

Official Title: A Phase IIB, Randomized, Double-Blind, Placebo-Controlled, Multicenter, Dose-Ranging Study to Assess the Efficacy and Safety of MSTT1041A in Patients with Uncontrolled Severe Asthma

NCT Number: NCT02918019

Document Date: SAP Version 2: 5 April 2019

STATISTICAL ANALYSIS PLAN

TITLE: A PHASE IIB, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER, DOSE-RANGING STUDY TO ASSESS THE EFFICACY AND SAFETY OF MSTT1041A IN PATIENTS WITH UNCONTROLLED SEVERE ASTHMA

PROTOCOL NUMBER: GB39242

STUDY DRUG: MSTT1041A (RO7187807)

VERSION NUMBER: 2

IND NUMBER: 129714

EUDRACT NUMBER: 2016-001549-13

SPONSOR: Genentech, Inc.

PLAN PREPARED BY: [REDACTED]

DATE FINAL: Version 1: 20 May 2017

DATE AMENDED Version 2: See electronic date stamp below

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

Name	Reason for Signing	Date and Time (UTC)
[REDACTED]	Company Signatory	05-Apr-2019 08:05:56

CONFIDENTIAL

This is a Genentech, Inc. document that contains confidential information. Nothing herein is to be disclosed without written consent from Genentech, Inc.

STATISTICAL ANALYSIS PLAN AMENDMENT

RATIONALE

The Statistical Analysis Plan (SAP) has been amended primarily to align with recently amended Protocol version 4. Major changes, along with a rationale for each change, are summarized below:

- The initial timepoint for measurement of efficacy endpoints has been specified as baseline. This is to accommodate the different baseline definitions in efficacy endpoints deemed appropriate and clinically relevant. The definition of baseline for each endpoint is the randomization visit (Week 2) unless otherwise specified in Section 4.
- The exploratory efficacy endpoint of change in nighttime symptoms has been removed, as this endpoint was included in error in the previous protocol version. Data from the nighttime symptom questions of the Asthma Daily Symptom Diary (ADSD) will be used in separate analyses to explore psychometric properties of the ADSD (Section 2.1).
- Type I Error management plan has been changed to implement a fixed sequence method. The revised testing hierarchy has been specified in Section 4.4.

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

TABLE OF CONTENTS

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE.....	2
1. BACKGROUND	5
2. STUDY DESIGN	5
2.1 Outcome Measures	7
2.1.1 Primary Efficacy Endpoint.....	7
2.1.2 Secondary Efficacy Endpoints.....	7
2.1.3 Exploratory Efficacy Endpoints	8
2.1.4 Safety Endpoints	8
2.1.5 Pharmacokinetic Endpoint.....	9
2.1.6 Exploratory Biomarker Endpoints	9
2.1.7 Exploratory Health Status Endpoints	9
2.2 Determination of Sample Size	9
2.3 Analysis Timing and Unblinding	9
2.4 Data Included in Week 54 Data Cut	10
3. STUDY CONDUCT	10
3.1 Randomization Issues	10
3.2 Data Monitoring	11
4. STATISTICAL METHODS	11
4.1 Analysis Populations	11
4.1.1 Modified Intend-to-Treat Population	11
4.1.2 Safety-Evaluable Population.....	12
4.1.3 Run-In Safety-Evaluable Population.....	12
4.2 Definition of Baseline.....	12
4.3 Analysis of Study Conduct.....	13
4.4 Analysis of Treatment Group Comparability	13
4.5 Efficacy Analysis.....	13
4.5.1 Primary Efficacy Endpoint.....	16
4.5.2 Secondary Efficacy Endpoints.....	17
4.5.3 Exploratory Efficacy Endpoints	22
4.5.4 Sensitivity Analyses.....	24

4.5.5	Subgroup Analyses	26
4.5.6	Biomarker Analyses.....	27
4.6	Pharmacokinetic and Pharmacodynamic Analyses	28
4.7	Safety Analyses	29
4.7.1	Exposure to Study Medication.....	30
4.7.2	Adverse Events	30
4.7.3	Laboratory Data	31
4.7.4	ECG Results.....	31
4.7.5	Vital Signs.....	32
4.8	Missing Data.....	32
4.9	Interim Analyses	32
5.	REFERENCES	33

LIST OF TABLES

Table 1	Fixed Sequence Testing Hierarchy	15
Table 2	Cutoff Values for Summary of Vital Signs	32

LIST OF FIGURES

Figure 1	Study Schema.....	7
----------	-------------------	---

LIST OF APPENDICES

Appendix 1	Protocol Synopsis	34
Appendix 2	Schedule of Assessments.....	46
Appendix 3	Schedule of Additional Pharmacokinetic Samples for Selected Patients.....	50
Appendix 4	Scoring of the St. George's Respiratory Questionnaire (SGRQ).....	51
Appendix 5	Definition of Asthma Exacerbation for Inclusion Criteria and Reporting of Adverse Events	56

1. BACKGROUND

The purpose of this document is to provide details of the planned analyses for Study GB39242. The primary focus of this Statistical Analysis Plan (SAP) is on analyses of complete data from the first 54 weeks of the study, which includes the 2-week single-blind placebo run-in as well as the 52-week double-blind treatment period. Such analyses are described in the main body of this document and will be performed after all patients have completed 54 weeks in the study or discontinued early, and all corresponding data have been entered into the database, reviewed, and verified. The 16-week follow-up period of the study may be ongoing at the time of the 54-week analysis. No changes to the SAP will be allowed at the time of or subsequent to the 54-week primary analysis.

The analyses specified in this document supersede the high-level analysis plan described in the protocol.

2. STUDY DESIGN

Study GB39242 is a Phase IIb, randomized, placebo-controlled, double-blind, multicenter, multi-arm study of MSTT1041A compared with placebo as add-on therapy in patients with severe, uncontrolled asthma who are receiving medium- or high-dose inhaled corticosteroids (ICS) therapy and at least one of the following additional controller medications: long-acting β -agonist, leukotriene modifier, long-acting muscarinic antagonist, or long-acting theophylline preparation. Patients must have evidence of uncontrolled disease consisting of an Asthma Control Questionnaire-five items (ACQ-5) score of ≥ 1.5 and at least one symptom of asthma that is not controlled (nighttime awakening ≥ 1 time/week and/or short-acting rescue therapy use > 2 days/week). Patients requiring use of systemic corticosteroids (oral, intravenous, or intramuscular) or biologic therapy (e.g., anti-Immunoglobulin E [IgE] or anti-interleukin-5) at screening were excluded from the study.

The study consists of a 2- to 4-week screening period, a 2-week single-blind placebo run-in period, a 52-week double-blind treatment period, and a 16-week safety follow-up period concluding with the end-of-study visit at Week 70 (see [Figure 1](#)). At the run-in visit (Week 0), scheduled for approximately 2 weeks prior to the randomization visit (Week 2), patients who met enrollment criteria for the run-in period received one single-blind dose of placebo to allow for evaluation of unexpected variability in asthma control. At the randomization visit, patients underwent further assessments to determine eligibility for randomization to the double-blind treatment period. The 52-week double-blind treatment period begins with the randomization visit (Week 2) and ends with the Week 54 visit. Patients who experienced unexpected variability in asthma control during the run-in period, as demonstrated by change in forced expiratory volume in 1 second (FEV₁) and/or fractional exhaled nitric oxide (FeNO), were not eligible for

randomization into the 52-week double-blind treatment period. Additional details are provided in the study protocol.

Eligible patients were randomized in a 1:1:1:1 ratio to receive MSTT1041A at one of three doses (70, 210, or 490 mg) or placebo. Enrollment caps were utilized to ensure adequate power for biomarker subgroup analysis based upon blood eosinophil status at visit 1 (≥ 300 , < 300 cells/ μ L). The randomization was conducted such that approximately 30 patients per arm will have eosinophil-high status and approximately 95 patients per arm will have eosinophil-low status. Study drug is administered as four SC abdominal injections at the randomization visit (Week 2), Week 6, and every 4 weeks thereafter through Week 50.

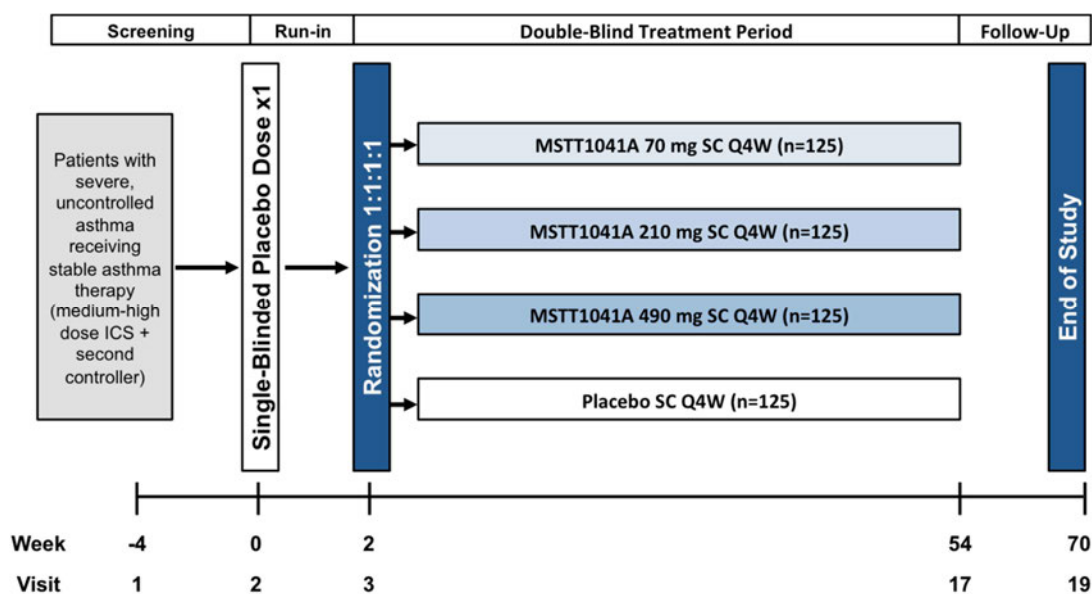
Safety and efficacy, as well as pharmacokinetic (PK) and pharmacodynamic (PD) measures, are assessed throughout the double-blind treatment period, as described in the protocol and detailed in the Schedule of Assessments (see [Appendix 2](#)). The primary efficacy endpoint is the incidence of asthma exacerbations during the 52-week double-blind treatment period (Week 2 through Week 54).

All patients will be followed for 20 weeks after the last dose of study treatment. The 20 weeks include 4 weeks after the final dose in the double-blind treatment period and 16 weeks in the safety follow-up period. Assessments for the safety follow-up period are detailed in the Schedule of Assessments (see [Appendix 2](#)).

All patients who discontinue study drug early during the double-blind treatment period are asked to continue with the study assessments in the double-blind treatment period and then complete the safety follow-up period. Patients who are unable or unwilling to continue with assessments in the double-blind treatment period complete an early termination visit and then enter the safety follow-up period.

The study design is depicted in [Figure 1](#).

Figure 1 Study Schema



ICS=inhaled corticosteroid; Q4W=every 4 weeks; SC=subcutaneous.

2.1 OUTCOME MEASURES

2.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint is as follows:

- Incidence of asthma exacerbations from baseline through Week 54, with asthma exacerbation defined as new or increased asthma symptoms (wheezing, coughing, dyspnea, chest tightness, and/or nighttime awakenings due to these symptoms) that result in one or both of the following:
 - Hospitalization or an emergency department visit with administration of systemic corticosteroids treatment
 - Treatment of systemic corticosteroids for ≥ 3 days or a long-acting depot corticosteroid preparation with a therapeutic effectiveness of ≥ 3 days

2.1.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are as follows:

- Absolute change in pre-bronchodilator FEV₁ (liters) from baseline to Week 54
- Time to first asthma exacerbation during the 52-week double-blind treatment period
- Achievement of improvement Standardized Asthma Quality-of-Life Questionnaires (AQLQS) score, defined as an increase of ≥ 0.5 points from baseline to Week 54
- Achievement of improvement in ACQ-5 score at Week 54, defined as a decrease of ≥ 0.5 points from baseline to Week 54
- Absolute change in patient-reported use of short-acting rescue therapy from baseline to Week 54

- Proportion of weeks without patient-reported asthma-related nighttime awakenings from baseline through Week 54
- Absolute change in patient-reported daytime asthma symptom severity as measured by the Asthma Daily Symptom Diary (ADSD) from baseline to Week 54

2.1.3 Exploratory Efficacy Endpoints

The exploratory efficacy endpoints are as follows:

- Incidence of asthma exacerbations from baseline through Week 54, with asthma exacerbation defined as in [Appendix 5](#)
- Incidence of severe asthma exacerbations from baseline through Week 54, with severe asthma exacerbation defined as asthma symptoms requiring hospitalization or resulting in death attributed to asthma
- Incidence of asthma exacerbations from randomization visit through Week 70
- Relative change in pre-bronchodilator FEV₁ (liters) from baseline to Week 54
- Absolute change in pre-bronchodilator FEV₁ (percentage predicted) from baseline to Week 54
- Absolute change in pre-bronchodilator FEV₁ (liters) from baseline to Week 70
- Achievement of improvement in St. George's Respiratory Questionnaire (SGRQ) score, defined as a decrease of ≥ 4 points from baseline to Week 54
- Achievement of improvement in ACQ-7 score, defined as a decrease of ≥ 0.5 points from baseline to Week 54
- Clinician's global impression of change (CGIC) in patient's asthma symptoms, as assessed through use of the CGIC, from baseline to Week 26 and Week 54
- Incidence of asthma exacerbations from baseline through Week 54 within each of the eosinophil-high (≥ 300 cells/ μ L) and eosinophil-low (< 300 cells/ μ L) groups
- Absolute change from baseline to Week 54 in pre-bronchodilator FEV₁ (liters) within each of the eosinophil-high (≥ 300 cells/ μ L) and eosinophil-low (< 300 cells/ μ L) groups
- Incidence of asthma exacerbations from baseline through Week 54 by *IL1RL1* genotype
- Absolute change in pre-bronchodilator FEV₁ (liters) from baseline to Week 54 by *IL1RL1* genotype

2.1.4 Safety Endpoints

The safety endpoints are as follows:

- Incidence and severity of adverse events, with severity determined through use of the World Health Organization-Adverse Drug Reaction Terminology (WHO-ART)
- Change from baseline in vital signs, ECGs, and clinical laboratory results
- Incidence of anti-drug antibodies (ADAs)

2.1.5 Pharmacokinetic Endpoint

The PK endpoint is as follows:

- Serum concentration of MSTT1041A at specified timepoints

2.1.6 Exploratory Biomarker Endpoints

The key exploratory biomarker endpoints are as follows:

- Relative change in FeNO and blood eosinophils from baseline through Week 54
- Absolute levels of the soluble form of ST2-(sST2) in blood from baseline through Week 70

2.1.7 Exploratory Health Status Endpoints

The exploratory health status endpoint is as follows:

- Change in health status utility as assessed by the EuroQol 5-Dimension Questionnaire (EQ-5D-5L) from baseline to Week 54

2.2 DETERMINATION OF SAMPLE SIZE

A total of 502 patients were randomly allocated in a 1:1:1:1 ratio to receive one of three doses of MSTT1041A or placebo. This sample size provides approximately 80% power to detect a 40% reduction in the annualized asthma exacerbation rate (AER) between one MSTT1041A arm and the placebo arm, assuming 0.63 exacerbations per patient per year in the placebo arm, no Poisson over dispersion, a 15% dropout rate, and a two-sided significance level (α) of 0.05.

Enrollment caps are utilized to ensure adequate power for biomarker subgroup analysis based upon blood eosinophil status at visit 1. Approximately 30 eosinophil-high patients (≥ 300 cells/ μ L) per arm and approximately 95 eosinophil-low patients (< 300 cells/ μ L) per arm are randomly allocated. The sample size provides approximately 80% power to detect a 50% reduction in the annualized AER for the subgroup of eosinophil-high patients assuming 1.0 exacerbations per patient per year in the placebo arm, a 15% dropout rate, a two-sided significance level of 0.15, and 30 patients in each treatment arm having eosinophil-high status. The sample size also provides approximately 67% power to detect a 35% reduction in the annualized AER for the subgroup of eosinophil-low patients, assuming 0.5 exacerbations per patient per year in the placebo arm, no Poisson over dispersion, a 15% dropout rate, a two-sided significance level of 0.15, and 95 patients in each treatment arm having eosinophil-low status.

2.3 ANALYSIS TIMING AND UNBLINDING

Analysis of data from the 52-week double-blind treatment period of the study, defined as the time of randomization (Week 2) through the Week 54 visit, will be performed when all patients either (i) have completed the Week 54 visit, (ii) have discontinued treatment early from the 52-week double-blind treatment period of the study but completed follow-up visits up to 52-weeks post-randomization, or (iii) have withdrawn from the

study prior to Week 54, and all data from the first 54 weeks of the study are in the database and have been reviewed.

At the time of the primary analysis, the Sponsor personnel who are analyzing data from the first 54 weeks of the study will be unblinded to treatment assignment. This includes personnel directly involved in the statistical analyses and programming activities as well as Sponsor personnel from other functions (e.g., Clinical Science, Safety Science, Clinical Pharmacology, and Regulatory Affairs) who will be involved in assessing, summarizing, and interpreting the data.

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

Aggregate results of the primary analysis, summarized by treatment arm and potentially biomarker group, may be reported to the public before completion of the study. However, patients and study site personnel will remain blinded to individual treatment assignment until after the study is completed (after all patients have either completed the safety follow-up period or discontinued early from the study), the database is locked, and the study analyses are final.

2.4 DATA INCLUDED IN WEEK 54 DATA CUT

The following data will be included in the Week 54 data cut:

- **For patients randomized to the 52-week double-blind treatment period:** All screening, run-in and post-baseline data with a clinical date (i.e., administration/assessment/onset/start date) on or before the date of randomization to Week 54 will be included.
- **For patients not randomized to the 52-week double-blind treatment period:** All screening and run-in data with a clinical date (i.e., administration/assessment/onset/start date).

The Week 54 data cut will include all data regardless of the type of study visit at which it was collected. This may include data collected at unscheduled visits, dosing termination visits, early termination visits, or safety follow-up visits, if the visit date was on or before the Week 54 data cutoff date.

3. STUDY CONDUCT

3.1 RANDOMIZATION ISSUES

Following successful completion of the run-in period, at the Week 2 visit, patients were randomly allocated to one of four treatment arms (MSTT1041A 70 mg, MSTT1041A 210 mg, MSTT1041A 490 mg, or placebo) in a 1:1:1:1 ratio through an interactive voice or web-based response system (IxRS). Randomization was stratified by blood

eosinophil status at visit 1 (< 150, ≥150 to <300, ≥300 cells/μL), number of documented asthma exacerbations (as defined in [Appendix 5](#)) in the previous 12 months (1-2, ≥3), total daily ICS dose at visit 1 (< 1000 μg, ≥ 1000 μg of fluticasone proportionate or equivalent), and country. A dynamic randomization method is used to obtain an approximately 1:1:1:1 ratio between the four treatment arms and within each stratum.

During both the run-in and the 52-week double-blind treatment periods, the IxRS will make study-drug kit assignments. Patient randomization and the study-drug kit assignments will be verified on an ongoing basis by an external and independent data coordinating center (iDCC). The iDCC independently reviews the logs to ensure that randomization and kit assignments are conducted correctly by the IxRS.

3.2 DATA MONITORING

An independent Data Monitoring Committee (iDMC) will monitor safety and study conduct on an ongoing basis. Members of the iDMC are external to the Sponsor and follow procedures that are detailed in a Charter that outlines the iDMC roles and responsibilities. The iDMC will meet approximately every 3 months to review unblinded safety and study conduct data prepared by the iDCC. In addition, the iDMC or the Sponsor may request ad hoc reviews at any time to address potential safety concerns.

Safety monitoring reviews include unblinded evaluation of all adverse events, serious adverse events, adverse events of special interest, major protocol deviations, ECG, and laboratory data. If the iDMC deems a benefit-risk assessment necessary, the iDMC may also review unblinded efficacy data. The iDMC may recommend stopping the study early for safety reasons. However, the iDMC may not recommend stopping the study early for positive efficacy or solely for futility. Formal stopping guidelines for assessing safety or the balance between the risks and benefits related to continuing the study will not be provided by the Sponsor. The iDMC will use its collective judgement to recommend early termination of the study if the data indicate an unacceptable safety profile. Details will be provided in the iDMC Charter.

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATIONS

Three analysis populations are defined for this study: the modified intend-to-treat (mITT) population, the safety-evaluable population, and the run-in safety-evaluable population.

4.1.1 Modified Intend-to-Treat Population

The mITT population will include all patients who were randomly allocated and received at least one dose of study drug (MSTT1041A or placebo) during the 52-week double-blind treatment period (Week 2 through Week 54). For analyses based on this population, patient treatment groups will be defined according to the treatment that was assigned at randomization (placebo, MSTT1041A 70 mg, MSTT1041A 210 mg, or MSTT1041A 490 mg).

4.1.2 Safety-Evaluable Population

The safety-evaluable population will include all patients who were randomly allocated and received at least one dose of study drug (MSTT1041A or placebo) during the 52-week double-blind treatment period (Week 2 through Week 54) with treatment groups defined according to the actual treatment received.

Because each patient can receive up to 13 doses of study drug during the 52-week double-blind treatment period, the treatment received will be defined as the study treatment (placebo, MSTT1041A 70 mg, MSTT1041A 210 mg, or MSTT1041A 490 mg) most frequently given to the patient during the double-blind treatment period. In the event that this is not unique, treatment received will be defined as the highest of the most frequently given dose levels (e.g., a patient who receives 1 dose of placebo and 1 dose of MSTT1041A 70 mg would be assigned to the MSTT1041A 70 mg treatment group for analyses based on the safety-evaluable population).

It is expected that most patients will receive their assigned treatment at all dosing timepoints and that only a small number of dosing errors, if any, will occur. Nonetheless, each dosing error will be reported as a major protocol deviation and reviewed in conjunction with the corresponding patient's safety data to assess for any impact on the overall safety conclusions.

4.1.3 Run-In Safety-Evaluable Population

The run-in safety-evaluable population will include all patients who received the single-blind placebo dose during the run-in period (Week 0 treatment), regardless of whether or not the patient is randomly allocated into the double-blind treatment period. For safety analyses based on this population, only events that occur during the 2-week run-in period will be included. As placebo should be the only study drug administered during the run-in period, all patients will be grouped together for analyses based on this population.

In the rare event that a patient mistakenly receives MSTT1041A rather than placebo during the run-in visit, each dosing error will be reported as a major protocol deviation and reviewed in conjunction with the corresponding patient's safety data to assess for any impact on the overall safety conclusions.

4.2 DEFINITION OF BASELINE

Baseline is defined as the last available pre-treatment value taken on or before the day of randomization (Week 2 visit) for all assessments, with the exception of the following efficacy endpoints: patient-reported use of short-acting rescue therapy, and patient-reported daytime asthma symptom severity as measured by Asthma Daily Symptom Diary (ADSD). For these outcome measures, details for derivation of baseline values can be found in [Section 4.5.2](#).

Baseline measurements will be used for the summary of demographic characteristics, as well as for all change-from-baseline analyses of efficacy, safety, and PD endpoints (including change from baseline in FEV1, CGIC, blood eosinophil count, and FeNO, and improvement from baseline in AQLQ(S), ACQ-5, SGRQ, ACQ-7).

For endpoints that are defined in terms of change from baseline or improvement from baseline, patients who do not have a baseline measurement will be excluded from the analyses.

4.3 ANALYSIS OF STUDY CONDUCT

The number of patients randomized will be tabulated by region, country, study site, and treatment arm. Patient disposition (the number of patients receiving the single-blind placebo dose during run-in, randomly allocated, receiving at least one dose of study drug during the double-blind treatment period, completing study treatment during the double-blind treatment period, and completing study visit assessments through Week 54) and time on study will be tabulated by treatment arm. Reasons for premature discontinuation from study treatment and reasons for premature discontinuation from the study prior to Week 54 will be summarized. Eligibility criteria deviations, dosing errors, and other major protocol deviations will be summarized.

4.4 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographic and baseline characteristics including but not limited to age, sex, race, ethnicity, blood eosinophil count, number of exacerbations in the prior year, concomitant asthma medication use, pulmonary function, and asthma control (as measured by the ACQ-5), will be summarized for the mITT population by treatment arm using descriptive statistics. Exposure to study drug (number of study drug treatments and duration of treatment) will be summarized for the double-blind treatment period by treatment arm.

4.5 EFFICACY ANALYSIS

Efficacy analyses will be conducted on the mITT population (see Section 4.1.1), with patients grouped according to the treatment assigned at randomization.

Efficacy summaries for the 52-week double-blind treatment period will include data from patients who discontinued study drug early but continued with study assessments and may include data collected at unscheduled visits, early termination visits, or safety follow-up visits, if the visit date occurred within 52-weeks after the randomization visit. For patients completing the Week 54 visits after the schedule Week 54 windows, their Week 54 visits will be considered out-of-window and the corresponding data will be treated as missing.

Comparisons of Interest

Comparisons of efficacy will be performed between each MSTT1041A dose level and the placebo group.

Thus, there will be three comparisons:








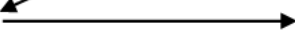
- MSTT1041A 70 mg with placebo
- MSTT1041A 210 mg with placebo
- MSTT1041A 490 mg with placebo

Hypothesis testing and estimation of treatment effects will be performed with regression models that are fit using data from all four treatment arms and which include three indicator variables to characterize the three treatment comparisons described above.

Type I Error Management

Type I error will be controlled using a fixed sequence testing procedure at the 5% level of significance (two-sided) for primary and secondary efficacy endpoints. For each endpoint the testing procedure will proceed with comparisons from highest to lowest MSTT1041A dose level versus placebo (see [Table 1](#)) in the order specified below.

Table 1 Fixed sequence testing hierarchy

		MSTT1041A		
Endpoint		490 mg	210 mg	70 mg
Primary	Incidence of asthma exacerbation from baseline through Week 54			
Secondary	1. Time to first asthma exacerbation during the 52-week double-blind treatment period			
	2. Absolute change in pre-bronchodilator FEV1 (liters) from baseline to Week 54			
	3. Achievement of improvement in ACQ-5 score at Week 54, defined as a decrease of ≥ 0.5 points from baseline to Week 54			
	4. Absolute change in patient reported daytime asthma symptom severity as measured by the Asthma Daily Symptom Diary (ADSD) from baseline to Week 54			
	5. Proportion of weeks without patient-reported asthma-related nighttime awakenings from baseline through Week 54			
	6. Absolute change in patient-reported use of short-acting rescue therapy from baseline to Week 54			
	7. Achievement of improvement Standardized Asthma Quality-of-Life Questionnaires (AQLQS) score, defined as an increase of ≥ 0.5 points from baseline to Week 54			

The testing procedure will be conducted at a two-sided significance level of 5%. Each hypothesis will be formally tested only if the preceding one is significant at 5% level.

No control of type I error will be applied for the exploratory endpoints, including sensitivity, subgroup, and biomarker analyses. Efficacy analyses will not involve formal comparisons between the three MSTT1041A treatment arms (70, 210, and 490 mg).

Covariate Adjustment

Unless otherwise noted, analyses of efficacy endpoints (primary, secondary, and exploratory) will be adjusted for each of the following covariates:

- Blood eosinophil level at visit 1 ($< 150, \geq 150$ to $< 300, \geq 300$ cells/ μ L)
- Number of asthma exacerbations requiring use of systemic corticosteroids within the 12 months prior to study entry (1–2, ≥ 3 events)
- Total daily ICS dose at visit 1 ($< 1000\mu\text{g}, \geq 1000 \mu\text{g}$ of fluticasone propionate or equivalent)
- Geographic region (North America [United States, Canada], Latin America [Argentina, Mexico, Peru], Central and Eastern Europe [Bulgaria, Czech Republic, Poland, Romania, Russia, Serbia, Ukraine], Western Europe plus Rest of World [Australia, Belgium, France, Germany, New Zealand, South Africa, South Korea]).

No missing data of these baseline covariates are expected. Stratification data collected in the IxRS system, rather than on the electronic Case Report Forms, will be used.

4.5.1 Primary Efficacy Endpoint

The primary efficacy objective for this study is to evaluate the efficacy of MSTT1041A compared with placebo on the basis of the following endpoint:

Incidence of asthma exacerbations from baseline through Week 54, with asthma exacerbation defined as new or increased asthma symptoms (wheezing, coughing, dyspnea, chest tightness, and/or nighttime awakenings due to these symptoms) that result in one or both of the following:

- Hospitalization or emergency department visit with administration of systemic corticosteroid treatment
- Treatment with systemic corticosteroids for ≥ 3 days or a long-acting depot corticosteroid preparation with a therapeutic effectiveness of ≥ 3 days

The estimand of interest, i.e., the mITT estimand, is the reduction in the rate of asthma exacerbations over 52 weeks (for MSTT1041A relative to placebo) for all randomized patients who received at least one dose of study drug, regardless of adherence to the assigned study treatment or to the protocol. Because the protocol specifies that patients who discontinue study drug early continue with study assessments, an analysis of all observed exacerbation data will lead to a reasonable estimate of the mITT estimand. However, some data may be missing because of early study discontinuation, and sensitivity of the primary analysis to missing data assumptions will be assessed by sensitivity analyses per Section 4.5.4.

For each patient, the time period at risk will be defined as follows:

- From the date of randomization to date of the actual Week 54 visit for patients who completed the double-blind treatment period

- From the date of randomization to the date of last assessment during the double-blind treatment period for patients who complete or discontinue from the study early

All protocol-defined asthma exacerbations with an onset date during this time period at risk will be included in the primary analysis. For each patient, the time at risk (in years) will be computed as the duration of this time period (last day minus first day [in days] plus 1) divided by 365.25. Unadjusted rates of asthma exacerbations will be estimated by dividing the total number of protocol-defined asthma exacerbations observed during the time at risk by the total time at risk (in years) for each treatment arm.

The rate of asthma exacerbations will be compared between each active treatment arm and the placebo arm with the use of a Poisson regression model with over-dispersion. The primary analysis will be on the basis of observed exacerbations, with no imputation for premature study discontinuation. The Poisson regression model will include terms for treatment arm (three indicator variables, corresponding to the three treatment arms) and the baseline covariates described in Section 4.5. Patient time at risk will be used as an offset term in the model. The dispersion parameter will be estimated by Pearson's χ^2 statistic divided by its degrees of freedom.

Rates of asthma exacerbations adjusted for specified baseline covariates (see Section 4.5) will be estimated for each treatment arm with the use of the Poisson regression model. The ratio of exacerbation rates (for each MSTT1041A arm compared with the placebo arm) will be estimated from the Poisson regression model along with associated 95% CIs and p-values. The absolute difference in the exacerbation rates (for each MSTT1041A arm compared with the placebo arm, expressed as events per year) will also be provided, along with the associated 95% CIs.

See Sections 4.5.4, 4.5.5, and 4.5.6 for sensitivity, subgroup, and additional biomarker analyses that will be performed for the primary efficacy endpoint, respectively.

4.5.2 Secondary Efficacy Endpoints

Time to First Asthma Exacerbation

A Cox proportional hazards regression model will be used to estimate the hazard ratio comparing each MSTT1041A treatment arm with the placebo arm with respect to time to the first asthma exacerbation. The model will include treatment arm and baseline covariates as described in Section 4.5. Patients who do not experience a protocol-defined asthma exacerbation during the 52-week double-blind treatment period will be censored at the end of their time at risk (see Section 4.5.1). Point estimates, 95% CIs, and p-values for the treatment effect (MSTT1041A vs. placebo) will be calculated for each MSTT1041A dose level based on the model.

Pre-Bronchodilator Forced Expiratory Volume in 1 Second (Liters)

The absolute change in pre-bronchodilator FEV₁ from baseline will be summarized by treatment arm and timepoint. The difference in mean absolute change in pre-bronchodilator FEV₁ at Week 54 from baseline between each MSTT1041A treatment arm and the placebo arm will be estimated using a mixed model for repeated measures (MMRM). The model will use absolute change from the randomization visit as the response variable and include terms for treatment arm, study visit, treatment arm by study visit interaction, baseline FEV₁ (as a continuous variable) as well as its interaction with study visit, and the baseline covariates as described in Section 4.5. Study visit will be included as a categorical variable. An unstructured covariance matrix will be specified to model the within-subject errors. Point estimates, 95% CIs, and p-values for the treatment effect (MSTT1041A vs. placebo) will be calculated for each MSTT1041A dose level on the basis of the model for all study visits, including Week 54.

In addition to evaluating the treatment benefit at Week 54, the absolute change from baseline in pre-bronchodilator FEV₁ over time will be characterized in order to establish the time of onset of sustained FEV₁ effect. Considering each dose level separately, if a statistically significant difference between MSTT1041A and placebo is observed at Week 54, then nominal tests for significance will be performed sequentially at each prior planned timepoint for which FEV₁ was assessed (working backward), continuing until the p-value for a timepoint is ≥ 0.05 . This procedure will be used to determine the earliest timepoint at which significance is observed at that timepoint and at all subsequent timepoints up to Week 54 (comparing MSTT1041A with placebo), and considered the time of onset of sustained FEV₁ improvement.

See Sections 4.5.4, 4.5.5, and 4.5.6 for sensitivity, subgroup, and additional biomarker analyses, respectively, that will be performed for the secondary endpoint of absolute change in pre-bronchodilator FEV₁ from the randomization visit at Week 54.

Asthma Quality of Life Questionnaire (Standardized)

Asthma-specific health-related quality of life will be measured by the AQLQ(S) scores. With this scale, higher scores indicate better health-related quality of life and 0.5 points is considered the minimally important difference (MID).

The AQLQ(S) domain scores will be calculated from the individual question responses as follows:

$$\text{Symptoms} = \sum (\text{Items } 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 29, 30) / 12$$

$$\text{Activity Limitation} = \sum (\text{Items } 1, 2, 3, 4, 5, 11, 19, 25, 28, 31, 32) / 11$$

$$\text{Emotional Function} = \sum (\text{Items } 7, 13, 15, 21, 27) / 5$$

$$\text{Environmental Stimuli} = \sum (\text{Items } 9, 17, 23, 26) / 4$$

The overall health-related quality of life-score will be calculated from the individual question responses as follows and will be summarized by treatment arm and timepoint for each biomarker group.

$$\text{AQLQ(S) Overall score} = \frac{\sum (\text{Items 1 to 32})}{32}$$

The proportion of patients who achieve a clinically meaningful improvement in AQLQ(S) (improvement of ≥ 0.5 points from baseline) at Week 54 will be summarized by treatment group and timepoint. AQLQ(S) at randomization visit (Week 2) will be used as the baseline AQLQ(S) score. A logistic regression model, with terms for treatment group, baseline AQLQ(S), and the baseline covariates described in Section 4.5 will be used to compare MSTT1041A with placebo. For patients without a Week 54 assessment of AQLQ(S), the Week 26 score will be used. Patients without a post-randomization assessment (Week 26 or Week 54) will be excluded from the analysis. Point estimates, 95% CIs, and p-values for the treatment difference (MSTT1041A vs. placebo) will be calculated for each MSTT1041A dose level.

Considering each dose level separately, if a statistically significant difference between MSTT1041A and placebo is observed at Week 54, then a test for significance will be performed at Week 26. This procedure will be used to determine the earliest timepoint at which significance is observed (comparing MSTT1041A with placebo), and considered the time of onset of sustained AQLQ(S) improvement.

Asthma Control Questionnaire

Asthma control, as measured by the ACQ, will be assessed by asking patients to recall their experience with asthma during the previous week. The full ACQ consists of seven questions: five questions related to symptoms (i.e., nighttime awakening, asthma symptoms upon awakening in the morning, activity limitation, shortness of breath, and wheezing frequency), one question about use of short-acting rescue therapy, and one question about FEV₁ (to be completed by site staff). The items are scored on a scale ranging from 0 (totally controlled) to 6 (extremely poorly controlled). The items are equally weighted. For the ACQ, higher scores indicate worse control, and 0.5 points is considered the MID. The ACQ can be scored two ways: 1) the mean of the five symptom-related items (ACQ-5); or 2) the mean of all 7 items (ACQ-7).

The proportion of patients who achieve a clinically meaningful improvement in asthma control based on the ACQ-5 (decrease of ≥ 0.5 points from baseline) will be analyzed using the same methodology as described for achievement of improvement from baseline at Week 54 in AQLQ(S). ACQ-5 at randomization visit (Week 2) will be used as the baseline ACQ-5 score.

Asthma Rescue Medication Use

Asthma rescue medication use as recorded in the patient's daily electronic diary (e-Diary) will be used for analysis. For each patient, baseline asthma rescue medication use will be defined as the proportion of days with rescue medication use over a 14-day period

prior to the Week 0 visit (i.e. during screening). Patients must have two consecutive compliant weeks during screening in their asthma rescue medication use to have a baseline score calculated. During screening, a patient is compliant in a week if they have at least 5 of 7 days of non-missing entries of asthma rescue medication use. The baseline proportion of days with asthma rescue medication use will first be calculated using the two consecutive screening compliant weeks prior to the run-in visit. However, if the condition of two consecutive screening compliant weeks does not hold in the 14 days prior to the run-in visit, the window will be shifted back by 7 days until a period of two consecutive screening compliant weeks is identified. As one of the inclusion criteria for patients to enter the run-in period is completion of the e-Diary for at least 5 of 7 days during each of 2 consecutive weeks during the screening period, i.e., patients having two consecutive screening compliant weeks, missing baseline score is not expected. In the rare event that a patient does not have a valid baseline score, the patient will be excluded from this analysis.

Although rescue medication is to be recorded each morning and each evening, any day on which at least one entry is collected will be included in analyses. For each patient and monthly post-baseline assessment timepoint (i.e., Week 6, Week 10, ...Week 54), the post-baseline asthma medication use for that timepoint will be defined as the proportion of days with rescue medication use over the 28 days on or prior to the timepoint (inclusive). Patients must have recorded their asthma rescue medication use at least once a day for at least 14 days during a 28-day interval to have a score calculated for the respective timepoint. Missing entries will be excluded from the calculation. For example, if a patient has only recorded rescue medication use for 20 of the 28 days, then the proportion of days will be calculated out of the 20 days with an entry.

The absolute change from baseline in the proportion of days with asthma rescue medication use will be summarized by treatment arm and timepoint. The absolute difference in mean change in proportion of days of rescue medication use from randomization at Week 54 between each treatment arm and the placebo arm will be estimated on the basis of an MMRM analysis. The model will use absolute change from baseline as the response variable and include terms for treatment arm, study visit, treatment arm by study visit interaction, baseline asthma rescue medication use (as a continuous variable) as well as its interaction with study visit, and the baseline covariates described in Section 4.5. Study visit will be included as a categorical variable. An unstructured covariance matrix will be specified to model the within-subject errors. Point estimates, 95% CIs, and p-values for the treatment difference (MSTT1041A vs. placebo) will be calculated for each MSTT1041A dose level based on the model for all study visits including Week 54.

Asthma-Related Nighttime Awakenings

Nighttime awakenings, as recorded each morning in the patient's daily e-Diary, will be used for analysis. For each patient, the proportion of weeks without a nighttime

awakening from baseline through Week 54 will be calculated. The randomization visit (Week 2) will be used as baseline.

From the randomization visit, the subsequent weeks will be counted by a 7-day window until Week 54 is reached. For each 1-week period after randomization visit, if there are 4 or more morning diaries with missing entries, the 1-week period is considered a non-compliant week and the corresponding value will be considered as missing. These missing values will not be imputed, and these missing weeks will not be included in the denominator when calculating the proportion of weeks. However, if there are >3 morning diaries with missing entries during a 1-week period but one of the available entries for the 1-week period reports a nighttime awakening, this 1-week period will be included in the calculation.

The proportion of weeks without a nighttime awakening due to asthma will be summarized by treatment arm. The absolute difference in mean proportion between each MSTT1041A dose level and the placebo arm will be estimated on the basis of an analysis of variance model. The model will use the proportion of weeks without a nighttime awakening due to asthma as the response variable and include terms for treatment arm and the baseline covariates described in Section 4.5. Point estimates, 95% CIs, and p-values for the treatment difference (MSTT1041A vs. placebo) will be calculated for each MSTT1041A dose level.

Daytime Asthma Symptom Severity

Daytime asthma symptom severity will be assessed by patient ratings of asthma symptoms each evening, as recorded in the ADSD on their e-Diary.

The daily daytime symptom scores are calculated each day as an average of the six core evening diary symptoms: difficulty breathing, wheezing, shortness of breath, chest tightness, chest pain and cough. Each question is scored from the scale of 0 to 10, in which higher scores indicate worse symptoms. The daytime symptom score for each timepoint is then calculated as a weekly average of the 7 individual average daily scores. Daily scores will only be calculated if data for 4 or more items per day are available.

Daily scores will be averaged over the previous week prior to each scheduled visit (i.e., 7 days prior to Week 54 visit). For each week after randomization visit, if there are >3 days with missing scores out of the 7 days, the week is considered a non-compliant week and the average score for the week will be considered as missing. These missing values will not be imputed but will be handled directly by the MMRM model.

The baseline value will be derived based on a 2-week screening period prior to the Week 0 visit. Patients must have two consecutive compliant weeks during screening to have a baseline score calculated. During screening, a patient is compliant in a week if they have at least 5 of 7 days of non-missing scores. The baseline score will first be calculated using the two consecutive screening compliant weeks prior to the run-in visit.

However, if the condition of two consecutive screening compliant weeks does not hold in the 14 days prior to the run-in visit, the window will be shifted back by 7 days until a period of two consecutive screening compliant weeks is identified. Patients without a 'valid' baseline score will be excluded from the analysis.

The absolute change from baseline in daytime symptom severity scores will be summarized by treatment arm and timepoint. The absolute difference in mean change in daytime symptom severity scores from baseline at Week 54 between each treatment arm and the placebo arm will be estimated on the basis of an MMRM analysis. The model will use absolute change from baseline as the response variable and include terms for treatment arm, study visit, treatment arm by study visit interaction, baseline symptom score as well as its interaction with study visit, and the baseline covariates described in Section 4.5. Study visit will be included as a categorical variable. An unstructured covariance matrix will be specified to model the within-subject errors. Point estimates, 95% CIs, and p-values for the treatment difference (MSTT1041A vs. placebo) will be calculated for each MSTT1041A dose level based on the model for all study visits including Week 54.

4.5.3 Exploratory Efficacy Endpoints

Asthma Exacerbation as Defined in Appendix 5

The incidence of asthma exacerbations, defined as in [Appendix 5](#), will be analyzed with the use of the same methodology as described for the primary endpoint (see Section 4.5.1).

Severe Asthma Exacerbations

The incidence of severe asthma exacerbations, defined as asthma symptoms requiring hospitalization or resulting in death attributed to asthma, will be analyzed with the use of the same methodology as described for the primary endpoint (see Section 4.5.1).

Asthma Exacerbations through Week 70

The incidence of severe asthma exacerbations from randomization through Week 70 will be analyzed with the use of the same methodology as described for the primary endpoint (see Section 4.5.1).

Relative Change in Pre-Bronchodilator Forced Expiratory Volume in 1 Second (Liters)

For each patient and timepoint, the relative change in pre-bronchodilator FEV₁ will be defined as the absolute change from the baseline in FEV₁ (liters) divided by the FEV₁ (liters) at baseline. The relative change in pre-bronchodilator FEV₁ from the baseline will be analyzed using the same methodology as described for the secondary endpoint of absolute change in pre-bronchodilator FEV₁ from baseline to Week 54, except that relative change from baseline will be the response variable.

Pre-Bronchodilator Forced Expiratory Volume in 1 Second (Percentage Predicted)

The absolute change in pre-bronchodilator FEV₁ (percentage predicted) from baseline to Week 54 will be analyzed using the same methodology as described for the secondary endpoint of absolute change in pre-bronchodilator FEV₁ from baseline to Week 54.

Pre-Bronchodilator Forced Expiratory Volume in 1 Second (Liters) at Week 70

The absolute change in pre-bronchodilator FEV₁ from baseline to Week 70 will be analyzed using the same methodology as described for the secondary endpoint of absolute change in pre-bronchodilator FEV₁ from baseline to Week 54.

St. George's Respiratory Questionnaire

The SGRQ is a 50-item respiratory-specific quality-of-life questionnaire initially developed and validated for use in chronic obstructive pulmonary disease. It includes questions that assess the impact of disease on symptoms, activity, and functionality. The symptom scale assesses the severity of respiratory symptoms, the activity scale examines impairment in patient activity as a result of respiratory symptoms, and the impact scale evaluates effects of respiratory symptoms on overall function and well-being. Each scale is scored from 0 to 100, and a total score represents the weighted average of these three components.

Each item in the SGRQ has specific weight (lowest possible weight=0, highest=100). Scores for the three scales and total are calculated using the assigned weights. The numerator and denominator for the total and each scale are comprised of individual item weights, minus specific scale weights. See [Appendix 4](#) for weighting, and numerator/denominator calculation details.

Total and scale scores will be calculated as follows:

$$\text{Total Score} = 100 * ([\sum \text{item weights}] / (\text{maximum total weight}))$$

$$\text{Symptom Score} = 100 * ([\sum \text{item 1,2,3,4,5,6,7,8 weights}] / (\text{maximum symptom total weight}))$$

$$\text{Activity Score} = 100 * ([\sum \text{item 11,15 weights}] / (\text{maximum activity total weight}))$$

$$\text{Impact Score} = 100 * ([\sum \text{item 9,10,12,13,14,16,17 weights}] / (\text{maximum impact total weight}))$$

The maximum number of missing values that can be tolerated for each scale are as follows:

- Symptoms: 2
- Activity: 4
- Impacts: 6

Missing item weights are subtracted from each scale's total weight for calculating that scale's score.

The proportion of patients who achieve a clinically meaningful improvement in the SGRQ (defined as a decrease of ≥ 4 points from baseline) will be analyzed using the same methodology as described for the secondary endpoint of achievement of improvement from baseline to Week 54 in AQLQ(S). SGRQ at randomization visit (Week 2) will be used as the baseline score.

Asthma Control Questionnaire

The achievement of improvement from randomization visit in ACQ-7 score at Week 54, defined as a decrease of ≥ 0.5 points, will be analyzed using the same methodology as described for the secondary endpoint of achievement of improvement from baseline to Week 54 in AQLQ(S). ACQ-7 at randomization visit (Week 2) will be used as the baseline score.

Clinician Global Impression of Change

The CGIC is a single-item assessment of the clinician's impression of a patient's change in asthma symptoms since beginning the 52-week double-blind treatment period. Change in asthma symptoms is rated on a 7-point scale ranging from very much worse to very much improved.

CGIC outcomes will be summarized on the basis of the proportion of patients with an outcome at each of the possible levels (i.e., "very much worse" to "very much improved") by treatment arm and timepoint. The proportion of patients with a "very much improved" or "much improved" CGIC score at Week 54 will be compared between each MSTT1041A arm and the placebo arm with the use of a logistic regression model with terms for treatment arm and the baseline covariates as described in Section 4.5. For patients without a Week 54 assessment, the Week 26 assessment will be used. Patients who have missing CGIC scores at both Weeks 26 and 54 will be excluded from analysis.

Biomarker Subgroup Analyses

Details about efficacy analyses by eosinophil and *IL1RL1* genotype subgroups are presented in Section 4.5.6.

4.5.4 Sensitivity Analyses

Sensitivity analyses will be performed to evaluate the robustness of the primary analysis results.

- **Parametric Model.** The primary efficacy endpoint will be re-analyzed using a negative binomial regression model, using the same covariates as specified for the primary analysis with Poisson regression. All other elements of the analysis will remain the same as in the primary analysis.

The following analyses will be performed to assess the sensitivity of results to missing data assumptions:

- **Missing Data Assumptions.** The primary endpoint (incidence of asthma exacerbations) and all secondary efficacy endpoints (except for time to first asthma exacerbation) found to be significant for a given MSTT1041A arm will be re-analyzed assuming that missing data is missing not at random, rather than missing at random (MAR), via tipping point sensitivity analyses implemented using multiple imputation. In general, imputation models will account for the same baseline covariates as used in the earlier described analysis of the endpoint and post-baseline outcome data when possible. Multiple imputed datasets will be grouped by a range of missing data assumptions and analyzed according to the same analysis models specified in Section 4.5, and the results will be combined for each assumption using Rubin's rules (Barnard and Rubin 1999). These analyses will be implemented as follows:
 - For continuous secondary endpoints that assess absolute changes from baseline to Week 54 (e.g., for pre-bronchodilator FEV1), imputation under the MAR assumption will be used to generate multiple datasets with imputed values. It will be assumed that the pattern of missing data in the outcome to be analyzed is monotone. In the event that the missing data do not follow a monotone pattern, multiple imputation using Markov chain Monte Carlo methods will be used to partially impute data under MAR, thereby converting the dataset to a monotone missing pattern. After imputation to achieve a monotone missing pattern, the remaining missing values will be imputed using a regression-based approach in a sequential manner with the use of univariate models, separately in each treatment group. For each visit, a regression model based on all patients who have observations (imputed or observed) available up to and including this visit will be fitted and used to impute missing values at this visit. All regression-based imputation models will include the baseline covariates described in Section 4.5 and prior observed or imputed outcomes.
 - Tipping-point analysis will be implemented by adding (or subtracting) a constant Δ to the post-dropout MAR imputed values; a separate Δ will be applied to each treatment arm (each MSTT1041A dose level and placebo). A range of Δ will be used to adjust post-dropout imputed values in the MSTT1041A and placebo arms independently based on a grid of Δ adjustments. For each endpoint, the size of the Δ will be chosen on the basis of the estimated treatment effect for a given MSTT1041A arm compared with placebo from the analyses specified in Section 4.5.2. The range will be chosen to include cases in which dropouts from the MSTT1041A arm have worse outcomes than dropouts from the placebo arm. For each endpoint, the resulting point estimates of treatment effect and corresponding p-values under each pair of Δ for each MSTT1041A arm and placebo comparison will be tabulated.
 - For the primary endpoint (incidence of asthma exacerbations), tipping point analysis via multiple imputation of event counts assuming a range of Δ will be performed. Event counts will be imputed at the patient level for the unobserved period between the last study visit and Week 54 for patients who discontinued study assessments

early. An event count corresponding to each unobserved period will be imputed by first fitting a negative binomial generalized linear model to the observed data and sampling from the posterior distribution of the model parameter estimates. These sampled parameter estimates along with the patient-specific observed covariate values, observed exacerbation counts, and length of unobserved interval will be used in conjunction with a random number generator to impute an event count for the unobserved time period, which is added to the observed event count (Keene et al. 2014). These augmented datasets will be analyzed according to the model described in Section 4.5.1 and the results combined using Rubin's rules.

- In the tipping point sensitivity analysis, the underlying event rate used to impute unobserved event counts for a given arm will be adjusted via a multiplier delta; a separate delta will be applied to each arm. A range of delta values will be used to adjust the underlying event rate used for imputation in the MSTT1041A and placebo arms independently based on a grid of delta adjustments. The delta values will be chosen on the basis of the treatment effect (rate ratio) observed for a given MSTT1041A arm compared with placebo from the analyses described in Section 4.5.1. The range of delta values will include cases where dropouts from the MSTT1041A arm are assumed to have worse outcomes than dropouts from the placebo arm. For each endpoint, the resulting point estimates of treatment effect and corresponding p-values under each pair of deltas for each MSTT1041A and placebo comparison will be tabulated.

Also, for the secondary endpoint (absolute change from baseline in FEV₁), a cumulative responder plot which displays, by treatment arm, the proportion of patients who achieve each specified absolute change in pre-bronchodilator FEV₁ from baseline at Week 54 for a range of possible FEV₁ changes will be provided. For this plot, patients with missing Week 54 data will be assumed to be non-responders.

4.5.5 Subgroup Analyses

Exploratory subgroup analyses of the primary efficacy endpoint will be performed to evaluate the consistency of the primary analysis results across pre-specified subgroups defined by demographic and baseline characteristics. The following subgroups will be analyzed with respect to the primary efficacy endpoint (incidence of asthma exacerbations) and for the secondary efficacy endpoint (absolute change in pre-bronchodilator FEV₁ from randomization at Week 54):

- Age (18 to <50 years, 50 to <65 years, ≥65 years)
- Sex (male, female)
- Baseline ICS total daily dose (<1000 µg, ≥1000 µg of fluticasone propionate or equivalent)
- Number of asthma exacerbations in the 12 months prior to study entry (1 event, 2 events, ≥3 events)
- Geographic region (North America, Latin America, Eastern Europe, Western Europe plus Rest of World)

Biomarker subgroup analyses are specified in Section 4.5.6.

A Poisson regression model similar to that specified for the primary analysis (Section 4.5.1) will be used for each subgroup analysis based on data subset for the patient subgroup of interest. Baseline covariates included in the primary analysis but no longer relevant given the subgroup of interest will be excluded from the model. A similar approach will be used for the secondary efficacy endpoint based on FEV₁. If convergence problems with the statistical models arise due to a small number of patients per subgroup, the analysis may be simplified by combining some of the subgroups or by excluding baseline stratification variables from the model. The estimated treatment effects (MSTT1041A vs. placebo) and corresponding 95% CIs from the models will be displayed graphically for each MSTT1041A dose level and each level of the subgroups specified.

4.5.6 Biomarker Analyses

Additional biomarker analyses will be performed to further evaluate the treatment benefit by biomarker-defined patient subpopulations. Two biomarkers will be considered individually: blood eosinophils and *IL1RL1* genotype.

Based on the primary efficacy endpoint (incidence of asthma exacerbations) and the secondary efficacy endpoint (absolute change in pre-bronchodilator FEV₁ from the randomization visit at Week 54), the treatment benefit will be assessed for the following pre-specified biomarker-defined patient subpopulations:

- **Blood eosinophil level at visit 1:** Eosinophil-high will be defined as ≥ 300 cells/ μ L and eosinophil-low will be defined as < 300 cells/ μ L at visit 1. It is estimated that approximately 25% of enrolled patients will be classified as eosinophil-high.
- ***IL1RL1* genotype:** Patient subsets will be defined by *IL1RL1* genotype, using the Single Nucleotide Polymorphisms (SNP) at rs10206753. Genotype-positive will be defined as those patients who are homozygous for the common variant (i.e., TT). Genotype-negative will be defined as patients who are not homozygous for the common variant (i.e., CT or CC). It is estimated that approximately 50% of enrolled patients will be classified as genotype-positive.

For the primary efficacy endpoint, a Poisson regression model similar to that specified for the primary analysis (Section 4.5.1) will be used for each subgroup analysis based on data subset for the biomarker subgroup of interest. Baseline covariates included in the primary analysis but no longer relevant given the subgroup of interest will be excluded from the model. A similar approach will be used for the secondary efficacy endpoint based on FEV₁. Point estimates for the treatment benefit (MSTT1041A vs. placebo) within each biomarker patient group as defined above will be presented along with 95% CIs and p-values.

Similarly, the appropriateness of the chosen cutoff values to define the biomarker-high/genotype-positive versus biomarker-low/genotype-negative patient group

for the primary analysis (e.g., 300 cells/ μL for eosinophils) will be assessed by performing additional analyses for the primary efficacy endpoint (incidence of asthma exacerbations) and the secondary efficacy endpoint (absolute change in pre-bronchodilator FEV₁ from randomization at Week 54) with the use of pre-specified alternative subgroups:

- **Alternative Subgroups for Blood Eosinophils:** Patients will be categorized into blood eosinophil subgroups defined as ≥ 150 cells/ μL and < 150 cells/ μL at visit 1.
- **Alternative Subgroups for *IL1RL1* Genotype:** Patients will be categorized into one of three genotype subgroups: TT, CT, or CC.

Covariate adjustment for these analyses will follow the principles as described in Section 4.5. All other elements of the analysis will remain the same as in the primary analysis. If convergence problems with the statistical models arise due to a small number of patients per subgroup, the analysis may be simplified by excluding baseline stratification variables from the model. Point estimates for the treatment benefit (MSTT1041A vs. placebo) within each biomarker patient subgroup will be presented as defined above, along with associated 95% CIs and p-values.

Additional exploratory analyses that will be performed include:

- **Prognostic Effect:** The prognostic effect of blood eosinophils and *IL1RL1* genotype will be evaluated by assessing the relationship between baseline biomarker levels and the observed rate of protocol-defined exacerbations based on patients who received treatment with placebo.
- **Treatment by Biomarker Interaction:** This study was not designed to compare the reduction in asthma exacerbation rate associated with MSTT1041A treatment between the biomarker-high/genotype-positive and the biomarker-low/genotype-negative patient groups. However, a comparison of the treatment by biomarker interaction for each of the individual biomarkers (as categorized above) will still be performed. In addition, the interaction of treatment by eosinophil level, treated as a continuous variable, will be explored.
- **Association between Blood Eosinophils and *IL1RL1* Genotype:** The relationship between blood eosinophils and *IL1RL1* genotype will be evaluated graphically by displaying blood eosinophil level at visit 1 by genotype.
- **Relationship with Demographics and Disease Characteristics:** The relationship between each biomarker (blood eosinophils and *IL1RL1* genotype) and demographics and disease characteristics (including but not limited to age, sex, race, geographic region, and the number of asthma exacerbations in the 12 months prior to study entry) will be evaluated.

4.6 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Several candidate PD biomarkers have been identified (e.g., FeNO, blood eosinophils, serum sST2, serum IL-5, plasma LIGHT, and serum amphiregulin) and will be measured to assess the effect of MSTT1041A on these biomarkers. Descriptive summaries of the

PD biomarker values at each visit, as well as changes from the randomization visit (absolute or percent change as appropriate), will be summarized by treatment arm.

The analyses of all PD endpoints will be based on the safety-evaluable population (see Section 4.1.2).

Serum MSTT1041A concentrations are measured at the following timepoints during the double-blind treatment period: Weeks 2, 6, 10, 14, 26, 38, 50, 54, and 70 (Predose, $C_{min,Wk4}$, $C_{min,Wk8}$, $C_{min,Wk12}$, $C_{min,Wk24}$, $C_{min,Wk36}$, $C_{min,Wk48}$, $C_{min,Wk52}$, follow-up concentration). The difference in the weeks for the concentration and the visit week is due to fact that study drug is given on Week 2 of the study, which is the start of the double-blind treatment period. In a subset of patients (approximately 20 per group), additional PK samples will be collected following first dose and at Week 26 for characterization of absorption and accumulation of MSTT1041A following SC administration. These concentrations will be tabulated and summarized using descriptive statistics (mean, standard deviation, coefficient of variation, median, minimum, and maximum) by treatment group and timepoint based on the safety-evaluable population (see Section 4.1.2). While treatment groups will be defined according to the actual treatment received, a review of dosing information including dosing errors, if any, and of the PK data for patients with unusual dosing patterns will be conducted to assess for any potential impact on the results.

Additional PK analyses of the double-blind treatment period data may be conducted as appropriate. Population PK modeling may be performed to characterize inter-individual variability, which may be reported separately from the clinical study report. Exploratory exposure-response analyses will be performed as appropriate, and may be reported separately from the clinical study report.

4.7 SAFETY ANALYSES

Safety will be assessed through the summary of adverse events, serious adverse events, adverse events of special interest, laboratory test results (hematology and serum chemistry, urinalysis), incidence of antibodies against MSTT1041A, ECG findings, and vital signs. Safety summaries for the 52-week double-blind treatment period will include outcomes as described in Section 2.1.4. This may include data collected at unscheduled visits, early termination visits, or safety follow-up visits.

Safety outcomes will be summarized based on the safety-evaluable population with patients grouped according to the actual treatment received (see Section 4.1.2). Safety summaries will be presented by treatment arm for all treated patients.

Safety assessments will also be conducted on the run-in safety-evaluable population, which consists of all patients who receive a placebo dose during the run-in period regardless of whether or not the patient is randomly allocated into the double-blind treatment period (see Section 4.1.3). These safety analyses will consider adverse

events that occur during the 2-week run-in period. Events that occur during this period will be followed until resolved.

4.7.1 Exposure to Study Medication

Exposure to study drug (number of study drug administrations and duration of treatment) will be summarized for the 52-week double-blind treatment period by treatment arm. Duration of treatment will be defined based on the difference (in days) between the dates of the first and last dose of study drug during the 52-week double-blind treatment period (Week 2 to Week 50 treatments) plus 29 days.

A listing of study drug administration with information on dosing errors during the placebo run-in and double-blind treatment periods will be provided.

4.7.2 Adverse Events

Verbatim descriptions of treatment-emergent adverse events will be coded using the latest version of MedDRA. A treatment-emergent adverse event is defined as any new adverse event or any worsening of an existing condition with an onset date on or after the first study drug administration date. Summaries of treatment-emergent events will be provided for each of the following categories:

- All adverse events
- All adverse events by most extreme severity
- Adverse events assessed as related to study treatment by the investigator
- All serious adverse events
- All suspected unexpected serious adverse events
- Adverse events that are a result of an error in study drug administration
- Adverse events leading to discontinuation of study treatment
- Adverse events leading to study discontinuation
- Adverse events of special interest

The adverse events of special interest will be identified as follows. Summaries will include tabulations by Medical Dictionary for Regulatory Activities (MedDRA) terms and listings, as appropriate.

- Cases of potential drug-induced liver injury that include an elevated alanine aminotransferase (ALT) or aspartate aminotransferase (AST) in combination with either an elevated bilirubin or clinical jaundice, as defined by Hy's law.
- Suspected transmission of an infectious agent by the study drug, as defined below
Any organism, virus, or infectious particle (e.g., prion protein transmitting transmissible spongiform encephalopathy), pathogenic or non-pathogenic, is considered an infectious agent. A transmission of an infectious agent may be suspected from clinical symptoms or laboratory findings that indicate an infection in

a patient exposed to a medicinal product. This term applies only when a contamination of the study drug is suspected.

- Potential major adverse cardiac events (MACE) will be identified as follows.

Death due to cardiovascular causes, non-fatal myocardial infarction, non-fatal stroke or transient ischemic attack (TIA), unstable angina or chest pain requiring hospitalization, coronary revascularization, and congestive heart failure (CHF) requiring hospitalization.

In addition, patient deaths and the primary cause of death will be listed, as well as any cases of pregnancy, anaphylaxis, anaphylactoid, and hypersensitivity.

4.7.3 Laboratory Data

Descriptive summaries of laboratory values at screening, run-in, randomization, and during the 52-week double-blind treatment period, including changes from the randomization visit, will be generated for hematology, chemistry, and urinalysis parameters. The proportion of patients with values outside the normal upper and lower limit at each visit will be calculated overall and by change from the randomization visit. In addition, selected lab parameters (including, but not limited to, AST, ALT, alkaline phosphatase, total bilirubin, hemoglobin, and absolute neutrophil count) will be summarized by grade using the WHO grading scale (2003).

The number and percentage of patients with positive serum antibodies against MSTT1041A at baseline and during the 52-week double-blind treatment period will be tabulated. The relationship between ADA status and safety, efficacy, PK, and biomarker endpoints will be explored.

4.7.4 ECG Results

The number and percentage of patients with abnormal ECG findings at screening, run-in (Week 0), randomization (Week 2), Week 4, Week 26, and Week 54 will be summarized.

In addition to the primary investigator, a separate agency (Cardiocore through Vitalograph) will review all ECG's independently. ECG findings that change from normal to abnormal or an abnormality to a different abnormality will be analyzed and summarized.

For corrected QT interval using Fridericia's method (QTcF), PR, and QRS intervals as well as heart rate, the values at each timepoint as well as changes from randomization will be summarized descriptively. The difference in mean QTcF change from randomization between each MSTT1041A arm and placebo and the corresponding 2-sided 90% CI will be summarized for Week 26 and Week 54.

The proportion of patients who experience a QTcF value >450, >480, or >500 ms for males or >470, >490, or >510 ms for females at a given visit, or an increase from randomization in QTcF of >30 and >60 ms will be summarized. The proportion of

patients who experience a QTcF value > 450 ms for males or > 470 ms for females, or a QTcF increase from baseline of > 30 ms at any time during the study will be summarized.

In the event that a patient has more than three valid ECG records per visit, only the last three readings (by timestamp) will be analyzed. For example, if a patient has four valid ECG records for the screening visit, only readings two through four will be analyzed, and the first reading will be ignored.

4.7.5 Vital Signs

Descriptive summaries of vital sign measurements and changes from randomization will be generated.

For the vital sign parameters of heart rate, respiration rate, systolic blood pressure, and diastolic blood pressure, the proportion of patients with a high or low value based on the cutoffs specified in Table 2 and the proportion of patients with a change from baseline (increase or decrease) of the amount specified in Table 2 will be summarized by visit.

Table 2 Cutoff Values for Summary of Vital Signs

Vital Sign	Units	Value at Visit		Change from Baseline	
		Low	High	Decrease	Increase
Heart rate	beats/min	< 55	> 100	≥ 20	≥ 20
Respiration rate	breaths/min	–	> 17	≥ 10	≥ 10
Systolic blood pressure	mmHg	≤ 90	≥ 140	≥ 20	≥ 20
Diastolic blood pressure	mmHg	–	≥ 90	≥ 15	≥ 15

4.8 MISSING DATA

Missing efficacy data will not be imputed for the primary efficacy endpoint (incidence of asthma exacerbations) or for the secondary efficacy endpoint absolute change in pre-bronchodilator FEV₁ from randomization visit at Week 54. Instead, missing data will be accounted for in the model as described in Section 4.5.1 and Section 4.5.2. Analyses to explore the sensitivity of results to missing data assumptions will be conducted as described in Section 4.5.4.

Missing data for other secondary and exploratory endpoints will be handled as described in Sections 4.5.2 and 4.5.3.

4.9 INTERIM ANALYSES

No efficacy interim analyses are planned or were conducted.

5. REFERENCES

Barnard J and Rubin DB. Small-sample degrees of freedom with multiple imputation. *Biometrika* 1999; 86:948–55.

Keene ON, Roger JH, Hartley BF, et al. Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharmaceut Statist* 2014;13:258–64.

WHO Toxicity Grading Scale For Determining The Severity Of Adverse Events.
[Accessed 8 May 2017]. Available from
http://www.icssc.org/Documents/Resources/AEManual2003AppendicesFebruary_06_2003%20final.pdf