for 12 DU). For the gender subgroups, the treatment difference to placebo was numerically larger for males (point estimate -1.59 for 12 DU) than for females (point estimate -0.82 for 12 DU). Since the trial was not powered to show efficacy in the specific subgroups, the observed differences might not be interpretable or conclusive.

6.2.11.4 Dropouts and/or Discontinuations

Please refer to section 6.2.10.1.3.

6.2.11.5 Exploratory and Post Hoc Analyses

N/A

6.2.12 Safety Analyses

The safety analysis was performed on the safety population. A total of 579 (58%) of the subjects in the overall trial population reported AEs during the trial. More subjects reported AEs in the 2 active treatment groups (i.e., 63% and 67% of the subjects in the 6 DU and 12 DU group, respectively) compared to placebo (46% of the subjects).

Most of the AEs reported in the active groups were assessed as Investigational medicinal product (IMP)-related, with 59% and 67% of the AEs assessed as possibly related to the treatment in the 6 DU and 12 DU groups, and 29% of the AEs assessed as possibly related in the placebo group. The majority of all IMP-related AEs were mild or moderate in severity. This pattern applied to all 3 treatment groups. The most frequently reported IMP-related AEs (defined as AEs reported in ≥ 2 % of the subjects in at least one of the active treatment groups) were local reactions in mouth and throat such as oral pruritus, throat irritation, and oedema mouth. These were mostly reported within the first 1-2 days after the first IMP intake, with very few onsets at later time points.

A total of 30 severe AEs were reported by 27 (3%) of the subjects during the trial. Nine of the severe AEs (reported by 8 subjects) were IMP-related. More IMP-related severe AEs were reported in actively treated subjects, with 3 (<1%) subjects reporting 3 severe events and 5 subjects (2%) reporting 6 severe events in the 6 DU and 12 DU groups, respectively. No severe IMP-related AEs were reported in the placebo group.

6.2.12.1 Methods

Safety analysis was performed mainly with descriptive statistics.

6.2.12.3 Deaths

No deaths occurred during the trial.

6.2.12.4 Nonfatal Serious Adverse Events

SAEs were reported by 12 subjects during the trial: 8 subjects from the placebo group and 4 subjects from the 6 DU group. No SAEs were reported in the 12 DU group. The investigators assessed all SAEs as being unlikely related to the treatment.

6.2.12.5 Adverse Events of Special Interest (AESI)

N/A

6.2.12.6 Clinical Test Results

N/A

6.2.12.7 Dropouts and/or Discontinuations

Thirty (30) subjects (3%) discontinued the trial due to 50 AEs. More subjects discontinued the trial in the active groups compared to placebo: 10 (3%) subjects from the 6 DU group, 13 (4%) subjects from the 12 DU group, and 7 (2%) subjects from the placebo group discontinued due to AEs.

6.3 Study P003: A Phase IIb, Randomized, Placebo-Controlled, Dose-Finding Clinical Trial to Study the Safety and Efficacy of MK-8237 using an Environmental Exposure Chamber in Subject with House Dust Induced Allergic Rhinitis/Rhinoconjunctivitis

6.3.1 Objectives

Primary objective:

To evaluate the dose-related efficacy of MK-8237 sublingual HDM tablet versus placebo in the treatment of HDM-induced rhinitis based on the average total nasal symptom score (TNSS), determined during the chamber challenge session at Week 24.

Key Secondary objectives:

- To evaluate the onset of action of MK-8237 versus placebo in the treatment of HDM-induced rhinitis based on the average TNSS during chamber sessions at Week 8, 16, and 24.
- To evaluate the dose response of MK-8237 versus placebo in the treatment of HDM-induced rhinitis based on the average TNSS during chamber sessions at Week 8, 16, and 24.
- To evaluate the efficacy of MK-8237 versus placebo in the treatment of HDM-induced rhinoconjunctivitis based on the average total symptom score (TSS) [sum of TNSS and TOSS (total ocular symptom score)] during the chamber session at Week 24.

Safety Objective:

To evaluate the safety and tolerability of MK-8237 (6 and 12 DU) dosed daily over a 6-month treatment period in adults with HDM-induced allergic rhinitis/rhinoconjunctivitis (ARC).

6.3.2 Design Overview

This was a Phase IIb, randomized, double-blind, placebo-controlled, parallel-group, dose finding clinical trial to study the safety and efficacy of MK-8237 using an Environmental Exposure Chamber (EEC) in adult subjects with house dust mite induced AR/ARC, conducted at a single site in conformance with Good Clinical Practice (GCP). A total of 120 subjects were planned to be randomized in a 1:1:1 ratio to receive MK-8237 12 DU, MK-8237 6 DU, or placebo sublingually once daily for approximately 24 weeks. The study assumed a dropout rate of 17%, with a plan for approximately 100 evaluable subjects for the primary analysis at Week 24.

The screening period was approximately 6 weeks, consisting of the screening visit and HDM screening challenge at Day -28. Subjects were required to meet a pre-specified threshold of 6 out of 12 points for total nasal symptom score at any time during the first 2 hours of the screening challenge. Randomization occurred on Day 1 (Visit 3) with on-site dosing of the first dose of study medication. The treatment period was 24 weeks in duration, consisting of Visits 4 through 9. The treatment period included three chamber challenges: Week 8 (Visit 5), Week 16 (Visit 7), and Week 24 (Visit 9). Monthly study visits (without chamber challenge) occurred at Week 4 (Visit 4), Week 12 (Visit 6), and Week 20 (Visit 8). A follow-up visit (Week 26, Visit 10) was performed approximately 2 weeks following the completion of the treatment period.

6.3.3 Population

Adult subjects (≥ 18 years of age) with a clinical history of house dust mite-induced allergic rhinitis/rhinoconjunctivitis of ≥ 1 year duration, with or without asthma, who demonstrated a positive skin prick test response and positive specific immunoglobulin E (IgE) reactivity to house dust mite (*Dermatophagoides [D.] pteronyssinus* and/or *D. farinae*), forced expiratory volume in 1 second (FEV1) $\geq 70\%$ of predicted at screening and randomization, and a total nasal symptom score ≥ 6 of 12 within the first 2 hours of the screening environmental exposure chamber session.

6.3.4 Study Treatments or Agents Mandated by the Protocol

MK-8237 12 DU, 6 DU, and placebo

6.3.6 Sites and Centers

This trial was conducted at one trial site, located in Vienna, Austria.

6.3.7 Surveillance/Monitoring

N/A

6.3.8 Endpoints and Criteria for Study Success

Primary endpoints

Average total nasal symptom score (TNSS) during the last 4 hours of the chamber challenge at Week 24. TNSS was the sum of 4 nasal symptoms (runny nose, blocked nose, sneezing, itchy nose), each rated on a scale from 0 (no symptoms) to 3 (severe symptoms).

Key Secondary Endpoints

- Average TNSS during the last 4 hours of the chamber challenge at Weeks 8 and 16.
- Average total symptom score (TSS) during the last 4 hours of the chamber challenge at Week 24. TSS was the sum of the 4 nasal symptoms comprising TNSS and the 2 ocular symptoms (itchy eyes and watery eyes) comprising the total ocular symptom score TOSS, with each symptom rated on a scale from 0 (no symptoms) to 3 (severe symptoms).

Safety Endpoints

- Tier 1 safety endpoints: Proportion of subjects with pre-specified local application site adverse events (AEs) (ear pruritus, mouth edema, oral pruritus, throat irritation); proportion of subjects with any systemic allergic reactions; and events treated with epinephrine.
- General safety endpoints: Proportion of subjects with any AEs, drug-related AEs, serious AEs, serious drug-related AEs, who died, and who discontinued due to an AE. Timeframe: 6 months

6.3.9 Statistical Considerations & Statistical Analysis Plan

- Hypotheses and Statistical Methods for Efficacy Endpoints

The efficacy analyses were performed on the full analysis set (FAS). In addition, the Per-Protocol (PP) population was used for supportive analysis. The average TNSS at Week 24 was analyzed using the analysis of covariance (ANCOVA) approach with treatment and baseline TNSS as covariates. The least squares (LS) means, treatment differences, and the associated 95% confidence intervals (CI) were estimated based on the ANCOVA model. Treatment difference relative to placebo was calculated as a percentage based on the LS means, with the associated 95% CI calculated based on the bootstrap re-sampling approach. The normality assumption of the parametric ANCOVA model was checked using the Shapiro-Wilk test and inspection of the Q-Q plot; no severe violation was observed. Two supportive analyses were performed for the primary endpoint. The first one was based on the PP population using the same ANCOVA as the primary analysis. The second one was the longitudinal data analysis (LDA) model based on the FAS population. The LDA model used the average TNSS obtained at screening, Week 8, 16, and 24 as response, and included treatment, session, and treatment-by-session interaction as fixed effects and subject as random effect. An unstructured covariance matrix was used to model the correlation among repeated measurements. The key secondary endpoints were analyzed on the FAS population using a similar ANCOVA approach as that of the primary endpoint.

- Sample size determination

This study targeted to enroll ~40 subjects per treatment group, with up to 120 subjects in total. Assuming a discontinuation rate of ~17%, there would be about 100 evaluable subjects for primary evaluation at Week 24. With 33 subjects per treatment group, this study would have 80% power to detect (2-sided, $\alpha=0.050$) a difference of 2.2 units in the average TNSS at Week 24 between the treatment and placebo arms, with an assumed SD of 3.2, and a detectable difference of 1.7 units with an assumed SD of 2.4. The between-treatment difference of 2.0 units in the average daily symptom scores was observed in the Stallergenes grass sublingual therapy study conducted at the same chamber. Further, the between-treatment differences, ranging from 0.8 to 1.5 units, were observed in the Phase III field trials of grass and ragweed AIT where the use of rescue medication was allowed.

- Multiplicity

The multiplicity of the primary analysis, where two doses (12 DU and 6 DU) are compared with placebo on the average TNSS, would be controlled using a step-down procedure. The hypothesis testing on the lower dose (6 DU) would be performed only when statistical significance is demonstrated for the higher dose (12 DU). For the

analyses of the Key Secondary and Other Secondary Endpoints, multiplicity was controlled only within each endpoint for testing the two doses in the same manner as described for the primary analysis.

Reviewer Comments: *In order to allow drawing firm conclusion about treatment benefits relevant to both the primary and secondary endpoints, proper multiplicity control is needed, not only for multiple comparisons within each endpoint but also for multiple endpoints to be tested, which includes the primary and key secondary endpoints.*

- Definitions of analysis populations
 - Full Analysis Set (FAS) population: all randomized subjects who received at least one dose of study treatment, and had at least one post-randomization observation for the analysis endpoint.
 - Per-Protocol population: PP excluded subjects due to important deviations from the protocol that may substantially affect the results of the primary efficacy endpoint.
Safety population: the All-Subjects-As-Treated (ASaT) population included all randomized subjects who took at least one dose of study medication. Subjects were analyzed in the treatment group corresponding to the actual treatment received.

- Missing data handling

The applicant planned to use the last observation carried forward (LOCF) method to impute missing observations within the chamber session at Week 24 as the sensitivity analysis for the primary efficacy endpoint. However, no data were missing on the primary endpoint from subjects who participated in the chamber session at Week 24. Therefore, the LOCF analysis was not performed. All analyses were performed based on the observed data only.

6.3.10 Study Population and Disposition

6.3.10.1 Populations Enrolled/Analyzed

Of the 153 subjects screened for inclusion, 29 subjects were excluded and 124 were randomized. Of the 124 randomized subjects, all (100%) subjects took at least one dose of study medication and were included in the safety analysis. One subject did not meet eligibility due to negative serum IgE result at screening. A total of 42 subjects received 12 DU MK-8237, 41 subjects received 6 DU, and 41 subjects received placebo. The FAS population had a total of 119 subjects, including 40 in the 12 DU group, 39 in the 6 DU group, and 40 in the placebo group. The PP population had a total of 115 subjects, including 39 in the 12 DU group, 37 in the 6 DU group, and 39 in the placebo group.

6.3.10.1.1 Demographics

The majority of subjects were within the age group of 18 to 50 years old across the treatment groups, ranging from 92.9% to 100%. Also the majority of subjects were white (race) and non-Hispanic/Latino (ethnicity) across the treatment groups, ranging from 88.1% to 92.7%. A lower percentage of subjects were male versus female in the 6 DU

group (26.8%), as compared with 54.8% in the 12 DU group and 58.5% in the placebo group.

6.3.10.1.2 Medical/Behavioral Characterization of the Enrolled Population

N/A

6.3.10.1.3 Subject Disposition

Of the 153 subjects screened for inclusion, 29 subjects were excluded and 124 were randomized. The majority of subjects excluded were due to not meeting the screening chamber challenge minimum nasal symptom score requirement (12 of 29 subjects; 42.9%) or inability to adhere to dose and visit schedule (12 of 29 subjects; 42.9%). Of the 124 randomized subjects, 106 (85.5%) subjects completed the protocol-specified, double-blind treatment period, while 18 subjects (14.5%) discontinued the investigational treatment early. The primary reasons for study discontinuation were AEs (9 subjects, 7.3% overall) and withdrawal by subject (8 subjects, 6.5%).

6.3.11 Efficacy Analyses

6.3.11.1 Analyses of Primary Endpoint(s)

The primary endpoint analysis on the FAS showed that statistical significance was demonstrated for the average TNSS at Week 24 between the respective treatment group (6 DU or 12 DU) and placebo group for testing the null hypothesis that there was no difference between the active treatment group and the placebo group (Table 11). The treatment difference relative to placebo with 95% CI was -26.6% (-39.6%, -11.2%) for the 6 DU group and -48.6% (-60.2%, -35.3%) for the 12 DU group (Table 11).

The ANCOVA analysis based on the Per-Protocol (PP) population showed similar results (data not shown). The normality check on the primary endpoint was performed with no severe violation of the normality assumption detected.

Table 11 Analysis of Average TNSS during Chamber Session at Week 24 (Full Analysis Set)

Treatment	N	Baseline Mean (SD)	Average TNSS at Week 24 Mean (SD)	Average TNSS at Week 24 LS Mean (95% CI)	Treatment difference (95% CI) p-value	Treatment difference relative to placebo (%) (95% CI) ¹
MK-8237 12 DU	36	7.74 (1.98)	3.83 (2.67)	3.83 (2.94, 4.72)	-3.62 (-4.85, -2.39) p-value < 0.001	- 48.6 (-60.2, -35.3)
MK-8237 6 DU	36	8.08 (1.82)	5.59 (2.65)	5.47 (4.55, 6.39)	- 1.98 (-3.24, -0.72) p-value = 0.003	- 26.6 (-39.6, -11.2)
Placebo	34	7.32 (1.61)	7.31 (2.69)	7.45 (6.57, 8.33)		

N = Number of subjects included in the analysis at Week 24; SD = Standard deviation; LS Mean = Least square mean; CI = Confidence Interval.

Note: 1. Treatment difference relative to Placebo was calculated by (MK-8237 - Placebo)/Placebo * 100%, confidence intervals were calculated by the bootstrap method.

Source: adapted from Table 11-2 in the P003 Clinical Study Report

Reviewer Comments: *The reviewer’s analysis obtained similar numerical results. The reviewer performed 500 runs of bootstrap analysis on the FAS to assess the potential impact of fluctuation of bootstrapping process on the estimation of the upper bound of the 95% CI of relative treatment effect (12 DU vs. Placebo). The analysis showed that the estimate of the upper bound ranged from -35.5% to -34.3%. The median was -35.0%, which was close to the applicant’s estimate.*

6.3.11.2 Analyses of Secondary Endpoints

The first key secondary endpoint was the average TNSS over the last 4 hours of the chamber challenge at Week 16. This analysis showed a statistically significantly lower TNSS for both MK-8237 12 DU and 6 DU groups, when compared to the placebo group at Week 16 (Table 12). The treatment effect was more pronounced in the 12 DU group than in the 6 DU group.

The second key secondary endpoint was the average TNSS over the last 4 hours of the chamber challenge at Week 8. The analysis showed statistically significant difference (p=0.007) between the 12 DU group and placebo group, while the difference between the 6 DU group and the placebo group was not statistically significant (p=0.198) (Table 12).

The third key secondary endpoint was the average TSS (ocular plus nasal) over the last 4 hours of the chamber challenge at Week 24. As shown in Table 12, the analysis showed a statistically significantly lower TSS for both MK-8237 12 DU and 6 DU groups when compared to the placebo group at Week 24.

Table 12 Analysis of Average TNSS during chamber session at Week 16, Average TNSS during chamber session at Week 8, and Average TSS during chamber session at Week 24 (Full Analysis Set)

Key secondary analysis	Treatment	N	Baseline Mean (SD)	Average TNSS at Week 24 Mean (SD)	Average TNSS at Week 24 LS Mean (95% CI)	Treatment difference (95% CI) p-value	Treatment difference relative to placebo (%) (95% CI) ¹
Average TNSS at Week 16	MK-8237 12 DU	39	7.76 (1.91)	4.83 (2.47)	4.82 (4.07, 5.56)	-2.08 (-3.14, -1.03) p-value<0.001	-30.1 (-42.3, -16.8)
	MK-8237 6 DU	36	8.05 (1.87)	5.81 (2.57)	5.67 (4.83, 6.50)	-1.23 (-2.36, -0.11) p-value=0.032	-17.8 (-31.1, -2.6)
	Placebo	38	7.42 (1.62)	6.76 (2.40)	6.90 (6.13, 7.67)		
Average TNSS at Week 8	MK-8237 12 DU	40	7.82 (1.92)	5.37 (2.57)	5.34 (4.53, 6.15)	-1.37 (-2.34, -0.39) p-value=0.007	-20.4 (-33.3, -6.8)
	MK-8237 6 DU	39	8.06 (1.86)	6.29 (2.00)	6.16 (5.55, 6.78)	-0.54 (-1.38, 0.29) p-value=0.198	-8.0 (-19.3, 4.0)
	Placebo	39	7.38 (1.59)	6.55 (1.95)	6.71 (6.13, 7.28)		

Key secondary analysis	Treatment	N	Baseline Mean (SD)	Average TNSS at Week 24 Mean (SD)	Average TNSS at Week 24 LS Mean (95% CI)	Treatment difference (95% CI) p-value	Treatment difference relative to placebo (%) (95% CI) ¹
Average TSS at Week 24	MK-8237 12 DU	36	9.77 (3.00)	4.41 (3.74)	4.43 (3.20, 5.66)	-4.84 (-6.59, -3.09) p-value<0.001	-52.2 (-65.0, -37.0)
	MK-8237 6 DU	36	10.34 (2.85)	6.73 (3.16)	6.62 (5.48, 7.77)	-2.65 (-4.35, -0.95) p-value=0.003	-28.6 (-41.8, -11.6)
	Placebo	34	9.41 (2.44)	9.18 (3.91)	9.27 (7.98, 10.57)		

N = Number of subjects included in the analyses. SD = Standard deviation; LS Mean = Least squares mean; CI = Confidence Interval.

Note: 1. Treatment difference relative to Placebo was calculated by (MK-8237 - Placebo)/Placebo * 100%, confidence intervals were calculated by the bootstrap method.

Source: adapted from Table 11-3, 11-4, and 11-5 in the P003 Clinical Study Report

Reviewer Comments: *Multiplicity was controlled only for multiple comparisons within each endpoint, but not for multiple endpoints. Since the analysis of average TNSS at Week 8 resulted in a non-significant treatment effect (6 DU vs. placebo), caution is needed for drawing firm conclusion from the secondary endpoint analyses to avoid inflation of Type 1 error.*

6.3.11.3 Subpopulation Analyses

The majority of subjects were within the age group of 18 to 50 years old across the treatment groups. Also, the majority of subjects were white across groups. The applicant performed subgroup analysis of the average TNSS at Week 24 by gender (male/female). The subgroup analysis by gender showed a similar trend favoring treatment with MK-8237 over placebo for both genders.

6.3.11.4 Dropouts and/or Discontinuations

Please refer to 6.3.10.1.3.

6.3.11.5 Exploratory and Post Hoc Analyses

N/A

6.3.12 Safety Analyses

The majority of subjects (85.5%) in this study experienced one or more AEs. AE rates in the two active treatment groups (90.5% in the 12 DU group and 87.8% in the 6 DU group) were higher than that in the placebo group (78.0%). Also, 76.2% and 61.0% of subjects in the 12 DU and 6 DU group reported drug-related adverse events as compared to 4.9% in the placebo group.

6.3.12.1 Methods

Safety parameters or adverse experiences of special interest that were identified a priori constitute Tier 1 safety endpoints that were subject to inferential testing for statistical significance, with p-values and 95% confidence intervals for between-group comparisons. Other safety parameters were considered Tier 2 or Tier 3. Tier 2 parameters

were assessed via point estimates with 95% confidence intervals for between-group comparisons; only point estimates by treatment group were provided for Tier 3 safety parameters.

6.3.12.3 Deaths

No death was reported in the study.

6.3.12.4 Nonfatal Serious Adverse Events

There was only one serious adverse event reported during the study. One subject in the placebo group experienced a serious adverse event of vertigo.

6.3.12.5 Adverse Events of Special Interest (AESI)

A total of 74%, 61%, and 2% of subjects in the 12 DU, 6 DU, and placebo groups, respectively, experienced one or more local application site reactions (ASRs). The majority of local ASRs were assessed as drug related by the investigator (74% in the 12 DU group, 59% in the 6 DU group, and 2% in the placebo group). Local allergic reactions in the mouth and throat occurred commonly.

There were no AEs of anaphylaxis and/or investigator-diagnosed systemic allergic reactions (including anaphylactic shock, anaphylactic reactions, or anaphylaxis) reported, and no AEs treated with epinephrine reported during the study. One subject (12 DU) had urticaria (not treated) and another subject (12 DU) had rhinorrhea (treated with antihistamine) assessed as drug related by the investigator; both were mild in intensity.

The proportion of subjects who experienced an AE was higher in asthma subjects (100%) than in non-asthma subjects (88%) in the 12 DU group. The proportion of subjects who experienced drug-related AEs was also higher in asthma subjects (100%) than in non-asthma subjects (69%).

6.3.12.6 Clinical Test Results

N/A

6.3.12.7 Dropouts and/or Discontinuations

Overall, nine subjects (3 [7.1%] in the 12 DU group and 6 [14.6%] in the placebo group) discontinued the study due to an adverse event. No subject was discontinued due to a drug-related adverse event.

7. INTEGRATED OVERVIEW OF EFFICACY

7.1 Indication #1 Immunotherapy for the treatment of house dust mite (HDM)-induced allergic rhinitis, with or without conjunctivitis in adults 18 through 65 years of age.

7.1.1 Methods of Integration

The applicant conducted combined analyses based on two Phase III AR/C trials, P001 and MT-06, considering similar trial design, eligibility criteria, demographics, and

primary endpoint across the trials. In P001, subjects were required to meet the symptom score on 5 of 7 consecutive calendar days while withholding allergy relieving medications. One major difference between the studies was that in MT-06, subjects were required to meet the symptom score, while using allergic rhinitis symptom relieving medication, on at least 8 days of the 15-day baseline period. Hence, the applicant indicated that subjects from MT-06 were potentially suffering from more severe AR/C than those in P001, resulting in slightly higher efficacy.

7.1.2 Demographics and Baseline Characteristics

A total of 2138 subjects between the ages of 12 and 85 with HDM AR/C were randomized into the trials. Overall, the treatment groups were relatively balanced with respect to baseline demographics and characteristics.

7.1.4 Analysis of Primary Endpoint(s)

The primary pooled efficacy analysis was performed based on observed data only using an ANCOVA model. An analysis with the missing data accounted for by multiple imputations was conducted using an ANCOVA model. Additionally, the pooled dataset was analyzed using a non-parametric approach. The details of the methods are described in the footnotes of Table 13.

As shown in Table 13, the average TCRS during the last 8 weeks of treatment was improved in the MK-8237 group compared to the placebo group. The treatment difference relative to placebo was -17.4% (95% CI, -23.4%, -10.9%) based on the ANCOVA analysis. This trend of improvement by MK-8237 was demonstrated by the ANCOVA analysis with multiple imputation, where the treatment difference relative to placebo was -13.2% (95% CI, -19.5%, -7.0%), and by the nonparametric analysis where the relative treatment difference was -19.6% (95% CI, -25.4%, -13.2%).

Table 13 Analysis of average Total Combined Rhinitis Score during the last 8 weeks of treatment using ANCOVA, ANCOVA with multiple imputation, and nonparametric analysis - combined Phase III AR/C Trials (MT-06, P001) Full Analysis Set

Analysis	Treatment	N	Baseline Mean (SD)	Last 8 weeks Mean (SD)	Last 8 weeks LS Mean (95% CI)	Last 8 weeks Median (LQ, UQ)	Treatment difference (95% CI) p-value	Treatment difference relative to placebo (%) (95% CI)
ANCOVA with observation ¹	MK-8237 12 DU	850	7.94 (1.71)	5.34 (4.01)	4.47 (4.21,4.73)	--	-0.94 (-1.31, - 0.57) p-value<0.001	-17.4 (-23.4, -10.9)
	Placebo	918	7.96 (1.72)	6.24 (4.24)	5.41 (5.14, 5.68)	--	--	--
ANCOVA with multiple imputation ²	MK-8237 12 DU	1058	7.94 (1.72)	5.34 (4.01)	5.00 (4.71, 5.29)	--	-0.76 (-1.14, -0.38) p-value<0.001	-13.2 (-19.5, -7.0)
	Placebo	1079	7.93 (1.74)	6.24 (4.24)	5.76 (5.48, 6.05)	--	--	--

Analysis	Treatment	N	Baseline Mean (SD)	Last 8 weeks Mean (SD)	Last 8 weeks LS Mean (95% CI)	Last 8 weeks Median (LQ, UQ)	Treatment difference (95% CI) p-value	Treatment difference relative to placebo (%) (95% CI)
Non-parametric approach ³	MK-8237 12 DU	850	7.94 (1.71)	5.34 (4.01)	--	4.50 (2.30, 7.50)	- 0.90 (-1.25, -0.52) p-value<0.001	-19.6 (-25.4, -13.2)
	Placebo	918	7.96 (1.72)	6.24 (4.24)	--	5.60 (3.10, 8.70)		

N = Number of subjects included in the analysis; SD = Standard deviation; LS Mean = Least square mean; CI = Confidence Interval; LQ = Lower Quartile; UQ = Upper Quartile

Note:

1. Analysis via analysis of covariance (ANCOVA) with square root transformed endpoint value as response, trial, treatment, and baseline asthma status as fixed effects and square root transformed baseline endpoint score as a covariate, and adjusted for different error variation for each treatment group. Back transformed LS means, treatment difference and the associated 95% CI were reported. Treatment difference relative to Placebo based on LS means was calculated by $(\text{MK-8237} - \text{Placebo})/\text{Placebo} * 100\%$, confidence intervals were calculated by the bootstrap method using 10,000 iterations.
2. Missing data in both treatment groups were imputed by sampling from the observed data of the endpoint in the placebo group using unrestricted random sampling with replacement. Rubin’s multiple imputation strategy was used and each missing value was replaced with 50 set of plausible values. Each imputed dataset was analyzed by the analysis of covariance model with square root transformed endpoint value as response, trial, treatment, asthma status as fixed effects, square root transformed baseline endpoint value as covariate, and adjusting for different error variation for each treatment group. Back transformed LS means, treatment difference and the associated 95% CI were reported. Treatment difference relative to Placebo was calculated by $(\text{MK-8237} - \text{Placebo})/\text{Placebo} * 100\%$, confidence intervals were calculated by applying the Fieller’s theorem.
3. The 95% confidence interval for median difference was based on the Hodges-Lehmann estimator. Treatment difference relative to Placebo based on medians was calculated by $(\text{MK-8237} - \text{Placebo})/\text{Placebo} * 100\%$, confidence intervals were calculated by the bootstrap method using 10,000 iterations.

Source: adapted from Table 2.7.3-allergy 28, Table 2.7.3-allergy 29, and Table 2.7.3-allergy 30 in the Summary of Clinical Efficacy

Reviewer Comments: *The pooled analysis with study P001 and MT-06 showed a similar trend in efficacy with the individual studies. However, as indicated by the applicant, the enrollment eligibility criteria were different between study populations, e.g. whether or not allergy relieving medications were allowed, which might have resulted in elevated heterogeneity between the two study populations. Therefore, the results of the pooled analysis need to be interpreted with caution.*

7.1.5 Analysis of Secondary Endpoint(s)

Post-hoc analyses of the secondary endpoints of average rhinitis DSS and average rhinitis DMS during the last 8 weeks of treatment were performed on the combined Phase III AR/C trials. The analysis of the average rhinitis DSS in the combined MT-06 and P001 trials showed that the treatment difference of MK-8237 relative to placebo was -15.9% (95% CI - 21.9%, -9.5%). The analysis of the average DMS in the combined MT-06 and P001 trials showed that the treatment difference of MK-8237 relative to placebo was -17.0% (95% CI, -32.3%, -1.7%).

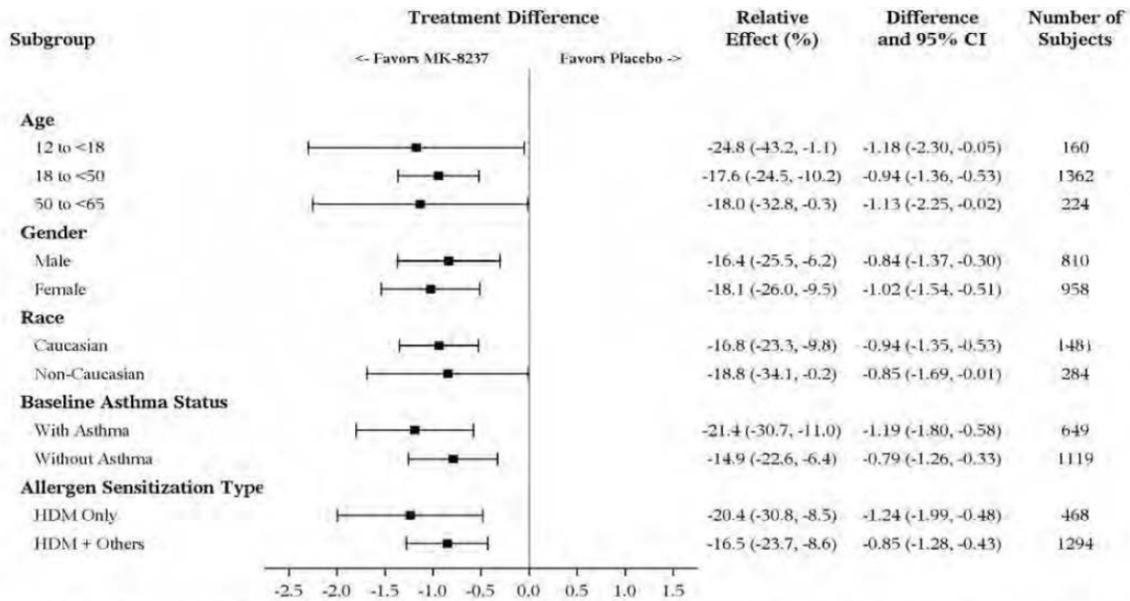
7.1.6 Other Endpoints

NA

7.1.7 Subpopulations

Subgroup analyses of average TCRS during the last 8 weeks of treatment included age group, gender, race, ethnicity, asthma status, allergen sensitization, and geographic region. MK-8237 generally demonstrated similar treatment effect across the various subgroups (Figure 4).

Figure 4 Analysis of average Total Combined Rhinitis Score during the last 8 weeks of treatment by demographic subgroups - combined Phase III AR/C Trials (MT-06, P001; Full Analysis Set)



Source: Figure 2.7.3-allergy: 5 in the Summary of Clinical Efficacy

Efficacy across geographic regions was examined by comparing the average TCRS during the last 8 weeks of treatment in P001 (North America) with MT-06 (Europe). Only observed data in the FAS population were used to compare the primary endpoint in the two trials. In P001, representing North America, the treatment difference of MK-8237 relative to placebo was -17.2% (95% CI -25.0%, -9.7%). In MT-06, representing Europe, the treatment difference of MK-8237 relative to placebo was -18.1% (95% CI -27.6%, -7.7%).

7.1.10 Additional Efficacy Issues/Analyses

NA

7.1.11 Efficacy Conclusions

The analysis based the pooled P001 and MT-06 showed a trend of improvement by MK-8237 treatment as compared to the placebo. Note that the pooled analyses are considered

to be exploratory in nature. Additionally, the results of the pooled analyses need to be interpreted with caution.

8. INTEGRATED OVERVIEW OF SAFETY

8.1 Safety Assessment Methods

The integrated safety data are presented for 3 data pools:

- 1) All Combined Phase II and III Trials (MT-02, MT-04, P003, MT-06, P001);
- 2) Combined Phase II and III Allergic Rhinitis/Allergic Rhinoconjunctivitis (AR/C) Trials (P003, MT-06, P001) and;
- 3) Combined Phase II and III Asthma Trials (MT-02, MT-04).

All subjects enrolled in the Phase II and III trials (MT-02, MT-04, P003, MT-06, and P001) are included in the most comprehensive pool, herein referred to as the All Combined Phase II and III Trials. Subjects who completed these trials received treatment for at least 6 months and up to 18 months. All subjects in this pool were required to be HDM allergic with AR/C symptoms and with or without asthma. All subjects in this pool were required to have evidence of IgE sensitivity, determined by skin prick test (SPT) and serum IgE to at least one of the major species of HDM (*D. farinae* or *D. pteronyssinus*).

The applicant also characterized the safety profile of MK-8237 in the subgroup of subjects with a primary diagnosis of HDM-induced AR/C and those with a primary diagnosis of HDM-induced asthma. All subjects enrolled in P001, MT-06, and P003 are included in the AR/C pool, referred to as the Combined Phase II and III AR/C Trials, while all subjects enrolled in MT-02 and MT-04 are included in the asthma pool, referred to as the Combined Phase II and III Asthma Trials.

8.2 Safety Database

8.2.1 Studies/Clinical Trials Used to Evaluate Safety

- Pooled Phase II and Phase III Studies (MT-02, MT-04, P003, MT-06, and P001);
- Pooled Phase II and III AR/C Trials (P003, MT-06, P001)
- Pooled Phase II and Phase III Asthma Studies (MT-02, MT-04)

8.2.2 Overall Exposure, Demographics of Pooled Safety Populations

- All Combined Phase II and III Trials: Overall, from the 5 pooled trials, 2,187 subjects received at least one dose of MK-8237. Subjects were treated for up to 550 days (approximately 18 months) with a median of 336 days and a mean of 305 days. Among the subjects treated with MK-8237, 1381 received at least one dose of 12 DU and 806 received at least one dose of 6 DU.
- Combined Phase II and III AR/C Trials: From the 3 pooled AR/C trials, 1,475 subjects received at least one dose of MK-8237. Subjects were treated for up to 406 days (approximately 13.5 months) with a median of 299 days and a mean of 266.9

days. Among the subjects treated with MK-8237, 1099 received at least one dose of 12 DU and 376 received at least one dose of 6 DU.

- Combined Phase II and III Asthma Trials (MT-02, MT-04): 712 subjects received at least one dose of MK-8237. Subjects were treated for up to 550 days (approximately 18 months) with a median of 412.5 days and a mean of 384.1 days. Among the subjects treated with MK-8237, 282 received at least one dose of 12 DU and 430 received at least one dose of 6 DU.

8.3 Caveats Introduced by Pooling of Data Across Studies/Clinical Trials

N/A

8.4 Safety Results

8.4.1 Deaths

No deaths were reported in all combined Phase II and III Trials (MT-02, MT-04, P003, MT-06, and P001).

8.4.2 Nonfatal Serious Adverse Events

Across the 5 Phase II and III trials, serious adverse events were reported in 68 subjects: 37 on MK-8237 (17 on 12 DU and 20 on 6 DU) and 31 on placebo. Among those, 66 subjects reported serious adverse events meeting the International Conference on Harmonization (ICH) definition of seriousness (15 on 12 DU, 20 on 6 DU, and 31 on placebo).

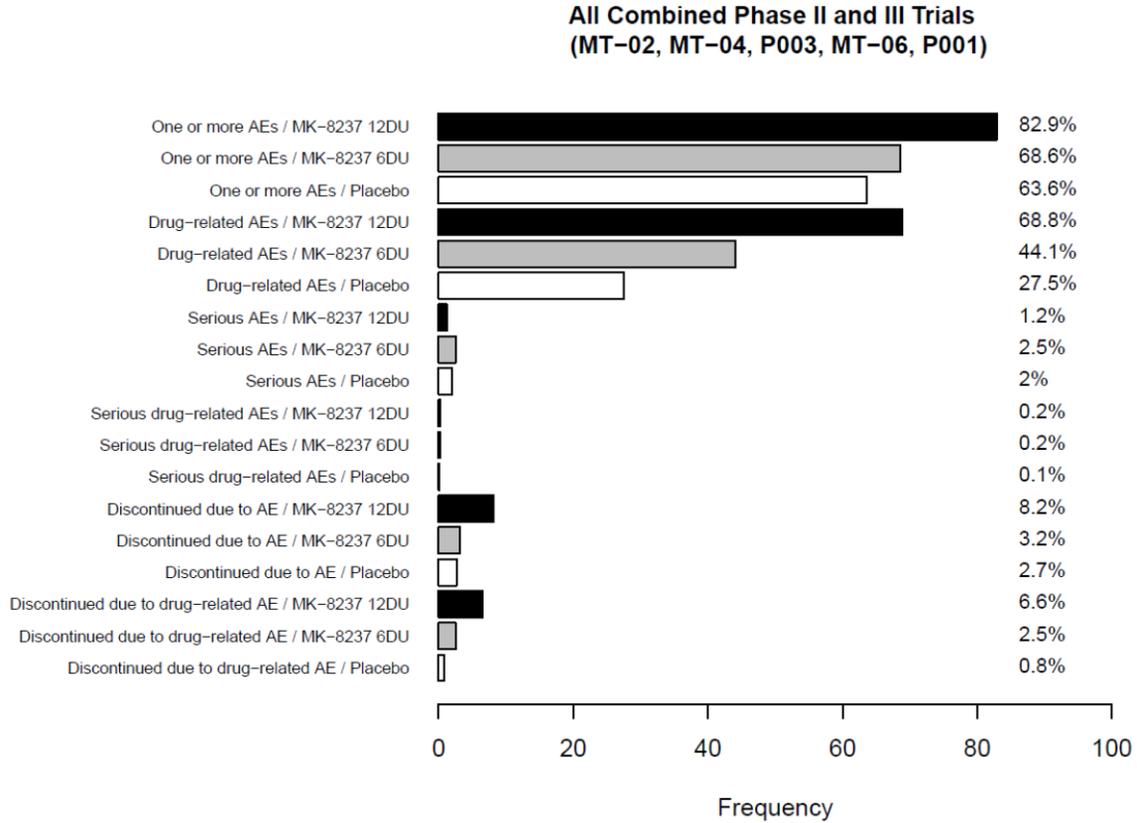
8.4.3 Study Dropouts/Discontinuations

Across the 5 Phase II and III trials, 139 (6.3%) subjects receiving MK-8237 (113 [8.2%] on 12 DU and 26 [3.2%] on 6 DU) discontinued from a trial because of an adverse event, as compared to 41 (2.7%) on placebo. A numerically greater proportion of subjects in the MK-8237 12 DU group discontinued from a trial due to an adverse event as compared to the 6 DU or placebo groups. The majority of events leading to discontinuation were adverse events associated with treatment. The most frequently reported adverse events that resulted in discontinuation were throat irritation, oral pruritus, swollen tongue, mouth swelling, and ear pruritus for the MK-8237 groups (12 DU and 6 DU) and asthma, dyspnea, and urticaria for the placebo group. The majority of discontinuations occurred within approximately 28 days of the first dose of study drug.

8.4.4 Common Adverse Events

- All Combined Phase II and III Trials: The reported occurrence of adverse events was higher for the MK-8237 12 DU group (82.9%) compared to the 6 DU (68.6%) and placebo groups (63.6%) (Figure 5). There were more drug-related adverse events reported in the MK-8237 groups (12 DU [68.8%] and 6 DU [44.1%]) as compared to the placebo group [27.5%], with the highest percentage of subjects reporting drug-related adverse events in the MK-8237 12 DU group. The most frequently reported drug related adverse events were events localized in or around the mouth and throat.

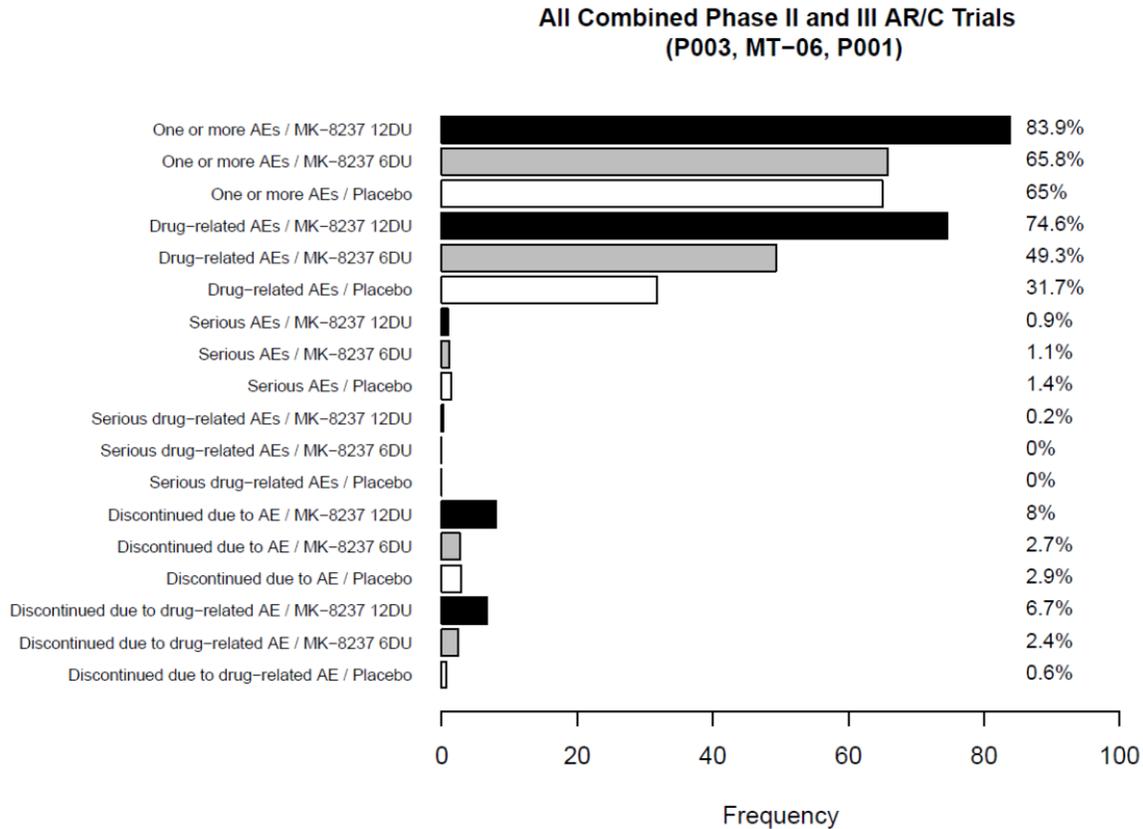
Figure 5 Adverse event summary for All Combined Phase II and III Trials (MT-02, MT-04, P003, MT-06, and P001)



Source: Generated by the reviewer based on Table 2.7.4: 12 in the Summary of Clinical Safety

- Combined Phase II and III AR/C Trials: The reported occurrence of adverse events was higher for the MK-8237 12 DU group (83.9%) compared to the 6 DU (65.8%) and placebo groups (65.0%) (Figure 6). There were more drug-related adverse events reported in the MK-8237 groups (12 DU [74.6%] and 6 DU [49.3%]) as compared to the placebo group [31.7%], with the highest percentage of subjects reporting drug-related adverse events in the MK-8237 12 DU group.

Figure 6 Adverse event summary for All Combined Phase II and III AR/C Trials (P003, MT-06, and P001)



Source: Generated by the reviewer based on Table 2.7.4: 13 in the Summary of Clinical Safety

- **Combined Phase II and III Asthma Trials:** The reported occurrence of adverse events was higher for the MK-8237 12 DU group (78.7%) and the 6 DU (71.0%) as compared to placebo groups (59.8%). There were more drug-related adverse events reported in the MK-8237 groups (12 DU [46.1%] and 6 DU [39.4%]) as compared to the placebo group [16.2%], with the highest percentage of subjects reporting drug-related adverse events in the MK-8237 12 DU group.

The applicant performed demographic subgroup analysis on All Combined Phase II and III Trials.

- **Age:** The majority (82.3%) of subjects were 18 to < 50 years of age. Subjects 12 to <18, 50 to <65, and ≥65 years of age comprised 5.6%, 10.9%, and 1.2% of the overall pool, respectively. The reported frequency of adverse events was higher for the MK-8237 groups (12 DU and 6 DU) as compared to the placebo group in all age groups. Overall, the safety profiles were similar across all age groups.
- **Gender:** Numerically more females reported adverse events across all treatment groups. For both genders, a higher proportion of subjects in the MK-8237 groups reported adverse events as compared to the placebo group.

- Race and ethnicity: The majority of subjects were white and not of Hispanic or Latino ethnicity. Due to the small number of subjects of other races or ethnicities, no meaningful conclusions can be drawn.

8.4.5 Clinical Test Results

Overall, for the trials with baseline and end-of-trial laboratory test data, evaluation of the laboratory values by treatment group revealed no clinically relevant changes over the course of any of the trials.

8.4.6 Systemic Adverse Events

In All Combined Phase II and III Trials, seven subjects (0.2%) reported systemic allergic reactions, 4 (0.3%) on MK-8237 12 DU and 3 (0.2%) on placebo. Only 1 of the subjects on MK-8237 had a systemic allergic reaction assessed by the investigator as drug related.

8.4.7 Local Reactogenicity

The most frequently reported adverse events for MK-8237 are those adverse events categorized as local application site reactions. In the All Combined Phase II and III Trials, after the first dose of study drug, the median time to onset of any local application site reaction ranged from 1 to 7 days in the MK-8237 12 DU group and from 1 to 176 days in the 6 DU group, as compared to placebo from 1 to 6 days.

In the All Combined Phase II and III Trials, five non-serious adverse events of severe local swelling in 5 subjects (0.2%) were reported for subjects receiving MK-8237 12 DU.

8.4.8 Adverse Events of Special Interest

N/A

9. ADDITIONAL STATISTICAL ISSUES

N/A

10. CONCLUSIONS

10.1 Statistical Issues and Collective Evidence

During the End of Phase II meeting, the applicant and CBER agreed that the pivotal North American study (Study P001) would be designed to demonstrate that, for the primary efficacy endpoint – average Total Combined Rhinitis Score (TCRS) during the last 8 weeks of treatment, the upper bound of the 95% CI for the treatment difference of MK-8237 12 DU relative to placebo is lower than or equal to -10%. As a result, the study showed that the difference between MK-8237 12DU and placebo was statistically significantly different from zero ($p < 0.001$); the percent treatment difference relative to placebo was -17.2% (95% CI, -25.0%, -9.7%). The upper bound of the 95% CI of relative treatment effect, calculated from the Full Analysis Set (FAS) using the bootstrap method, slightly missed the predefined threshold of -10%. The reviewer performed 500

runs of bootstrap analysis on the FAS to assess the potential impact of fluctuation of bootstrapping process on the estimation of the upper bound of the 95% CI of relative treatment effect (Figure 2). The analysis showed that the estimate of the upper bound ranged from -8.70% to -10.20%. The median was -9.68%, which was very close to the applicant's estimate. Additionally, the applicant's sensitivity analysis based on the Per-Protocol (PP) population showed that percent treatment difference relative to placebo was -16.7% (95% CI, -25.5%, -8.5%), using the bootstrap method. Other sensitivity analyses reported that the upper bounds of the 95% CIs of percent relative treatment difference ranged from -6.5% to -8.8%.

The applicant also provided studies MT-06 (P015) and P003 as additional evidence. The European Phase III field study MT-06 showed that the difference between MK-8237 12DU and placebo in average TCRS during the last 8 weeks of treatment was statistically significantly different from zero ($p < 0.01$); the percent treatment difference of MK-8237 12DU relative to placebo was -16.1% (95% CI, -25.8%, -5.7%) based on the FAS with multiple imputation, or -18.1% (95% CI, -27.6%, -7.7%) based on the FAS with observations. The upper bound of the 95% CI of relative treatment difference did not meet the criterion for U.S. registration. The Phase II EEC study P003 showed that the difference between MK-8237 12DU and placebo in the average total nasal symptom score (TNSS) determined during the chamber session at Week 24 was statistically significantly different from zero ($p < 0.001$); the percent treatment difference of MK-8237 12DU relative to placebo was -48.6% (95% CI, -60.2%, -35.3%).

The reviewer defers to the Medical Officer for further consideration of the totality of the evidence.

10.2 Conclusions and Recommendations

Efficacy: In the pivotal North American study (Study P001), the primary efficacy endpoint was close to but slightly missed the pre-defined success criterion. As additional evidence, the Phase II EEC study P003 showed substantial treatment difference between MK-8237 12DU and placebo in the average total nasal symptom score (TNSS).

Safety: In the pivotal North American Study P001, there were more subjects with adverse events (AEs), drug-related AEs, discontinuations due to AEs, and discontinuations due to drug-related AEs in the MK-8237 12 DU group than in the placebo group. The integrated safety analysis showed a similar trend in the safety findings.

The reviewer defers to the medical officers for further consideration of the totality of evidence for efficacy and safety of this product.