

STATISTICAL ANALYSIS PLAN

A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY OF
THE EFFICACY, SAFETY, AND TOLERABILITY OF SERLOPITANT FOR
THE TREATMENT OF PRURITUS IN ADULTS AND ADOLESCENTS WITH
A HISTORY OF ATOPIC DERMATITIS (ATOMIK)

MTI-103

Sponsored by:

Menlo Therapeutics Inc.
200 Cardinal Way, 2nd Floor
Redwood City, CA 94063

Version: 2.0

Date Approved: 07 February 2018

Confidentiality Statement:

This document is a confidential communication of Menlo Therapeutics Inc. As such, the recipients agree not to disclose or reproduce, without prior written approval, this document and any attachments, except to appropriate Institutional Review Boards, Ethics Committees, representatives of the US Food and Drug Administration, other regulatory agencies or as otherwise required by applicable laws or regulations.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS.....	4
1 STUDY DESCRIPTION	5
1.1 Introduction.....	5
1.2 Objectives	5
1.3 Study Design.....	5
1.4 Method of Assigning Subjects to Treatment Groups.....	7
1.5 Treatment Blinding	7
1.6 Decision Rule and Sample Size	7
2 STATISTICAL METHODS.....	8
2.1 Populations Analyzed	8
2.2 Study Drug Dosing and Compliance	8
2.3 Study Endpoints	9
2.3.1 Endpoint Data Capture.....	10
2.3.2 Itch Numeric Rating Scale	10
2.3.3 Static Patient Global Assessment of Itch Severity.....	10
2.3.4 Patient Global Impression of Change in Itch Severity.....	10
2.3.5 ItchyQoL.....	10
2.3.6 5-D Pruritus Scale.....	11
2.3.7 Actigraphy.....	12
2.3.8 Investigators' Global Assessment of Atopic Dermatitis Severity	12
2.3.9 Safety Parameters.....	13
2.3.10 Pharmacokinetic Measurements	13
2.4 Study Day and Visit Windows.....	13
2.5 Statistical Assessment of the Study Objectives	14
2.5.1 Primary Analysis.....	14
2.5.2 Secondary Analyses	15
2.6 Confounding Event	15
2.7 Handling of Missing Data.....	15
2.8 Sensitivity Analyses, Subgroups and Covariates.....	16
2.9 Safety Analyses.....	17
2.10 Interim Analysis.....	17
3 SUMMARY TABLES, LISTINGS, AND FIGURES.....	18
3.1 General Conventions.....	18
3.2 Definition of Baseline	18
3.3 Clinical Study Subjects	18
3.3.1 Subject Disposition and Analysis Populations.	18
3.3.2 Demographics and Baseline Characteristics.....	19

3.3.3	Concomitant Medications and Procedures.....	19
3.3.4	Medical History	20
3.3.5	Clinical Study Treatment	20
3.4	Analysis of Efficacy Endpoints	20
3.4.1	Calculation Examples for Daily Measures	20
3.4.2	WI-NRS	21
3.4.3	ItchyQol	22
3.4.4	Additional and Exploratory Endpoints	22
3.5	Analysis of Safety Endpoints.....	23
3.5.1	Adverse Events	23
3.5.2	Clinical Laboratory Evaluations	23
3.5.3	Vital Signs.....	23
3.5.4	Electrocardiogram Results	23
3.6	Pregnancy.....	23
3.7	Psychometric Analyses	24
3.8	Pharmacokinetic Analyses	24
4	REFERENCES	25

LIST OF TABLES

Table 1	Schedule of eDiary and Actigraphy Assessments	6
Table 2	Schedule of Visit Activities	6
Table 3	Analysis Windows	13
Table 4	Daily Windows	14

LIST OF ABBREVIATIONS

Abbreviation	Definition
AD	Atopic dermatitis
AE	Adverse Event
AI-NRS	Average-Itch Numeric Rating Scale
ANOVA	Analysis of Variance
BOCF	Baseline Observation Carried Forward
CMH	Cochran Mantel Haenszel
CRF	Case report form
ECG	Electrocardiogram
eDC	Electronic data capture
eDiary	Electronic diary
ETD	Early Treatment Discontinuation
FAS	Full Analysis Set
iDMC	Independent data monitoring committee
ICF	Informed Consent Form
IGA-AD	Investigators' Global Assessment of atopic dermatitis severity
Itch NRS	Itch Numeric Rating Scale
LOCF	Last Observation Carried Forward
NK₁-R	Neurokinin-1 receptor
PGIC	Patient Global Impression of Change
PK	Pharmacokinetics
SAE	Serious Adverse Event
sPGA	Static Patient Global Assessment
TEAE	Treatment-Emergent Adverse Events
WI-NRS	Worst-Itch Numeric Rating Scale

1 STUDY DESCRIPTION

1.1 Introduction

In this study serlopitant is being studied for the treatment of pruritus (itching) in adults and adolescents with a history of atopic dermatitis. Atopic dermatitis (AD) is an inflammatory, chronically relapsing and intensely pruritic skin disease referred to as “the itch that rashes.” Pruritus is the “cardinal symptom” in AD ([Lewis-Jones 2006](#)).

Serlopitant is a small molecule, highly selective neurokinin-1 receptor (NK₁-R) antagonist that is administered orally. NK₁-R stimulation has been shown to be an important pathway for pruritus perception ([Ständer 2015](#)). Evidence suggests that inhibition of this pathway results in decreased pruritus, see protocol for details. Prior to initiation of this study, serlopitant was evaluated in over 1,000 subjects across thirteen Phase 1 studies and four Phase 2 studies.

1.2 Objectives

Study MTI-103 contains primary and secondary objectives.

The primary objective is to assess the efficacy of serlopitant for the treatment of pruritus in adults and adolescents with a history of atopic dermatitis.

The secondary objectives of this study are as follows:

- To assess the safety and tolerability of repeated oral doses of serlopitant in adults and adolescents with a history of atopic dermatitis
- To assess the psychometric properties of Worst-Itch Numeric Rating Scale (WI-NRS)

1.3 Study Design

This is a double-blind, randomized, placebo-controlled study. Approximately 450 subjects will be randomized in a 1:1:1 ratio to receive daily oral doses of serlopitant 1 mg, 5 mg, or placebo for 6 weeks. The study will consist of three periods (screening, treatment, follow-up), for a total study period of approximately 12 weeks.

During the screening period, subjects will be provided with an electronic device for recording electronic diary (eDiary) assessments throughout the study. This eDiary will be used to capture efficacy endpoint data and treatment information (see [Table 1](#)). At the Baseline visit, eligible subjects will be randomized, and study drug dispensed beginning with a loading dose of 3 tablets to be taken at bedtime that same day (Study Day 1). Starting on Study Day 2, subjects will take one tablet per day at bedtime. Treatment will continue for 6 weeks. The primary efficacy endpoint will be assessed during Week 6.

After completion of the 6-week treatment period, all subjects will enter a 4-week follow-up period. Subjects who discontinue treatment early at any time during the treatment period will

have an Early Treatment Discontinuation (ETD) visit within 7 days after their last dose of study drug in addition to a follow-up visit 4 weeks after the end of treatment or at Week 6, whichever is later.

An abbreviated schedule of activities for the study can be found in [Table 2](#) with greater details available in the protocol.

Table 1 Schedule of eDiary and Actigraphy Assessments

Device	Assessment	Timing of Assessments
eDiary	WI-NRS / AI-NRS	Once daily for duration of the study
	ItchyQoL / sPGA	At Screening, Baseline visit, Days 15 (± 2), 29 (± 3), 43 (± 3) and 28 days (± 3) after the ETD visit
	PGIC	Days 15 (± 2), 29 (± 3), 43 (± 3) and 28 days (± 3) after ETD visit
Actigraphy watch	Scratching / Sleep	Nighttime monitor for duration of the study

Table 2 Schedule of Visit Activities

Examination	Screening	Baseline Day 1	Week 1 Day 8	Week 2 Day 15	Week 4 Day 29	Week 6 Day 43 / ETD	Follow-up ETD + 28
Informed consent	X						
Medical history	X						
I/E criteria	X	X					
Randomization		X					
Physical exam	X			X		X	X
Vital signs	X	X		X	X	X	X
ECG	X			X			X
Labs	X			X		X	X
PK				X	X	X	
IGA-AD	X	X		X	X	X	X
5-D Pruritus Scale	X	X		X	X	X	X
Concomitant medications Procedures/ Non-Medication Therapy		X	X	X	X	X	X
Adverse events		X	X	X	X	X	X

1.4 Method of Assigning Subjects to Treatment Groups

Subjects will be randomized to serlopitant 1 mg, 5 mg, or placebo in a 1:1:1 ratio. Stratified permuted block randomization will be used. Randomization will be stratified by sex (male, female) and race (white, non-white). An Interactive Web Response System will be used to perform the randomization.

1.5 Treatment Blinding

This is a double-blind study. The sponsor will not have access to unblinded data during the study and there is no expectation that the study data (e.g. adverse events, dosing, or lab results) will unblind the sponsor, investigator or subjects. Greater details concerning study blinding can be found in a separate blinding plan.

1.6 Decision Rule and Sample Size

The significance level for this study is 5% (one-sided). If the one-sided p-values are less than 5% then the null hypothesis of no treatment effect is rejected. If the one-sided p-values are additionally less than 2.5%, then the serlopitant-based regimen will have met the generally accepted level of evidence required to demonstrate efficacy.

This study contains two active dosing arms. A stepdown procedure will be used to control for multiplicity. This procedure will start by testing the primary endpoint for the 5 mg treatment group and then proceed to the 1 mg group should statistical significance (one-sided p-values < 0.05) be reached for the 5 mg vs placebo comparison.

Statistical significance for the key secondary endpoints will be dependent upon statistical significance being reached for the primary endpoint at the given dose level (i.e. a gatekeeper within dose arm will be used for the secondary endpoints). As there are multiple secondary endpoints a stepdown procedure will be used starting with the WI-NRS 4 point responder rate endpoint, then the ItchyQoL.

The target sample size of 450 randomized and dosed subjects (150 per group) has been determined based upon a 1:1:1 allocation of subjects to treatment groups, an assumed common treatment effect for the 1 mg and 5 mg arms, and a 5% one-sided alpha level. One-hundred and fifty subjects per group results in 90% power for each test individually assuming a treatment effect (placebo change from baseline - serlopitant change from baseline) of -1 and a standard deviation of 2.9.

The sample size calculations have been performed in PASS 13 using a t-test. The primary analysis will control for the stratification factors. It is expected that this unstratified power estimate will under-estimate the true power as it does not take the variance reduction resulting from stratification into account ([Matts and Lachin 1988](#)).

2 STATISTICAL METHODS

2.1 Populations Analyzed

Four analysis populations will be defined. These populations are:

- Screened Population – All subjects who provide informed consent and attend the screening visit.
- Randomized Population – All subjects who are randomized. Subjects will be analyzed within the treatment group to which they are randomized.
- Full Analysis Set (FAS) – Subset of subjects in the randomized population who received at least one dose of study medication. Subjects will be analyzed within the treatment group to which they are randomized.
- Safety Population – Subset of subjects who received at least one dose of study medication. This population will be analyzed based upon the actual treatment received. If a dosing error occurs and the subjects received treatment at more than one dose level, where placebo is considered a dose level, they will be summarized in the dose group which represents the highest dose received.

2.2 Study Drug Dosing and Compliance

Study drug usage (medication taken yes/no) on each study day will be recorded within the eDiary. Tablet counts will also be recorded in the case report form. Each data source will be independently used to summarize tablets used and compliance. In this way two measures of dosing and compliance will be available. The calculation of these two measures will be identical and hence any difference between them will be reflective of the extent the subject reported dosing data in the eDiary and tablet counts disagree.

The total number of tablets used will be summarized along with the duration of treatment and compliance. The duration of treatment (last dose date – first dose date + 1) will be based on dosing information provided within the CRF. Compliance will be calculated as,

$$\text{Compliance}(\%) = \frac{\text{Tablets Used}}{\text{Last Dose Date} - \text{First Dose Date} + 3}$$

The number of tablets used will be based upon the results provided within the eDiary and the recorded tablet counts within eDC. Should a subject not return a bottle, for compliance calculations it will be assumed they returned the average number of tablets that were returned in their other bottles. Subjects with no returned bottles will have missing tablet count compliance. The denominator for the compliance calculation includes a + 3 to account for the loading dose received on Study Day 1. This compliance calculation implies a subject who takes treatment for one week but takes all expected tablets for that week will be 100% compliant with treatment.

2.3 Study Endpoints

The study endpoints are listed below, with greater detail concerning these endpoints provided in [Sections 2.3.2 to 2.3.10](#).

Primary Efficacy Endpoint

The primary efficacy endpoint is the change in WI-NRS from baseline to Week 6.

Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints are as follows:

- WI-NRS 4-point responder rate at Week 6
- Change in ItchyQoL from baseline to Week 6

Additional Endpoints

Additional secondary efficacy endpoints include the following:

- WI-NRS 3-point responder rates
- Change in Average-Itch NRS (AI-NRS) and AI-NRS responder rates
- Change in 5-D Pruritus Scale
- Change in Static Patient Global Assessment of Itch Severity (sPGA)
- Change in Patient Global Impression of Change in Itch Severity (PGIC)
- Change in number of night-time scratching events per hour from baseline to Week 6

Safety endpoints include the following:

- Incidence of treatment-emergent adverse events (TEAEs) and serious adverse events (SAEs)
- Changes in clinical laboratory parameters following study drug exposure
- Changes in vital sign and electrocardiogram (ECG) parameters following study drug exposure
- Plasma concentrations of serlopitant and metabolites

Exploratory endpoints include the following:

- Change in Investigators' Global Assessment of atopic dermatitis severity (IGA-AD)
- Change in sleep efficiency
- Change in mean activity during the sleep period

2.3.1 *Endpoint Data Capture*

The study endpoints will be captured within five electronic systems, the electronic data capture (eDC system) , subject eDiary, subject actigraphy devices, central laboratories (ACM Global) and centralized ECG assessments. The adverse events, vital signs, and IGA-AD will be captured in the CRF. The eDiary will capture the WI-NRS, ItchyQoL, AI-NRS, sPGA, and PGIC. The actigraphy devices will record night-time scratching, sleep efficiency, and mean activity during the sleep period during each night the subjects is on study.

2.3.2 *Itch Numeric Rating Scale*

The Itch NRS is an instrument for measurement of itch intensity that asks subjects to rate the intensity of their itch on an 11-point scale ranging from 0 (no itch) to 10 (worst itch imaginable) during a 24-hour recall period. Both WI-NRS and AI-NRS during a 24-hour recall period will be captured. The daily NRS results will be summarized at Weeks 2, 4, 6 and follow-up as outlined in [Section 2.4](#). A 4-point responder is a subject who had at least a 4-point reduction in score between the visit and baseline. A 3-point responder is similarly defined.

2.3.3 *Static Patient Global Assessment of Itch Severity*

The sPGA assesses overall itch severity by asking the severity of itchiness in the past 7 days on a 5-point scale as none (0), mild (1), moderate (2), severe (3), and very severe (4).

2.3.4 *Patient Global Impression of Change in Itch Severity*

The PGIC in Itch Severity is a single item measure that assesses the change in overall itch severity since the Baseline visit on a 7-point scale from (very much better) to (very much worse).

2.3.5 *ItchyQoL*

ItchyQoL is a 22-item pruritus-specific instrument that measures the degree to which pruritus affects quality-of-life. The responses to the items are Never (1), Rarely (2), Sometimes (3), Often (4) and All the Time (5). This impact is quantified into 3 subscales (symptom, function, emotion) and an overall score. A higher score corresponds to a more adverse impact. The overall score is the average of the 22 items. If an item is missing the total score will be missing. The score for each subscale is the average score for the items that make up the subscale. The items that make up the subscales are,

- Symptom subscale score (items 1-6):
 - My itchy skin condition bleeds,
 - My skin hurts because of my itchy skin condition,
 - My itchy skin condition burns or stings,
 - I get scars from my itchy skin condition,

- I need to scratch my itchy skin condition,
- Temperature/seasonal changes aggravate my itchy skin condition
- Function subscale score (items 7-13):
 - I spend a lot of money treating my itchy skin condition,
 - My itchy skin condition makes it hard to work or do what I enjoy,
 - My itchy skin condition affects my interaction with others,
 - My itchy skin condition affects how well I sleep,
 - My itchy skin often makes it difficult to concentrate,
 - My itchy skin condition limits the types of clothes I can wear,
 - My itchy skin condition forces me to buy special soaps, detergents, and lotions.
- Emotion subscale score (items 14-22):
 - I am frustrated by my itchy skin condition,
 - I am embarrassed by my itchy skin condition,
 - My itchy skin condition drives me crazy/nuts,
 - My itchy skin condition makes me feel angry or irritable,
 - My itchy skin condition makes me feel depressed or sad,
 - I worry about what other people think about me because of my itchy skin condition,
 - I worry that the itching will last forever,
 - I feel self-conscious because of my itchy skin condition,
 - My personality has changed because of my itchy skin condition.

2.3.6 5-D Pruritus Scale

The 5-D Pruritus Scale assesses the five domains of degree, duration, direction, disability and distribution of pruritus. Subjects rate their symptoms over the preceding 2-week period. The scores of each of the five domains are achieved separately and then summed together to obtain a total 5-D score (Elman 2010) resulting in a range of scores from 5 to 25 (no to most severe pruritus). If any one item is missing the total score will be missing.

The single-item domains, duration, degree and direction, are captured on a likert scale. The score for these domains is equal to the value recorded for the domain (i.e. 1 to 5). The disability domain includes four items that assess the impact of itching on: sleep, leisure /social activities, housework/errands and work/school. The score for the disability domain is achieved by taking the highest score on any of the four items. For the distribution domain, the number of affected body parts is given a point total (0–2 = 1, 3–5 = 2, 6–10 = 3, 11–13 = 4, 14–16 = 5).

2.3.7 Actigraphy

The actigraphy assessments requires each subject to wear the actigraphy devices, which are the size of a medium size wrist watch, on each wrist at night (during sleep), with the option to wear them during the waking hours as well. The waking hours data will be captured but not directly analyzed. The devices are returned to the study site where the data are downloaded and sent for analysis. Subjects will be permitted to not wear the devices up to three nights over any seven-day period (i.e. up to three nights per week).

The 3-dimensional acceleration data (m/sec^2) from the actigraphy device is analyzed to discriminate scratching from other nighttime events. These scratching results will include

1. Scratching Endpoints

- a. scratching time per sleep period (the time of the first and last itching event),
- b. number of scratching events per sleep period
- c. the itching events per hour (*secondary endpoints*) and
- d. the intensity of itching per hours (seconds of itching per hour).

2. Sleep Endpoints

- a. sleep duration (end of sleep – start of sleep),
- b. total sleep time (time subject was asleep during sleep period, does not include wake time during sleep period),
- c. sleep efficiency (*exploratory endpoint*: sleep time/sleep duration),
- d. Wake After Sleep Onset (WASO) and
- e. the mean activity during the sleep period (*exploratory endpoint*).

For the actigraphy endpoints to mirror the techniques used in development of the actigraphy tool as outlined in Moreau et al. (2017), the analysis will be limited to sleep periods between 30 minutes and 18 hours in length.

2.3.8 Investigators' Global Assessment of Atopic Dermatitis Severity

The IGA-AD is an instrument used to assess the overall severity of atopic dermatitis at a given time point, as determined by the investigator. It consists of a 6-point scale ranging from 0 (clear) to 5 (very severe). An IGA-AD responder is defined as a subject with a score of “clear” or “almost clear” and at least a 2-grade reduction from baseline score.

2.3.9 Safety Parameters

Vital signs will include measurements of heart rate, sitting blood pressure, respiration rate, and temperature. A standard 12-lead ECG will be performed and centrally read. From this central read the following parameters will be available: RR, PR, QRS, QT, HR, QTcB, QTcF, and the “interpretation” (Normal or Abnormal). If the result is considered abnormal, the specific findings will also be provided. Physical examination abnormalities identified as preexisting before treatment will be recorded as medical history. Laboratory results identified as clinically significant and abnormal before treatment will be recorded as medical history or associated with an ongoing historical event. Clinically significant new or worsened abnormalities that appear after treatment will be recorded as AEs.

Vital signs, ECG parameters and relevant interpretations, and laboratory results will be listed.

2.3.10 Pharmacokinetic Measurements

Blood samples for pharmacokinetic (PK) measurements will be collected as outlined in [Table 2](#).

2.4 Study Day and Visit Windows

Summary tables will report data based upon the protocol scheduled time points (Weeks 2, 4, 6 and Follow-up). Assessment will be assigned to these time points based upon the study day they are performed on. The derivation of study day depends upon whether the event occurred prior to or after treatment. This results in the first day of treatment being on day 1 and the day prior to the first day of treatment being day -1. For events that occur prior to the first day of treatment the following definition is used,

$$\text{Study day} = \text{date of assessment} - \text{date of first treatment}.$$

For events that occur on or after the first day of treatment, a plus one is added,

$$\text{Study day} = \text{date of assessment} - \text{date of first treatment} + 1.$$

The analysis windows used to report non-daily endpoints as well as ItchyQoL, sPGA and PGIC are outlined in [Table 3](#). If more than one assessment is available within the range the assessment closest to the target day is reported for the analyses window. If two observations exist with the same distance to the target day, the first observation is used.

Table 3 Analysis Windows

Visit	Range	Target Day
Baseline	Day \leq 1	Day 1
Week 2	Day 2 to Day 21	Day 15
Week 4	Day 22 to Day 35	Day 29
Week 6	Day 36 to Day 57	Day 43
Week 10	> Day 57	Day 71

In addition to the analysis windows, data will be summarized at Follow-Up. The Follow-Up assessment will be the last assessment available that is at least seven days after the end of treatment. Should no post treatment assessments be available the subject will not have a follow-up assessment. While a Week 10 analysis window is defined, data tables will be limited to Baseline, Week 2, 4, 6 and Follow-up.

eDiary and actigraphy daily results (WI-NRS, AI-NRS, night-time scratching, sleep efficiency and mean activity during the sleep) will be summarized at Baseline, Week 2, 4, 6 and Follow-up. The seven days leading up to the time point (see [Table 4](#) range), will be used to create the summary measures unless the subject does not have at least seven results. In this case as many observations as are available within the complete data range will be used and additional data from the extended range (see [Table 4](#)) will be used until seven data points are achieved or the end of the extended range is reached. An example of this algorithm can be found in [Section 3.4.1](#).

Table 4 Daily Windows

Visit	Range (complete data)	Extended Range
Baseline	Day -7 to -1	Day -11 to -8
Week 2	Day 8 to Day 14	Day 15 to Day 18
Week 4	Day 22 to Day 28	Day 29 to Day 32
Week 6	Day 36 to Day 42	Day 43 to Day 46

An additional ‘Follow-Up’ timepoint will also be created. This timepoint includes the last 7 post treatment observations that are at least 7 days after the end of treatment.

For the WI-NRS and AI-NRS endpoints results from within the windows provided in [Table 4](#) and at Follow-up may not be used based upon the confounding logic outlined in [Section 2.6](#).

2.5 Statistical Assessment of the Study Objectives

The efficacy endpoints will be summarized within the FAS using descriptive statistics by time point and treatment. For the primary and key secondary endpoints these statistics will include 95% confidence intervals for the treatment difference. Testing will also be used for the primary and key secondary endpoints.

2.5.1 Primary Analysis

The difference in the primary efficacy outcome measure (WI-NRS change from baseline to Week 6) between treatment groups will be tested using an analysis of variance (ANOVA) model controlling for the stratification factors. A main effects model with an interaction term (treatment by stratification factors) and a Type II hypothesis will be used. More precisely this ANOVA model uses the following as a stratified adjusted measure of the change from baseline for a given treatment group:

$$\Delta = \sum_{j=1}^4 \frac{n_{1j}n_{2j}}{n_{1j}+n_{2j}} \Delta_j ,$$

where j represents the stratification cells, n is the sample size for treatment 1 and 2 and Δ_j is the change measure for stratum j . Should a site make an error during randomization and select the wrong stratification factor (e.g. select the White stratification cell for a non-White subject) the test will use how the subject was randomized. All other summaries of race (e.g. baseline characteristics or subgroup analyses) will be based upon the correct race group and not the as randomized group.

Conceptually the hypotheses being tested for this study are:

$$H_0: \Delta_{Placebo} \geq \Delta_{Serlopitant} \quad H_a: \Delta_{Placebo} < \Delta_{Serlopitant}$$

where $\Delta_{Placebo}$ is the placebo change from baseline to Week 6 and $\Delta_{Serlopitant}$ is the similar measure for serlopitant. Each serlopitant treatment group (1 mg and 5 mg) will be compared to placebo separately.

The primary endpoint will utilize the analysis windows outlined in [Section 2.4](#), the confounding rules outlined in [Section 2.6](#) and the missing data rules outlined in [Section 2.7](#).

2.5.2 Secondary Analyses

The differences between treatment groups for the key secondary endpoints will be tested using the same ANOVA model as used for the primary endpoint (change in ItchyQoL) or a Cochran Mantel Haenszel (CMH) test (WI-NRS 4-point responder rate) controlling for the 'as randomized' stratification factors. A CMH test will also be conducted for the WI-NRS 3-point responder rate, AI-NRS 4-point responder rate, and AI-NRS 3-point responder rate.

2.6 Confounding Event

The use of excluded therapies / procedures or clinically significant changes in the dose or frequency of allowed adjunctive therapies can result in a change in pruritus. In the statistical plan the use of excluded therapies / procedures or clinically significant changes in the dose or frequency of allowed adjunctive therapies is called a confounding event. Short treatment (up to 3 days) of an excluded medication for indications other than atopic dermatitis or pruritus (see protocol for specific medications) is not considered a confounding event. Should a confounding event occur, the cause of changes in pruritus can no longer be clearly attributed to the randomized treatment. As such, the results from the WI-NRS/AI-NRS following this confounding event (i.e. on the day the confounding event occurred or after) will not be included in the analysis.

Confounding events will be identified by a blinded clinical review of the concomitant medication and procedures data prior to database lock.

2.7 Handling of Missing Data

Summary statistics will generally be reported based upon observed data. Should a determination of treatment period (on treatment, pre-treatment) be required for adverse events or concomitant medication but the corresponding date is missing, or is a partial date,

the event/medication will be considered on treatment unless the portions of the data that are available indicate this is not possible.

If a subject fails to complete their eDiary for a week or more, or a confounding event occurs the WI-NRS and AI-NRS endpoints, including the primary endpoint, may be missing. In this case, their change from baseline value will be imputed. The subject's data prior to confounding will be used. From that point forward their trajectory will be imputed based upon the trajectory of their treatment group. The imputation approach first determines the average change for each time point (i.e. Weeks 2, 4, 6) for each treatment group. The difference between weeks n-1 and n will be considered the change for Week n. These change measure will be used to 'fill in' the missing weeks between the last completed week (pre-confounding) and Week 6. Hence, a placebo subject who receives excluded therapy in Week 4 will have their Week 6 WI-NRS change from baseline imputed as $\Delta_4 + \Delta_{(4,6)}$, where Δ_4 is the observed change from baseline result for the subject up to Week 4 and $\Delta_{(4-6)}$ is the placebo change from Week 4 to 6 (i.e. the average placebo decrease from Week 4 to 6).

Missing data imputation will only be used for the WI-NRS and AI-NRS endpoints.

2.8 Sensitivity Analyses, Subgroups and Covariates

Descriptive statistics for the change in WI-NRS at Week 6 will be provided for the subgroups indicated below. Forest plots of the difference and confidence interval for this difference will also be created. The subgroups include,

- Sex
- Age Group (< 18 years, 18 to 64 years, > 64 years)
- Race
 - All race groups with less than 30 subjects will be combined with 'Other'
 - A White versus Non-White comparison will be made
- Baseline WI-NRS
- Baseline IGA-AD
- Use of allowed concomitant medications as outlined in Section 5.7.1 of the protocol
- Treatment completion (completed treatment, did not complete treatment).

To investigate the impact of missing data / confounding sensitivity analyses will be performed for the WI-NRS endpoint. These methods include,

1. Multiple imputation, MI, methods will be used to recalculate the p-values at Week 6.
2. The summary statistics will be produced where the missing data is imputed using
 - a. Data after the confounding event are excluded with no imputation

- b. Last Observation Carried Forward (LOCF),
- c. Baseline Observation Carried Forward (BOCF).

2.9 Safety Analyses

Safety endpoints will be summarized with descriptive statistics. All safety summaries and analyses will be performed using the safety population.

Prior and Concomitant medications will be coded by the World Health Organization (WHO) Drug Dictionary version 2012/Q4. Adverse Events and Medical History will be coded using MedDRA version 17.1 (or later).

2.10 Interim Analysis

An independent Data Monitoring Committee (iDMC) will monitor unblinded safety data. The iDMC may make recommendations to the Sponsor to continue or terminate the study based upon the outcomes of the data reviews. The iDMC is not tasked with stopping the study early for efficacy or for stopping for futility. Greater detail concerning the iDMC can be found in the iDMC charter.

3 SUMMARY TABLES, LISTINGS, AND FIGURES

3.1 General Conventions

Unless otherwise stated, summary statistics including the number of subjects (n), mean, standard deviation, median, minimum and maximum will be presented for continuous variables. Generally, the minimum and maximum values will be presented to the same decimal precision as the raw values, the mean and median values to one more, and the standard deviation, to two more decimal places than the raw values. For categorical variables, per category, the absolute counts (n) and percentages (%) of subjects with data. Percentages will be presented to one decimal place.

For AEs, medical history (prior and concomitant diseases) and concomitant medications reported on a per-subject basis, the denominator for the percentage calculation will be the number of subjects at risk in each treatment group. A subject will be considered at risk if the subject is in the analysis set and in the subgroup of interest.

All p-values will be rounded to 4 decimal places. If a p-value is less than 0.0001 it will be displayed as “< 0.0001”, a p-value rounding to 1 will be displayed as “> 0.9999”.

Unless otherwise specified all summaries will be performed by treatment group, all efficacy analyses will be based upon the FAS population and all safety will be based upon the safety population. The safety and demographic tables will include a total serlopitant (1 mg and 5 mg doses) and a total column (total serlopitant and placebo) along with the treatment columns.

3.2 Definition of Baseline

Unless specified elsewhere, baseline is defined as the last available measurement prior to administration of study drug. For events that occur on the same day as the first administration of study drug (i.e. assessments performed at the Baseline visit) the assessment will be classified as pre-treatment and hence baseline. Baseline for the daily measures recorded in the eDiary and actigraphy device is defined in [Table 4](#).

3.3 Clinical Study Subjects

3.3.1 Subject Disposition and Analysis Populations.

The number of subjects randomized, treated, and discontinued early, will be summarized within all randomized subjects. The reason for early termination will be summarized. Additionally, a summary of subjects present at each visit as determined by the presence of a vital signs result, will be provided.

A listing of entry criteria that were not met will be produced. The listed criteria will include the criteria language. A table of the violated entry criteria for randomized subjects will be produced as well as a table that summarizes the screen failure reasons for screen failure subjects.

The number of subjects in each analysis population will be summarized.

3.3.2 *Demographics and Baseline Characteristics*

Demographic and baseline characteristics will be summarized descriptively by treatment group. This will include the following items:

1. Demographic variables

- a. Age
- b. Race
- c. Ethnicity
- d. Sex
- e. Height
- f. Weight
 - Baseline WI-NRS
 - Presence or absence of active AD
 - Baseline IGA-AD
 - Age at AD diagnoses
 - AD severity relative to prior 3 and 12 months
 - Family history of atopy (eczema, asthma, allergic rhinitis)
 - Prior Treatments of AD (e.g. antibiotics, antihistamines, corticosteroids, biologics, immunomodulatory agents)

3.3.3 *Concomitant Medications and Procedures*

Medications and procedures with a stop date before the treatment dosing date will be considered prior medications/procedures. Medications/procedures with a start or stop date on or after the treatment dosing date will be considered concomitant. All medications/procedures marked as ongoing are concomitant.

A medication/procedure with an incomplete stop date will be considered concomitant if:

1. Month is missing and year is equal to or after the year of treatment dosing date
2. Day is missing and year is equal to the year of the treatment dosing date and month is equal to or after the month of the treatment dosing date.

All concomitant medications will be summarized by treatment group. All medications, including prior medications, will be provided in listings.

A table and listing will be produced to summarize confounding events as defined in [Section 2.6](#).

A table and listing will be produced to summarize the allowed concomitant medications (antihistamines, topical steroids, topical calcineurin inhibitors, emollients).

A table and listing will be produced to summarize topical adjunctive therapies.

Procedures and non-medication therapies will be listed. A listing of confounding procedures / non-medication therapies will be provided.

3.3.4 Medical History

Medical history will be tabulated by system organ class, preferred terms and treatment group.

3.3.5 Clinical Study Treatment

The total tablets used, duration of treatment (Last dose date – first dose date + 1), number of subjects with dosing errors and compliance (see [Section 2.2](#)) will be summarized. Tablets used will be calculated by summing the number of days the subject indicates medication was used and based upon tablet counts reported in the CRF. For the first treatment day of the diary-based calculation, if a subject indicates they used treatment it will be assumed three tablets were taken. The kit numbers provided in the eDC system will be used to identify dosing errors. Should an error occur the tablet counts provided in the eDC system will be used to determine treatment assignment for the safety population.

3.4 Analysis of Efficacy Endpoints

3.4.1 Calculation Examples for Daily Measures

Below are examples of how the eDiary and actigraphy endpoints will be calculated based upon windows outlined in [Table 4](#) and rules outlined in [Section 2.4](#).

1. Example 1 (complete data)
 - a. The subject has results for all seven days leading up to the visit (e.g. $x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}$).
 - b. The average for Week 2 is $(x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14})/7$.
2. Example 2 (incomplete data, completed in the extended range)
 - a. The subject has results for less than seven days in the original range (e.g. $x_{22}, x_{24}, x_{25}, x_{27}, x_{28}$) but have enough in the extended range to reach seven results overall (e.g. x_{29}, x_{31}). In this example the subject is missing diary data on days 23, 26, and 30.
 - b. The average for Week 4 is $(x_{22} + x_{24} + x_{25} + x_{27} + x_{28} + x_{29} + x_{31})/7$.

3. Example 3 (incomplete data)

- a. The subject has results for less than seven days in the original range (e.g. x_{40}, x_{42}) and does not have enough in the extended range to reach seven results overall (e.g. x_{44}, x_{48}, x_{49}). In this example the subject is missing diary data on days 36, 37, 38, 39, 41, 43, 45, 46, 47 and 49.
 - The average for Week 6 is $(x_{40} + x_{42} + x_{44} + x_{48} + x_{49})/5$.

4. Example 4 (confounding event)

- a. The subject has results for all seven days leading up to the visit but has a confounding event part way through window (e.g. $x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}$ all exist but excluded medication or therapy was received on day 11).
- b. The average for Week 2 is $(x_8 + x_9 + x_{10})/3$.

5. Example 5 (no data)

- a. The subject does not complete the diary after day 10 or had a confounding event before day 10.
- b. The average for Weeks 4 and 6 are missing.

3.4.2 **WI-NRS**

The WI-NRS endpoints (result, change from baseline, percent change from baseline, 3-point and 4-point responders) will be summarized at the primary time point, Week 6, as well as at Week 2, 4 and Follow-up. The summary measures will include 95% Wald confidence intervals for the treatment difference. The baseline WI-NRS will also be summarized, though no confidence interval will be produced. A graph of the mean WI-NRS results and change from baseline results by timepoint and treatment will be produced as well as a cumulative distribution function plot (CDF plot) of the Week 6 results.

The WI-NRS scores for days -7 to 14 will be plotted by averaging the available data across subjects within treatment group and study day. These results will be summarized in a table along with change from baseline where baseline is defined in the identical manner used for the weekly measures, see [Table 4](#).

The summary measures for the WI-NRS change from baseline results and the 4-point responder rate will be produced using four endpoint derivation methods, one primary method and three sensitivity measures. The difference between these methods is how the weekly measures are calculated. The primary derivation method uses all reported data prior to use of excluded therapy or changes to allowed adjunctive therapies and the missing data imputation rules outlined in [Section 2.7](#). Sensitivity analyses as defined in [Section 2.8](#) will also be conducted.

Testing will be used for the change from baseline and 4-point responder rates endpoints at Week 6. The test used for the 4-point responder rate (primary endpoint) is a CMH test controlling for the 'as randomized' stratification factors. The 'as randomized' stratification factors are based upon the stratification cell the subject was randomized to. The differences

between treatment groups for the change from baseline measure will be tested using an ANOVA model controlling for the stratification factors. A main effects model with an interaction term (treatment by stratification factors) and a Type II hypothesis will be used. Each serlopitant treatment group (1 mg and 5 mg) will be compared to placebo separately and the multiplicity rules outlined in [Section 1.6](#) will be used to interpret these p-values.

‘Sensitivity’ p-values will be constructed using multiple imputation (MI) methodology. This approach imputes randomly generated values instead of the fixed value. Using the example from [Section 2.6](#) (i.e. a subject had data up to Week 4 but is missing their Week 6 results: $\Delta_4 + \Delta_{(4,6)}$), at least 1,000 random realizations of $\Delta_{(4,6)}$ will be generated using code similar to the following,

```
proc mi nimpute=1000;  
class trt;  
monotone reg(x);  
var trt x;
```

From these realizations of $\Delta_{(4,6)}$ the change from baseline to Week 6 will be calculated. Hence, there will be at least 1,000 random realizations of the responder status. The p-value from the MI analysis will be calculated using the SAS software procedure `mianalyze`.

Finally, the change in WI-NRS results will be summarized within the subgroups outlined in [Section 2.8](#). Forest plots based upon the same subgroups will also be produced.

3.4.3 *ItchyQol*

The ItchyQol endpoint will be summarized at Baseline, Week 2, 4, 6 and Follow-up. The summary measures will include 95% Wald confidence intervals of the treatment difference. In addition, a graph of the mean score by treatment and timepoint will be produced along with a graph of change from baseline. The difference between serlopitant 1 mg and 5 mg individual versus placebo at Week 6 will be tested with the identical ANOVA model as used for the change from baseline WI-NRS endpoint.

3.4.4 *Additional and Exploratory Endpoints*

The additional and exploratory endpoints (AI-NRS, sPGA, IGA-AD, 5-D Pruritus Scale, Actigraphy endpoints) will be summarized at baseline, Week 2, 4, 6 and Follow-up. The result as well as change from baseline will be summarized. The PGIC endpoint is a change measure so only the results at each time point will be summarized. As the 5-D Pruritus Scale was introduced into the study part way through the study, only subjects with a baseline value will be included in the summary table.

For the IGA-AD endpoint a separate analysis will be performed where the result is dichotomized to responder/non-responder. This measure will be summarized at Week 2, 4, 6 and Follow-up and be limited to subjects with a mild or greater assessment at baseline.

3.5 Analysis of Safety Endpoints

3.5.1 Adverse Events

AEs and SAEs will be recorded from the first study drug administration through the Follow-up visit. A treatment-emergent AE (TEAE) is an AE which started on or after the date of the first treatment. An overview of AEs, which includes subject incidence of TEAEs, treatment-related TEAEs, serious TEAEs, TEAEs by toxicity grade and TEAEs leading to permanent discontinuation of treatment. For TEAEs presented by toxicity grade, the worst grade during the clinical study will be presented for each subject.

The subject incidence of TEAEs, treatment-related TEAEs and TEAEs leading to permanent discontinuation of treatment will be summarized by system organ class and preferred term. Treatment-emergent AEs will also be summarized in a table by toxicity grade. For TEAEs presented by toxicity grade, the worst grade for each event during the clinical study will be presented for each subject.

All AEs will be presented as a listing by subject. A separate listing of pruritus and atopic dermatitis adverse events (as identified by the investigator) will be presented. TEAEs leading to treatment discontinuation will be provided in a separate listing.

SAEs will be listed and summarized in a similar manner to TEAEs.

3.5.2 Clinical Laboratory Evaluations

Summary statistics for actual values and for changes from baseline will be tabulated for clinical laboratory results by analysis window. Shifts from baseline clinical laboratory values based upon the normal range will be tabulated. These tables will utilize the normal ranges provided for the individual sample.

3.5.3 Vital Signs

The observed data at baseline and change from baseline for each measurement day will be summarized with descriptive statistics.

3.5.4 Electrocardiogram Results

The overall ECG assessment (abnormal or normal) will be summarized by treatment and analysis window. The QTcF values will be summarized by treatment and analysis window as both the QTcF value and change from baseline. A summary of the number of subjects with a value > 450, 480 and 500 and the numbers of subjects with change from baseline values > 30 and 60 by treatment and analysis window will be produced.

3.6 Pregnancy

A listing of pregnancy tests results will be produced.

3.7 Psychometric Analyses

A psychometric assessment of the WI-NRS will be conducted outside of this analysis plan.

3.8 Pharmacokinetic Analyses

Plasma concentrations will be summarized by time-point and treatment.

4 REFERENCES

The following literature references cited in this document are available upon request.

1. Lewis-Jones S. Quality of life and childhood atopic dermatitis: the misery of living with childhood eczema. *Int J Clin Pract.* 2006;60(8):984-92.
2. Ständer S, Luger TA. NK-1 antagonists and itch. *Handb Exp Pharmacol.* 2015;226:237-55. doi: 10.1007/978-3-662-44605-8_14.
3. PASS 13 Power Analysis and Sample Size Software (2014). NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/pass.
4. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Control Clin Trials.* 1988;9(4):327-44.
5. Elman S, Hynan LS, Gabriel V, and Mayo MJ. The 5-D Itch Scale: a New Measure of Pruritus. *Br J Dermatol.* 2010;162:587-593.
6. Moreau A, Anderer P, Ross M, Cerny A, Almazan T, Peterson B. Detection of nocturnal scratching movements in patients with atopic dermatitis using accelerometers and recurrent neural networks. *IEEE J Biomed Health Inform.* 2017;162:587-593.

APPROVAL SHEET

Product:	Serlopitant
Protocol Number:	MTI-103
SAP Version:	2.0
Version Date:	07 February 2018

The individuals signing below have reviewed and approve this statistical analysis plan.

 _____ Statisticians	_____ Date
---	---------------

 _____ Senior Vice President, Clinical Development	_____ Date
---	---------------

 _____ Clinical Operations Manager	_____ Date
--	---------------