A Randomized, Double-Blind, Placebo-controlled Study Of The Efficacy, Safety, And Tolerability Of Serlopitant For The Treatment Of Chronic Pruritus Of Unknown Origin

Protocol Version/Date: Version 1.0 / 30 Jan 2020

NCT number: NCT03841331

STATISTICAL ANALYSIS PLAN

A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY OF THE EFFICACY, SAFETY, AND TOLERABILITY OF SERLOPITANT FOR THE TREATMENT OF CHRONIC PRURITUS OF UNKNOWN ORIGIN

MTI-117

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

Version: 1.0
Date Approved: 30Jan2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse event
ALT	Alanine aminotransferase
AMH	Anti-Mullerian hormone
ANOVA	Analysis of variance
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Classification
CMH	Cochran Mantel Haenszel test
CNS	Central Nervous System
CPUO	Chronic Pruritus of Unknown Origin
ECG	Electrocardiogram
eCRF	Electronic case report form
eDiary	Electronic diary
ESS	Epworth Sleepiness Scale
FSH	Follicle-stimulating hormone
HIV	Human immunodeficiency virus
IRT	Interactive Response Technology
ITT	Intent-to-Treat
IWRS	Interactive Web Response System
LDH	Lactate dehydrogenase
LH	Luteinizing hormone
LOCF	Last Observation Carried Forward
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NRS	Numeric Rating Scale
PD	Pharmacodynamics
PGIC	Patient Global Impression of Change
PK	Pharmacokinetics
SAE	Serious adverse event
sPGA	Static Patient Global Assessment
TEAE	Treatment-emergent adverse event
TSH	Thyroid stimulating hormone
VAS	Visual Analog Scale
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 with chronic pruritus of unknown origin. Pruritus, or itch, is defined as an unpleasant sensation that elicits the desire or reflex to scratch (Ikoma 2006). Chronic pruritus may be associated with a variety of underlying conditions, including both dermatological and systemic diseases. However, in 8-15% of patients with chronic pruritus, an underlying etiology cannot be identified (Weisshaar 2012).

Serlopitant is a small molecule, highly selective neurokinin-1 receptor (NK_1-R) antagonist that is administered orally. NK_1-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 1,500 subjects.

1.2 Objectives

Study MTI-117 contains primary and secondary objectives.

- The primary objective of this study is to assess the efficacy of repeated oral doses of serlopitant in adult subjects with chronic pruritus of unknown origin.
- The secondary objective of this study is to assess the safety and tolerability of repeated oral doses of serlopitant in adult subjects with chronic pruritus of unknown origin.

1.3 Study Design

This is a double-blind, randomized, placebo-controlled study. Subjects will be randomized in a 1:1 ratio to receive daily oral doses of serlopitant 5 mg or placebo for 10 weeks. After completion of the treatment period or early discontinuation of study drug treatment, all subjects will enter a 5-week follow-up period.

The study will consist of three periods, for a total study period of 18 weeks:

- Screening period: 3 weeks
- Treatment period: 10 weeks
- Follow-up period: 5 weeks

Informed consent may occur prior to the Screening visit. During the screening period, subjects will be assessed for conditions associated with chronic pruritus. If the investigator deems that additional assessments are necessary to rule out specific conditions the subject may not be randomized until these are completed and the pruritus is still considered to be of unknown origin.

Subjects will be provided an electronic diary (eDiary) at the Screening visit and are to complete the diary daily until the end of study participation. See Section 2.3.1 for more detail. At the Baseline visit (Day 1), subjects will take a loading dose (3 tablets taken orally) at the site. Starting on Day 2, subjects will take one tablet per day until Week 10 or until the study drug is discontinued.

The primary efficacy endpoint is the Worst-Itch Numeric Rating Scale (WI-NRS) 4-point responder rate at Week 10. This endpoint is based upon the WI-NRS results entered into the eDiary. Other efficacy endpoints are captured within the case report forms. Safety endpoints are captured within the case report form or central laboratories.

An abbreviated schedule of activities for the study can be found in Table 1, with greater details of the schedule available in the protocol.

Table 1 Schedule of Visit Activities

Examination	Screening	Baseline	Week 2	Week 4	Week 6	Week 10	F/U
Demographics	X						
Paper WI-NRS	X						
Daily WI-NRS	X						
WI-VAS / ESS / sPGA		X	X	X	X	X	X
PGIC			X	X	X	X	X
ECG	X			X			X
Vital signs	X	X	X	X	X	X	X
Medical history / prior and current medications	X	X					
Physical exam	X	X		X	X	X	X
Concomitant medications / emollients		X	X	X	X	X	X
Hematology/Chemistry	X			X		X	X
Endocrine/Reproductive	X					X	X
Serology, Serum IgE, Iron, Other labs	X						
Urine pregnancy test	X	X		X		X	X
PK blood draw						X	
Dispense / collect study drug		X	X	X	X	X	
AEs/SAEs	X						

1.4 Method of Assigning Subjects to Treatment Groups

Subjects will be randomized to serlopitant 5 mg or placebo in a 1:1 ratio. Stratified permuted block randomization will be used. Randomization will be stratified by the subject's reported WI-NRS score for the 1-week period prior to the Baseline visit (7 to < 8.5, 8.5 to 10), and by the subject's age at the Baseline visit (18 to < 50 years, ≥ 50 years). Randomization of subjects < 50 years of age will be capped at 25% of the total randomized population. This cap may be increased up to 33% of the total randomized population at the Sponsor's discretion. An Interactive Web Response System will be used to perform the randomization.

1.5 Treatment Blinding

This is a double-blind study. Unless required for safety reporting, 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. The investigator can unblind individual subjects for emergency medical purposes. 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-value is less than 5% then the null hypothesis of no treatment effect is rejected. If the one-sided p-value is additionally less than 2.5%, then the serlopitant-based regimen will have met the generally accepted level of evidence required to demonstrate efficacy.

The target sample size of 200 randomized and dosed subjects (100 per group) has been determined based upon a 1:1 allocation of subjects to treatment groups and a 5% one-sided alpha level. A sample size of 200 subjects results in at least 90% power assuming a placebo 4-point responder rate of CCI% and serlopitant 4-point responder rate of CCI% at Week 10.

The sample size calculations have been performed in PASS 13 ("PASS 13 Power Analysis and Sample Size Software" 2014) and used a Chi-Squared 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

Five analysis populations will be defined. These populations are:

- Screened Population – All subjects who provide informed consent.
- Randomized Population – All subjects who are randomized. Subjects will be analyzed within the treatment group to which they are randomized.
- Screen Failures – All screened subjects who are not randomized.
- Full Analysis Population (FAP) – Subset of subjects in the randomized population who were dispensed 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 both placebo and serlopitant they will be summarized in the serlopitant treatment group.

2.2 Study Drug Dosing and Compliance

Study drug accountability (i.e. how many tablets were dispensed and returned, date of first and last dose) will be recorded in the case report form. Additionally, subjects will record daily dosing within the eDiary. The tablet counts reported within the CRF will be used as the dosing and compliance data source. From these data the duration of treatment, number of tablets used, and compliance with treatment will be determined. The duration of treatment is

$$last\ dose\ date - first\ dose\ date + 1.$$

Compliance will be calculated as,

$$Compliance(\%) = \frac{Tablets\ Used}{Treatment\ End\ date - Treatment\ Start\ date + 3}.$$

The number of tablets used will be based upon the counts within the electronic data capture (eDC). For each bottle, the difference between the tablets dispensed and returned minus any tablets lost will be the tablets used for that bottle. The total number of tablets used will be the sum across bottles. Should a subject not return a bottle, it will be assumed that they returned the average number of tablets that were returned in their other bottles. Subjects with no returned bottles will have missing tablet count and compliance (i.e. no effort will be made to impute values).

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.

The eDiary dosing data will be used by study site personnel to monitor dosing in real time but will not be used for dosing and compliance calculations. The days of dosing as measured by the eDiary will be summarized.

2.3 Study Endpoints

The study endpoints are listed below, with greater detail concerning these endpoints provided in Sections 2.3.2 to 2.3.8.

Primary Efficacy Endpoint

The primary efficacy endpoint is the WI-NRS 4-point responder rate at Week 10.

Secondary Efficacy Endpoints

The key secondary efficacy endpoints are as follows:

- WI-NRS 4-point responder rate at Weeks 2 4, 6, and 8
- WI-NRS 3-point responder rate at Weeks 2, 4, 6, 8, and 10
- Change from baseline in WI-NRS at Weeks 2, 4, 6, 8, and 10
- Change from baseline in daily WI-NRS scores through Week 2
- Change from baseline in WI-VAS at Weeks 2, 4, 6, and 10

Safety Endpoints

Safety endpoints include the following:

- Incidence of TEAEs and SAEs
- Changes from baseline in clinical laboratory parameters following study drug exposure
- Changes from baseline in vital sign and ECG parameters following study drug exposure
- Plasma concentrations of serlopitant and metabolites

Exploratory Endpoints

The exploratory endpoints include the following:

- Change from baseline in the Epworth Sleepiness Scale (ESS) to Weeks 2, 4, 6, and 10
- Change from baseline in Static Patient Global Assessment of Itch Severity (sPGA) to Weeks 2, 4, 6, and 10
- Patient Global Impression of Change in Itch Severity (PGIC) at Weeks 2, 4, 6, and 10

2.3.1 Endpoint Data Capture

The study endpoints will be captured within the eDC system, from central laboratories, within the IRT system, or a subject eDiary (see Table 2). The eDiary will be used to capture the WI-NRS data daily. From these measures the primary endpoint will be created. This system will also calculate and provide two measures to the study sites that are required for eligibility determinations: eDiary compliance and baseline 1-week average WI-NRS score.

Table 2 Data Capture Systems

Examination	eDC	Central Laboratory	IRT system	eDiary
Demographics	X			
Paper WI-NRS	X			
Daily WI-NRS				X
WI-VAS / ESS / sPGA	X			
PGIC	X			
ECG	X	X		
Vital signs	X			
Medical history / prior medications	X			
Physical exam	X			
Concomitant medications	X			
Laboratory Results		X		
Urine pregnancy test	X			
PK blood draw		X		
Dispense / collect study drug	X			
AEs/SAEs	X			
Randomization			X	

2.3.2 Worst Itch - Numeric Rating Scale

The Itch NRS is a validated, self-reported, instrument for measurement of itch intensity. It uses a 24-hour recall period and asks subjects to rate the worst intensity of their itch on an 11-point scale ranging from 0 (no itch) to 10 (worst itch imaginable). Higher scores indicate greater itch intensity during that 24 period. During the study, WI-NRS assessments will be reported by the subject via eDiary. The daily NRS results will be summarized. The daily results will be averaged to create weekly measures outlined in Section 2.4. A 4-point responder is a subject who had at least a 4-point reduction in score between the week and baseline. A 3-point responder is similarly defined. The intercurrent rules outlined in Section 2.6 will be applied to the responder rate endpoints.

A ‘paper WI-NRS’ will be captured at screening to be used as part of the eligibility criteria. This measure will not be used within the efficacy calculations.

2.3.3 Worst Itch - Visual Analog Scale

The Itch VAS is a validated, self-reported, instrument for measurement of itch intensity. It uses a 24-hour recall period and asks subjects to rate the worst intensity of their itch on a 100-mm horizontal line ranging from 0 mm (no itch) to 100 mm (worst itch imaginable). Higher scores indicate greater itch intensity. The VAS measurement will be summarized in centimeters. WI-VAS assessments will be reported by the subject via a paper form administered at study visits.

2.3.4 Epworth Sleepiness Scale

The ESS is a quality of life instrument intended to measure daytime sleepiness by use of a very short questionnaire (8 items each scored from 0-3). Each of the eight items have four possible responses: no chance of dozing (0), slight chance of dozing (1), moderate chance of dozing (2), high chance of dozing (3). The ESS total score is the sum of the responses to each item and can range from 0 to 24. If an individual item is missing, the total score will be missing. Higher scores indicate greater daytime sleepiness.

The total score can be interpreted as follows (http://epworthsleepinessscale.com/about-the-ess/):

- 0-10: Normal Daytime Sleepiness
- 11-12: Mild Excessive Daytime Sleepiness
- 13-15: Moderate Excessive Daytime Sleepiness
- 16-24: Severe Excessive Daytime Sleepiness

2.3.5 Static Patient Global Assessment of Itch Severity

The Static Patient Global Assessment of Itch Severity (sPGA) is designed to assess overall itch severity. Each subject is asked to rate the severity of his/her itchiness in the past 7 days on a 5-point scale as none, mild, moderate, severe, and very severe. Higher scores indicate greater itch severity.

2.3.6 Patient Global Impression of Change in Itch Severity

The Patient Global Impression of Change in Itch Severity (PGIC) is a single item used to assess the change in overall itch severity since the baseline visit. Each subject will rate the change in his/her itch severity on a 7-point scale from (very much better) to (very much worse).

2.3.7 Safety Parameters

The following safety parameters are captured within this study.

2.3.7.1 Adverse Events and Serious Adverse Events

The following is a brief summary of terms associated with adverse events. Greater details concerning these terms can be found in the protocol.

An AE is any untoward medical occurrence associated with the use of a drug in humans, whether or not it is considered drug related. An AE is considered treatment emergent, TEAE, if it began on the day of first treatment or after. For AEs that occur on the first day of dosing, the AE will be considered TEAE if the study site indicates that the event occurred after the first study drug dosing. An AE is considered "serious" if it results in death, is a life-threatening AE, results in hospitalization or prolongation of hospitalization, persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions, a congenital anomaly/ birth defect or is an important medical event.

AEs and SAEs will be recorded from the first study drug administration through the end of the follow-up period. After informed consent but prior to initiation of study drug, only SAEs considered by the investigator to be caused by a protocol-mandated intervention will be collected.

AEs will be graded for severity using the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) v5.0 to describe the maximum intensity of the adverse event. As such, severity will be reported as Mild, Moderate, Severe, Very Severe, Life Threatening or Disabling, Death Related.

An investigator's causality assessment is the determination of whether there exists a reasonable possibility that the investigational product caused or contributed to an AE. Causality of an AE will be assessed by the investigator using the following terms: Likely Related and Likely Unrelated.

Adverse events and medical history will be coded using MedDRA version 21.0.

2.3.7.2 Other Safety Endpoints

- Vital signs:
  - Systolic Blood Pressure, Diastolic Blood Pressure, Heart Rate, Respiration Rate, Height, Weight, Temperature and presence of a clinically significant abnormal values.
- 12-Lead Electrocardiogram:
  - RR, PR, QRS, QT, HR, QTcB, QTcF, and the ECG interpretation. A central read cardiologist assessment of normal vs abnormal will be available as well as an investigator assessment of clinically significant vs not clinically significant.
- Physical examination abnormalities identified as preexisting before treatment will be recorded as medical history. Abnormalities identified after treatment will be recorded as adverse events.

- Laboratory samples will include:
  - Hematology, Chemistry, Immunochemistry and Urine Pregnancy.

2.3.8 Pharmacokinetic Measurements

There will be one PK sample at the Week 10 visit. The plasma concentrations of serlopitant and metabolites M1/M1a, M2/M2a and M3 will be determined.

2.4 Study Day and Visit Windows

Summary tables will report data based upon the protocol scheduled time points (Baseline, Weeks 2, 4, 6, 10 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,

Study day = date of assessment - date of first treatment.

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

Study day = date of assessment - date of first treatment + 1.

The analysis windows used to report non-daily endpoints 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
Screening	< Day – 7	First
Baseline	Day -7 to Day 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. The baseline of an endpoint (see Section 3.2) may not always occur at the baseline visit.

The eDiary WI-NRS assessments will be summarized at Baseline, Week 2, 4, 6, 10, 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 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[a]	Day -7 to -1 or Day -6 to -1, Day 1	Day -11 to -8 or Day -10 to -7
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
Week 8	Day 47 to Day 53	Day 54 to Day 57
Week 10	Day 64 to Day 70	Day 60 to Day 63

[a] If the diary was completed prior to dosing on Day 1 then the baseline range includes the day of dosing (Day 1) otherwise the baseline range does not include Day 1.

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.

2.5 Statistical Assessment of the Study Objectives

The efficacy endpoints will be summarized within the FAP 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. No multiplicity adjustment will be used, and the p-values reported for the key secondary endpoints will be descriptive in nature.

2.5.1 Primary Analysis

The difference in the primary efficacy outcome measure (WI-NRS 4-point responder rate at Week 10) between treatment groups will be summarized by treatment group. The difference in responder rates between treatment arms will be tested using a Cochran-Mantel-Haenszel (CMH) test controlling for the 'as randomized' stratification factors. Conceptually, the hypotheses being tested are:

$$H_0: \pi_p \geq \pi_s \qquad H_a: \pi_p < \pi_s$$

where π_p is the placebo responder rate at Week 10 and π_s is the similar measure for serlopitant.

The primary endpoint will utilize the analysis windows outlined in Section 2.4, the intercurrent rules outlined in Section 2.6, and the missing data rules outlined in Section 2.7.

2.5.2 Secondary Analyses

Secondary endpoints will be summarized with descriptive statistics by time point and treatment. These summary measures will include baseline, result at the time point, and change from baseline as is appropriate for the endpoint. Missing data imputation will be limited to the WI-NRS endpoint.

The differences between treatment groups for the WI-NRS 4-point at Weeks 2, 4, 6, and 8, and 3-point responder rates at Weeks 2, 4, 6, 8, and 10 will be tested using the CMH test identical to the one used for the primary endpoint.

The change from baseline to Weeks 2, 4, 6, 8, and 10 for the WI-NRS will be tested using an analysis of variance (ANOVA) model with treatment group and stratification factor as fixed effects. A separate model/test will be performed at each time-point. [redacted]

CCI

[redacted]

[redacted]

The change in WI-VAS to Weeks 2, 4, 6, and 10 will be tested using a similar ANOVA model.

2.6 Intercurrent Event

Intercurrent events include withdrawal from the study due to lack of efficacy or rescue therapy use for treatment of worsening pruritus. Rescue therapy will be identified by the site who indicate if medications were taken to treat worsening of pruritus.

Subjects who withdraw from the study early following stopping treatment due to an intercurrent event will have WI-NRS datapoints following the events set to the larger of baseline and the last WI-NRS value prior to the event. For rescue therapy this value will be imputed from the start of rescue therapy to 4 weeks after the end of this therapy. This approach will result in subjects being defined as non-responders following the intercurrent event.

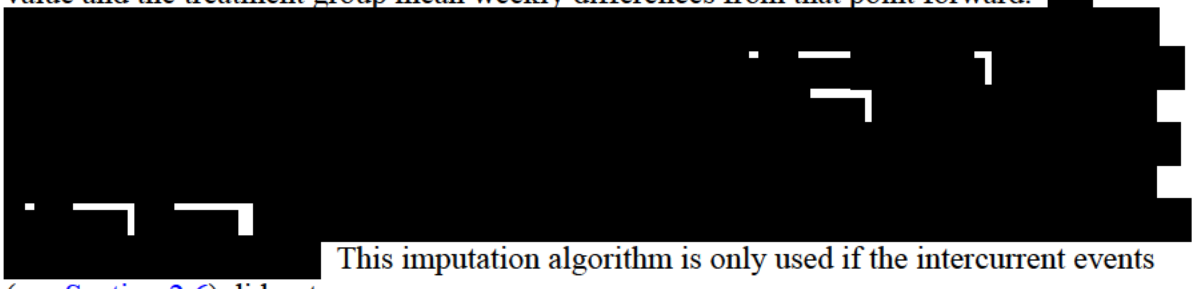
The missing data rules outlined in Section 2.7 will not be used for the results following an intercurrent event.

2.7 Handling of Missing Data

Summary statistics will generally be reported based upon observed data. Should a determination of treatment period (on treatment or pre-treatment) be required for adverse events or concomitant medication but the corresponding date is missing or partial, the event/medication will be considered on treatment unless the portions of the date that are available indicate this is not possible.

If a subject fails to complete their eDiary for a week or more, the WI-NRS endpoints, including the primary endpoint, may be missing. In this case, the WI-NRS change from baseline value will be imputed. The approach uses the subject's last week with data. 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 between analysis weeks (see Table 4) for each treatment group. The change between analysis weeks is the difference between results at adjacent analysis Weeks (e.g. results at Week 10 – results at Week 8). Any missing change from baseline results will be imputed by combining the last known value and the treatment group mean weekly differences from that point forward. CCI



This imputation algorithm is only used if the intercurrent events (see Section 2.6) did not occur.

For the responder endpoints, the same imputation will be performed and the responder status determined based upon the change from baseline.

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

2.8 Sensitivity Analyses, Subgroups and Covariates

Descriptive statistics for the WI-NRS responder rate at Week 10 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
  - ($\leq$ 64 years, > 64 years)
  - ($<$ 50 years, $\geq$ 50 years).
- Race - all race groups with less than 30 subjects will be combined with 'Other'

- Baseline WI-NRS (< 8.5, >= 8.5)
- Duration of chronic pruritus (< 1 yr, 1-5yrs,>5 yrs)

Additionally, the WI-NRS responder rates at Week 10 will be summarized for subject with and without the following post baseline events. These are post-baseline factors and may be correlated with treatment outcomes.

- Study Disposition
  - Completed study
  - Did not complete the study
- Ever require / did not require rescue therapy for treatment of worsening pruritus
- Treatment completion (completed treatment, did not complete treatment).

To investigate the impact of missing data, missing WI-NRS data will be derived using the method of Markov Chain Monte Carlo (MCMC) multiple imputation. Imputation will be conducted within each treatment group independently, so the pattern of missing observations in one treatment group cannot influence missing value estimations in another. The MCMC imputation will use the observed results at each timepoint. For each imputation process, 1000 imputations will be performed. To mimic the original missing data algorithm, the continuous values will be imputed and the responder status will be determined. The intercurrent events rules (see Section 2.6) will continue to be used. A CMH test, identical to the test used for the primary endpoint, will be used based upon the imputed responder status.

2.9 Safety Analyses

Safety endpoints will be summarized with descriptive statistics. All safety summaries and analyses will be performed using the safety population. Vital signs, laboratory, and ECG results will be summarized by timepoint and treatment. Adverse events and concomitant medications/procedures will be summarized overall.

Prior and concomitant medications and procedures will be coded by the World Health Organization (WHO) Drug Dictionary version 2018/Q3. Medications and procedures will be considered concomitant unless they ended prior to treatment.

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 and concomitant medications will be 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 FAP population, and all safety analyses will be based upon the safety population.

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, hence baseline. Baseline for the daily measures that are recorded in the eDiary are 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 who attended each visit as determined by the presence of a vital signs result.

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:

- Age – continuous and categories (≤ 64 years, > 64 years) and (< 50 years, ≥ 50 years).
- Race
- Ethnicity
- Sex
- Height
- Weight
- Baseline WI-NRS – continuous and categories (< 8.5, >= 8.5)
- Screening WI-NRS
- Baseline WI-VAS
- Length of chronic pruritus – continuous and categories (<1 yr, 1-5 yrs, >5 yrs)
- Location of pruritus
- Randomization strata
- Protocol version screened under

The reproductive measures (e.g. child bearing potential, surgical sterilization procedure, method of contraception, female subject specific medical history) will be listed but otherwise not tabulated.

The serology, serum IgE, iron, other labs and additional tests sites perform to confirm entry criteria results will be listed but otherwise not tabulated.

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 and procedures will be summarized by treatment group. All medications/procedures, including prior medications/procedures, will be provided in listings.

A table will be produced to summarize rescue medications used for intercurrent events defined in Section 2.6.

3.3.4 Medical History

Medical history will be tabulated by system organ class, preferred terms and treatment group and provided in a listing. The female specific medical history will be listed separately from other medical history.

3.3.5 Clinical Study Treatment

The total tablets used, duration of treatment, and compliance (see Section 2.2) will be summarized. The days of dosing as measured by the eDiary will be summarized. All eDC study drug data will be listed, including lost tablets. The days of dosing from the eDiary will be listed, but the daily results will not.

3.4 Analysis of Efficacy Endpoints

3.4.1 Calculation Examples for Daily Measures

Below are examples of how the eDiary WI-NRS weekly results 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_{39}, x_{40}, x_{42}) and does not have enough in the extended range to reach seven results overall (e.g. x_{44}, x_{46}). In this example the subject is missing diary data on days 36, 37, 38, 41, 43 and 45.
    - The average for Week 6 is $(x_{39}+x_{40}+x_{42}+x_{44}+x_{46})/5$.

4. Example 4 (no data)
    a. The subject does not complete the diary from day 60 onwards.
    b. The average for Week 10 is missing and the missing data rules (Section 2.7) are used.

3.4.2 WI-NRS

The WI-NRS endpoints (result, change from baseline, 3-point and 4-point responders) will be summarized at the primary time point (Week 10) as well as at Weeks 2, 4, 6, 8 and Follow-up. The summary measures will include 95% Wald confidence intervals for the treatment difference and a p-value for the treatment difference (for the change from baseline and responder rate measures). The baseline WI-NRS will also be summarized, but no confidence interval will be produced. A graph of the mean WI-NRS results and mean change from baseline results and the 4-point responder rate (baseline included with a rate of zero) by timepoint and treatment will be produced.

The WI-NRS scores for days 1 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. Daily WI-NRS results will be summarized in a table based upon availability of the data regardless of any intercurrent events.

The test used for the 3 and 4-point responder rates at Weeks 2, 4, 6, 8 and 10 is a CMH test controlling for the 'as randomized' stratification factors. The 'as randomized' stratification factors are based upon the stratification cell that 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. The estimated treatment effect from this test will be a weighted average based upon the equation provided in Section 2.5.

The WI-NRS 4-point responder rates at Week 10 will be summarized within the subgroups and separately by the post baseline factors outlined in Section 2.8. Forest plots based upon the same groups will also be produced.

A missing data sensitivity analysis will be conducted using the multiple imputation method outlined in Section 2.8. This analysis will include summary statistics by time points for the 4-point WI-NRS responder rate and the continuous WI-NRS measures as well as testing for the 4-point WI-NRS responder endpoint at Week 10. The following steps will be followed:

1. Using the daily eDiary data, the windows in Table 4 and the intercurrent rules outlined in Section 2.6 calculate the weekly WI-NRS values. Within this will be unknown values (i.e. no data within the Table 4 windows)

2. The missing WI-NRS values will be filled in using the MCMC method to generate at least 1000 datasets. [redacted]

3. From each dataset the responder status will be determined.

4. The resulting point estimates (Mantel Haenszel treatment effect and standard error) and the responder rate p-value from the data will be calculated using the SAS software procedure Freq.

5. These results will be combined with the SAS software procedure mianalyze.

3.4.3 WI-VAS

The WI-VAS endpoint will be summarized with descriptive statistics at Baseline, Weeks 2, 4, 6, 10 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. The differences between treatment groups for the change from baseline measure will be tested using an ANOVA model similar to the WI-NRS endpoint.

3.4.4 Exploratory Endpoints

The ESS endpoint will be summarized at Baseline, Weeks 2, 4, 6, 10 and Follow-up. The total score and change from baseline in the total score will be summarized. The difference in average scores between treatment group will be summarized.

The sPGA endpoint will be summarized at Baseline, Weeks 2, 4, 6, 10 and Follow-up. The categorical result (e.g. Mild, Moderate, …) will be summarized at each timepoint for each treatment group.

The PGIC will be summarized at Weeks 2, 4, 6, 10 and Follow-up. The categorical result will be summarized at each timepoint.

3.5 Analysis of Safety Endpoints

3.5.1 Adverse Events

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. TEAEs leading to treatment discontinuation will be provided in a separate listing.

3.5.2 Clinical Laboratory Evaluations

Clinical laboratory results will be summarized by type of laboratory assessment (hematology, chemistry, endocrine), timepoint, and treatment. Summary statistics for actual values and changes from baseline will be tabulated by analysis window. Shifts from baseline clinical laboratory values based upon the normal range will be tabulated for each post baseline sample timepoint. 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 positive pregnancy tests results will be produced.

3.7 Pharmacokinetic Analyses

Plasma concentrations at week 10 will be summarized. The bioanalytical laboratory will provide the plasma concentration data for serlopitant and 3 metabolites in ng/mL. Plasma concentrations will be converted to nmol/L (nM) by the Statistician using the following equation:

> Plasma serlopitant concentration in nM = [(plasma serlopitant concentration in ng/mL) / (MW of 555.54 ng/nmol)] x (1000 mL/L)

The same conversion will be followed for each metabolite concentration using the MW of CCI ng/nmol. Results will be summarized using Geometric Mean and Coefficient of variation (CV). The formulas used are shown below. Below BLQ values will be set to ½ of the BLQ values (i.e., CCI ng/mL = CCI nM).

Geometric Coefficient of Variation Formula:

$$\text{Geo CV} = \sqrt{exp^{(\ln(GeoSD))^2} - 1} \times 100\%$$

Geometric Mean Formula:

$$\text{Geo Mean} = \sqrt[N]{y_1 \times y_2 \times y_3 \times \ldots \times y_N} = exp^{\left[\frac{\ln(y_1) + \ln(y_n)}{n}\right]}$$

Geometric Standard Deviation Formula:

$$\text{Geo SD} = exp^{(SD(\ln(y_1)\ln(y_2)\ln(y_3)\ldots\ln(y_x)))}$$

4 REFERENCES

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

Ikoma A, Steinhoff M, Ständer S, et al. The neurobiology of itch. *Nat Rev Neurosci* 2006;7:535-547.

PASS 13 Power Analysis and Sample Size Software (2014). NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/pass.

Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. Control Clin Trials. 1988;9(4):327-44.

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.

Weisshaar E et al. European guideline on chronic pruritus. *Acta Derm Venereol* 2012;92:563-81.

APPROVAL SHEET

Product:	Serlopitant
Protocol Number:	MTI-117
SAP Version:	1.0
Version Date:	30 Jan 2020

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

PPD

30 Jan 2020
Date

PPD

30 Jan 2020
Date

PPD

30 Jan 2020
Date