Official Title: A Phase 2b/3, Randomized, Double-Blind, Placebo-Controlled, 2-Arm, Efficacy and Safety Study in Prurigo Nodularis (PN) With Nalbuphine ER Tablets for Pruritus Relief Through Itch Scratch Modulation (PRISM Study)

NCT Number: NCT03497975

Document Date: Statistical Analysis Plan (FINAL Version 2.1) : 1 June 2022

Statistical Analysis Plan (SAP)

Protocol Title:	A Phase 2b/3, Randomized, Double-Blind, Placebo-Controlled, 2-Arm, Efficacy, and Safety Study in Prurigo Nodularis with Nalbuphine ER Tablets for Pruritus Relief Through Itch Scratch Modulation (PRISM Study)
Protocol Number:	TR11
Protocol Version, Date	Version 7.0, 21 July 2021
ICON ID:	4318/0001
Document Version, Date:	FINAL Version 2.1, 1 June 2022

Prepared by:
ICON Clinical Research Services

On behalf of:
Trevi Therapeutics, Inc.

SIGNATURE PAGE

Prepared at ICON Clinical Research by:

[redacted] Ph.D. [redacted] Date

Reviewed at ICON Clinical Research by:

[redacted] Biostatistics Date

Approved at Trevi Therapeutics, Inc. by:

[redacted] Ph.D. [redacted] Date

[redacted] Pharm. D Date

[redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ADaM	Analysis Dataset Model
AE	Adverse Event
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CI	Confidence Interval
CP	Conditional Power
CRF	Case Report Form
CSR	Clinical Study Report
DBL	Database Lock
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
IA	Interim Analysis
ICH	International Conference on Harmonisation
ICE	Intercurrent Events
IGA-PN	Investigator Global Assessment-Prurigo Nodularis
IRT	Interactive Response Technology
ItchyQoL	Itchy Quality of Life
ITT	Intent To Treat
IWRS	Interactive Web Response System
LOCF	Last Observation Carried Forward
MAR	Missing at Random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model for Repeated Measurements
MVN	Multivariate Normal

OC	Observed Cases
OR	Odds Ratio
PAS	Prurigo Activity Score
PBI-P	Patient Benefit Index, Pruritus
PCS	Potentially Clinically Significant
PMH	Past Medical History
PNQ	Patients Need Questionnaire
PN	Prurigo Nodularis
PP	Per Protocol
PROMIS	Patient-Reported Outcomes Measurement Information System
PT	Preferred Term
Q1	Lower Quartile
Q3	Upper Quartile
QLS	Quality of Life Scale
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SDTM	Study Data Tabulation Model
SI	Standard International
SOC	System Organ Class
SOP	Standard Operating Procedure
SOWS	Subjective Opiate Withdrawal Scale
SSP	Study Specific Procedure
TEAE	Treatment Emergent Adverse Event
TFLs	Tables, Figures and Listings
ULN	Upper Limit of Normal
WI-NRS	Worst Itch Numerical Rating Scale
WOCF	Worst Observation Carried Forward

I. INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide detailed descriptions of the statistical methods, data derivations, and data displays for study protocol TR11 "A Phase 2b/3, Randomized, Double-Blind, Placebo-Controlled, 2-Arm, Efficacy, and Safety Study in Prurigo Nodularis with Nalbuphine ER Tablets for Pruritus Relief Through Itch Scratch Modulation (PRISM Study)" dated 14 March 2019 for final analysis. The table of contents and templates for the TFLs will be produced in a separate document.

Any deviations from this SAP will be described and justified in the Clinical Study Report (CSR).

The preparation of this SAP has been based on International Conference on Harmonization (ICH) E9 and ICH E9[1] guidelines.

All data analyses and generation of TFLs will be performed using SAS 9.4® or higher.

II. STUDY OBJECTIVES

A. Primary objective

To evaluate the effect of Nalbuphine Extended-Release tablets (NAL ER) on itch as assessed by the percentage of Responders ('response' is defined as a ≥ 4-point reduction in the 7-day average Worst Itch Numerical Rating Scale [WI-NRS]).

B. Secondary objectives

The secondary objectives of this study are designed to establish proof of principle and are as follows:

Key secondary objectives:

- To evaluate the effect of NAL ER on itch-related quality of life as assessed by the ItchyQoL total score.
- To evaluate the effect of NAL ER on Prurigo Nodularis (PN) skin lesions as assessed by the Prurigo Activity Score (PAS) Question 5a.
- To evaluate the effect of NAL ER on sleep as assessed by the PROMIS Sleep Disturbance Short Form 8a.

Other secondary objectives:

- To evaluate the effect of NAL ER on itch as assessed by the mean change in WI-NRS.
- To evaluate the benefit to subjects of NAL ER using the Patient Benefit Index, pruritus version (PBI-P).
- To evaluate the safety and tolerability of NAL ER.
- To assess the PK of nalbuphine and its metabolites.

III. STUDY DESIGN

A. General study design

This is a randomized, double-blind, placebo-controlled, 2-arm study, with an open-label extension period following double-blind treatment, to investigate NAL ER tablets' anti-pruritic efficacy and

safety. Subjects will be randomized to NAL ER (2-week titration followed by 162-mg BID for 12 weeks) or matching placebo (14 weeks duration), with the primary endpoint evaluation at Week 14. If permanent discontinuation of investigational product occurs anytime between Day 1 and the Week 14 visit, the subject must complete the off-treatment visit and procedures, and must return to the clinic for ALL remaining study evaluation visits up to and including the Week 14 visit. During the open-label extension, subjects who received NAL ER will continue on NAL ER for a total of 38 additional weeks (total treatment duration 52 weeks including titration) and subjects who received placebo will crossover to NAL ER for a total of 38 weeks (including titration). Upon discontinuation of investigational product, all subjects will complete a 2-week off-treatment safety follow-up period, regardless of when and why the subject discontinued study treatment unless they withdraw consent. In addition, a final follow-up itch and safety assessment will be made via phone call 4 weeks after the last dose of investigational product. The total planned study duration from randomization through the last phone contact is 56 weeks.

Approximately 240 subjects with diagnosed PN will be randomized (1:1) to one of the following treatment arms:

Arm 1: Blinded active titration over 2 weeks (to achieve a final NAL ER dose of 162 mg) followed by 162-mg BID for 12 weeks with continuation into the extension titration period (blinded), and the open-label fixed-dose period, followed by a 2-week off-treatment safety follow-up period and a final phone contact, for itch and safety assessment, 4 weeks after investigational product discontinuation.

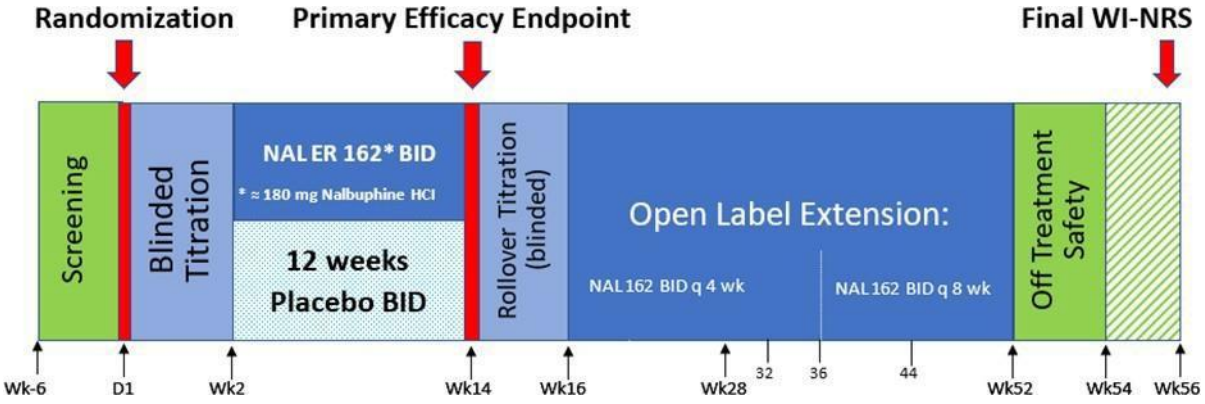
Arm 2: Placebo "titration" over 2 weeks followed by placebo BID for 12 weeks with continuation into the extension titration period (blinded; to achieve a final NAL ER dose of 162 mg) and the open-label fixed-dose period (NAL ER 162 mg), followed by a 2-week off-treatment safety follow-up period and a final phone contact, for itch and safety assessment, 4 weeks after investigational product discontinuation.

An interim sample size re-estimation will occur once 50% (120) of the subjects have either completed the Week 14 primary endpoint assessment or terminated the study early. A maximum of 360 subjects total may be enrolled, dependent upon the results of the re-estimation.

After Study Week 28, an elective dose reduction to NAL ER 108 mg BID may be considered for subjects who achieve a confirmed WI-NRS $\leq$ 3 and visible lesion healing (as assessed by an improvement of at least 1 category in the PAS lesion healing activity score). The decision to dose reduce is at the discretion of the Investigator and the subject, but once this dosing has started, no re-escalation will be permitted.

The study design schematic is presented in Figure 1.

Figure 1 Study Design



BID: twice daily; D = day; NAL ER: Nalbuphine 162-mg extended release; WI-NRS: Worst Itch Numerical Rating Scale; W =week.

B. Randomization and blinding

Randomization will be performed by an IVRS/IWRS. Upon confirmation of eligibility by the study site, subjects will be randomized in a 1:1 ratio to either Arm 1 or Arm 2.

C. Study treatments and assessments

Double-blind Treatment Period

Following randomization, the subject will receive a 2-week blister card for the titration segment of the double-blind period. Each card will contain 3 tablets for the AM dose and 3 tablets for the PM dose. The tablets will be combinations of either 27 mg/54 mg/placebo tablets, or placebo tablets only, depending on the arm assigned. At the end of titration visit, subjects will be given one 70 count bottle for the start of the fixed-dose segment of the double-blind period containing either placebo tablets or 162-mg NAL ER tablets, depending on the arm assigned. Further bottles will be dispensed at subsequent study visits. At each visit at which investigational product is to be dispensed, enough investigational product should be supplied to ensure that a sufficient number of tablets are available for dosing until the next visit, allowing for visit windows and/or the possibility of a missed visit.

Open-Label Period

Subjects will receive a 2-week blister card for the titration segment of the open-label period. Each card will contain 3 tablets for the AM dose and 3 tablets for the PM dose except for the first day (only PM dose is included). The tablets will be combinations of either 27 mg/54 mg/placebo tablets for subjects in Arm 2 who received placebo during double-blind treatment and are crossing over to NAL ER, or a card containing 3 active 54-mg NAL ER tablets for the AM dose and three 54-mg NAL ER tablets for the PM dose for subjects in Arm 1 who are continuing with NAL ER. At the end of titration visit, subjects will be given one 70-count bottle for the start of the open-label fixed-dose period containing 162-mg NAL ER tablets. Further bottles will be dispensed at subsequent study visits. At each visit at which investigational product is to be dispensed, enough investigational product should be supplied to ensure that a sufficient number of tablets are available for dosing until the next visit, allowing for visit windows and/or the possibility of a missed visit.

Down-titration: At the regularly scheduled study visits for Week 28, Week 32, and Week 36, an elective dose reduction to a 108 mg NAL ER BID dose may be considered for subjects who achieve a confirmed WI-NRS ≤ 3 **and** visible lesion healing as assessed by an improvement of at least 1 category in the PAS healing activity score, Item 5b. For example, if "0-24%" of PN lesions at baseline were assessed as 'healed', then a 1-category change means that "25-50%" of current Week 28 PN lesions are now evaluated as 'healed'. The WI-NRS scores at and after Week 24 will be available to site staff via the Trialogics portal in order to assess eligibility for this down titration. The decision to dose reduce is at the discretion of the Investigator and with the agreement of the subject, but once this dosing has started, no re-escalation is permitted. Dose reduction is not permitted prior to Study Week 28 or after Study Week 36.

The schedule of assessments is presented in Table 1 and Table 2.

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period														
Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Informed consent[13]	X													Refer to Table 2 for schedule of open-label extension period
Medical, neurological, and Prurigo Nodularis history	X													
Physical examination	X									X				
Brief neurological assessment	X	X			X								X	
Vital signs[3]	X	X			X			X	X	X			X	
Height, weight, and BMI	X													

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period														
Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Central ECG[4] (triplicate)	X				X			X		X			X	

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period														
Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Clinical Laboratory assessments[5]	X	X			X			X		X				
Tolerability Intervention			X	X	X	X	X				X	X		
Urinalysis	X	X												
Urine pregnancy[5]	X	X			X			X	X	X			X	
Inclusion/exclusion criteria[1]	X	X												
Randomization via IVRS/IWRS		X												

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period														
Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Blood for PK assessment		X			X			X	X	X			X	

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period														
Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Record and assess AEs, concomitant medications, and therapies (including restricted and prohibited medications) [6]	X	X	X	X	X	X	X	X	X	X	X	X	X	
e-diary[7] – Dispense and train/review use	X													
e-diary – retrieve and review use and entry compliance[7]	X	X			X			X	X	X			X	
Dispense Subject	X													

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period														
Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Medication Log and Subject Symptom Log[6]														

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period														
Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Retrieve/review/re-dispense Subject Medication Log and Subject Symptom Log[6]		X			X			X	X	X			X	
ItchyQoL[7]	X	X						X	X	X				
PAS and IGA-PN	X	X							X	X				
PBI-P[7]	X	X							X	X				
Sleep Scale (PROMIS)[7]	X	X						X	X	X				
WI-NRS[7]	X	X	X	X	X	X	X	X	X	X			X	
Photography[8]		X								X				

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period

Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Dispense investigational product		Card[12]			B			B	B	Card			B[9]	

Table 1. Schedule of Assessments – Double-blind, Placebo-controlled Period														
Study Period →	**Screening and Washout[1,2]**	**Double-blind Titration Period[10]**				**Double-blind Fixed-dose Period[10]**					**Extension Titration Period**			**Open-label Fixed-dose Period**
						W3D1 to W14D7					**Transition to Open-label**			**W17, D1 to W52, D7**
Assessment ↓	**Screening and Washout**	**Baseline[1]**	☎	☎	**End of Titration**	☎	☎	**Repeating visit q4wk**			☎	☎	**End of Titration**	**Repeating extension visits**
Visit Week	**-6 to -1**	**1**	**1**	**1**	**2[9]**	**3**	**3**	**6**	**10**	**14**	**15**	**15**	**16[11,]**	**Weeks 17 to 52**
Study Day (±2 days*; Day 14 and Day 112 -2 days only)[14]	**Up to Day -1**	**1**	**3**	**7**	**14***	**15**	**21**	**42**	**70**	**98**	**101**	**105**	**112**	**Days 113 to 364**
Dispense Instructions for Taking TR11 Study Medication[12]		X			X					X			X	
Retrieve investigational product and containers					NA[9]			B + Card	B	B			NA[11]	
Investigational product accountability					X			X	X	X			X	
Investigational product - review dosing compliance					X			X	X	X			X	

AE: adverse event; **B**: bottle; **BMI**: body mass index; **Card**: blister card; D: Day; **ECG**: electrocardiogram; **e-diary**: electronic diary; **IGA-PN**: Investigator Global Assessment-Prurigo Nodularis; **IWRS**: Interactive Web Response System; **LFT**: liver function test; **NA**: not applicable; **PAS**: Prurigo Activity Score; **PBI-P**: Patient Benefit Index, pruritus version; **PK**: pharmacokinetic; **PRO**: patient-reported outcome; **q4wk**: every 4 weeks; **SHBG**: sex hormone---binding globulin; **SOWS**: Subjective Opiate Withdrawal Scale; **TSH**: thyroid stimulating hormone; **W**: Week; **WI-NRS**: Worst Itch – Numerical Rating Scale; **WOCBP**: woman of childbearing potential; ☎: telephone call.

1. If clinically indicated, **screening procedures** may be completed over multiple visits (such as needed for washout of medications intended for anti- pruritic treatment) as long as they occur within the screening period. However, if the subject meets all inclusion/exclusion criteria, they should be randomized as soon as possible after meeting study requirements, but no later than 6 weeks after the initial screening visit. Except in cases of medication washout, the baseline (randomization) visit can take place in as little as 7 days after all screening procedures have been performed, the required testing results have been obtained, and eligibility has been verified.
2. If a **medication washout** is required, WI-NRS collection should begin following completion of the washout.
3. **Vital signs** are to be obtained after the subject has been seated for at least 5 minutes.
4. **ECGs** are to be performed in triplicate (3 serial ECGs at least 1 minute apart) after the subject has been in the supine position for at least 5 minutes. ECGs are to be transmitted to be read centrally by a core ECG laboratory. The average of the QTcF values from the triplicate ECGs will be performed and presented in a report to be used as the final eligibility assessment for the subject.
5. **Laboratory Evaluations:** Hematology and serum chemistry, LFTs, specified endocrine tests (at screening and Week 14 only: TSH, with reflex free T4 if above Upper Limit Normal (ULN) or below Lower Limit Normal (LLN), free testosterone [the fraction not bound to SHBG or albumin] and SHBG for males), and serum pregnancy test if WOCBP (regardless of sexual status at screening and baseline); urine pregnancy test to be performed by the local laboratory and confirmed to be negative prior to dispensing investigational product. TSH values above the upper limit of normal (ULN) should be repeated at the next study visit (or approximately 1 month later; whichever occurs first). If TSH remains above ULN, a Free T4 will be performed on the existing laboratory sample.
6. **AEs** are to be continually assessed starting with the signing of the informed consent. Concomitant medications should be collected from 14 days priorto the signing of consent. During the screening visit, subjects must be instructed to record any **new AEs or concomitant medication use** on the **Subject Symptom Log and Subject Medication Log** and to bring the logs to every visit (including the telephone contacts). Site to retrieve and review these logs at each subsequent visit prior to re-dispensing them or issuing new logs to the subject. Events of special interest should trigger the collection of additional information using the **AESI worksheet** and documented under the **AESI section of the Adverse Event eCRF**. Event details and both **subject and investigator narratives** should be thoroughly documented in the AESI worksheet and completed under the AESI section of the Adverse Event eCRF.
7. During screening, the **e-diary** must be dispensed to minimally allow for **7 contiguous days of WI-NRS collection** prior to the baseline visit (collection during this period provides formal study data required for randomization). The **e-diary administers the WI-NRS, medication compliance, and SOWS** when applicable. Other PROs (Itchy QoL, PBI-P, and PROMIS sleep scale) are collected at select visits as described in. PROs administered during screening, but prior to Day -7, are to familiarize the subject to the instruments only, and not for the purpose of data collection. If multiple visits occur during the screening period, collection of these data (WI-NRS and PROs) should begin at the visit prior to the Week 1 baseline visit. Throughout the double-blind period, the subject is to record the WI-NRS data daily; daily WI-NRS data collection via the e-diary or in the Trialogics web portal ends at the Week 14 visit.
8. **Photography** to be performed at selected sites only. See the Canfield Quick Reference Guide and the User Reference Manual for photography procedures as applicable.
9. The **Double-blind Titration blister card** will be **reviewed, but not be retrieved at this visit**. See the Pharmacy Manual and Instructions for TakingTR11 Study Medication for additional instructions.
10. **Subjects who discontinue investigational product for any reason** (other than withdrawal of consent) prior to the Week 14 visit must be asked to complete the Off-Treatment visit, the Last Visit, and the End of Study telephone call and to return to the site for their study schedule planned Week 14 visit at study Day 98 (see Table 2)

11. The **Open-Label Titration blister card** will be **reviewed, but not retrieved**, at this visit. See the Pharmacy Manual and Instructions for Taking TR11 Study Medication for additional instructions.
12. Site to dispense the Instructions for Taking TR11 Study Medication as applicable to the subject's dosing period (i.e., Double-blind titration blister card or Double-blind fixed dose bottle). Note: Although both AM and PM doses are present on Day 1, only the PM dose will be taken (both Arms).
13. At the time of informed consent, **subjects identified as taking a concomitant short-acting benzodiazepine should be provided with the Benzodiazepine Patient Information Sheet**, informing them of safety information and warnings regarding the use of NAL ER in combination with this class of drug. Likewise, any subject who is subsequently prescribed short-acting benzodiazepines **during the course of the study** should also be provided with this information sheet.
14. At study Day 14 and Day 112, the **titration card** only contains 14 days of dosing. Subjects should be scheduled to come in on the target date or up to 2 days prior to ensure not to run out of investigational product.
15. **Height** and **BMI** will be obtained at the screening visit only. Weight will be obtained at the screening visit and subsequently according to the schedule.

DocUUID : 7e92afa8-f525-4d40-81e8-2b78d5ce783b

Table 2 Schedule of Assessments – Open-label Extension Period											
Study Period →			**Open-label Fixed-dose Period**						**Washout and Safety Follow-up Period**		**Last Contact for WI-NRS**
			Repeating visit q4wk or q8wk								
Assessment ↓	☎	☎	**W17 to W52**						**Off-treatment[1]**	**Last Visit[1]**	**End of Study Phone call[1]**
Visit Week	**17**	**17**	**20**	**24**	**28[2]**	**32[2]**	**36[2],**	**44**	**52**	**54**	**56**
Study Day (±2 days)	**113**	**120**	**140**	**168**	**196**	**224**	**252**	**308**	**364**	**378**	**392**
Physical examination									X	X	
Brief neurological assessment									X	X	
Vital signs[3]			X	X	X	X	X	X	X	X	
Central ECG[4]			X		X				X		
Clinical Laboratory assessments [5]					X			X	X	X	
Blood for PK assessment			X		X	X	X	X	X		
Urinalysis					X				X		
Urine pregnancy (WOCBP)			X	X	X	X	X[3]	X	X	X	

Table 2 Schedule of Assessments – Open-label Extension Period											
Study Period →			**Open-label Fixed-dose Period**						**Washout and Safety Follow-up Period**		**Last Contact for WI-NRS**
			Repeating visit q4wk or q8wk								
Assessment ↓	☎	☎	**W17 to W52**						**Off-treatment[1]**	**Last Visit[1]**	**End of Study Phone call[1]**
Visit Week	**17**	**17**	**20**	**24**	**28[2]**	**32[2]**	**36[2],**	**44**	**52**	**54**	**56**
Study Day (±2 days)	**113**	**120**	**140**	**168**	**196**	**224**	**252**	**308**	**364**	**378**	**392**
Record and assess AEs[6], concomitant medications, and therapies (including restricted and prohibited medications)	X	X	X	X	X	X	X	X	X	X	X
Tolerability intervention	X	X			X		X				
e-diary[7] – retrieve and/or review entry compliance			X	X	X	X	X	X	X	X	
Retrieve/review/re-dispense Subject Medication Log and Subject Symptom Log[8]			X	X	X	X	X	X	X	X	
ItchyQoL[7]			X	X	X		X	X	X		
PAS and IGA-PN			X	X	X	X	X	X	X		
PBI-P[7]			X		X		X	X	X		

Table 2 Schedule of Assessments – Open-label Extension Period											
Study Period →			**Open-label Fixed-dose Period**						**Washout and Safety Follow-up Period**		**Last Contact for WI-NRS**
			Repeating visit q4wk or q8wk								
Assessment ↓	☎	☎	**W17 to W52**						**Off-treatment[1]**	**Last Visit[1]**	**End of Study Phone call[1]**
Visit Week	**17**	**17**	**20**	**24**	**28[2]**	**32[2]**	**36[2],**	**44**	**52**	**54**	**56**
Study Day (±2 days)	**113**	**120**	**140**	**168**	**196**	**224**	**252**	**308**	**364**	**378**	**392**
Sleep Scale (PROMIS)[7]			X	X	X		X	X	X		
WI-NRS[7]			X	X	X	X	X	X	X	X	X
Photography[8]					X				X		
Administer 14-day SOWS[9]									X		
Retrieve 14-day SOWS[9]										X	
Dispense investigational product			B	B	B	B	BB	BB			
Retrieve investigational product and containers			B + Card	BB	B	B	B	BB	BB (B)[10]		
Investigational product - accountability			X	X	X	X	X	X	X		
Investigational product - review dosing compliance			X	X	X	X	X	X	X		

AE: adverse event; B: bottle; BID: twice daily; Card: blister card; ECG: electrocardiogram; e-diary: electronic diary; IGA-PN: Investigator Global Assessment-Prurigo AE: adverse event; B: bottle; BID: twice daily; Card: blister card; ECG: electrocardiogram; e-diary: electronic diary; IGA-PN: Investigator Global Assessment-Prurigo Nodularis; LFT: liver function test; NA: not applicable; PAS: Prurigo Activity Score; PBI-P: Patient Benefit Index, pruritic versions; PK: pharmacokinetic; PRO: patient-reported outcome; q4wk: every 4 weeks; q8wk: every 8 weeks; SHBG: sex hormone-binding globulin; SOWS: Subjective Opiate Withdrawal Scale; TSH: thyroid stimulating hormone; W: Week; WI-NRS: Worst Itch – Numerical Rating Scale; WOCBP: woman of childbearing potential; ☎: telephone call.

1. **Off-Treatment visit** occurs upon stopping investigational product; regardless of reason. Subjects who elect to go off-treatment prior to 52 weeks of treatment (**premature discontinuation** of investigational product) should be asked to complete the Off-Treatment visit, the 2-week SOWS evaluation, the Last Visit (occurs 14 days after off-treatment visit date), and the End of Study telephone call.
2. Assess **eligibility for down titration** based on criteria of confirmed WI-NRS ≤ 3 and 1 category improvement in PAS lesion healing. If both criteria are met, may discuss the option for a permanent dose reduction to 108 mg BID. Note: if criteria are met, this decision is at the discretion of the Investigator and with the agreement of the subject, but once dose reduction has been initiated no re-escalation is permitted.
3. **Vital signs** are to be obtained after the subject has been seated for at least 5 minutes.
4. **ECGs** are to be performed in triplicate (3 serial ECGs at least 1 minute apart) after the subject has been in the supine position for at least 5 minutes. ECGs are to be transmitted to be read centrally by a core ECG laboratory. The average of the QTcF values from the triplicate ECGs will be performed and presented in a report to be used as the final eligibility assessment for the subject.
5. **Laboratory Evaluations:** Hematology and serum chemistry, LFTs, specified endocrine (TSH all subjects with reflex free T4 if above Upper Limit Normal (ULN) or below Lower Limit Normal (LLN), free testosterone [the fraction not bound to SHBG or albumin] and SHBG for males) and pregnancy if WOCBP (regardless of sexual status); urine pregnancy test to be performed by the local laboratory and confirmed to be negative prior to dispensing investigational product. TSH values above the upper limit of normal (ULN) should be repeated at the next study visit (or approximately 1 month later; whichever occurs first). If TSH remains above the ULN, a Free T4 will be performed on the existing laboratory sample).
6. **AEs** are to be continually assessed starting with the signing of the informed consent. Subjects should continue to record any new AEs or concomitant medication use on the Subject Symptom Log and Subject Medication Log. Site staff are required to remind subjects to bring the logs to each visit (including the telephone contacts). Logs should be reviewed by site staff at every visit prior to re-dispensing (or issuing new logs as needed) to subjects. Final retrieval of the logs, for filing in source records, should occur at the last clinic visit. **Events of special interest should trigger the collection of additional information using the AESI worksheet and documented under the AESI section of the Adverse Event eCRF**. Event details and both subject and investigator narratives should be thoroughly documented in the AESI worksheet and completed under the AESI section of the Adverse Event eCRF.
7. **PROs** (Patient Reported Outcomes) are to be administered by site staff or e-diary as described in the Trialogics User Manual. WI-NRS data will also be collected from the subject via the Trialogics web portal at each clinic visit (beginning at Week 20).
8. **Photography:** To be performed at certain sites only. See the Canfield Quick Reference Guide and User Reference Manual for photography procedures as applicable.
9. The 14-day **SOWS** is to be completed, on the e-diary, at home by the subject. Site staff must change the subject's status to "Off-treatment" on the Trialogics web portal to trigger the launch of the SOWS. Subjects must be instructed to enter SOWS data daily for the 14 days following discontinuation of investigational product regardless of when this occurs (i.e. Week 52 visit or Premature Discontinuation). Data should be reviewed by site staff during the 14-day collection period in order to assess the need for any withdrawal intervention. The e-diary is to be retrieved at the last clinic study visit.

IV.

V. STUDY ENDPOINTS

A. Primary Efficacy Endpoint

The primary efficacy endpoint is the difference between the percentage of "Responders" at Week 14 for the NAL ER treatment arm versus the placebo arm. A "Responder" is defined as a subject with a ≥ 4-point decrease in the 7-day average WI-NRS from baseline to Week 14.

B. Secondary Efficacy Endpoints

Key secondary efficacy endpoints include the following:

- The mean change in ItchyQoL from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage of prurigenous lesions with excoriations/crusts (item 5a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- The mean change in sleep disturbance (PROMIS Sleep Disturbance Short Form 8a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.

Other secondary efficacy endpoints include the following:

- The mean change in 7-day average WI-NRS from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage of healed lesions (item 5b) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage number of lesions (item 2) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- Change in IGA-PN as assessed by the percentage of subjects having a 1-category improvement in activity for the NAL ER treatment arm versus the placebo arm.
- Change in IGA-PN as assessed by the percentage of subjects having a 1-category improvement in stage for the NAL ER treatment arm versus the placebo arm.
- The proportion of subjects having a PBI-P score of ≥ 1 at Week 14 for the NAL ER treatment arm versus the placebo arm.

C. Pharmacokinetics Endpoint

Nalbuphine plasma concentration (and metabolites as needed).

D. Safety Endpoints

Safety will be assessed based on adverse events (AEs), clinical laboratory measurements, central cardiac core laboratory read-12-lead ECG, vital signs, brief neurological assessments, and physical examinations.

Subjects will also complete the Subjective Opiate Withdrawal Scale (SOWS) on a daily basis for the 2 weeks following the last dose of investigational product, whenever that occurs and regardless of the reason (unless consent is withdrawn).

An independent Data Safety Monitoring Board will periodically review safety data.

5. SAMPLE SIZE AND POWER

The planned sample size of approximately 240 subjects (120 per group) is based on assumed responder percentages for placebo and NAL ER of 25% and 45%, respectively. These responder percentages represent estimates based on the Week 10 data from the TR03 study. Power is set at 90% with a 2-sided significance level of 0.05.

5.1 Responder percentage assumptions

In the TR03 study, all those who achieved a 50% reduction in WI-NRS itch also achieved at least a 4-point reduction in itch from baseline, matching the responder definition for the current study. Results showed a response rate of 6 of 18 (33.3%) for the NAL ER 162 mg treatment group versus 4 of 22 (18.2%) for placebo. Of note, the placebo response in TR03 falls on the lower end of the expected range for endpoints that reflect PRO endpoints in neurologically-mediated diseases such as pain or psychiatric disease. Studies in these areas frequently show placebo responses in the 20% to 30% range and sometimes as high as 40%. The role of placebo response in dermatology has been increasingly documented and recognized in recent years. Given the small size of the TR03 dataset, and the variability of placebo response in comparable disease states, the estimate of placebo response for TR11 was set at 25%.

The NAL ER 162 mg responder rate is estimated at 45% for TR11, versus the 33.3% observed in TR03 for the following reasons: First, the TR03 endpoint was set at Week 10, and it is expected that the additional 4 weeks of active therapy will likely increase the response rate in this treatment arm. Second, experience from the TR03 study indicated that with close follow-up and guidance during the titration period, subject drop-out due to early symptoms associated with opiate initiation can be decreased, although not eliminated. Given that the completer's analysis in TR03 demonstrated a response rate of 50% (again acknowledging the small size of the cohort), increasing the number of subjects completing the study is expected to result in a somewhat higher all-treated response rate in the active arm than was seen in TR03.

5.2 Calculations

Approximately 30 - 45 sites are expected to participate in the current study. In SAS PROC POWER, the LOGISTIC setting was used to approximate the power for a chi-square test comparing the 2 treatment groups with respect to response, without site in the model, and a CMH

test comparing the 2 treatment groups with respect to response, with site in the model. In a model without site, the sample size was n=234 (117 per treatment group). The model with sites included was originally considered with the following assumptions: 30 sites of varying sizes (6 sites with 1% of the total sample each, 6 with 2%, 6 with 3%, 6 with 4%, 3 with 5%, 1 with 6%, 1 with 8% and 1 with 11%); and varying responder rates among sites (one fixed as having the lowest, and the next lowest having an odds ratio of 1.05 compared to the lowest and so on, until the one with the highest, which has an odds ratio of [1.05]29 ≈ 4.12 compared to the lowest; this variation among sites is perhaps larger variation among sites than one would expect). In the PROC POWER runs, this site effect was specified in fixed order, and the site size was randomly scrambled. In 10 runs, the sample size needed for 90% power ranged from 239 to 242, averaging 240.7. Hence, the final target sample size is taken to be 240. With a sample size of 240, the study will have 80% power to detect a difference if the NAL ER 162 mg responder rate is as low as 41.8%.

To address the concern of the reliability of the estimates of treatment effectiveness from the previous studies, an adaptive mid-course sample size re-estimation procedure will be performed when 50% of the subjects have either completed the Week 14 primary endpoint assessment or terminated from the study early. The data for the re-estimation of sample size will be unblinded to an independent statistician and the sample size may be increased up to the predefined maximum allowable sample size if the conditional power is within the promising zone; therefore, this re-estimation will not inflate the type I error.

Given that the interim analysis is to be performed at 50% of the initial sample size (120 subjects), the targeted power is 90%, and the maximum allowable sample size is set at 1.5 times the initial sample size (360 subjects), the conditional power cut-off value (CPmin) is 41%[2]. If the CP is between 41% and 90%, the number of subjects per treatment arm will be increased, up to the maximum allowable sample size for this study (360), to recover the targeted power of 90%.

Since sample size re-estimation occurs only when the interim conditional power falls in the pre-specified “promising” range, and the study will not stop for efficacy regardless of the conditional power, the overall alpha will be protected, and the final analysis will be carried out using conventional tests, without the need for weighing the stage 1 and 2 results or adjusting the alpha value [2]

6. ANALYSIS POPULATIONS

6.1. Modified Intention-To-Treat population (mITT)

The mITT population will consist of all randomized subjects who have received at least a single dose of the investigational product. The mITT population will be used for all efficacy analyses. In the event of a discrepancy between the randomized treatment and the actual treatment received, subjects will be analysed according to the randomization treatment.

6.2. Safety population (Safety)

The safety population will consist of all randomized subjects who have received at least a single dose of the investigational product. The safety population will be used in all safety analyses. In the event of a discrepancy between the randomized treatment and the actual treatment received, subjects will be analyzed according to the treatment they actually received.

6.3. Per-Protocol population (PP)

The per-protocol (PP) population will include all subjects in the mITT population without any major protocol deviations that could have influenced the validity of the data for the primary efficacy variable.

The deviations can include, but are not limited to: e.g. inclusion criteria 1-6 and exclusion criteria 1-3, 16, 17, and 19-25 must have been satisfied.

- Key inclusion/exclusion criteria not satisfied.
- Use of prohibited concomitant medications.
- Inadequate investigational product compliance (<50% compliant), which will be determined before breaking the blind.

The subjects to be excluded from the PP population will be identified in a blinded fashion and documented in a memo prior to the database lock and unblinding.

7. STATISTICAL CONSIDERATIONS AND ANALYSIS

7.2. Derived Variables

7.2.1. Baseline

The baseline WI-NRS value is calculated as the arithmetic mean of the WI-NRS values (minimum of 5 required) taken for eligibility review by site at the time of randomization. The Trialogics original average of the WI-NRS will be used as the baseline for all analysis. For all other efficacy and safety variables, baseline is defined as the last non-missing evaluation (including repeated and unscheduled assessments) taken prior to or on the first dose of study medication.

7.2.2. Changes from Baseline

For WI-NRS, the change from baseline during the double-blind period is described in Section 7.3.1. For quantitative measurements other than WI-NRS, change from baseline will be calculated as:

Change from baseline = Test Value at Visit X – Baseline Value

If either value is missing, then the change from baseline will also be missing.

7.2.3. Date of First Dose of Study Drug

The date of randomization will be utilized for the date of first study drug dose, and will be referred to as the start date of study drug.

For the analysis referring to the open-label period, the date of the first administration of the extension titration medication will be considered as the start of this period.

7.2.4. Date of Last Administration of Study Drug

The date of last administration of the study drug in the double-blind period is derived as the last date when a dose of study drug was administered and recorded on the eCRF page of “Exposure (EX)” in that period.

The date of last administration of the study drug in the open-label period is derived as the last date when a dose of study drug was administered and recorded on the eCRF page of “Exposure (EX)” in that period.

The date of the last administration of the study drug for the study is derived as the last date when a dose of the study drug was administered and recorded on the eCRF page of "Exposure (EX)" whenever that occurs.

7.2.5. Study Day

The study day (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption, etc.) will be calculated using the start date the of double-blind study drug (Study Day 1) as the origin. Subsequent study days will be calculated as (date of assessment) – (start date of study drug) + 1.

For visits (or events) that occur prior to study day 1, the study day is defined as (date of visit (or events) – date of study day 1). There is no study day 0. Therefore, the study day displayed on the listing will be negative.

7.2.6. Exposure and Compliance

Duration of exposure (days) = (Date of last administration of study drug - Date of first administration of study drug) +1.

Compliance (%) = Number of tablets subject actually took/ Number of tablets subject was expected to take multiplied by 100.

Duration and compliance will be calculated for the double-blind and open-label periods, separately.

7.2.7. Visit Window Definitions

Windowing is applicable to efficacy and safety endpoints, except for WI-NRS and SOWS for reporting purposes.

	BL*	Week 2	Week 6	Week 10	Week 14 **	Week 16	Week 20	Week 24	Week 28	Week 32	Week 36	Week 44	Week 52	Week 54	Week 56
Scheduled Day	1	14	42	70	98	112	140	168	196	224	252	308	364	378	
ItchyQoL	Day 1		22 to 56	57 to 84	85 to (titration start day-1)		titration start day to 154	155 to 182	183 to 224		225 to 280	281 to 336	>=337		
PAS and IGA-PN	Day 1			29 to 84	85 to (titration start day-1)		titration start day to 154	155 to 182	183 to 210	211 to 238	239 to 280	281 to 336	>=337		
PBI-P	Day 1			29 to 84	85 to (titration start day-1)		titration start day to 168		169 to 210	211 to 238	239 to 280	281 to 336	>=337		
Sleep Scale	Day 1		22 to 56	57 to 84	85 to (titration start day-1)		titration start day to 154	155 to 182	183 to 224		225 to 280	281 to 336	>=337		
Vital Signs	Day 1	8 to 28	29 to 56	57 to 84	85 to (titration start day-1)	titration start day to 126	127 to 154	155 to 182	183 to 210	211 to 238	239 to 280	281 to 336	337 to 371	>=372	
Clinical lab	Day 1	8 to 28	29 to 70		71 to (titration start day-1)				106 to 252			253 to 336	337 to 371	>=372	
Urinalysis	Day 1								99 to 280				>=281		

* If a Day 1 value is missing, then the last screening value is used as the baseline value.

** In the Week 14 column, the end of the window applies to subjects who go on to enter the open-label period. For subjects who don't enter the open-label period, the end of the window is 14 days later than that displayed in the column.

Note: ECG windowing will be established by ERT, who will perform ECG analyses.

If two or more observations fall into the same window, the observation closest to the scheduled day will be used in the analysis. If there is a tie, the later observation will be used.

7.3. Efficacy Derived Variables

7.3.1. WI-NRS

The WI-NRS is a PRO instrument, designed to quantify the intensity of worst itching experienced during a 24-hour period. The scale is a set of boxes, one for each number, from 0 (no itching) to 10 (worst possible itching). The NRS is a widely used instrument recommended by the IFSI for quantifying itch intensity as well as a useful instrument for grouping subjects into categories of itch intensity described as mild, moderate, or severe[3]: The itch NRS has been investigated in subjects with chronic pruritus of a variety of origins, and a high reliability and concurrent validity was found[3,4].

In this study, during the double-blind period, subjects will be asked to record the WI-NRS value daily, at approximately the same time each day, usually in the late afternoon or early evening. All WI-NRS information will be entered into the e-diary. During the open-label period, only a single value will be collected for the WI-NRS at each study visit. This will be collected using the same electronic diary format via web portal.

The primary endpoint depends on the change from baseline to Week 14. The calculation of the Week 14 value is described below.

For the 2-, 6- and 10-week visits, the WI-NRS will be calculated as the arithmetic means of the WI-NRS values on Study Days 8 through 14 (Week 2), 36 through 42 (Week 6), and 64 through 70 (Week 10), respectively, regardless of when the Weeks 2, 6 and 10 visits actually occur.

The Week 14 WI-NRS is calculated as follows:

- Case 1: Subject's Week 14 visit is on or after Study Day 98 (target day)
  - Use all values in the period Study Day 92 through Study Day 98
    - If there are ≥ 5 values, calculate the arithmetic mean of them
    - If there are < 5 values, calculate an arithmetic mean based on all of them plus any available values in the range from Study Day 99 till titration begins that will bring the total to 5 values, but do not use any values collected after titration begins. If this range has more values than necessary to bring the total to 5 values, use values closest to Study Day 98. If there are still less than 5 values, continue to use any available values in the first 3

days before Day 92 that will bring the total to 5 values. If this range has more values than necessary to bring the total to 5 values, use values closest to Study Day 92

- Case 2: Subject's Week 14 visit is before Study Day 98
  - Use all values from Study Day 92 to the Study Day of the Week 14 visit
    - If there are ≥ 5 values, calculate the arithmetic mean of them
    - If there are < 5 values, calculate an arithmetic mean based on all of them plus any available values in the first 3 days before Day 92 that will bring the total to 5 values. If this range has more values than necessary to bring the total to 5 values, use values closest to Study Day 92

A "Responder" is defined as a subject with a ≥ 4-point decrease in the 7-day average WI-NRS from baseline to Week 14. Baseline is as described in Section 7.2.1.

7.3.2. ItchyQoL Total Score (ItchyQoL)

The ItchyQoL consists of 22 pruritus-specific items measuring how pruritus affects subjects' QoL in the area of symptoms related to the itch condition (6 questions), functional limitations (7 questions), and emotions (9 questions). The subject scores each question never = 1, rarely = 2, sometimes = 3, often = 4, all the time = 5. The ItchyQoL must be administered at the site under supervision by trained and authorized study staff, strictly adhering to administration instructions.

The ItchyQoL total score will be obtained as the sum of the 22 items. For each subscale (symptoms, functional limitations, emotions) of ItchyQoL, if one item within a subscale is missing, the median of remaining items within that subscale will be used to impute the missing value. If more than 1 item within a subscale is missing, that subscale and the ItchyQoL total score will be set to missing.

7.3.3. PBI-P

Before therapy, the subject fills in the PBI-P questionnaire as to the individual importance of treatment objectives. During and after therapy, the subject completes a matched "on-treatment" questionnaire and rates the extent to which the treatment objectives have been achieved. The instrument consists of 27 multiple choice questions that can be answered "not at all", "somewhat", "moderately", "quite", and "very". The PBI-P must be administered at the site under supervision by trained and authorized study staff, strictly adhering to administration instructions[4].

Part 1: Patient Needs Questionnaire (PNQ). The PNQ consists of the "treatment goal" items. Each item has 5 categories: not at all (0), somewhat (1), moderately (2), quite (3), and very (4). PNQ is analysed descriptively. The analysis consists of the following:

- means and standard deviations of all items ("does not apply to me" is coded as 0 for this single-item analysis of the PNQ)
- for each item the percentage of the response "does not apply to me"
- for each item the percentage of responses with high agreement (quite or very)
- percentage of missing values

Part 2: Patient Benefit Questionnaire (PBQ). The PBQ consists of the "benefit" items measuring the extent to which the treatment goals have been achieved. It is analysed descriptively. The analysis consists of the following:

- means and standard deviations of all items (("did not apply to me") is treated as missing here)
- for each item the percentage of the response "did not apply to me"
- for each item the percentage of responses with high agreement (quite or very)
- percentage of missing values

The PBI-P global score may only be computed if the subject has provided valid data on importance (PNQ) and benefit (PBQ) for at least 75% of the respective treatment goals, that is, for at least 21 of the 27 items. The responses "does/did not apply" and "question answered" are considered valid values when counting the number of non-missing responses.

The PBI-P global score is computed for each subject according to the following algorithm.
Identify the subjects who have 21/27 needs answered at baseline & 21/27 of the benefits questions answered at each visit. For this assessment 0 & 5 are considered non-missing.

PBI-P Global Score:

Individual PNQ ('Patient Needs' or 'Importance') scores are in the SDTM data set QS.
(QS.QSTESTCD.PBIP101 to QS.QSTESTCD.PBIP127).

Individual PBQ ('Patient Benefit' or 'Benefit') scores are in the SDTM data set QS.
(QS.QSTESTCD.PBIP201 to QS. QSTESTCD.PBIP227).

1. Determine individual baseline PNQ scores.

2. Count PNQ scores, where individual score is from 1 to 4, as PNQ count.
3. Sum PNQ scores, where individual score is from 1 to 4, as PNQ sum.
4. Merge PBQ score at visit with corresponding baseline PNQ scores, creating item pairs. (e.g., merge PBIP101 with PBIP201) by subject-visit.
5. For each item pair, if both PNQ and PBQ score is 1 to 4, multiply PNQ*PBQ as calculated raw score.
6. Sum calculated raw scores as PBQ sum.
7. If PNQ count >= 21, calculate and output PBI-P Global Score as (PBQ sum/PNQ sum).

However, for the score calculation, both "does/did not apply" and "question unanswered" will be treated as missing values. The global score will be calculated using only those items pairs (i.e., importance item and benefit item) for which the subject has given a response other than "does/did not apply" in both PNQ and PBQ. That is, importance items for which the corresponding benefit item has been rated as "does/did not apply" or "not at all" (or vice versa) will not be included in the sum of needs (i.e., the denominator in the algorithm).

The interpretation of data includes the distribution characteristics of PBI-P (e.g., mean and standard deviation) and the proportion of subjects with PBI-P ≥ 1, PBI-P ≥ 2, PBI-P ≥ 3. Subjects with PBI-P ≥ 1 are considered as having at least minimum patient-relevant treatment benefit (based on pilot studies).

7.3.4. Prurigo Activity Score (PAS)

The PAS consists of 5 quantitative and qualitative measurements related to the examination of the skin: Type; number; distribution; quantitative number of lesions in a representative body part; and activity. Prurigo lesion activity is recorded as a stage (0 to 4), based on the percentage of overall lesions with the relevant characteristic.

Three types of PAS responders are defined, one for each of the following items: Prurigenous lesions with excoriations/crusts (item 5a); Healed lesions (item 5b); Number of lesions (item 2). For each of these, a responder is defined as a subject who has at least 1-category improvement in the relevant item from baseline to Week 14. For all three, a patient who does not have Week 14 measurement is a non-responder.

7.3.5. Investigator Global Assessment – Prurigo Nodularis (IGA-PN).

The IGA-PN collects an Investigator Global Assessment of the status of the PN skin lesions. The IGA-PN uses a 5-category scale (scoring 0 to 4) for assessing the status of 2 aspects of the disease: The Activity (amount of excoriation and crusting associated with the prurigo lesions), and the Stage (the quantitative presence and proportion of flattening of the lesions). An IGA-PN responder for Activity or Stage is defined as a subject who has at a least 1-category improvement in the respective score from baseline to Week 14.

7.3.6. Sleep Scale

There is no assessment scale that is targeted specifically to evaluate sleep in the PN population. The Patient Reported Outcomes Measurement Information System (PROMIS) Sleep Disturbance Short Form 8a questionnaire has been developed as a general tool for assessing sleep in the context of clinical trials. It consists of 8 open-ended statements about the subject's sleep over the past 7 days, with 5 options for completing the statement. There is 1 broad sleep quality question with options for completing with: "very poor", "poor", "fair", "good", and "very good". The remaining 7 questions can be answered with: "not at all", "a little bit", "somewhat", "quite a bit", and "very much". The sleep scale score is estimated per algorithm below.

Each question usually has five response options ranging in value from one to five. To find the total raw score for a short form with all questions answered, sum the values of the response to each question and use Table 1 in the Appendix to translate the total raw score into a T-score for each participant. The 95% confidence interval around the observed score is estimated as (T-score +/- (1.96*SE).

7.4. Handling of missing data and outliers

7.4.1. Missing data analysis methods

For the primary efficacy endpoint, missing values at week 14 will be set to missing. The data will be analysed using the mixed model technique. The analysis assumes that the missing data is missing at random (MAR). Thus, it provides an unbiased estimate of the treatment effect that would have been observed had all subjects continued on treatment until Week 14. It therefore assumes that the response for withdrawn subjects will follow the trajectory of the respective treatment after discontinuation. In order to assess the robustness of the results to the MAR assumption, a sensitivity analysis will be conducted under

the assumption that the data is missing not at random (MNAR). The details of this analysis are described in

For the change from baseline in ItchyQoL and PROMIS, all observed data up to and including Week 14 will be used without using imputation methods for missing data. However, in sensitivity analyses, two multiple imputation (MI) methods will be used: Control based multiple Imputation and Tipping Point Analysis approaches

7.4.2. Handling of missing or incomplete dates

Unless otherwise stated, other than for partial dates, missing dates will not be imputed and will be treated as missing, as specified below. The imputed date will be included in the analysis datasets, but not displayed in the listings.

Concomitant Medications Start/End Date

Partial missing start/end date to concomitant medications will be imputed as follows:

Start date of concomitant medications

- If only the day of the month is missing, use the first day of the month to replace the missing part;
- If both the day and the month are missing, January 1st will be used to replace the missing part;
- If day, month and year are all missing, use a date one day before the first date of study drug.

End date of concomitant medications

- If only day is missing, use the last day of the month
- If both day and month are missing, use the last day of the year
- If day, month and year are all missing, assign 'continuing' status to stop date

Adverse Events Start/End Date

Missing or partial missing date to the onset date of AE will be imputed as below:

- If the AE onset date is completely missing, the AE start date will be imputed as the first date of study drug;
- If the AE onset date is partially missing, then
    - If both the year and the month are available, and they correspond to the year and month of the first dosing date, then the AE start date will be imputed as the first dosing date;
    - If both the year and the month are available, and they do not correspond to the year and month of the first dosing date, then the AE start date will be imputed as the 1st day of the month;

- If only the year is available, and it corresponds to the year of the first dosing date, then the AE start date will be imputed as the first dosing date;
- If only the year is available, and it does not correspond to the year of the first date of study drug, then the AE start date will be imputed as the January 1st of the year.

Adverse event end date will be imputed as below for the partial date only.

- If both the year and the month are available, AE end date will be imputed as the last day of the month;
- If only the year is available, AE end date will be imputed as the December 31st of the year.

  If the imputed AE end date is after the death date, then the date of the death will be imputed as the AE end date.

8. STATISTICAL METHODS AND ANALYSI8

8.1 General Statistical Conventions

All statistical procedures will be completed using SAS 9.4® or higher.

Unless otherwise stated, all statistical testing will be two-sided and will be performed using a significance (alpha) level of 0.05, and two-sided 95% confidence intervals (CI) will be provided when relevant.

Continuous variables will be summarized using descriptive statistics, including number of subjects (n), mean, median, standard deviation (SD), minimum and maximum. For efficacy variables with subscales, the total score and subscales will be summarized.

For categorical variables, summaries will include counts of subjects and percentages. Percentages will be rounded to one decimal place.

Unless otherwise stated all summaries will be presented by treatment group.

All subject data, including those derived, will be presented in individual subject data listings. Unless otherwise stated, unscheduled visit results will be included in date/time chronological order, within subject listings only. All listings will be sorted by treatment group, subject ID, date/time and visit/study day. Subject's sex and age will be stated on each listing. Unless otherwise stated, data listings will be based on All Subjects Randomized Set.

8.2 Subject disposition

Subject disposition information will be tabulated for all screened subjects, including rows for all screened subjects, randomized subjects (total and by whether or not treated), and screen failures (total and by reason).

A separate table will describe disposition for all randomized subjects, by treatment group and overall. The table will provide rows for the number of randomized subjects and the number of treated subjects (i.e., the mITT population), as well as the numbers and percentages (based on the mITT population) of subjects who completed the double-blind period, who entered the open-label period, who completed the entire study, and who withdrew early (by period). The primary reason for early withdrawal will also be tabulated.

The number and percentage (using the mITT population as the denominator) of subjects in each analysis set will also be tabulated.

By-subject disposition listings will be provided separately for screened subjects who were not in the ITT population and for subjects in the mITT population.

Treatment Misallocations:

If a subject was randomized but took incorrect treatment throughout the double-blind period, they will be reported under the randomized treatment group for all efficacy analysis, but will be reported under the treatment they actually received for all safety analysis. However, if a subject took the incorrect treatment for only part of the time during the double-blind period, they will be reported under the randomized treatment group for all efficacy and safety analysis. The subjects who took incorrect treatment intermittently will be identified, and Ad Hoc analyses may be performed.

Protocol deviations

All key protocol deviations (categorized) will be summarized by treatment group and overall, using the mITT population. Subjects excluded from the mITT population to yield the Per-Protocol population will be listed, with reason for exclusion, treatment group, age, sex and race. A by-subject listing will include the inclusion/exclusion criteria violated at Screening and at Baseline Visits as well as other protocol deviations identified, based on data recorded on the eCRF and/or protocol deviation logs from ICON Medical (on Randomized subjects).

8.3 Demographics and baseline characteristics

8.3.1 Demographics

Age, height, weight, BMI and other continuous demographic variables at baseline will be summarized descriptively using the mITT population. Sex, race, ethnicity and other categorical variables will be summarized using frequency counts and percentage for the mITT Population. By-subject listings will also be provided.

8.3.2 Baseline and disease characteristics (Prurigo Nodularis History)

The following categorical disease baseline characteristics will be summarized using frequency counts and percentages for the mITT population:

- When did the subject's Prurigo Nodularis start?
- If less than 3 months, did it start greater than 6 weeks ago?
- The quality of the itch
- What increases the itching?
- Scratching Behavior 1
- Scratching Behavior 2

- How does the subject perform the scratching?
- Previous treatments for Prurigo Nodularis

Data in Part 2 of the Prurigo Nodularis History represent Past Medical History (PMH) that may represent a risk factor for subsequent evolution of PN. Analyses of key efficacy endpoints will be performed for subgroups of sufficient size (See Section 8.6.2). These subgroups will combine subjects with relevant past medical histories based on the general understanding of the mechanism that may drive the association with itch. The following broad subgroups are anticipated, although the final grouping will be established based on the actual PMH data available prior to unblinding. Of note, the immune system may contribute to the itch component across these various categories and diseases, and if the understanding of PN itch evolves during the timeframe of the TR11 trial, these categories may shift.

- Subjects with a past history of *dermatologic diseases* associated with itch (atopic dermatitis, nummular or discoid eczema, venous stasis skin disorder, Epidermolysis Bullosa Acquisita, or other skin disease depending on type);
- Subjects with a past history of *metabolic diseases* associated with itch, or with other primary diseases thought for which itch induction is thought to act through a metabolic pathway, such as hepatic impairment (Diabetes, Diabetic or Non-Diabetic Kidney Disease, Chronic Anemia, Hypo- or Hyperthyroidism, Hepatitis B or C presumably mediated via hepatic dysfunction);
- Subjects with a non-dermatologic history of *atopic disease* (gluten enteropathy/Celiac disease, food allergy, environmental allergy, malabsorption/nutritional deficiencies);
- Subjects with a past history of neurological disease (peripheral neuropathy of hands/feet; other peripheral neuropathy; radiculopathy/sciatica; degenerative disc disease; degenerative vertebral bony disease; notalgia paresthetica; brachioradial neuropathy; trigeminal neuropathy; history of herpes simplex [severe] or herpes zoster infection);
- Other: subjects with a past history of various miscellaneous diseases expected to be of lower frequency (gastric ulcer disease, HIV infection, Haematologic malignancy, solid organ malignancy).

These conditions will be summarized using frequency counts and percentages using the mITT population, with the following categories for each condition: Yes, resolved; Yes, not resolved; No. Depending on

the completeness of the dates of onset and resolution, durations may be summarized with descriptive statistics.

By-subject listings will also be provided.

8.3.3 Neurological history

Neurological history will be summarized using frequency counts and percentages for the mITT population.

By-subject listings will also be provided.

8.3.4 Medical history

A summary of medical history will be presented by system organ class (SOC) and preferred term (PT) using Medical Dictionary for Regulatory Affairs® (MedDRA) Version 21.1 or higher. By-subject listings will also be provided.

8.3.5 Prior and concomitant medications

Medications used in this study will be coded using the latest available version of the World Health Organization Drug Dictionary Enhanced (WHODDE).

Prior medications: those medications taken (i.e., dosing documented) during the last 14 days prior to the first dose of study drug.

Concomitant medications (double-blind period): those medications taken (i.e., dosing documented) at any time from the first dose of study medication date through the end of the subject's double-blind participation.

Concomitant medications (open-label period): those medications taken (i.e., dosing documented) at any time from the start of the open label titration period (i.e., day of first dose of open label titration study drug supplies) through the date of the last dose of study medication.

Prior medications and concomitant medications will be summarized descriptively using frequency tables by Anatomical Therapeutic Chemical' (ATC) class and generic name by treatment group for the mITT population and presented separately for the following groups:

- Prior medications;
- Concomitant medications taken during the double-blind treatment period;
- Concomitant medications taken during the open-label period;
- Medications taken during the first 14 days after the date of the last dose of study medication.

Details for imputing missing or partial start and/or stop dates of medication are described in

8.4 Extent of exposure

8.4.1 Treatment duration

The duration of double-blind period study drug (in days) will be calculated as: last dose date in the double-blind period – first dose date in the double-blind period + 1 day, regardless of study drug interruption.

The duration of the open-label period study drug (in days) will be calculated as: last dose date in the open-label period – first dose date in the extension titration period + 1 day, regardless of study drug interruption.

Duration of overall study drug (in days) will be calculated as: last dose date – first dose date + 1 day, regardless of study drug interruption.

Study drug exposure will be summarised by treatment group for the safety population using descriptive statistics. By-subject listings will also be provided.

8.4.2 Treatment compliance

Compliance in the double-blind period (%) = Number of double-blind period tablets subject actually took/ Number of double-blind period tablets subject was expected to take*100. In the double-blind titration period, a subject is expected to take 3 tablets in the AM and 3 tablets in PM for two weeks, with the exception of Day 1, on which they take only the PM dose. In the double-blind fixed-dose period, a subject is expected to take 1 tablet each in the AM and PM. The number of double-blind period tablets a subject was expected to take is the sum of {6 x [last dose date of double-blind titration – first dose date of double-blind titration +1] – 3} and {2 x [last dose date of double-blind fixed-dose – first dose date of double-blind fixed-dose +1]}. The number of double-blind period tablets a subject actually took will come from the pill counts from the eCRF, i.e., the number of tablets dispensed - the number of tablets returned from the blister cards and bottles for the double-blind period.

Compliance in the open-label period (%) = Number of tablets subject actually took during open-label period/ Number of tablets subject was expected to take during open-label period*100. The compliance is calculated similarly to the double-blind period. The number of tablets a subject was expected to take is the sum of {6 x [last dose date of extension titration – first dose date of extension titration +1] – 3} and {2 x [last dose date of open-label fixed-dose period – first dose date of open-label fixed-dose period +1]}.

Number of tablets a subject actually took during open-label period is calculated by the number of tablets dispensed - the number of tablets returned from the blister cards and bottles for the open-label period.
Study drug compliance will be summarized by treatment group, with the number of subjects (n), mean, standard deviation, standard error, median, minimum, and maximum. They will also be summarized in categories "< 50% compliant" and "≥ 50% compliant" using frequency tables. Also, the categories of "< 60% compliant" and "≥60% compliant" will be provided by count and percentage. By-subject listings will also be provided.
Study drug compliance summaries will be based on the Safety Population.

8.4.3 Treatment dose reduction

The count and percentage of subjects who reduced their dose per protocol (at weeks 28, 32 or 36) will be summarised by prior treatment group on the Safety Population.

8.5 Efficacy analyses

8.5.1 Analysis methods

8.5.1.1. Logistic Regression Analysis

The primary efficacy endpoint (and other binary efficacy endpoints) "responder type" efficacy endpoints will be analyzed with a logistic regression model with baseline score as a covariate and site (USA versus other) as a fixed effect. The logistic regression model will provide estimated response rates, standard error, odds ratio (OR), two-sided 95% confidence interval for OR, and p-values at Week 14.

8.5.1.2 Mixed model repeated measures analysis

A mixed model repeated measures analysis using 'direct likelihood' will be performed for continuous efficacy outcomes. The SAS procedure PROC MIXED will be used. The preferred model will include the fixed effects of treatment, visit, treatment-by-visit interaction, and the baseline value of the endpoint variable as a covariate. An unstructured matrix for the variance-covariance will be used. The denominator degrees of freedom will be calculated according to the Kenward-Roger method.
In case of non-convergence of the preferred model or memory issues, the following back-up models are defined:

1. The same as the preferred model, but the Kenward-Roger method will be replaced by Satterthwaite approximation.

2. The same as the preferred model, but with handling of the selection of the covariance structure for the final model in a hierarchical fashion, with the order being Toeplitz, AR (1), and compound symmetry, respectively.

The second back-up model will only be used if the first back-up model does not converge or has memory issues.

The model will provide least-squares mean estimates, standard errors, two-sided 95% confidence intervals and p-values (where applicable) for mean change at all visits between treatments (NAL ER vs. placebo).

8.6. Multiplicity

In order to control for Type, I error (0.05), a fixed sequence testing procedure will be used on the primary and 3 key secondary endpoints. All endpoints will be tested (NAL ER treatment arm versus the placebo arm) at the 0.05 level of significance, following a pre-specified order. As soon as one endpoint assessment is found to be non-significant, subsequent endpoints will not be assessed. The rank order of testing is specified below.

- Time to onset of effect on pruritus as measured by proportion of participants with an improvement (reduction) in WI-NRS by ≥4 from baseline during the 14-week treatment period.
- Change in Investigator Global Assessment-Prurigo Nodularis (IGA-PN) as assessed by the percentage of subjects having a 1-category improvement in activity.
- The mean change in 7-day average WI-NRS from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage of prurigenous lesions with excoriations/crusts (item 5a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage of healed lesions (item 5b) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage number of lesions (item 2) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- The mean change in ItchyQoL from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.

- The mean change in sleep disturbance (PROMIS Sleep Disturbance Short Form 8a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.
- The proportion of subjects having a PBI-P score of ≥ 1 at Week 14 for the NAL ER treatment arm versus the placebo arm.

8.6 Analysis of Primary Efficacy Endpoint

Hypothesis:

The null hypothesis is that there is no treatment group difference (NAL ER vs. placebo) for the primary efficacy endpoint and the alternative hypothesis is that a treatment group difference exists.

Statistical Decision Rules

Unless otherwise indicated, all tests will be 2-sided, and differences resulting in a p-value <0.05 will be considered statistically significant.

The primary efficacy endpoint is the difference between the percent responders at Week 14 for the NAL ER treatment arm versus the placebo arm.

A responder is defined as a subject with a ≥ 4-point decrease in the 7-day average WI-NRS from baseline to Week 14.

The proportions of responders in NAL ER and placebo will be compared using a logistic regression model, with baseline WI-NRS as a covariate and country as fixed effect, as described below. Missing Week 14 WI-NRS data will be set to missing.

Sample SAS code:

```
proc genmod data=winrsa descending; where avisitn ge 3;
class usubjid avisitn trt posite ;
model resp = trt avisitn avisitn*trt posite base
/ dist=binomial;
repeated subject=usubjid / withinsubject=avisitn
type=un;
lsmeans trt*avisitn /cl ilink diff oddsratio;
Ods output diffs=diff lsmeans=lsm ;
run;
```

*** Estimate Response rate and SE**;

Recoding the resp to a new resp variable

```
data winrsa; set winrsa;
if resp = 0 then respnw = 2;
else if resp= 1 then respnw = 1;
run;
```

Using proc freq to create the binomial dataset;

```
proc sort data = winrsa; by avisitn trt;
proc freq data = winrsa; by avisitn trt;
tables respnw/ out = FreqCount  binomial ;
exact binomial;
ods output binomial=bin;
run;
```

The variable *_bin_* is the response rate for resp = 1;
The variable for SE is E_bin

*** Estimate Odds ratio and CI;**
These are provided in the diff dataset and can be output directly.

run;

Sensitivity analysis

To assess the impact of the above "treatment failure" rule, the missing data strategy will employ the control-based multiple imputation (MI) procedure and delta-based multiple imputation (or tipping point analysis) to assess the sensitivity of the study conclusions to the missing response data. Multiple imputation provides a better estimate of the effect of missing data on the variance of the estimates and is therefore less likely to result in false positive inferences. Control based imputation restricts imputing responses for the active arm participants to responses in the control arm. This process will be done sequentially for each time point with missing responses. The process is described by Ratitch & O'Kelly (2011)[6] and summarized below. These methods will be implemented in SAS9.4® using PROC MI and PROC MIANALYZE and are based on pattern mixture models that assume monotone drop out.
In order to meet the monotone assumption, if there are intermittent missing data points, these will be first imputed using a Monte Carlo Markov Chain Method (MCMC). Please refer to Appendix.
Intermittent missing data is typically a very small percentage of overall missing and is not expected to affect or bias inference. The MI procedure will be used for subjects with missing Week 14 WI-NRS data and/or who withdraw prematurely prior to Week 14.

8.6.1.1 Control Based Imputation.

The monotone missing data is sequentially imputed using the responses of the control group only. This is done sequentially for each time point. In order to ensure that only the control group data is used for active arm subjects with responses missing starting at time t1, the data set is subset to include all control subjects and only those active arm subjects who have missing values at time t. This is repeated for each time point going forward – please refer to Appendix for details.

PROC MI is then used to impute all missing values for t1 in the treatment arm using the standard PMM imputation approach. Please refer to Appendix for details.
The data set is combined with the rest of the data and a new subset is created for active arm recipients with responses missing for time t2. This is repeated for each subsequent time point.

Finally, each fully imputed data set is analyzed using the appropriate procedure, e.g., Proc Mixed, and then the results are collated using PROC MIANALYZE.

After imputing the missing values, logistic regression will be performed 50 times using imputed datasets. The estimates will be combined using PROC MIANALYZE procedure.

8.6.1.2. Tipping point analysis.

The tipping point analyses is to assess how much the missing data would have to differ from the imputed data that is based on a MAR assumption in order to reverse the conclusions of the trials. This can be done using PROC MIANALYZE. Tipping point analyses are typically used to ensure that missing data imputation does not create false positive inferences.

this time, we do not know if the current study will result in an inference that suggests treatment benefits. The TPA description below describes the derivation for treatment benefit and absence of treatment benefit. The primary decision to be made in the tipping point analysis is determining the value of delta). In this case, we have used the standard deviation of the efficacy measure of the subjects in the treatment group who completed all visits.

Non-monotonic data will be imputed as above. The Monotonic missing data will be the result of subjects dropping out of the study. For the primary endpoint, we are only concerned with responders (decrease ≥ 4 points in the WINRS scale) and non-responders at week 14. So in this case we will be looking at the dropouts in the stable double-blind phase.

Determining the delta for the tipping point analysis would follow the following steps. Impute missing values one visit at a time as above for control subjects; Please refer to the Appendix for details.

1. Estimate the mean scores and standard deviations (SD) in those who completed all visits;
2. Use the standard deviation of the completers as the initial delta value for the dropouts;
3. If the week 2 value is missing in the original data, then in the TPA, the imputed value for week 2 will be modified by adding the SD of the completer's week 2 value to the imputed value. Assuming that the scores in completers indicate better outcomes than dropouts, i.e. mean score in completers < mean score in dropouts, the TPA score will first assume worse outcomes in dropouts. The delta value will therefore be added to the mean score of completers to impute missing scores for the dropouts. Since the data set has been converted to monotonic missing data, if week 2 values are missing, week 4, week 6, week 10 & week 14 will also be missing & have been imputed assuming MAR. Therefore, each of these values will be increased by the SD of the corresponding week for the completers. This will be repeated for each set of missing values, i.e.,

if the week of dropout is week 4, then week 4, week6, week 10 & week 14 imputed values will be modified.

4. After we run the models adding the SD to the imputed values, we will also subtract the delta value to assess how inference might change if dropouts had better outcomes than completers and the delta values will be subtracted from the mean scores of completers to impute values for the dropouts.

8.6.1.3. Sensitivity Analysis for Subjects who take Prohibited Medications

This analysis is to assess the impact of concomitant medications that might affect the primary endpoint. The medications listed below are considered inter-current events (ICEs) confounding with the efficacy outcomes.

- μ-opioid receptor agonists (e.g., codeine, oxycodone, hydromorphone, oxymorphone, morphine, fentanyl, tramadol, butorphanol, pentazocine, meperidine, methadone)
- Opioid receptor antagonists naltrexone and nalmefene
- Antihistamines (topical or systemic)
- Topical calcineurin inhibitors
- Topical corticosteroids
- Topical capsaicin
- Anti-convulsant class drugs
- Cannabinoid agonists
- Topical doxepin
- Thalidomide or methotrexate
- Cyclosporin A
- Biologics with an immune suppressive effect*
- Investigational drug products*
- Cryosurgery
- UV-therapy

Efficacy data on subjects who experience an ICE will be set to missing at the designated time points after the start of such events. The analysis will provide an answer to the question that is crucial to individual subjects; "If I take this study medication as part of my treatment regimen, without adding any further medications that may impact the underlying disease, what response (response is defined as a ≥ 4-point reduction in the 7-day average WI-NRS) might be anticipated after 14 weeks?"

Sensitivity analysis using missing data handling approaches will also be performed as outlined in the Appendix.

In addition, the primary endpoint will be analyzed using the PP population.

8.6.2. Subgroup analyses

Additional subgroup summaries will be performed for the primary endpoint to determine whether differences exist in primary endpoint results between subgroups. These subgroup summaries will be carried out using the subjects from the mITT.

The list of potential subgroups (with applicable definitions in parentheses) includes, but is not necessarily limited to, the following:

- Age (18-39, 40-64, 65+ years)
- Sex (Female vs. Male)
- History of diseases associated with, or risk factors for, prurigo nodularis, as described in
    - Dermatologic diseases;
    - Conditions associated with metabolic itch (e.g., liver, renal or endocrine);
    - Atopy (including atopic dermatitis alone, if the numbers are large enough);
    - Neurologic conditions associated with itch;
    - Other.

These medical history subgroup analyses will be performed if the smaller of the 2 groups (with or without the condition) has at least 30 subjects. Additional post-hoc subgroup evaluations may be performed if the initial observations warrant it.

For each subgroup variable, the frequency and percentage of responders for each category will be displayed by treatment group. The difference in response rate between the treatment groups and 95% CI will also be presented. These subgroup analyses will be considered exploratory, as they do not control for multiplicity.

8.6.3 Analysis of secondary efficacy endpoints

All secondary endpoints will be analyzed using the mITT population.

MMRM without imputation for missing data:

The following secondary efficacy endpoints will be analyzed using a mixed model for repeated measures (MMRM):

- The mean change in 7-day average WINRS from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- The mean change in ItchyQoL from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- The mean change in sleep disturbance (PROMIS Sleep Disturbance Short Form 8a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo Arm.

The analysis of change from baseline to week 14 will be performed utilizing all observed data for visits up to and including Week 14, without imputation for missing data. The model used for this analysis will include the fixed effects of treatment, visit, treatment-by-visit interaction, and baseline value. An unstructured variance-covariance matrix will be used in the analyses. The comparison of interest is the treatment difference at Week 14. The model will provide least-squares mean estimates, standard errors, two-sided 95% confidence intervals and p-values for mean change at all time points between treatments (NAL ER vs. placebo) for ItchyQoL (total score and subscales: symptoms, functional limitations, emotions) and PROMIS. The same statistics will be provided for WI-NRS, except for the p-values.
Sample SAS code:

```
PROC MIXED data=xxxx;
  CLASS visit treatment subjid;
  MODEL CHG = treatment visit treatment*visit baseline/solution DDFM=KENWARDROGER;
  REPEATED visit / type=un subject=subjid;
  LSMEANS treatment*visit / pdiff CL;
  ODS output Diffs=outdata1 LSMeans=outdata2;
RUN;
```

Sensitivity analyses: MMRM with imputation for missing data

We will use the imputation methods described above for the primary efficacy endpoints to assess the sensitivity of missing data on inference regarding the secondary endpoints. Since these are continuous measures, we will use the delta methods described above.

Summary for Categorical Endpoints:

The count and percentage will be provided for the following endpoints:

- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage of prurigenous lesions with excoriations/crusts (item 5a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;

- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage of healed lesions (item 5b) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage number of lesions (item 2) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- Change in IGA-PN as assessed by the percentage of subjects having a 1-category improvement in activity from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- Change in IGA-PN as assessed by the percentage of subjects having a 1-category improvement in stage from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- The proportion of subjects with PBI-P ≥ 1 at Week 14 for the NAL ER treatment arm versus the placebo arm;
- The proportion of subjects with PBI-P ≥ 2 at Week 14 for the NAL ER treatment arm versus the placebo arm;
- The proportion of subjects with PBI-P ≥ 3 at Week 14 for the NAL ER treatment arm versus the placebo arm.

In addition, shift tables from baseline for improved ≥ 1, no change, worsened ≥ 1, and discontinued at Week 14 will be presented for PAS (item 5a), PAS (item 5b), PAS (item 2), IGA-PN for Activity, and IGA-PN for Stage.

Logistic Regression for Categorical Endpoints

The logistic regression model used for the primary endpoint will also be used for several secondary efficacy endpoints. Subjects with missing Week 14 measurement are treated as non-responders. Response rate, standard error, OR between the treatment groups, 95% confidence interval and p-values will be provided. The endpoints are:

- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage of prurigenous lesions with excoriations/crusts (item 5a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- Change in IGA-PN as assessed by the percentage of subjects having a 1-category improvement in activity from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm (95% CI only, with no p-value).

Sensitivity analysis for PAS (item 5a)

The sensitivity analysis to assess the impact of concomitant medications use for the primary endpoint will also be used for PAS (item 5a).

For open-label period visits, descriptive statistics will be displayed by treatment and overall for all secondary variables.

8.7 Safety analyses

Safety summaries will be conducted on the safety population (treated subjects) and will be performed for all safety variables specified below:

- Adverse events;
- Laboratory tests;
- Vital signs;
- Central Electrocardiograms (ECG);
- Brief neurological assessments;
- Physical Examinations;
- SOWS.

All safety data will be summarized by treatment group.

The safety summaries of changes from baseline to a specific time point in safety variables (e.g., laboratory parameters, vital signs, and ECG) will only include subjects from the safety population who have data available for both the baseline and the time point under consideration unless otherwise specified. No statistical tests will be performed.

8.7.1 Adverse events

All adverse events (AEs) will be classified by SOC and PT according to the MedDRA Version 21.0 or higher.

In summaries by SOC and PT, adverse events will be sorted by decreasing frequency within each SOC and PT according to the total group. In summaries by PT, AEs will be sorted by decreasing frequency according to the total group.

Treatment-emergent adverse events (TEAEs) are defined as AEs that started or worsened in severity on or after the first double-blind dose of study medication.

Double-blind period treatment-emergent adverse events (TEAEs) are defined as AEs that started or worsened in severity on or after the first double-blind dose of study medication and prior to the first extension titration dose of study medication in the open-label period.

Open-label period treatment-emergent adverse events (TEAEs) are defined as AEs that started or worsened in severity on or after the first extension titration dose of the study medication.

All TEAEs will be summarized by period (double-blind period and open-label period) and treatment group, presenting the number and percentage of subjects having at least one TEAE and the number of TEAEs occurring in each SOC and PT. A subject with multiple occurrences of an AE will be counted only once in the corresponding AE category.

The following AE summaries will be performed by period:

- An overall summary of the number and percentage of subjects with the specified AEs, the number of episodes in each TEAE category will be presented;
- Exposure-adjusted AE incidence rate (per 100 subject-years) for TEAEs, regardless of study drug relationship by SOC and PT;
- Number and percentage of subjects with TEAEs, regardless of study drug relationship by SOC and PT;
- Number and percentage of subjects with TEAEs suspected to be drug-related by SOC and PT;
- Number and percentage of subjects with CTC grade 3 or higher TEAEs, regardless of study drug relationship by SOC and PT;
- Number and percentage of subjects with CTC grade 3 or higher TEAEs suspected to be drug-related by SOC and PT;

TEAEs starting after the first dose of study medication with a missing severity will be classified as severe. If a subject reports a TEAE more than once within that System Organ Class (SOC)/ Preferred Term (PT), the AE with the highest severity will be used in the corresponding severity summaries.

Relationship, as indicated by the Investigator, is classified as “Not related”, “Unlikely related”, “Possibly related”, “Probably related” and “Definitely related” (increasing strength of relationship). TEAEs with a

missing relationship to study medication will be regarded as "Probably related" to study medication. If a subject reports the same AE more than once within that SOC/PT, the AE with the strongest relationship to study medication will be used in the corresponding relationship summaries. Relationship will be summarized as "Related" and "Unrelated". Adverse events with investigator indicated drug-event relationship of "Not related" and "Unlikely related" are classified as "Unrelated" with the study medication, while adverse events with drug-event relationship documented as "Possibly related", "Probably Related" and "Definitely related" are considered as "Related" with the study Medication.

Deaths, Serious and Other Significant Adverse Events

TEAEs leading to Death are those events with an "Outcome" recorded as "Fatal" on the Adverse Events page of the eCRF. The number of subjects with TEAEs leading to death will be summarized by SOC and PT. A listing of all TEAEs leading to death will be provided.

Serious adverse events (SAEs) are those events with a response of "Yes" for the item "Was AE Serious" on the AE eCRF form. Summaries of incidence rates (frequencies and percentages) of SAEs by SOC and PT will be prepared. A by-subject data listing of all SAEs will be provided.

Actions taken with study drug are dose discontinuation and interruption from the eCRF page of Adverse Events. TEAEs leading to drug adjustment will be summarized by SOC and PT for study treatment discontinuation and interruptions.

The following summaries of TEAEs will be provided:

- TEAEs leading to deaths, by SOC and PT;
- Serious TEAEs, regardless of study drug relationship, by SOC and PT;
- Serious TEAEs suspected to be drug-related, by SOC and PT;
- TEAEs leading to drug interruption by SOC and PT;
- TEAEs leading to study drug discontinuation by SOC and PT.

Listings to be provided are:

- AEs leading to Deaths (CRF Adverse Event page: AE outcome is Fatal.);
- Serious AEs;
- AEs leading to dose interruption and discontinuation, respectively;
- All deaths with date of event, study day, AEs reported around the time of death etc.

8.7.2 Clinical laboratory evaluations

All laboratory data will be summarized in the International System of Units (SI) and conventional units, respectively. Quantitative laboratory measurements reported as "< X", i.e. below the lower limit of quantification (BLQ), or "> X", i.e. above the upper limit of quantification (ULQ), will be converted to X for the purpose of quantitative summaries, but will be presented as recorded, i.e. as "< X" or "> X" in the listings. All laboratory assessments will be listed.

Descriptive statistics of absolute value and change from baseline (mean, SD, median, minimum, and maximum) for continuous assessments will be presented by windowed visits.

Laboratory Reference Ranges High/Low Criteria and Investigator's Assessment

Quantitative laboratory measurements will be compared with the relevant laboratory reference ranges in SI units and categorized as:

- Low: Below the lower limit of the laboratory reference range;
- Normal: Within the laboratory reference range (upper and lower limit included);
- High: Above the upper limit of the laboratory reference range.

Shift tables (Low, Normal, High) from baseline to worst post-baseline in lab parameters for each treatment group in the double-blind treatment period will be displayed.

Values outside the normal range will be flagged in the listings of individual subject data. Unscheduled visits will not be included in the analysis but will be listed.

Hepatobiliary Laboratory Abnormalities

Treatment emergent hepatobiliary laboratory abnormalities defined as values outside pre-specified thresholds will be summarized and listed by treatment and windowed visit. Treatment emergent is defined as any disorder/event recorded on the eCRF which started on or after the start date of treatment.

A summary of any treatment emergent hepatobiliary laboratory abnormalities according to different pre-specified thresholds will be made by treatment group:

- ALT or/and AST >3 x ULN and ≤5 x ULN;
- ALT or/and AST >5 x ULN;
- Total bilirubin >2 x ULN;
- (ALT and/or AST >3 x ULN) and Bilirubin > 2 x ULN;

- Hy's law laboratory criteria: (ALT and/or AST >3 x ULN) and Bilirubin > 2 x ULN and ALP<2 x ULN.

ULN = Upper limit of normal range.

8.7.3 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The parameters collected are the following: systolic and diastolic blood pressure (mmHg), body temperature (°C), heart rate (beats/min) and respiratory rate (breaths/min). The observed values and change from baseline will be summarized by descriptive statistics at each windowed visit. By-subject listings will also be provided.

8.7.4 Brief Neurological Assessment

All neurological assessment data and abnormalities will be listed (including pre-treatment and post-treatment results). The count and percentage will be provided by treatment group and visit.

8.7.5 Physical examinations

All physical examination data and abnormalities will be listed (including pre-treatment and post-treatment results). The count and percentage of normal vs. abnormal assessments will be provided by treatment group and visit.

8.7.6 Electrocardiograms

Electrocardiogram data (e.g., heart rate, PR, QTcF intervals, etc.) will be presented in listings by subject and collection date/time. A complete ECG assessment will be analyzed and reported in a separate ECG report from ERT.

8.7.7 SOWS

The Subjective Opiate Withdrawal Scale (SOWS) is a self-administered scale for grading opioid withdrawal symptoms. It contains 16 symptoms whose intensity the subject rates on a scale of 0 (“not at all”) to 4 (“extremely”). In this study, subjects will complete SOWS daily for 14 days, starting at the off-treatment visit through the washout and safety follow-up period visit to the last visit on the study. Any subjects who are discontinued from investigational product will complete the SOWS daily for the 14 days following the last dose of investigational product (unless consent is withdrawn).

SOWS total scores and the individual item scores will be summarized descriptively by day (there are 14 days of SOWS) and treatment.

Categories of SOWS total score (mild: ≤10; moderate: 11 – 20; and severe: 21 – 30; and > 30) will be provided by count and percentage by day and treatment.

By-subject listings will also be provided.

8.7.8 Pharmacokinetic analysis

Investigational product plasma concentration data (nalbuphine and metabolites) will be listed by collection time, as applicable. Details will be summarized in a bioanalytical report generated by Covance Labs and appended to the CSR. Additional PK/PD analyses may be performed and presented in a separate document and may be appended to the CSR.

8.8 Primary analysis

The primary study analyses will be performed when the last subject has completed the double-blind treatment period. This will be documented in a full clinical study report. Final analyses and an addendum clinical study report will be performed when the last subject has completed the Week 56 phone call.

8.8.1 Interim analysis (Adaptive Sample Size Re-estimation)

The sample size re-estimation will be performed by one unblinded study statistician, who will not have a decision-making role in the trial; no one else, including Sponsor or clinical team members, will see the data presented in the Data Monitoring Committee (DMC). The communication plan from the DMC is to report one of the following statements based on the conditional power results: 1) "carry on"; 2) "increase sample size to XXX"; or 3) "stop the study due to unfavorable safety and/or efficacy concern." Details about the DMC and the data monitoring procedures are specified in the DMC charter. When 50% of the subjects have provided the Week 14 primary endpoint data or have terminated the study and the data are considered "lockable" by data management, the data file will be provided with access limited to the specified unblinded statistician in the DMC; the proportion of responders will be estimated; and the conditional power (CP) will be calculated. Given that the interim analysis is to be performed at 50% of the initial sample size (120 subjects), the targeted power is 90%, and the maximum allowable sample size is set at 1.5 times the initial sample size (360 subjects), the conditional power cut-off value (CPmin) is 41%[8]. If the CP is between 41% and 90%, the number of subjects per treatment arm will be increased, up

to the maximum allowable sample size for this study (120 × 1.5 = 180 each group), to recover the targeted power of 90%.

Since sample size re-estimation occurs only when the interim conditional power falls in the pre-specified "promising" range. The study will not stop for efficacy regardless of the conditional power; the overall alpha will be protected; and the final analysis will be carried out using conventional tests, without the need for weighing the stage 1 and 2 results or adjusting the alpha value.

9 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

The following changes were agreed upon

1. Sites were changed from random effects to region (US or Europe) which is being treated as a fixed effect.
2. LOCF was excluded from imputation methods.
3. The ranked outcome for adjustment for multiplicity was changed. The ItchyQoL measure & the PROMIS scale were excluded from the ranking and replaced with PAS (5a)
4. The analyses for prohibited medication were changed to set all efficacy outcomes to missing from the date of ingestion of the prohibited medication. Previously, missing outcomes were set = non-responders.
5. A new set of simple multiple imputation models were added to assess the impact of missing efficacy measures due to the use of prohibited medications.

10 REFERENCES

1. ICH Topic E9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96 – adopted March 1998).
2. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Statistics in Medicine 2011; 30(28):3267-84.
3. Miriam Kimel, Phd; Claudia Xeidler, MD; Paul Kwon, MD; et al (2020), Validation of psychometric properties of the itch numeric rating scale for pruritus associated with prurigo nodularis.
4. Blome C, Augustin M, Siepmann D, et al. Measuring patient-relevant benefits in pruritus treatment: development and validation of a specific outcomes tool. Br J Dermatol. 2009 Nov; 161(5):1143-8.
5. Ratitch & O'Kelly (2011), Implementation of Pattern Mixture Models Using SAS/STA Procedures.

11 VERSION HISTORY

Version	Date	Notes
Draft 1.0	27Aug2018	First draft base on Protocol V1.0
Draft 2.0	10Oct2018	Stable version after addressing Trevi review comments and incorporating changes in Protocol Amendment 1 (Protocol V2.0)
Final 1.0	15Mar2019	Final version after incorporating changes in Protocol Amendment 2 (Protocol V3.1)
Draft V1.4	19Feb 2022	Version above modified after several meetings with TREVI in 2021 & 2022 as clarified in comments above.
Draft V1.5	9Mar2022	Version above modified after several meetings with TREVI in 2021 & 2022 as clarified in comments above. V1.5 had modified section of TPA. The analysis has been refined here & an updated Guidance document added to the Appendix.
Draft V1.6	23Mar 2022	Version above modified after several meetings with TREVI in 2021 & 2022 as clarified in comments above. V1.5 had modified section of TPA. The analysis has been refined here & an updated Guidance document added to the Appendix. Proc GENMOD has been modified as well.
Final 2.0	25th March 2022	Modified rank order of the outcomes to address issues of multiplicity & efficacy outcomes for subjects who took Prohibited Medications
Final 2.1	01 June 2022	Changed Compliance cut off to 50%. Changed Efficacy Analysis from Proc Glimmix to Proc Genmod because of convergence issues. Reduced number of imputations from 1000 to 50 to increase efficiency.

12 APPENDIX

12.3 Programming Guidance Document for Multiple Imputation, Control –based Imputation & Tipping Point Analysis

Introduction.
This appendix describes three approaches to multiple imputation.

1. Simple Multiple Imputation;
2. Control- based Imputation;
3. Tipping Point Analysis.

All of these are being implemented in data sets that have repeated measures data collected from subjects at multiple time points.
The primary efficacy endpoint is defined as the difference between the percentage of "Responders" at Week 14 for the NAL ER treatment arm versus the placebo arm. A "Responder" is defined as a subject with a ≥ 4-point decrease in the 7-day average WI-NRS from baseline to Week 14**."**
Secondary endpoints are:

- The mean change in ItchyQoL from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- Change in PAS as assessed by the percentage of subjects having a 1-category improvement in the percentage of prurigenous lesions with excoriations/crusts (item 5a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm;
- The mean change in sleep disturbance (PROMIS Sleep Disturbance Short Form 8a) from baseline to Week 14 for the NAL ER treatment arm versus the placebo arm.

Multiple imputation is a missing data approach that assumes that the missing data is missing at random (MAR), i.e., that the probability of missing is not a function of the values that are missing but only depend on the observed responses.

The advantage of MI over single imputation is that it allows one to adjust the variance to account for the imputation of the missing data.

All of the data sets on which imputation is applied are longitudinal data sets, specifically the following ADaM datasets: ADEFF and ADQS. Please note: ADEFF will use 4 visits (AVISITN 2-5/ Week 2, 6, 10, 14)
ADQS will use 3 visits (AVISITN 3-5/Week 6, 10, 14).

The following is an explanation of the process.
This is a guidance document. It does not provide the exact code to be implemented but describes the processes & principles that need to be followed in order to implement these analyses.

There are five steps in simple MI process.

1. Check missing data patterns;
2. Impute non-monotonic / intermittent data – leaving only monotone missing data;
3. Use monotone regression to impute monotone missing data & create multiply imputed data sets;
4. Run the efficacy analyses (PROC GENMOD, PROC MIXED) on the multiply imputed data sets;
5. Use PROC MIANALYZE to combine the efficacy output to get a single estimate from all the imputed data sets.

Step 1. Checking missing data patterns

For this study, the missing data patterns will be checked by treatment and visits 3 to 5.
There are 3 steps to this process:

Create dataset for efficacy measure including visits from baseline to week 14

```
data mi_chk1; set datax;
where param="" and paramcd="" and dtype="" and 3 <= visit <=5;
run;
```

Transpose dataset to display each visit as one column:

```
proc sort data = mi_chk1; by treatment   id;
proc transpose data=mi_chk1 out=mi_chk2 prefix='';
by treatment  id;
id avisitn;
var aval;
run;
```

Use PROC MI with the NIMPUTE=0 option to create the "Missing Data Patterns" table for the specified variables:

```
ods output MissPattern=missp;
proc mi data=mi_chk2 NIMPUTE=0;
  by treatment;
  var visit3 – visit5;
run;
/*Add baseline to estimate change from baseline*/

proc sort data = mi_chk1; by id; data base; set mi_chk1 (Keep = id base country); by id; if first.id;
run;
proc sort data = mi_chk2; by id;
proc sort data = base; by id;
data mi_chk2; merge mi_chk2 base; by id;
run;
```

Step 2: **Impute non-monotonic / intermittent data – leaving only monotone missing data.**
Fill in the intermittent missing data using the following code.

```
proc sort data = mi_chk2; by treatment;
proc mi data=mi_chk2 seed=12345 NIMPUTE=50 out=MI;
by trt01pn;
mcmc impute=monotone;
var Visit3 – Visit5;
run;
```

Step 3. Use monotone regression to impute monotone missing data & create multiply imputed data sets.

Here the non-missing values for subjects in the control group are used to impute values for the subjects with missing values in the control group and similarly for the subjects in the active treatment group.

Visit 3 values are imputed using baseline value and treatment;
visit 4 values are imputed using baseline, treatment and visit 3 & so on.

```
proc sort data = mi; by _Imputation_;
proc mi data=mi out=mi_reg seed=54321 nimpute=1;
by _Imputation_;
var trt01pn base v3 v4 v5;
class trt01pn;
monotone regression;
run;
```

Convert the data set from a horizontal dataset, i.e. one row per subject to a vertical data set, i.e., multiple rows per subject.

```
Data mi_reg1; set mi_reg;
Visit = 1; WINRSA=base; output;
Visit = 2; WINRSA=v3; output;
visit = 3; WINRSA=v4; output;
visit = 4; WINRSA=v5; output;
drop v3 v4 v5;
run;
```

Step 4. Run the efficacy analyses (PROC GENMOD, PROC MIXED) on the multiply imputed data sets.

Create variables needed for the 50 data sets, e.g. RESP, Pooled Site, treatment coded as (0-Placebo,1-Treatment) etc.

1. **Fit the models on the 50 imputed data sets by _Imputation_.**

```
proc sort data = mi_reg1;
 by _imputation_ visit trt;
```

ods select none; /*suppress printing the output*/

```
proc genmod data=winrsa descending; where avisitn ge 3;
class usubjid avisitn trt posite ;
model resp = trt avisitn avisitn*trt posite base
/ dist=binomial;
repeated subject=usubjid / withinsubject=avisitn
type=un;
lsmeans trt*avisitn /cl ilink diff oddsratio;
Ods output diffs=diff lsmeans=lsm ;
run;
```

Step 5. Combine estimates of the 50 GENMOD models to create an overall estimate and a relative efficacy score.

Due to the finicky nature of PROC MIANALYZE, the suggestion is to produce output to one effect at a time. This can be programmed in a macro.

Run three separate instances of GENMOD to get the effect for the interaction first, then the treatment & then visits

MIANALYZE cannot handle too many effects and too many variables in the input data sets, so subset to the needed effects. For the interaction, choose the estimate for the same visit but different treatments. That is what the next two statements do.

```
data diff_mi_nw; set diff_mi; where visit = _visit and trt ne _          trt;run;
data diff_mi_nw2; set diff_mi_nw (keep = _imputation_  estimate  stderr effect); run;
```

The data set diff_mi_nw2 is then fed into the Proc MIANALYZE:

```
proc mianalyze parms=diff_mi_nw2; /*where _imputation_ le 4;
modeleffects  treatment*visit ;
run;
```

Since this is GENMOD the results should be presented as odds ratios (ORs). The final estimate & the SEs should be exponentiated to create the ORs.

Also fit the continuous ITCH score using Proc Mixed.

```
ods select none;
proc mixed data = mi_reg1; by _Imputation_ ;
class variables;
Model (Continuous endpoint) = Treatment, Baseline Vars, Time, Treatment * Time/solution
covb;
repeated ID/ type = cs subject = ID(treatment)
lsmeans Interaction Term/diff cl;
```

```
ods output Diffs=diff_mi LSMEANS=lsm_mi;
run;
ods select all;

Subset the data set to be input into MIANALYZE.
data diff_mi_nw; set diff_mi (keep = _imputation_
estimate stderr effect probt);
run;

proc mianalyze parms=diff_mi_nw;
modeleffects  effect of interest (e.g., treatment, Visit or interaction but to be done one at a time);
run;
```

Control-Based Imputation.
STEPS 1-4 ARE EXACTLY AS DESCRIBED ABOVE.

The imputation of missing values is done one visit at a time.
The difference between this process and the simple multiple imputation above is primarily in the data management step where a dataset is created with all control subjects and only those trt = NALER subjects who have missing values for visit, one visit at a time.

Below is the suggested approach.
In this case use v3, v4 & v5. After the intermittent missing data is filled in, one would expect all v3 to be filled in & only v4 & v5 to be missing.
This is also a way to check & make sure that the data is behaving as expected.

Step 1. Create two separate data sets, i.e., one with just control subjects and one with just NALER subjects;
Step 2. Create a subset NALER data to include only those who have missing values for visit 4;
Step 3. Set the Control data set and the dataset with V4 missing by _Imputation & usubjid.

Now, the dataset has all of the control subjects and only the NALER subjects with v4 missing in this data set.

Sample Program:

```
data control_mi; set mi; where trt = 0; run;
data trt_mi; set mi; where trt = 1;
data v4_trt_mi; set trt_mi; where v4 = .; run;
proc sort data = control_mi; by _imputation_ usubjid;
proc sort data = v4_trt_mi; by _imputation_ usubjid;
data mi_v4; set v4_trt_mi control_mi; by _imputation_ usubjid;
run;
```

Then use monotone regression to impute the v4 values.

```
proc sort data = mi_v4; by _Imputation_;
proc mi data=mi_v4 out=mi_reg_v4 seed=54321 nimpute=1;
by _Imputation_;
var  base v3 v4
monotone regression (v4);
run;
```

Repeat for v5.

Then merge the data set back to include the full treatment = NALER data set with the non-missing & imputed values together. Set the Trt = 1 & trt = 0 data sets by _Imputation_ & usubjid

2. **Convert the data set from a horizontal dataset, i.e. one row per subject to a vertical data set, i.e., multiple rows per subject**.

```
Data mi_reg1; set mi_reg;
Visit = 1; WINRSA=base; output;
```

```
Visit = 2; WINRSA=v3; output;
visit = 3; WINRSA=v4; output;
visit = 4; WINRSA=v5; output;
drop v3 v4 v5;
run;
```

Create required variables needed in these 50 data sets, e.g. RESP, Pooled Site, treatment coded as (0,1) etc.

Run Proc GENMOD or Proc Mixed
Run Proc MIANALYZE

Tipping Point Analysis.
In this analyses, the concept is to assess the level of bias introduced by the missing at random assumption in the first simple MI performed above. In this approach, subtract a pre-determined value from the imputed values for the actively treated group. this approach uses the standard deviation of the scores for the completers in the treated group. It could be a %change in the mean or a variety of other values depending on the expected effect. The suggested approach is below.

```
data mi_chk1; set adeff;
where param="WINRS Average" and paramcd="WINRSA" and dtype="" and 2 <=avisitn<=5;
run;
data mi_chk1; set mi_chk1;
if trt01pn = 1 then trt = 1;
else if trt01pn = 2 then trt = 0;
run;

proc sort data = mi_chk1; by trt  usubjid;
proc transpose data=mi_chk1 out=mi_chk2 prefix=v;
by trt  usubjid;
id avisitn;
var aval;
run;
   by trt ;
   var v2-v5;
run;
```

Next estimate the standard deviation of the WINRS score in the control group and active treatment group by visit among those subjects who completed all visits.

Then merge a data set with the delta value to the data set where values will be imputed.

Use proc univariate to obtain the mean values, stddev, range for the completers in the treated group in order to identify a delta value (SD in this case. Other approaches could use the median, or percentiles).

Suggested code is below.

```
proc sort data = mi_chk1; by usubjid;
data base; set mi_chk1 ( Keep = usubjid base country complfl);   by usubjid; if first.usubjid;run;
proc sort data = mi_chk2; by usubjid;
proc sort data = base; by usubjid;
data mi_chk2; merge mi_chk2 base; by usubjid;
run;
proc univariate data = mi_chk2; where complfl = 'Y'; by  trt;
var v2 v3 v5 v5;
output out=WINRSCOMP_uni mean=Meanv2 Meanv3 meanv4 meanv5
                         std=stdv2 stdv3 stdv4 stdv5
               range=range_v2 range_v3 range_v4 range_v5;
run;
```

```
proc sort data = WINRSCOMP_UNI; by trt;
data mi_chk3; merge mi_chk2 winrscomp_uni; by trt;
run;
```

Fill in the intermittent missing data using the following code.

```
proc sort data = mi_chk3; by treatment;
proc mi data=mi_chk3 seed=12345 NIMPUTE=50 out=MI;
  by trt01pn;
  mcmc impute=monotone;
  var Visit2 – Visit5;
run;
```

Create an id variable to identify which visits were imputed in the active treatment group.

```
data control_mi; set mi; where trt = 0; run;
data NALER_mi; set mi; where trt = 1;
Data NALER_MI; Set NALER_MI;
if V2 = . then Imp_v2 = 1; else IMP_V2 = 0;
if V3 = . then Imp_v3 = 1; else Imp_v3 = 0;
if v4 = . then Imp_v4 = 1; else Imp_v4 = 0;
if v5 = . then Imp_v5 = 1; else Imp_v5 = 0;
run;

proc sort data = control_mi; by trt _imputation_ usubjid;
proc sort data = NALER_mi; by trt _imputation_ usubjid;
data mi; set NALER_mi control_mi; by trt  _imputation_  usubjid;
run;
```

Imputing values in the monotone missing data using regression.
Visit 3 values are imputed using baseline value and treatment;
Visit 4 values are imputed using baseline, treatment and visit 3 & so on.

```
proc sort data = mi; by _Imputation_;
proc mi data=mi out=mi_reg seed=54321 nimpute=1;
by _Imputation_;
var trt01pn base v3 v4 v5;
class trt01pn;
monotone regression;
run;
```

1. **Tipping Point Analysis – Starts here**:

Take each group of dropouts – those who dropped out at v3 & subtracting (or adding) the SD of the completers group for that visit from the imputed value.

Then merge it to the original data set.

Then go to the next set of dropouts.

```
data mi_reg2_Impv3; set mi_reg2;
if Imp_v3 = 1
then do;

        /*Create a macro where this value is replaced by the following: (-.25SD, -0.5SD, -
        0.75stdev, -SD, 0.25SD, 0.5SD, 0.75SD, SD), thus reducing the imputed values by a
        quarter, a half, three-quarter and a whole SD. Then, on the other side add 0.25SD, 0.5SD,
        0.75SD*/

v3 = v3-stdv3;
v4 = v4-stdv4;
v5_= v5-stdv5; (each of these is done only for visits that were imputed in the previous step)
end;
run;
proc sort data = mi_reg2; by _Imputation_ usubjid;
proc sort data = mi_reg2_impv3; by _Imputation_ usubjid;
data mi_reg4; merge mi_reg2 mi_reg2_impv3; by _Imputation_ usubjid;
run;

data mi_reg4_Impv4; set mi_reg4;
if Imp_v4 = 1
then do;
v4 = v4-stdv4;
v5 = v5-stdv5;
end;
run;
proc sort data = mi_reg4; by _Imputation_ usubjid;
proc sort data = mi_reg4_impv4; by _Imputation_ usubjid;
data mi_reg5; merge mi_reg4 mi_reg4_impv4; by _Imputation_ usubjid;
run;

data mi_reg5_Impv5; set mi_reg5;
if Imp_v5 = 1
then do;
v5 = v5-stdv5;
end;
run;
proc sort data = mi_reg5; by _Imputation_ usubjid;
```

```
proc sort data = mi_reg5_impv5; by _Imputation_ usubjid;
data mi_reg6; merge mi_reg5 mi_reg5_impv5; by _Imputation_ usubjid;
run;
```

The result of the above steps is a modified horizontal data set with the modified imputed values. Now, proceed as before by converting this to a vertical data set & fitting GENMOD or MIXED.

2. **Convert the data set from a horizontal dataset, i.e. one row per subject to a vertical data set, i.e., multiple rows per subject**.

```
Data mi_reg1; set mi_reg;
Visit = 1; WINRSA=base; output;
Visit = 2; WINRSA=v3; output;
visit = 3; WINRSA=v4; output;
visit = 4; WINRSA=v5; output;
drop v3 v4 v5;
run;
```

3. **Create required variables in these 50 data sets, e.g. RESP, Pooled Site, treatment coded as (0,1) etc.**

4. **Fit the models on the 50 imputed data sets by _Imputation_.**

```
PROC GENMOD data=mi method= QUAD; by _imputation_;
   Class trt visit posite usubjid;
   Model Resp (descending) = trt visit  trt*visit posite base/ dist=binary
   link=logit /*ddfm = KR*/ s  ;
   random  intercept / subject = usubjid(trt) type = un g s;
   lsmeans  trt/cl ilink diff oddsratio;
   Ods output diffs=diff_mi lsmeans=lsm_mi parameterestimates=parms1 ;
run;
```

Combine estimates of the 50 GENMOD models to create an overall estimate and a relative efficacy score. Use PROC MIANALYZE as above. Also follow guidance re ORs provided earlier.

12.4 Sleep Disturbance Conversion Table.

Sleep Disturbance 8a		
Raw Score	T_Score	SE
8	28.9	4.8
9	33.1	3.7
10	35.9	3.3
11	38	3
12	39.8	2.9
13	41.4	2.8
14	42.9	2.7
15	44.2	2.7
16	45.5	2.6
17	46.7	2.6
18	47.9	2.6
19	49	2.6
20	51.2	2.5
21	52.2	2.5
22	52.2	2.5
23	53.3	2.5
24	54.3	2.5
25	55.3	2.5
26	56.3	2.5
27	57.3	2.5
28	58.3	2.5
29	59.4	2.5
30	60.4	2.5
31	61.5	2.5
32	62.6	2.5
33	63.7	2.6
34	64.8	2.6
35	66.1	2.7
36	67.5	2.8
37	69	3
38	70.8	3.2
39	73	3.5
40	76.6	4.4