Official Title

A Study to Assess the Safety and Efficacy of Nemolizumab (CD14152) in Subjects With Prurigo Nodularis (PN)

NCT Number: NCT03181503

Date: 21 May 2019

16.1.9 Documentation of statistical methods and interim analysis plans

Document included:

Statistical Analysis Plan Version 2.0, Dated 21 May 2019

PAREXEL International

Note to File

CRF Version and Formatting Issues Found in 16-May-2019 Audit Report #14121 for SAP V2.0

Summary:

CRF version: The SAP section 2 refers to **version 4 of the CRF** dated 08 Feb 2018, while it looks that the final CRF was version 6 (cf. above finding). This was corrected to version 6 dated 11 June 2018.

Following Formatting issues have been corrected in the SAP attachment to this NTF.

Formatting issues:

- Pages numbers in the ToC do not match the actual pages
- Schedule of assessment included in the SAP appendix 1 is cut (follow-up & early termination visits are not displayed - format issue) and not in line with protocol latest version 02 - schedule for CPK blood sampling: to be taken at V7/week 8 while required by the CSP (only for France).

This document has been signed electronically on the final page by the following:
Author Signatory

[redacted]

Author of the SAP

PAREXEL International

GALDERMA R&D, SNC

Protocol Number: RD.03.SPR.115828

A randomized, placebo-controlled, double-blinded, parallel group, multicenter study to assess the safety and efficacy of nemolizumab (CD14152) in subjects with prurigo nodularis (PN)

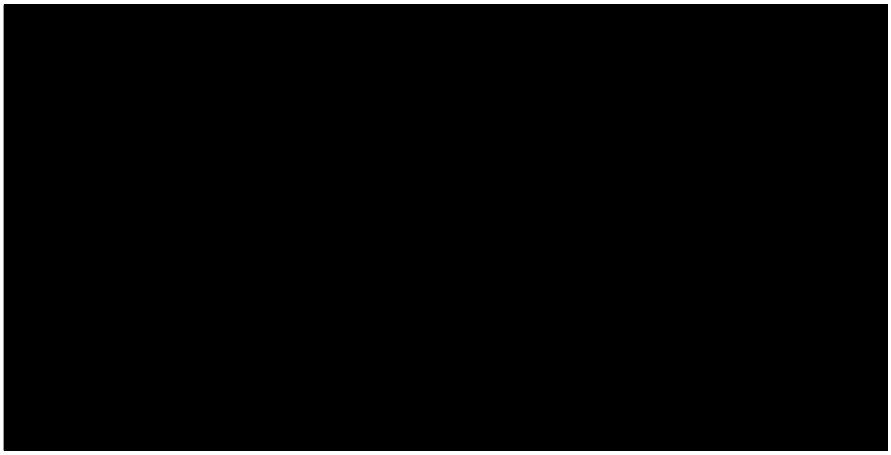
Statistical Analysis Plan

Version 2.0

PAREXEL Project Number: 235088

SPONSOR SIGNATURE PAGE

Approved by:



PAREXEL SIGNATURE PAGE

Signature below confirms that the Statistical Analysis Plan was developed in accordance with SOP-GDO-WW-019 and that it is approved for release.

This document has been approved and signed electronically on the final page by the following:

	Signatory
Author	[redacted]
	Signatory
Reviewer	[redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	*Term*
ADA	Anti-Drug Antibody
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
AUC	Area under the Curve
BLQ	Below the Limit of Quantification
BMI	Body Mass Index
bpm	Beats per Minute
CI	Confidence Interval
Cmax	Maximum Concentration
CMH	Cochran-Mantel-Haenszel
CPK	Creatinine Phosphokinase
CRO	Contract Research Organization
CSR	Clinical Study Report
dL	Deciliter
DLQI	Dermatology Life Quality Index
DPS	Dynamic Pruritus Scale
ECG	Electrocardiogram
eCRF	electronic Case Report Forms
EMA	European Medicines Agency
FDA	Food and Drug Administration
GGT	Gamma-Glutamyl Transpeptidase
HDL	High-Density Lipoprotein

Abbreviation	*Term*
hsCRP	High Sensitivity C-Reactive Protein
ICF	Informed Consent Form
ICH	International Conference on Harmonisation
IGA	Investigator's Global Assessment
IgE	Immunoglobulin E
IL	Interleukin
IRR	Injection-related Reaction
ITT	Intent-to-treat; intention-to-treat
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
KM	Kaplan-Meier
LDH	Lactate Dehydrogenase
LDL	Low-Density Lipoprotein
LOCF	Last Observation Carried Forward
LSmean	Least Squares Mean
MCV	Mean Cell Volume
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
mL	Milliliter
NCA	Non-Compartmental Analysis
NRS	Numeric Rating Scale
OC	Observed Case
PAS	Prurigo Activity Score
PCSV	Potentially Clinically Significant Value
PD	Pharmacodynamics
PDAP	PD Analysis Population
PEF	Peak Expiratory Flow
PK	Pharmacokinetics
PKAP	PK Analysis Population
PN	Prurigo nodularis

Abbreviation	*Term*
PP	Per-Protocol
PT	Preferred Term
Q1	Lower quartile
Q3	Upper quartile
Q4W	Every 4 weeks
QoL	Quality of Life
RA	receptor A
RBC	Red Blood Cell
ROC	Receiver Operating Characteristic
SAE	Serious Adverse Event
SAF	Safety (analysis population)
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class
TARC	thymus and activation-regulated chemokine
TCS	Topical Corticosteroids
TEAE	Treatment-emergent Adverse Event
TESAE	Treatment-emergent Serious Adverse Event
TLF	Tables, Listings, and Figures
ULN	Upper Limit of Normal
VRS	Verbal Rating Scale
WBC	White Blood Cell
WHODD	World Health Organization Drug Dictionary

1 AMENDMENTS FROM PREVIOUS VERSION(S)

The following is a change from the Statistical Analysis Plan (SAP) dated 17Aug2018:

- Added details of how to calculate pharmacokinetic parameters to the section 3.2.4 of the Version 1.0 of the SAP. The new section is 5.9.1 of this Amendment of the SAP (Version 2.0).

2 INTRODUCTION

This SAP describes the planned analysis and reporting for GALDERMA R&D, SNC, protocol RD.03.SPR.115828, entitled "A study to assess the safety and efficacy of nemolizumab (CD14152) in subjects with prurigo nodularis (PN)", All Countries Amendment Version 2.0 dated Apr 09, 2018.

Prurigo nodularis (PN) is characterized by the presence of symmetrically distributed multiple (up to hundreds), highly pruritic, hyperkeratotic, erosive or crusted nodules and papules (Hyde JN et al., 1909). This leads to an impaired quality of life and high burden due not only to the severe itch but also the chronic, skin lesions and lack of treatment options (Warlich B., et al 2015).

The goal of PN treatment is to break the itch-scratch cycle and allow the skin to heal. Treatment of chronic pruritus is still notoriously challenging and frustrating for both dermatologists and patients as in the majority of cases, the response is either limited or associated with severe adverse events with the current therapy options. There is no standardized or approved therapy for PN up-to-date and evidence from controlled studies is limited (Fostini AC et al., 2013). The difficulty in treating this disease is reflected in the wide range of treatments proposed in the literature.

Nemolizumab (CD14152) is a humanized anti-human IL-31 receptor A (RA) monoclonal antibody that inhibits the binding of IL-31 to IL-31RA and subsequent signal transduction. IL-31 has been suggested to be a key player in the development of pruritus. Recent clinical studies of nemolizumab in patients with atopic dermatitis showed a rapid onset of pruritus reduction within one week followed by disease improvement (Ruzicka T et al 2017).

The main objective of this study is to evaluate the efficacy, safety and pharmacokinetics (PK) of multiple subcutaneous doses of nemolizumab in the treatment of pruritus and lesions associated to PN.

The planned analyses identified in this SAP may be included in the clinical study report (CSR), regulatory submissions, or future manuscripts. Also, post hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post hoc, or unplanned, exploratory analyses performed, if included, will be clearly identified as such in the final CSR.

This SAP is based upon the following study documents:

- Study protocol, All Countries Amendment Version 2.0 (09 Apr 2018)
- Electronic Case Report Form (eCRF), Version 6.0 (11-June-2018).

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials. All work planned and reported for this SAP will follow internationally accepted guidelines, e.g., Ethical Guidelines for Statistical Practice published by the American Statistical Association and Code of Conduct published by the Royal Statistical Society, for statistical practice.

The reader of this SAP is encouraged to also read the clinical protocol, and other relevant documents for details on the planned conduct of this study. Other than the schedule of assessments which is provided in Appendix 1, operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analyses.

3 STUDY OBJECTIVES

3.1 Primary Objective

The primary objective is to assess the efficacy of nemolizumab compared to placebo in the treatment of pruritus in patients suffering from PN.

3.2 Secondary Objectives

The secondary objectives are:

Efficacy and safety:

- Evaluation of the safety of nemolizumab compared to its placebo in patients with PN.

- Evaluation of the efficacy of nemolizumab compared to its placebo in the treatment of *prurigo* lesions in patients with PN.
- Evaluation of nemolizumab effect compared to its placebo on quality of life in patients with PN

Pharmacokinetics (PK)

- Characterization of nemolizumab PK profile and exposure response relationship in patients with PN

Pharmacodynamics (PD)

- Evaluation of the effect of nemolizumab on biomarkers in patients with PN

Biophysical

- Evaluation of the efficacy of nemolizumab on scratching events and sleep improvement using actigraphy
- Evaluation of the efficacy of nemolizumab on lesions improvement using whole body images device only on equipped sites. This will be evaluated based on the availability of the data.

4 INVESTIGATIONAL PLAN

4.1 Overall Study Design and Plan

This is a randomized, placebo-controlled, double-blinded, parallel group, multicenter study to evaluate the safety and efficacy of nemolizumab in patients suffering from PN. Approximately 70 adult patients suffering from PN for at least 6 months with severe pruritus defined by the mean of the worst daily intensity of the NRS score ≥ 7 over the previous week at baseline will be randomized in the study. Subjects meeting inclusion/exclusion criteria will be randomized in a 1:1 ratio to nemolizumab or placebo. Each subject will receive three subcutaneous injections of 0.5mg/kg of nemolizumab or matching placebo. Injections will be administered every 4 weeks (at baseline, week 4 & week 8).

Subjects' participation in the study will be up to 22 weeks, including an up to 4-week screening period, a 12-week treatment period (last study drug injection at week 8) and a 6-week follow-up period (10 weeks after the last study drug injection corresponding to 5 half-lives of nemolizumab).

Assessments of efficacy, safety, PK, PD, and biophysical will be conducted throughout this clinical trial. Refer to Appendix 1 for the complete schedule of assessments.

4.2 Efficacy and Safety Variables

4.2.1 Efficacy Variables

4.2.1.1 Pruritus Numeric Rating Scale (NRS)

Pruritus NRS is a scale to be used by the subjects to report the intensity of their pruritus (itch) during the last 24 hours all along the study from first screening visit to the follow-up visit. Subjects will receive instructions on how to record their pruritus NRS scores on an electronic device and will complete the assessment once daily at home in the evening throughout the clinical trial (including the follow-up period).

Subjects will be asked the following questions in their local language:

- For average itch intensity: "On a scale of 0 to 10, with 0 being 'no itch' and 10 being 'worst itch imaginable', how would you rate your itch overall during the previous 24 hours?"
- For maximum (Peak) itch intensity: "On a scale of 0 to 10, with 0 being 'no itch' and 10 being 'worst itch imaginable', how would you rate your itch at the worst moment during the previous 24 hours?"

Weekly prorated average score will be calculated and used in the analysis. If a subject has less than 5 diary entries in a week, the weekly average scores will be set to missing. Refer to Section 5.2 for classification of diary data into analysis week/visit.

4.2.1.2 Verbal Rating Scale (VRS)

The VRS consists of a list of adjectives describing different levels of symptom intensity, to be used by the subjects to report the intensity of their pruritus (itch) over the last 24 hours. Subjects will receive instructions on how to record their pruritus VRS scores on an electronic device and will complete the assessment once daily in the evening from Screening visit 2 and throughout the clinical trial (including the follow-up period). In the eDiary, patients will mark on a 5-point VRS, the response that best describes their pruritus intensity in the last 24 hours.

Subjects will be asked the following questions in their local language:

- For average itch intensity: "On a scale of 0 to 4, with 0 being 'no itch' and 4 being 'very severe itch', how would you rate your itch overall during the previous 24 hours?"
- For maximum (Peak) itch intensity: "On a scale of 0 to 4, with 0 being 'no itch' and 4 being 'very severe itch', how would you rate your worst itch during the previous 24 hours?"

Weekly prorated average score will be calculated and used in the analysis. If a subject has less than 5 diary entries in a week, the weekly average scores will be set to missing. Refer to Section 5.2 for classification of diary data into analysis week/visit.

4.2.1.3 Dynamic Pruritus Score (DPS)

The 9-point DPS is a dynamic scale to be used by subjects to evaluate the change of their pruritus compared with an earlier time point (i.e. before injection on Day 1). The scale ranges from 0 (strongly worsened pruritus) to 8 ([almost] no pruritus anymore), including intermediate marks for slightly improved/worsened, moderately improved/worsened, and rather improved/worsened. Subjects will receive instructions on how to record their DPS score on an electronic device displaying the scale in their local language, and will complete the assessment 24, 48h and 72h after the 1st injection at baseline and at W4 before the second injection. This assessment will be performed by the subject based on his itch perception over the 24h hours preceding the baseline injection.

4.2.1.4 Investigator Global Assessment of Prurigo (IGA)

IGA is a 5-point scale ranging from 0 (clear) to 4 (severe) used by the investigator or trained designee to evaluate the severity of the disease. IGA corresponds to the overall assessment of the severity of prurigo including presence of crust and nodules or skin bleeding that will be evaluated at Baseline, Week 4, Week 8, Week 12, Week 18 & in case of Early termination visit / Unscheduled visit, if applicable.

4.2.1.5 Assessment of Prurigo Activity Score (PAS)

The evaluation of the disease will be performed by the Investigator using the Prurigo Assessment Scale (PAS) at baseline and Week 12 using the entire scale and at Weeks 4, 8, 18 and in case of Early termination visit / Unscheduled visit, if applicable using item 5 (number of lesions only) and 6 (excoriation/crusts and healed lesions).

4.2.1.6 Sleep Disturbance Numeric Rating Scale (NRS)

The sleep disturbance NRS is a scale to be used by the subjects to report the degree of their sleep loss related to PN. Subjects will receive instructions on how to record their sleep disturbance NRS scores on an electronic device and will complete the assessment once daily in the morning from Screening 2 (Day -7) to Week 4 visit (Day 29).

Subjects will be asked the following question in their local language:

> On a scale of 0 to 10, with 0 being 'no sleep loss related to signs/symptoms of PN' and 10 being 'I cannot sleep at all due to the signs/symptoms of PN', how would you rate your sleep last night?"

Weekly prorated average score will be calculated and used in the analysis. If a subject has less than 5 diary entries in a week, the weekly average scores will be set to missing. Refer to Section 5.2 for classification of diary data into analysis week/visit.

4.2.1.7 Dermatology Life Quality Index (DLQI)

DLQI is a validated 10-item questionnaire, covering domains including symptoms/feelings, daily activities, leisure, work/school, personal relationships and treatment. Subject will rate each question ranging from 0 (not at all) to 3 (very much), and the total score ranges from 0 to 30, with a higher score indicating a poorer quality of life. Please note that "Not relevant" is scored as 0, and Question 7 'Yes – prevented working or studying' is scored as 3. DLQI will be given only to the subset of subjects who fluently speak a language in which the questionnaire is presented (based on availability of validated translations in participating countries). This assessment will be performed by the subject on the provided electronic device at baseline, Week 4, and Week 12, and in case of Early termination visit / Unscheduled visit, if applicable.

For general inflammatory skin conditions, a change in DLQI score of at least 4 points is considered clinically important. Refer to Appendix 3 for details on scoring, data handling rules, and analysis on DLQI.

4.2.2 Efficacy endpoints

4.2.2.1 Primary endpoint

Percent change from baseline in NRS to week 4 (weekly average of the peak).

4.2.2.2 Secondary efficacy endpoints

- Absolute and percent change from baseline in weekly average of the peak and average pruritus NRS to each visit
- Absolute and percent change from baseline in weekly average of the peak and average VRS to each visit
- DPS at hours 24, 48, and 72 for baseline and Week 4 visits
- Absolute change from baseline in PAS Item 5 (number of lesions) to each visit

- PAS Item 6 (excoriation/crusts and healed lesions stages) at each visit
- IGA at each visit
- Proportion of subjects achieving IGA success (defined as IGA=0[clear] or IGA=1[Almost clear] with two-point improvement from baseline) at Week 12

4.2.2.3 Other efficacy endpoints

- Absolute change and shift from baseline in Dermatology Life Quality Index (DLQI) to each visitAbsolute and percent change from baseline in weekly average Sleep Disturbance Numeric Rating scale (NRS) up to Week 4
- Objective assessment of scratching events during the night and sleep duration using Actigraphy. Ratio 'scratching duration/duration' and 'scratching events/duration' up to week 4 are the study endpoints.

4.2.3 Safety Variables

Safety assessments will be conducted for all subjects at the screening visit (upon the signature of the informed consent form [ICF]) and at every subsequent visit.

Safety assessments include electrocardiogram (ECG), physical examination, vital signs, body weight, respiratory assessments (peak expiratory flow [PEF] measurement only for subjects with a medical history of asthma), laboratory safety tests, and AE recording (including serious AEs [SAE] and AEs of Special Interest [AESI]).

4.2.3.1 Electrocardiograms (ECG)

A 12-lead ECG will be performed at screening 1, baseline and Week 8, or early termination and/or unscheduled visits when applicable.

All abnormal ECG findings considered to be clinically significant by the investigator at the screening visit will be recorded in the Medical History form. Any clinically significant changes from the screening visit will be reported as AEs in the eCRF.

4.2.3.2 Physical examination and vital signs

Physical examination: The physical examination will be performed at screening 1, baseline, weeks 4, 8, 12, 18, or early termination and/or unscheduled visits when applicable.

All clinically significant abnormal findings at the screening visit will be recorded in the Medical History form. Any clinically significant changes from the screening visit will be recorded as an AE.

Vital signs: The vital signs will be evaluated at screening 1, baseline, weeks 1, 2, 4, 8, 12, 16, 18, or early termination and/or unscheduled visits when applicable.

Vital signs will include pulse rate, systolic and diastolic blood pressure (after the subject has been siting for at least 5 minutes). All abnormal values at the screening visit identified as clinically significant by the investigator will be recorded in the Medical History form. Any clinically significant changes from the screening visit will be recorded as an AE.

Height and Weight: Height will be measured at the baseline visit only, and weight will be measured at baseline, weeks 4, 8, and 12.

Any clinically significant weight changes from the screening visit will be recorded as an AE.

Blood pressure (systolic and diastolic), pulse rate, and weight values meeting criteria for potentially clinically significant values (PCSV) will be reported.

Criteria for PCSV for vital signs and weight are listed in Appendix 2.

4.2.3.3 Respiratory Assessments

At each visit, investigator or designee will perform a respiratory physical examination and ask all subjects whether they have experienced any signs/symptoms of asthma. In addition, for subjects reporting a medical history of asthma, PEF will be performed. PEF measurements should consist of 3 good efforts, with the best result documented.

Newly diagnosed asthma or worsening of asthma during the study will be reported as AESI.

4.2.3.4 Laboratory safety tests

The following laboratory safety tests will be performed:

- Hematology: white blood cell (WBC) count with differential count (including eosinophils), red blood cell (RBC) count, hemoglobin (Hb), hematocrit (hct), mean cell volume (MCV), and platelet count (Plt)
- Blood chemistry: sodium, potassium, calcium, chloride, glucose, urea, creatinine, aspartate aminotransferase (AST)/ alanine aminotransferase (ALT), alkaline phosphatase (ALP), total and direct bilirubin, creatinine phosphokinase (CPK), high sensitivity C-reactive protein (hsCRP), fibrinogen, gamma-glutamyl transpeptidase (GGT), lactate dehydrogenase (LDH), total protein, albumin, uric acid, total cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL) and triglycerides
- CPK isoenzyme test will be performed only if CPK is elevated to >2.5X Upper Limit of Normal (ULN).
- Urinalysis: blood, proteins, leukocytes, glucose, ketones, nitrites, bilirubin, urobilinogen, pH, and specific gravity

All clinically significant out-of-range laboratory values at the screening visit will be recorded in the medical history form. All clinically significant out-of-range laboratory values after the screening visit are to be reported as an AE if this abnormality was not present at the screening visit or is assessed as having worsened since the screening visit (i.e., changed significantly from the screening visit).

4.2.3.5 Adverse events

Adverse events (AE) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 19.0.

An AE will be considered a treatment-emergent AE (TEAE) if it started after the date of the first dose of study drug or it started before the date of the first dose of study drug and worsened after the date of the first dose. AEs occurring on the date of the first dose of study drug will be considered a TEAE if the onset of the AE is on or after the time of injection on that day. Where dates are missing or partially missing, AEs will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to suggest that the AE started prior to the first dose of study treatment.

The investigator will have to assess if there is a reasonable possibility of a causal relationship between the study drug and/or a study procedure (e.g. injection, blood sample collection) and an AE. Related AEs will be defined as study drug related AEs or study procedure related AEs.

The AESIs for this clinical trial have been defined as follows:

- Elevated ALT or AST (> 3 ULN) in combination with elevated bilirubin > 2 ULN, whether or not considered as related to the study drug by the investigator
- Elevated CPK (≥ 2.5 ULN), if considered as related to the study drug by the investigator
- Asthma or worsening of asthma
- IRR (local and systemic reactions, including hypersensitivity)
- Peripheral edema
- Skin or systemic infection
- Headache
- Exacerbation of AD, in subjects with medical history of AD

Missing date information for AEs

The following imputation rules only apply to cases in which the start date for AEs is incomplete (i.e., partial missing):

Missing Month and Day

- If the year of the incomplete start date is the same as the year of the first administration of the study drug, then the month and day of the first administration of the study drug will be assigned to the missing fields.
- If the year of the incomplete start date is before the year of the first administration of the study drug, *December 31* will be assigned to the missing fields.
- If the year of the incomplete start date is after the year of the first dose of double-blind IP, *January 1* will be assigned to the missing fields.

Missing Month Only

- If only the month is missing, then the day will be treated as missing and both the month and the day will be replaced according to the procedure above.

Missing Day Only

- If the month and year of the incomplete start date are the same as the month and year of the first administration of the study drug, then the day of the first administration of the study drug will be assigned to the missing day.
- If either the year of the incomplete start date is before the year of the date of the first administration of the study drug or if both years are the same but the month of the incomplete start date is before the month of the date of the first administration of the study drug, then the last day of the month will be assigned to the missing day.
- If either the year of the incomplete start date is after the year of the date of the first administration of the study drug or if both years are the same but the month of the incomplete start date is after the month of the date of the first administration of the study drug, then the first day of the month will be assigned to the missing day.

If the stop date is complete and the imputed start date as above is after the stop date, then the start date will be imputed by the stop date.

If the start date is completely missing and the stop date is complete, then the following algorithm will be used to impute the start date:

- If the stop date is after the date of the first administration of the study drug, then the date of the first administration of the study drug will be assigned to the missing start date.
- If the stop date is before the date of the first administration of the study drug, then the stop date will be assigned to the missing start date.

4.2.4 Pharmacokinetic and Anti-drug Antibody Assessments

Blood samples will be collected to determine the PK profile of nemolizumab and to assess anti-drug antibodies (ADA). The serum concentration of nemolizumab will be assessed at baseline, weeks 1, 2, 4, 8, 12, 16, 18, and in case of early termination or unscheduled visit for safety reasons. ADA will be assessed at baseline, weeks 4, 8, 12, 16 and 18, and in case of early termination or unscheduled visit for safety reason.

Results will be described in the bioanalytical report, which will be included as an appendix in the final clinical study report.

Nemolizumab serum concentrations will be determined using validated ELISA method by a sponsor representative. A bioanalytical plan describing the details of the bioanalytical work related to serum samples of nemolizumab assessment will be written before the beginning of the sample analysis.

4.2.5 Pharmacodynamic Assessments

Blood and skin samples (D-Squames and skin biopsies) will be collected according to schedule of assessments to investigate the effect of nemolizumab on selected biomarkers, including but not limited to microbiome, cytokine, chemokine IL-31, IL31-RA.

The PD analyses will be performed by a Sponsor representative. The results will be described in a PD report, which will be included in the final study report.

4.2.6 Biophysical Assessments

4.2.6.1 Actigraphy

High resolution actigraphy will be used in this study to evaluate scratching events and duration of sleep during the night. Philips will, in accordance with procedures, centrally analyze assessment data for scratching events and sleep parameters obtained from the actigraphy device and submit data to the CRO. Patients will wear two devices, one on each wrist, from D-7 and every day during the first 4 weeks of the study, by following the instructions provided by the CRO.

Sleep assessment evaluation will include but not be limited to the parameter 'sleep efficiency'.

Scratching assessment evaluation will include but not be limited to the 2 ratio 'scratching duration/duration' and 'scratching events/duration'.

4.2.6.2 Whole Body Imaging

This device will be used in equipped sites only and the photographs of the entire body will be taken according to the operational manual provided by the Sponsor at least at baseline and Week 12 (Day 85) and if it is possible at Week 4 (Day 29), Week 8 (Day 57), and Week 18 (Day 126). It is for illustration and as exploratory evaluation of the lesions improvement.

5 STATISTICAL METHODS

5.1 Data Quality Assurance

All tables, figures and data listings (TLFs) to be included in the report will be independently checked for consistency, integrity and in accordance with standard PAREXEL procedures.

5.2 General Presentation Considerations

Unless otherwise stated, descriptive summaries will be presented by treatment group. In all tables the treatment groups will be presented in the order of Placebo, Nemolizumab.

Continuous data will be summarized in terms of the mean, standard deviation (SD), median, minimum, maximum, lower quartile (Q1), upper quartile (Q3), and number of observations, unless otherwise stated. Where data are collected over time, both the observed data, change from

baseline and percent change from baseline when appropriate will be summarized at each time point. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, Q1, and Q3 will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to 2 more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point (n), frequency counts and percentages. Percentages will be presented to one decimal place. A percentage of 100% will be reported as 100%. Percentages will not be presented for zero counts. Percentages will be calculated using n as the denominator. If sample sizes are small, the data displays will show the percentages, but any textual report will describe frequencies only. Changes from baseline in categorical data will be summarized using shift tables where appropriate.

P-values greater than or equal to 0.001, in general, will be presented to three decimal places. However, if a p-value is only presented to four decimal places (by SAS) it will not be rounded again, but will be presented to four decimal places. P-values less than 0.001 will be presented as "<0.001", and p-values greater than 0.999 will be presented as ">0.999".

Confidence intervals (CIs) will be presented to one more decimal place than the raw data.

All report outputs will be produced using SAS® version 9.3 or a later version in a secure and validated environment. The outputs will be provided as individual TLFs in RTF as well as PDF formats and all TLFs in one PDF bundle.

All statistical tests will be two-sided and will be performed at the 5% level of significance, unless otherwise stated.

Baseline

'Baseline' is defined as the last available assessment prior to the first planned study drug. The date of the first planned study drug refers to date of the first actual administration of study drug at visit V3/Baseline/Day 1 if exists; otherwise, date of randomization.

For diary data (pruritus NRS, VRS, Sleep disturbance NRS), the baseline values will be derived from data collected during the 7 days prior to the first planned study drug. Baseline score will be the weekly prorated average of non-missing subject diary scores reported during the 7 days. A minimum of 5 daily scores out of the 7 days is required to calculate the weekly prorated average score.

Analysis Visit

Study day is relative to date of randomization. Day -1 is the day before randomization, and day 1 is date of randomization.

Assessments at early termination and unscheduled visits will be slotted based on the following visit window.

Period	Analysis Visit	Target day (relative to randomization)	Day ranges
Screening Period	Screening 1	-28	-28 to -8
	Screening 2	-7	-7 to -1
Treatment	Baseline	1	≤1
	Week 1	8	2-11
	Week 2	15	12-22
	Week 4	29	23-43
	Week 8	57	44-71
	Week 12	85	72-99
Follow-up	Week 16	113	100-119
	Week 18	126	>119

The diary reported during study will be classified into analysis visits as follows (i.e., during the 7 days prior to the target day of analysis visit. In this case, the target day of analysis visit is relative to date of the first planned study drug.):

- Day -7 to -1 = Baseline
- Day 1 to 7 = Week 1
- Day 8 to 14 = Week 2
- Day 22 to 28 = Week 4
- Day 50 to 56 = Week 8
- Day 78 to 84 = Week 12
- Day 106 to 112 = Week 16
- Day 120 to 126 = Week 18

These analysis visits will be used in the calculations for all week-based parameters collected on subject's diary (e.g., pruritus NRS, VRS, Sleep disturbance NRS, PAS, DPS, DLQI) Other non-diary data, such as , IGA, vital signs, body weight, clinical laboratory assessments, respiratory assessment/PEF, will be analyzed according to actual scheduled visits.

For efficacy and safety data (except clinical laboratory assessments), where two or more assessments (include both scheduled and unscheduled assessments) are available for the same visit interval, the one closest to the target visit date will be used for the summary and analyses. For clinical laboratory assessments, if repeated measurements are taken for either time point (scheduled visit), then the last measurement will be used for the value for that time point. Unscheduled labs will be included in listings, but not summaries of the data. All post baseline assessments, including repeated or unscheduled visits, will be used for potentially clinically significant value (PCSV) determinations.

5.3 Study Subjects

5.3.1 Disposition of Subjects

A clear accounting of the disposition of all subjects who enter the study will be provided, from screening to study completion.

The following subject disposition summaries will be produced:

- The number of subjects screened (Analysis population: Screened Subjects)
- Screen failures (Analysis population: Screened Subjects)
- Subjects Randomized by clinical site and overall (Analysis population: All Randomized Subjects)

The following summaries will be produced for ITT population:

- The number of subjects randomized
- The number and percentage of subjects treated (with at least one dose of study drug)
- The number and percentage of subjects randomized but not treated (not taken any dose of study drug)
- The number and percentage of subjects who temporarily discontinued the study treatment and reasons for discontinuation
- The number and percentage of subjects who permanently discontinued the study treatment and reasons for discontinuation
- The number of subjects who completed study treatment period (defined as completing week 12 visit)
- The number of subjects who completed the study
- The number of subjects who discontinued from the study and reason of discontinuation
- A Kaplan-Meier plot of the time to permanent study treatment discontinuation will be provided. Subjects who did not discontinue from the study treatment will be censored at their last dose of study drug (Analysis population: ITT).
- A Kaplan-Meier plot of the time to study discontinuation will be provided. Subjects who did not discontinue from the study will be censored at the date of study completion or the last visit date on the study (Analysis population: ITT).

- Time (days) to permanent discontinuation of study treatment by reasons for discontinuation will be displayed graphically in subjects having permanently discontinued study treatment.
- Time (days) to study discontinuation by reasons for discontinuation from study will be displayed graphically in subjects having discontinued study.

By-subject listings of eligibility details, randomization details, visit dates and withdrawal/study completion details (including reason for discontinuation and duration of treatment prior to discontinuation) will be provided.

5.3.2 Protocol Deviations

Deviations from the protocol including violations of inclusion/exclusion criteria will be assessed as 'minor' or 'major'. Major PDs are defined as those deviations from the protocol that are likely to have an impact on the subject's right, safety, well-being, and/or the validity of the data for analysis. Minor deviations include all deviations from the protocol excluding those considered as major. The impact of major protocol deviations on the efficacy and/or safety results will be investigated by assessing the robustness of the study results and conclusions to the choice of analysis population (see Section 5.4), both including and excluding data potentially affected by major protocol deviations. Major PDs that will lead to the exclusion of a subject from the per-protocol (PP) population will be identified. The final determination of major protocol deviations and the exclusion of subjects from each of the analysis populations will be made prior to database lock.

Refer to the study protocol deviation specification for more information.

The following protocol deviation summaries will be provided:

- Number and percentage of subjects with a major protocol deviation by type of deviation (Analysis population: ITT)
- Number and percentage of subjects with a major protocol deviation resulting in exclusion of subjects from PP analysis by type of deviation (Analysis population: ITT)

A by-subject listing of protocol deviations will be provided.

5.4 Analysis Populations

The following analysis populations will be used to analyze the data.

Intent-to-Treat Population (ITT):
The ITT population will consist of all randomized subjects.

Per Protocol Population (PP):
The PP Population is defined as comprising the ITT subjects who have no major protocol deviations that would have a significant effect on the efficacy of the study treatment.

The PP population will be finalized prior to the final database lock.

Safety Population (SAF):
The safety (SAF) population will comprise all subjects in ITT population who receive at least one dose of study drug.

PK Analysis Population (PKAP):
The PK analysis population will include all subjects in the SAF population who provide at least one post-baseline evaluable drug concentration value.

PD Analysis Population (PDAP):

The PD statistical analyses will be performed based on the PP population, and after exclusion of samples unable to produce reliable quantifications of cytokines, and/or immunohistological markers and/ or transcriptomic biomarkers, due to the material quantity and/or quality.

The efficacy summaries and analyses will be based on the ITT population, which is based upon the Intention-to-Treat (ITT) principle. In the event that a subject is allocated the incorrect study treatment as per the study randomization list, subjects will be summarized and analyzed 'as randomized' (i.e., by randomized treatment group). In the event that a subject is stratified incorrectly, 'randomized stratum' will be used rather than 'actual stratum'.

For the primary and selected secondary efficacy endpoints, a sensitivity analysis will be performed on the PP population to assess the robustness of the study conclusions to the choice of analysis population. Subjects will be included in the analysis according to the treatment 'as randomized'.

The safety summaries and analyses will be based on the SAF population. Randomized subjects will only be excluded if there is clear, documented evidence that the subject did not receive any study drug injection. In the event that a subject is allocated the incorrect study treatment as per the study randomization list, subjects will be summarized and analyzed 'as treated' (i.e., by allocated treatment group).

Upon database release, protocol deviation and analysis population outputs will be produced and will be reviewed by Sponsor. An analysis population classification meeting will be arranged to discuss the outputs and to decide which subjects and/or subject data will be excluded from certain analyses. Decisions made regarding the exclusion of subjects and/or subject data from analyses will be made prior to database lock and unblinding and will be documented and approved by Sponsor.

The following analysis population summaries will be provided:

- The number and percentage of subjects in each study population will be presented by study treatment group including total (Analysis population: ITT).

A by-subject listing of analysis population details will be provided. This listing will be presented by treatment group and will include: subject identifier, inclusion/exclusion flag for each population and reason for exclusion from each population. All randomized subjects will appear on this listing.

The following derived and computed variables will be produced:

Variable Description	Computation Methods, Notes, or Equation(s)
Flag of ITT subjects	'Y' if subjects randomized into the study; otherwise 'N'
Flag of SAF subjects	'Y' if subjects in ITT and date and/or time of the first dose injection on Study Drug Administration form is non-missing; otherwise 'N'
Flag of PP subjects	'Y' if subjects in ITT who complete 12-week Treatment Period with the exception of major protocol violators who have significant effect on the efficacy; otherwise 'N'
Flag of PKAP subjects	'Y' if subjects in the safety population who provided at least one post-baseline evaluable drug concentration value; otherwise 'N'
Flag of PDAP subjects	'Y' if subjects in the PP, and after exclusion of samples unable to produce reliable quantifications of cytokines, and/or immunohistological markers and/ or transcriptomic biomarkers, due to the material quantity and/or quality; otherwise 'N'

5.5 Demographic and Other Baseline Characteristics

5.5.1 Demographic and Baseline Characteristics

The following demographic and baseline characteristics will be summarized by descriptive statistics (Analysis population: ITT):

- Age at time of informed consent (in years, as a continuous variable)
- Age category (≤65, >65 years)
- Gender (Male, Female)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Hawaiian Native or Other Pacific Islander, Other)
- Country
- Weight at baseline (in kilograms)
- Height at screening (in centimeters)
- BMI (kg/m^2) calculated using weight and height at screening
  BMI (kg/m^2) = weight (kg) / height $(m)^2$
- Presence or absence of background of atopy
- Baseline average and peak pruritus NRS (weekly average during the 7 days prior to day 1, as a continuous variable)
- Purigo baseline Assessment (PAS)
- IGA
- Number of nodules on the entire body

Stratification factors include country and presence or absence of background of atopy. A summary table for presence or absence of background of atopy will be provided to show any discrepancies between what was reported through Interactive Voice Response System (IVRS)/ Interactive Web Response System (IWRS) versus eCRF data (at baseline visit). Country and background of atopy will be used as a cofactor in the primary and secondary efficacy endpoint analyses.

By-subject listings of demographic data and other baseline characteristics (as summarized above) will be provided.

5.5.2 Medical History

Medical history from the screening visit will be summarized by treatment group and overall. Medical history will be coded using the MedDRA version 19.0. The number and percentage of subjects experiencing at least one such diagnosis will be summarized by the MedDRA system organ class (SOC) and preferred term (PT) (Analysis population: ITT).

Summary (n, %) may be presented according to MedDRA groups for specific MH analyses (such as asthma, food allergy, allergic conjunctivitis, etc.).

By-subject listings of medical history will be provided.

5.5.3 Prior and Concomitant Therapies

Start and stop dates of medications or medical and surgical procedures will be compared to the date of first dose of study drug to allow them to be classified as either Prior or Concomitant. Medications or medical and surgical procedures that start and stop prior to the date of first dose of study drug will be classified as Prior.

If start and/or stop dates of medications or medical and surgical procedures are missing or partial, the dates will be compared as far as possible with the date of first dose of study drug. Medications or medical and surgical procedures will be assumed to be Concomitant, unless there is clear evidence (through comparison of partial dates) to suggest that they started and stopped prior to the first dose of study treatment. If there is clear evidence to suggest that the medication or medical and surgical procedure started and stopped prior to the first dose of study drug, they will be assumed to be Prior.

Missing date information for prior and concomitant medications

For prior and concomitant medications, incomplete (i.e., partial missing) start dates and/or stop dates will be imputed. When the start date and the stop date are both incomplete for a subject, the start date will be imputed first.

Incomplete start date

The following rules will be applied to impute the missing numeric fields for an incomplete prior or concomitant medication start date. If the stop date is complete (or imputed) and the imputed start date is after the stop date, then the start date will be imputed using the stop date.

Missing Month and Day

- If the year of the incomplete start date is the same as the year of the first administration of the study drug, then the month and day of the first administration of the study drug will be assigned to the missing fields.
- If the year of the incomplete start date is before the year of the first administration of the study drug, *December 31* will be assigned to the missing fields.
- If the year of the incomplete start date is after the year of the first administration of the study drug, *January 1* will be assigned to the missing fields.

Missing Month Only

- If only the month is missing, then the day will be treated as missing and both the month and the day will be replaced according to the procedure above.

Missing Day Only

- If the month and year of the incomplete start date are the same as the month and year of the first administration of the study drug, then the day of the first administration of the study drug will be assigned to the missing day.
- If either the year of the incomplete start date is before the year of the date of the first administration of the study drug or if both years are the same but the month of the incomplete start date is before the month of the date of the first administration of the study drug, then the last day of the month will be assigned to the missing day.
- If either the year of the incomplete start date is after the year of the date of the first administration of the study drug or if both years are the same but the month of the incomplete start date is after the month of the date of the first administration of the study drug, then the first day of the month will be assigned to the missing day.

If the stop date is complete and the imputed start date as above is after the stop date, then the start date will be imputed by the stop date.

Incomplete stop date

The following rules will be applied to impute the missing numeric fields for an incomplete prior or concomitant medication stop date. If the date of the last administration of the study drug is missing, then replace it with the last visit date in the imputations described below. If the imputed stop date is before the start date (imputed or non-imputed start date), then the imputed stop date will be equal to the start date.

Missing Month and Day

- If the year of the incomplete stop date is the same as the year of the last administration of the study drug, then the month and day of the last administration of the study drug will be assigned to the missing fields.
- If the year of the incomplete stop date is before the year of the last administration of the study drug, *December 31* will be assigned to the missing fields.
- If the year of the incomplete stop date is after the year of the last administration of the study drug, *January 1* will be assigned to the missing fields.

Missing Month Only

- If only the month is missing, then the day will be treated as missing and both the month and the day will be replaced according to the above procedure.

Missing Day Only

- If the month and year of the incomplete stop date are the same as the month and year of the last administration of the study drug, then the day of the last administration of the study drug will be assigned to the missing day.
- If either the year of the incomplete stop date is before the year of the date of the last administration of the study drug or if both years are the same but the month of the incomplete stop date is before the month of the date of the last administration of the study drug, then the last day of the month will be assigned to the missing day.
- If either the year of the incomplete stop date is after the year of the date of the last administration of the study drug or if both years are the same but the month of the incomplete stop date is after the month of the date of the last administration of the study drug, then the first day of the month will be assigned to the missing day.

5.5.3.1 Prior and concomitant medications

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHODD), version March 2016 E B2.

Prior and Concomitant medications will be summarized separately by Anatomical Therapeutic Chemical (ATC) level 2, level 3, and preferred name in frequency tables (Analysis population: ITT). Subjects with more than one medication in a given ATC level and preferred name will be counted only once in that category. The following summaries will be produced:

- Number and percentage of subjects who had Prior medications by ATC level 2, level 3, and preferred name
- Number and percentage of subjects who had Concomitant medications by ATC level 2, level 3, and preferred name

A by-subject listing of all prior and concomitant medications will be provided.

5.5.3.2 Prior and concomitant medical and surgical procedures

Medical and surgical procedures will be coded using the MedDRA version 19.0.

Prior and Concomitant medical and surgical procedures will be summarized separately by SOC and PT in frequency tables (Analysis population: ITT). Subjects with more than one procedure in a given SOC and PT will be counted only once in that category. The following summaries will be produced:

- Number and percentage of subjects who had Prior medical/surgical procedures by SOC and PT
- Number and percentage of subjects who had Concomitant medical/surgical procedures by SOC and PT

A by-subject listing of all prior and concomitant medical and surgical procedures will be provided.

5.5.3.3 Rescue therapies

Rescue treatments for pruritus could be added to the study drug from Day 29. In that case, all efficacy and safety assessments should be completed before starting the rescue treatments. The number and percentage of subjects initiating use of rescue therapies will be summarized overall and by preferred name (for rescue medications) or preferred term (for rescue procedures) for all subjects in the ITT population. Rescue medication will be summarized by topical and systemic groups. Subjects with more than one rescue therapy in a given preferred name or preferred term will be counted only once in that category.

By-subject listings of all rescue therapies will be provided.

5.6 Treatment Compliance

Subject compliance with study treatment will be assessed via relative injections volume (%) of study drug. Refer to Section 5.8.1 for more analyses.

Actual Total Volume (mL) is defined as the sum of the real volume injected according to the case report form records.
Prescribed Total Volume (mL) is defined as number of total injections Type equation here.x (Weight of the subject in kg x 0.02 ml/Kg)) .
Compliance Rate (%) is defined as

(Volume(mL) of actual injections / Volume of planned injections) x 100

A Subject is study medication compliance if the compliance rate in between 80% and 120%..

5.7 Efficacy Evaluation

Primary inference for efficacy analysis will be the percent change from baseline in NRS to week 4 (weekly average of the peak). This analysis will be based on the ITT population, and repeated using the PP population. Primary imputation method for any missing data including those who took rescue medication as failure will be the LOCF (last observation carried forward) approach. In addition, Multiple Imputation (MI) using the missing at random assumption will be conducted on ITT population:

5.7.1 Analysis and Data Conventions

This study is designed to test for superiority. Appropriate statistical tests to assess the superiority of nemolizumab should be performed, with the null hypothesis being: there is no difference between nemolizumab and placebo treatments and the alternative hypothesis: there is a difference between treatments in reducing the pruritus as major symptom of PN at week 4.

5.7.1.1 Multi-center Studies

For the purpose of the summaries and analyses, the term 'Center' will be used to define each investigator site. All sites within the country (France, Germany, Poland and Austria) will be pooled together in the efficacy analyses. If the country is found to be a significant factor in ANOVA or ANCOVA models the evaluation of the consistency of treatment effects across countries will be performed via subgroup analysis (see Section 5.7.1.6).

5.7.1.2 Adjustments for Covariates

The primary and secondary efficacy analyses will be adjusted for the following baseline covariates:

- Presence or absence of background of atopy (the stratification factor for randomization to the treatments),
- Country

Additional covariates will be used as follows:

- Absolute change from baseline to any visits of the weekly average of the peak and of the average pruritus NRS will use the same factors as the percent change but will also include baseline NRS as a covariate.
- Absolute change and shift from baseline to each visit in Dermatology Life Quality Index (DLQI) will include baseline DLQI, background of atopy, and country as covariates.

5.7.1.3 Handling of Missing Data

Use of rescue therapy: For the purpose of efficacy analysis, subjects receiving any rescue therapies will be considered as treatment failures. All efficacy data, except Observed Case (OC), will be set to missing after rescue medication is used. In OC analysis, observed data after subject has received rescue treatment will be included.

Diary data: If a subject has less than 5 diary entries in a week, the weekly prorated average scores will be set to missing. Otherwise the method of computing weekly score when data are missing will be mean replacement approach. i.e. calculation of the overall weekly score a given week as 7 times average of non-missing average of that week.

The following methods will be used to impute the missing values:

Continuous endpoints: To impute the missing values for the primary and secondary endpoints, the following imputation approaches will be carried out to assess the robustness of study conclusions with choice of imputation method.

For sensitivity analyses, continuous endpoints will be analysed using analysis of covariance (ANCOVA) including terms of treatment group, country, presence or absence of background of atopy (the stratification factor for randomization to the treatments) and the corresponding baseline value.

1. **Last Observation Carried Forward (LOCF):** The primary imputation method for any missing data will be the LOCF approach. The last observed post-baseline value will be used to replace missing values for continuous variables. For questionnaire-based variables, LOCF will be applied to missing individual questions first, and the value of the total (or sub-scale) score will be derived based on the imputed individual questions.

2. **Observed Case (OC):** No data will be imputed. The observed values are used in analysis, including assessments post rescue medication.

3. **Multiple imputation (MI) method under assumption missing at random (MAR):**

   Multiple Imputation (MI) using the Missing At Random (MAR) assumption will also be used. The MI procedure of the SAS system will be used to generate five sets of data with missing values imputed from observed data. It is expected that the pattern of missing data will be monotonic, with slight deviations being corrected by the Markov Chain Monte Carlo (MCMC) method of the MI procedure. Linear regression will be employed to model the missing NRS score, with the following covariates included in the imputation model: treatment and non-missing data from earlier timepoints.

   - Analysis phase: The imputed datasets will be analyzed using the methodology described for percent change from baseline in NRS score.
   - Pooling phase: The results from the analysis of the multiple imputed datasets will be combined by the MIANALYZE procedure of the SAS system.

Binary endpoints:

All missing values will be treated as a Non-Responder for the binary endpoints. If a subject withdraws from the study, all assessments after withdrawal will be considered as Non-Responder.

- LOCF, OC, MI under missing not at random assumption approaches will be used as sensitivity analysis to impute the missing values for the selected secondary endpoints as appropriate. For sensitivity analyses, binary endpoints will be analysed using a Cochran-Mantel-Haenszel (CMH) test stratified by, presence or absence of background of atopy and country (Australia, France, Germany, Poland, and US)

There will be no imputation for missing laboratory and vital sign data.

5.7.1.4 Multiple Comparisons/Multiplicity

There will be no adjustment for multiplicity.

5.7.1.5 Planned Analyses

Interim Analysis: No interim analyses are planned.

5.7.1.6 Examination of Subgroups

To evaluate the consistency of treatment effects, subgroup analyses will be explored for the primary and selected secondary efficacy endpoints based on:

- Age (18-65 or >65)
- Gender (Male, Female)
- Country (Austria, France, Germany, and Poland)
- Presence or absence of background of atopy (the stratification factor for randomization to the treatments)
- Baseline number of prurigo nodularis nodule (< median vs ≥ median), if exact counts are not available in the database then use (20-100 vs. >100) categories as classes.

Subgroup analyses will be explored for safety endpoints (e.g., AEs) based on:

- Age (18-65 or >65)
- Gender (Male, Female)

- Country (Austria, France, Germany, and Poland)
- Presence or absence of background of atopy (the stratification factor for randomization to the treatments)

If the number of subjects in a subgroup is too small, subgroups may be pooled for analyses.

Decisions regarding merging subgroups or eliminating some subgroups will be made at the blinded review before treatment unblinding and database lock.

Summary statistics will be provided for each of the subgroups by time points. The estimated mean change of treatment groups and 95% CI will be calculated and displayed graphically using forest plots for subgroup.

5.7.2 Primary Efficacy Endpoint

Pruritus Numeric Rating Score (NRS) is a scale to be used by the subjects to report the intensity of their pruritus (itch) during the last 24 hours all along the study from the first screening visit to the follow up visit.

The primary variable for the assessment of efficacy is the percent change from baseline in NRS at week 4 of average of the peak pruritus.

An analysis of variance (ANOVA) model will be used to test the difference between the nemolizumab and placebo treatments with percent change from baseline in NRS scores at week 4 being the dependent variable, and treatment, presence or absence of atopy and country terms as factors.

A 95% Confidence for the difference in LSmeans of treatments and p-value will be calculated based on contrast test statistics.

Other covariates may be explored and will be added into the model, if appropriate.

Sensitivity analysis: The robustness of the primary analysis will be investigated by performing the sensitivity analyses, as described in 5.7.1.3.

Subgroup analysis: The homogeneity of the treatment effect for a number of important subgroups will be investigated, as described in 5.7.1.6.

A by-subject listing of the primary efficacy data will be provided. In addition, a boxplot and line plots showing the mean percent change and 95% CI from baseline in NRS over time within each treatment group will be produced for ITT and PP populations.

5.7.3 Secondary Efficacy Endpoints

5.7.3.1 Assessment of pruritus and diary data by the Subject

Pruritus Numeric Rating Scale (NRS)

Absolute and percent change from baseline (Day 1) in weekly average of the peak and average pruritus NRS at each visit (Weeks: 1, 2, 4, 8, 12, 16, 18, E/T or unscheduled visit, if applicable) up to week 18.

The descriptive statistics for each time point and absolute change and percent change from baseline will be presented by each treatment group. Least square means, difference from baseline and 95% CI for that will also be provided at each assessment time point.

The percent and absolute change from baseline of peak and average NRS will be analyzed separately at each assessment timepoint using the ANOVA model described for the primary analysis above. For absolute change endpoint, the ANCOVA model, with the same factors as in the percent change model, will be performed, but with baseline NRS as a covariate.

A 95% CI for the treatment difference will be calculated for both absolute change and percent change of NRS from baseline and will be plotted by treatment group for ITT population.

Sensitivity analysis: Sensitivity analyses will be performed as described in 5.7.1.3.

Subgroup analysis: Subgroup analyses will be investigated as described in 5.7.1.6.

Verbal rating scale (VRS)

The VRS consists of a list of adjectives describing different levels of symptom intensity to be used by the subjects to report the intensity of their pruritus (itch) over the last 24 hours.

The intensity of subject's pruritus (itch) will be measured daily using the five-point scale, where 0 = no itch, 1 = mild itch, 2 = moderate itch, 3 = severe itch, 4 = very severe itch.

The efficacy endpoint for the VRS will be the absolute and percent change from baseline in weekly average of the peak and average VRS to each visit (Day 1, baseline, Week 1, 2, 4, 8, 12, 16, 18, E/T or unscheduled visit, if applicable). These endpoints will be compared with nemolizumab and placebo treatments using ANOVA model for the primary analysis described above separately at each assessment visit. For absolute change endpoint the ANOVA model will be performed, but with baseline VRS as an additional covariate.

A 95% CI for the treatment difference will be calculated for both absolute change and percent change of VRS from baseline and will be plotted by treatment group for ITT population.

Dynamic Pruritus Score (DPS)

The 9-point DPS is a dynamic scale to be used by subjects to evaluate the change of their pruritus compared with an earlier time point (i.e. before injection on Day 1).

The details of measuring scale and assessment time are given in section 4.2.1.3. The data will be analyzed using a Cochran-Mantel-Haenszel (CMH) test with ridit scores between nemolizumab and placebo, stratified by by background of atopy (presence or absence) and country.

5.7.3.2 Assessment of Prurigo by Physician

Investigator Global Assessment (IGA) of Prurigo

IGA is a 5-point scale ranging from 0 (clear) to 4 (severe) used by the investigator or trained designee to evaluate the severity of the disease. The details of measuring scale and assessment time are given in section 4.2.1.4.

IGA evaluation of the severity of at each post-treatment timepoint will be summarized as a continuous variable and as a categorical variable, given its ordinal nature by treatment group. The data will be analyzed using a Cochran-Mantel-Haenszel (CMH) test with ridit scores between nemolizumab and placebo, stratified by by background of atopy (presence or absence) and country.

The shift table will also be provided for overall investigator-rated evaluation change from baseline to Week 18 for each treatment group and overall.

Assessment of Prurigo Activity Score (PAS)

The evaluation of the disease will be performed by the Investigator using the Prurigo Assessment Scale (PAS) at baseline and Week 12 using the entire scale and at week 4, 8,18 and in case of Early termination visit / Unscheduled visit, if applicable, using item 5 (number of lesions only) and 6 (excoriation/crusts and healed lesions).

PAS evaluation of the severity of at each post-treatment timepoint will be summarized as a continuous variable and as a categorical variable, given its ordinal nature by treatment group. The data will be analyzed using a Cochran-Mantel-Haenszel (CMH) test with ridit scores between nemolizumab and placebo, stratified by by background of atopy (presence or absence) and country. The shift table will also be provided for overall investigator-rated evaluation change from baseline to Week 18 for each treatment group and overall.

5.7.4 Other Efficacy Endpoints

Dermatology Life Quality Index (DLQI)

DLQI will be assessed at baseline, weeks 4, 12, or early termination visit when applicable.

The DLQI score will be categorized as follows:

- 0-1 = no effect at all on subject's life
- 2-5 = small effect on subject's life
- 6-10 = moderate effect on subject's life
- 11-20 = very large effect on subject's life
- 21-30 = extremely large effect on subject's life

The DLQI can be analyzed under 6 sub-scales, namely, symptoms and feelings (question 1 and 2), daily activities (question 3 and 4), leisure (question 5 and 6), work and school (question 7), personal relationships (question 8 and 9), and treatment (question 10). For general inflammatory skin conditions, a change in DLQI score of at least 4 points is considered clinically important.

The primary study population for all QoL and productivity data will be the ITT Population. DLQI will be summarized by treatment group and visit. By-subject listings of these data should be provided.

Change from baseline in DLQI at each visit will be analyzed using ANCOVA with treatment group, baseline DLQI as a covariate, back ground of atopy and country as cofactors. In addition, the following analyses will be performed:

- A shift table (by visit) will be generated using the categories described above. Change from baseline in each sub-scale at each visit will be analyzed as the efficacy endpoint using ANCOVA.
- Proportion of subjects achieving a change in DLQI score of at least 4 points at each Visit up to week 12 will be analyzed using a CMH test with ridit scores between nemolizumab and placebo, stratified by by background of atopy (presence or absence) and country.

Number and percentage of subjects will also be provided for the following DLQI score categories at each visit for the above categories and analyze using CMH test stratified by background of atopy and by country with the redit scores.

Sleep disturbance Numeric Rating Scale (NRS)

The efficacy endpoint for the Sleep disturbance NRS will be the Absolute and Percent change from baseline in weekly average of Sleep disturbance NRS (Day 1, baseline), Week 1, Week 2, Week 4, and E/T or unscheduled visit, if applicable). These endpoints will be compared across nemolizumab and placebo treatment groups using ANOVA model for the primary analysis above separately at each assessment visit. For absolute change endpoint, the same ANOVA model, with the same factors as in the percent change model will be performed, but with baseline NRS as a covariate.

Summary statistics of objective assessment of scratching events during the night and sleep duration the two Ratios 'scratching duration/duration' and 'scratching events/duration' up to week 4 will be summarized.

A 95% CI for the treatment difference will be calculated for both absolute change and percent change of NRS from baseline and will be plotted by treatment group for ITT population.

5.8 Safety Evaluation

All safety summaries and analyses will be based upon the SAF population as defined in Section 5.4.

5.8.1 Extent of Exposure

The extent of treatment will be summarized by treatment group as follows:

- Duration of study drug (days)
- Cumulative dose of nemolizumab (mg/kg) received
- Relative dose volume (%) of nemolizumab
- Proportion of subjects compliant with study drug at each study drug administration visit by treatment group

5.8.2 Adverse Events

Analysis of severity: The severity of each AE will be summarized as assessed by the investigator (Mild, Moderate, Severe). Within the same MedDRA PT, only the most severe AE for each subject will be counted in tabulations by severity. Within a MedDRA SOC, subjects with more than one MedDRA PT will be counted only once for the most severe AE reported. AEs for which the severity is missing will be imputed to be Severe; this imputation will take place prior to determining the most severe AE within a SOC or PT for a given subject.

Analysis of causality: The relationship of each AE to the study drug and/or study procedure will be summarized as assessed by the investigator (Reasonable Possibility, No Reasonable Possibility). Within the same MedDRA PT, only the AE with the highest ranked relationship to study drug and/or study procedure for each subject will be counted in tabulations by causality. Within a MedDRA SOC, subjects with more than one MedDRA PT will be counted only once for the AE that is most related to study drug and/or study procedure. AEs for which the relationship to study drug is missing will be considered as Reasonable Possibility related. The imputation for a missing relationship will take place prior to determining the most related AE within a SOC or PT for a given subject.

High-level summary of TEAEs will be presented based on 1) all causalities, 2) study drug related, and 3) study procedure (including TCS) related. The table will include the number and percentage of subjects for the following categories:

- Number of TEAEs
- Subjects with TEAEs
- Subjects with treatment-emergent SAEs (TESAEs)
- Subjects with severe TEAEs
- TEAE with fatal outcome
- Subjects temporarily discontinued from study treatment due to TEAEs
- Subjects permanently discontinued from study treatment due to TEAEs
- Subjects discontinued from study due to TEAEs

The number and percentage of subjects reporting any TEAE will be summarized by SOC and PT.

- Summary of TEAEs by MedDRA SOC and PT based on 1) all causalities; 2) study drug related, and 3) study procedure (including TCS) related
- Summary of Severe TEAEs by MedDRA SOC and PT based on 1) all causalities; 2) study drug related, and 3) study procedure (including TCS) related
- Summary of the most common TEAEs (>=5%) by MedDRA SOC and PT based on all causalities
- Summary of TEAEs by MedDRA SOC and PT based on all causalities for subgroups as described in Section 5.7.1.6

AE summaries will be ordered in terms of decreasing frequency for SOC, and PT within SOC, in the treatment group (combined, nemolizumab, then placebo), and then alphabetically for SOC, and PT within SOC. If deemed as appropriate, a different order may be applied.

All AEs will be provided in a by-subject listing which will include both the term reported on the eCRF (verbatim term) and the PT and SOC to which it is coded. Relative start and stop days will be included along with the actual onset and resolution dates. Pre-treatment AEs will be listed separately.

5.8.3 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

Other significant TEAEs are those TEAEs reported as leading to permanent or temporary discontinuation of study treatment, leading to discontinuation of study, AESIs.

The following summaries will be produced:

- Summary of TESAEs by MedDRA SOC and PT based on 1) all causalities; 2) study drug related, and 3) study procedure (including TCS) related
- Summary of TEAEs associated with permanent discontinuation of study treatment by MedDRA SOC and PT based on 1) all causalities; 2) study drug related, and 3) study procedure (including TCS) related
- Summary of AEs associated with discontinuation of study participation by MedDRA SOC and PT based on 1) all causalities; 2) study drug related, and 3) study procedure (including TCS) related
- Summary of AESIs based on 1) all causalities; 2) study drug related

In addition, separate listings will be provided for the following:

- Deaths and other SAEs
- Permanent discontinuations of study treatment due to AEs
- Discontinuations of study participation due to AEs
- Temporary discontinuations of study treatment due to AEs
- AESIs

5.8.4 Clinical Laboratory Evaluation

All laboratory values will be reported in SI units.

Laboratory data (absolute values and change from baseline) will be summarized descriptively by visit and treatment group.

In addition, the number and percent of subjects below, within, and above the laboratory reference ranges and the number and percentage of subjects who met criteria of PCSV (see Appendix 5) will be summarized by treatment group. Shift tables will be generated using the reference ranges. The shift will also be illustrated by the spagetthi plots.

A by-subject listing of all laboratory data will be provided by treatment group, with abnormal values flagged. Laboratory reference ranges should also be listed.

5.8.5 Vital Signs, Weight, Physical Findings and Other Observations Related to Safety

5.8.5.1 Vital Signs and Weight

All vital signs and weight data (absolute values and change from baseline) will be summarized by visit and treatment group.

In addition, the number and percent of subjects with blood pressure (systolic and diastolic), pulse rate, and weight values meeting criteria for PCSV will be summarized by treatment group.

A by-subject listing of vital signs data will be provided, with potentially clinically significant values flagged.

5.8.5.2 Peak Expiratory Flow (PEF)

For subject with medical history of asthma, PEF measurements (absolute values and change from baseline for actual PEF and %predicted PEF) will be summarized by visit and treatment group.

Number and percentage of subjects who met the following criteria will be summarized.

- With %predicted PEF less than 80%
- With >=20% change (decrease) from baseline in PEF

A by-subject listing of respiratory assessments/PEF will be provided.

5.8.5.3 Electrocardiogram

A by-subject listing of ECG overall interpretation will be provided.

5.9 Pharmacokinetic and Pharmacodynamics Evaluations

5.9.1 Pharmacokinetic Parameter Calculation Methods

Derivation of PK parameters will be the responsibility of Quantitative Clinical Development (QCD), PAREXEL International. PK parameters will be calculated by non-compartmental analysis methods from the concentration-time data using Phoenix® WinNonlin® (WNL) Professional (Version 6.3 or higher). The following guidelines will be applied:

- Actual sampling times relative to dosing will be used in the calculation of all derived PK parameters. If the dosing time is not available for the calculation of actual sample time or the sample collection time is not available, the nominal time as recorded in the database will be used in the calculation. The listing of serum concentrations will note any instance in which a nominal time is presented in place of an actual sample time.
- BLQs at the beginning of a subject profile (i.e., before the first incidence of a measurable concentration) will be assigned to zero.
- BLQs at the end of a subject profile (i.e., after the last incidence of a measurable concentration) will be set to missing.
- Single BLQs which fall between two measurable concentrations will be set to missing.
- Consecutive BLQs which fall between measureable concentrations will be set to missing.
- Data from subjects with missing concentration values (missing samples) may be used if pharmacokinetic parameters can be estimated using the remaining data points.

Derivation of the pharmacokinetic parameters for nemolizumab will be performed using the following guidelines:

- C_{max} will be obtained directly from the concentration-time data.
- T_{max} is the time at which C_{max} is observed.
- T_{last} is the time of the last quantifiable concentration.
- λ_Z will be estimated at terminal phase by linear regression after log-transformation of the concentrations:
  - Only those data points that are judged to describe the terminal log-linear decline will be used in the regression.
  - A minimum number of three data points in the terminal phase will be used in calculating λ_z with the line of regression starting at any post-C_{max} data point (C_{max} should not be part of the regression slope). The adjusted correlation coefficient (R^2 adjusted) in general should be greater than 0.80. Any value < 0.80 will be excluded from the analysis.
  - The interval used to determine λz should be equal or greater than 1.5-fold the estimated half-life or otherwise flagged and excluded from the analysis.
- $t_{½}$ will be calculated as $\ln 2/\lambda_Z$.
- AUC is calculated by the linear up/log down method (linear method will be employed for all incremental trapezoids arising from increasing concentrations and the logarithmic method will be used for those arising from decreasing concentrations.

$$AUC_{0\text{-}t} = \int_0^t C(t)dt$$

$$AUC_{0\text{-}inf} = \int_0^t C(t)dt + \int_t^\infty C(t)dt = AUC_{0\text{-}t} + C_t/\lambda_z$$

- C_{last} is the last observed quantifiable concentration.
- $\%AUC_{ex}$ will be calculated as $(1 - [AUC_{0\text{-}t}/AUC_{0\text{-}inf}]) \times 100$. The $\%AUC_{ex}$ should not exceed 20% for each individual profile. If the $\%AUC_{ex}$ is more than 20%, the individual result should be flagged and mentioned in the report as well as all parameters depending on $AUC_{0\text{-}inf}$. If the $\%AUC_{ex}$ is greater than 30%, the value and all dependent parameters may be flagged and excluded, at the discretion of the PK scientist.

When appropriate, the following PK parameters will be determined for each subject:

From baseline to week 4:

- C_{max}: The observed peak drug concentration

- T_{max}: The time at which C_{max} occurs
- T_{last}: Time of last quantifiable concentration
- AUC_{0-4W}: Area under the concentration time curve from pre dose through 28 (4 weeks) days post dosing. AUC_{0-4W} will be calculated by mixed linear logarithmic trapezoidal method

From baseline to week 12:

- AUC0-12w: Area under the concentration time curve from pre dose through 12 weeks post dosing. AUC0-12w will be calculated by mixed linear logarithmic trapezoidal method. AUC0-12w included the serum drug concentration up to week 12 (i.e. the serum concentration collected at baseline, week 1, 2, 4, 8 and 12)

From baseline to week 18:

- AUC0-18w: Area under the concentration time curve from pre dose through 24 weeks post dosing. AUC0-18w will be calculated by mixed linear logarithmic trapezoidal method. AUC0-18w included the serum drug concentration up to week 18 (i.e. the serum concentration collected at baseline, week 1, 2, 4, 8, 12, 16 and 18)

- From baseline to week 18 (Early termination or Unscheduled visit, if applicable)
  - AUC0-t: area under the concentration time curve calculated by the mixed linear logarithmic trapezoidal method from T0 up to the sampling time corresponding to the last quantifiable concentration (Clast)

From week 12 to week 18 (or Early termination or Unscheduled visit, if applicable):

- $t_{1/2}$: the terminal half-life value ($t_{1/2}$) will be calculated using the equation ln2/k after the last drug injection (week 8).
- AUC_{0-t}: area under the concentration time curve calculated by the mixed linear logarithmic trapezoidal method from T0 up to the sampling time corresponding to the last quantifiable concentration (C_{last}).
- AUC_{0-inf}: Area under the plasma concentration-time curve calculated by the mixed linear-logarithmic trapezoidal method from T0 and extrapolated to time infinity as: $AUC_{0-inf} = AUC_{0-t} + C_{last} / k_{el}$. When the extrapolation represents more than 20%, AUC_{0-inf} and $t_{1/2}$ will not be reported.

At weeks 4, 8, 12, 16, 18 (or Early termination or Unscheduled visit if applicable):

- C_{trough}: The residual drug concentration
- Accumulation index: calculated with the serum nemolizumab trough concentration obtained 4 weeks after the first dose (week 4) and 4 weeks after last drug injection (week 18).

5.9.2 Pharmacokinetic Parameters and Anti-drug Antibody Analyses

The PK parameters derived using non-compartmental techniques will be regarded as primary endpoints for the PK analyses (see Section 5.9.1). Primary inference for all the PK parameters will be based on the PKAP population.

The concentration at each time point will be summarized as arithmetic mean, standard deviation, median, minimum, and maximum, number of BLQs (Below the Limit of Quantification). Pharmacokinetic parameters using geometric means will be compared to determine when steady state conditions are achieved during the treatment period. Geometric means and between-subject coefficients of variation (CVb) will be calculated for $\log_e$-transformed $AUC_{0\text{-}28d}$, C_{trough} and C_{max} where:

Geometric mean = exp(mean on log scale)

CV_b (%) = sqrt[exp(SD^2) - 1] x 100

– where SD is the standard deviation of the $\log_e$-transformed data.

Following $\log_e$-transformation, C_{trough} will be analyzed using analysis of variance (ANOVA), the model will include time and subject as a fixed effect. The residual variance from the model will be used to calculate point estimates and 90% CIs for the least squares means for each treatment formulation on the $\log_e$ scale. These estimates will be back transformed to give point estimates and 90% CIs on the original scale.

The potential relationship between plasma concentrations of nemolizumab and change in pruritus NRS, biomarkers or other indicators of disease activity will be explored using PK/PD modeling, as appropriate.

Incidence of positive ADA results will be summarized by treatment group (absolute occurrence and percent of subjects). Individual PK profiles of subjects with or without ADA will be plotted to explore ADA impact on systemic exposure levels.

5.9.3 Pharmacodynamic Analyses

Primary inference for all biomarker analyses based on blood and skin samples (D-Squames and skin biopsies) will be based on the observed cases. All biomarker variables (using logarithm transformation) will be summarized across treatments at each time point.

Any observations in trends within the Biomarker data may be explored with PK/PD correlations/models. All of this analysis will be considered exploratory. The outputs will be in a separate report.

The detailed analysis will be carried out by Biomarker group in GALDERMA, and will be reported separately.

5.10 Determination of Sample Size

For the current study, it is assumed that the common standard deviation is 35 and that the true difference between Nemolizumab and the placebo is 30% in terms of percent change from baseline in worst itching score on a NRS at Week 4.

With an effect size of (30/35=) 0.857, a power of 90% and a type I error of 5% two-sided; at least 30 subjects are needed per group. In order to maintain the power of the tests, for per-protocol population, in case of drop outs/major deviations at Week 4, the sample size will be increased to 35 subjects per group, i.e. 70 to be randomized.

5.11 Changes in the Conduct of the Study or Planned Analysis

Changes in the conduct of the study may be instituted through a protocol amendment. Planned analyses will be revised as appropriate. Similarly, planned analyses may be changed as a result of planned blinded data reviews. Changes will be finalized prior to database lock.

The ridit score is calculated combining data in the atopy and country stratum due to small sample size in each strata. The Placebo group is used as the reference group to calculate mean redit score of the nemolizumab group using the method describies in Uwawunkonye et. al. paper.

6 REFERENCES

Basra MK et al. Determining the minimal clinically important difference and responsiveness of the Dermatology Life Quality Index (DLQI): further data. Dermatology, 2015; 230(1):27-33.

Fostini AC, Girolomoni G, Tessari G. Prurigo nodularis : an update on etiopathogenesis and therapy. J. Dermatolg treat. 2013. 24(6):458-62.

Hyde JN, Montgomery FHA. Practical treatise on diseases of the skin for the use of students and practitioners. Lea & Febiger, Philadelphia, 1909. 174-175.

Ruzicka T, Hanifin JM, Furue M, Pulka G, Mlynarczyk I, Wollenberg A, Galus R, Etoh T, Mihara R, Yoshida H, Stewart J, Kabashima K; XCIMA Study Group. Anti-Interleukin-31 Receptor A Antibody for Atopic Dermatitis. N Engl J Med. 2017 Mar 2;376(9):826-835.

Warlich B, Osada N, Bruland P, Stumpf A, Schneider G, Dugas M, Pfleiderer B, Ständer S. Health-Related Quality of Life in Chronic Pruritus: An Analysis Related to Disease Etiology, Clinical Skin Conditions and Itch Intensity. Dermatology. 2015;231(3):253-9.

Uwawunkonye, EG and Anaene, OIC. A comparative study between ridit and modified ridit analysis. American Journal of Theretical and Applied Statistics. 2013; 2(6):248-254.

Appendix 1 Schedule of assessments

Table 2 Schedule of Assessments

Study period	Screening period (Day -28 to Day-1)		Treatment period						Observation period/ Follow up		Early termination (E/T) or Unscheduled visit (if applicable)
Visit	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	
Week	Screening 1	screening 2	Baseline	W1	W2	W4	W8	W12	W16	W18	
Day	D-28 to D-8	Day-7 [a]	Day 1	Day 8	Day 15	Day 29	Day 57	Day 85	D113	Day 126	
Visit window				±1 day	±2 days	±2 days	±2 days	±5 days	±5 days	±7 days	
ICF and when applicable country specific consent form	•										
Photo Consent and Biopsies consent form (if applicable)	•										
Demographics	•										
Medical history/ Previous therapies and procedures	•										
Inclusion/Exclusion criteria	•		•								
Concomitant therapies and procedures	•	•	•	•	•	•	•	•	•	•	•
Height			•								
Weight			•			•	•	•			
Blood sample for serology (HIV, Hepatitis B and C test)	•										
Blood samples for specific IgE	•										
Blood samples for TB test [b]	•										
SAFETY ASSESSMENTS											
Vital signs	•		•	•	•	•	•	•	•	•	•
Physical exam	•		•			•	•	•		•	•
Blood samples for hematology and biochemistry [c]		•	•			•		•			•
Blood sample for CPK [d]	•		•			•	•	•			•
Urinalysis		•	•			•		•			•
12-lead ECG	•		•				•				•
Pregnancy test [e]	Serum		Urine			Urine	Urine	Urine	Urine	Urine	Urine
Respiratory assessment [f]	•	•	•	•	•	•	•	•	•	•	•
Adverse Events	•	•	•	•	•	•	•	•	•	•	•
STUDY DRUG											
Randomization			•								
Subcutaneous study drug injection			•			•	•				

(a) Screening visit 2 must be performed at least 7 days prior to Day 1 visit. Eligible subjects according to safety assessments will be provided with the Actigraphy and ePro devices.

(b) In case of indeterminate result for TB test, the test should be repeated at Scr 2 in the local laboratory (only one retest is allowed). If the test is still indeterminate, the subject must not be included in the study

(c) Blood samples for hematology and biochemistry must be performed in fasting conditions

(d) In case elevated CPK > 1.5 ULN at screening 1, the test will be repetead at screening 2. If the result is still abnormal, the subject will not be included.V7/W8 CPK applicable only to France

(e) Only for female of childbearing potential. Serum pregnancy test to be performed at screening visit 1 and urine pregnancy test (UPT) for all other visits. UPT should have a sensitivity threshold of less than 25mIU/mL. There should be at least 14 days between the serum pregnancy test at screening and the UPT at baseline.

(f) Respiratory assessments for all subjects, but PEF will be measured only for subjects with medical history of asthma

Study period	Screening period (Day -28 to Day-1)		Treatment period						Observation period/ Follow up		Early termination (E/T) or Unscheduled visit (if applicable)
Visit	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	
Week	Screening 1	screening 2	Baseline	W1	W2	W4	W8	W12	W16	W18	
Day	D-28 to D-8	Day-7 (a)	Day 1	Day 8	Day 15	Day 29	Day 57	Day 85	D113	Day 126	
Visit window				±1 day	±2 days	±2 days	±2 days	±5 days	±5 days	±7 days	
EFFICACY/ PATIENTS REPORTED OUTCOMES											
NRS (g)	•	Once daily by the subject at home in the evening until Day 126									•
VRS (g)		Once daily by the subject at home in the evening until Day 126									•
Sleep disturbance NRS(h)		Once daily by the subject at home in the morning until Day 29									
DPS (i)			•			•					
DLQI			•			•		•			•
Prurigo Assessment Scale (PAS)			•			•	•	•		•	•
IGA			•			•	•	•		•	•
PK AND ADA ASSESSMENTS											
Blood sample for PK			•	•	•	•	•	•	•	•	•
Blood sample for ADA			•			•	•	•	•	•	•
PHARMACODYNAMIC ASSESSMENTS											
Blood sample for biomarkers			•			•		•			
D-Squames samples			•			•		•			
4-mm skin biopsies (j)			•					•			
Control of biopsies healing(k)					•				•		
BIOPHYSICAL ASSESSMENTS											
Photos of whole body (l)			•			•	•	•		•	
Actigraphy		Every day until Day 29									
Exit Form										•	

(g) NRS and VRS must be performed by the subject at home using the device provided by the site once daily in the evening from screening 2 to the last visit.

(h) NRS for sleep disturbance must be performed by the subject at home using the device provided by the site once daily in the morning from screening 2 to Day 29

(i) DPS must be performed by the subject at home following instruction provided by the Nurse on site and then subject should record DPS 24, 48 & 72 hrs after the 1st injection and at week 4 before the injection.

(j) Skin biopsies will be performed on lesional and non lesional skin at baseline and on lesional skin at week 12 only for patients consenting to have skin biopsies.

(k) When applicable, biopsies healing should be checked 2 weeks after skin biopsies

(l) Only for selected equipped sites – at least a Baseline & Week 12; Week 4, week 8 & week 18 if possible.

Appendix 2 Criteria for potentially clinically significant vital signs and weight

Pulse rate

≤50 bpm and decrease from baseline ≥20 bpm

≥120 bpm and increase from baseline≥20 bpm

To be applied for all positions (including missing).

SBP

≤95 mmHg and decrease from baseline ≥20mmHg

≥160 mmHg and increase from baseline ≥20 mmHg

To be applied for all positions (including missing).

DBP

≤45 mmHg and decrease from baseline ≥10 mmHg

≥110 mmHg and increase from baseline ≥10 mmHg

To be applied for all positions (including missing).

Weight

≥5% increase from baseline

≥5% decrease from baseline

Appendix 3 Dermatology Life Quality Index (DLQI)

The DLQI questionnaire is designed for use in adults, i.e. subjects over the age of 16. It is self-explanatory and can be simply handed to the subject who is asked to fill it in without the need for detailed explanation. It is usually completed in one to two minutes.

Scoring
The scoring of each question is as follows:

Response	Score
Very much	scored 3
A lot	scored 2
A little	scored 1
Not at all	scored 0
Not relevant	scored 0
Question unanswered	scored 0
Question 7: "prevented work or studying"	scored 3

The DLQI is calculated by summing the score of each question resulting in a maximum of 30 and a minimum of 0. The higher the score, the more quality of life is impaired. The DLQI can also be expressed as a percentage of the maximum possible score of 30.

****Please Note:** That the scores associated with the different answers should not be printed on the DLQI itself, as this might cause bias****

Meaning of DLQI Scores
0-1 = no effect at all on subject's life
2-5 = small effect on subject's life
6-10 = moderate effect on subject's life

11-20 = very large effect on subject's life
21-30 = extremely large effect on subject's life

Detailed analysis of the DLQI
The DLQI can be analyzed under six headings as follows:

Section	Questions	Score
Symptoms and feelings	Questions 1 and 2	Score maximum 6
Daily activities	Questions 3 and 4	Score maximum 6
Leisure	Questions 5 and 6	Score maximum 6
Work and School	Question 7	Score maximum 3
Personal relationships	Questions 8 and 9	Score maximum 6
Treatment	Question 10	Score maximum 3

The scores for each of these sections can also be expressed as a percentage of either 6 or 3.

Interpretation of incorrectly completed questionnaires

There is a very high success rate of accurate completion of the DLQI. However, sometimes subjects do make mistakes.

- If one question is left unanswered this is scored 0 and the scores are summed and expressed as usual out of a maximum of 30.
- If two or more questions are left unanswered the questionnaire is not scored.
- If question 7 is answered 'yes' this is scored 3. If question 7 is answered 'no' but then either 'a lot' or 'a little' is ticked this is then scored 2 or 1. If "Not relevant" is ticked, the score for Question 7 is 0. If it is answered 'no', but the second half is left incomplete, the score will remain 0.
- If two or more response options are ticked, the response option with the highest score should be recorded.
- If there is a response between two tick boxes, the lower of the two score options should be recorded.

- The DLQI can be analyzed by calculating the score for each of its six sub-scales (see above). When using sub-scales, if the answer to one question in a sub-scale is missing, that sub-scale should not be scored.

Minimal Clinically Important Difference of the DLQI

For **general inflammatory skin conditions** a change in DLQI score of at least 4 points is considered clinically important (Basra et al, 2015). This means that a subject's DLQI score has to either increase or decrease by at least 4 points in order to suggest that there has actually been a meaningful change in that subject's quality of life since the prior measurement of his/her DLQI scores (Refer to **Basra MK et al. paper).**

Appendix 4 Validation of the NRS responder analysis

To assess the appropriateness of responder thresholds in the NRS responder analysis (an improvement of weekly average pruritus peak NRS ≥4), an approach combines an anchor-based approach and ROC curves will be performed to identify the responder thresholds that best predict classification based on an external criterion. First, an external anchor or another external criterion that is appropriate to identify responders (e.g., Pruritus PCS) is used to classify all subjects into responder (at least 1 grade improvement from baseline in PCS) or non-responder groups. Then, to check for appropriateness, the relationship between the PCS criterion and the NRS measure (change from baseline in weekly average pruritus peak NRS) is examined through correlation analyses. Finally, the predictive accuracy of how well the specific NRS change scores relate to the classifications is evaluated using logistic regression analyses and graphically displayed as an ROC curve. Each point on the curve provides the sensitivity (true positives) versus one minus the specificity (false positives) trade-off for identifying responders for a specific unit of change on the NRS measure. The entire range of change scores and the likelihood classifications are provided within the one curve. A diagonal line is included in the figure as a reference. Evaluations that produce curves close to the diagonal line should be viewed with caution because the diagonal line represents chance classification.

The area under curve (AUC) and its 95% CI will be presented, along with percent concordant and discordant.

The analyses described above will be performed using all data across all visits and the data at week 24.

The following cumulative distribution function (CDF) plots will be produced to illustrate the between treatment group differences and within-subject change from baseline to Week 24.

- CDF plot illustrating the treatment differences from baseline to Week 24
- CDF plot of peak pruritus change score from baseline to Week 24 for all subjects by the different PCS response options at Week 24
- CDF plot of peak pruritus change score from baseline to Week 24 for all subjects by the PCS change scores (differences) between baseline and Week 24 (e.g. ± 3, ± 2, ± 1, and 0 point change)

In addition, the mean change in peak pruritus NRS from baseline to Week 24 will be calculated in the following subjects:

- An improvement of ≥ 1 in PCS from baseline to week 24
- 50-74%, 75-89%, or 90-100% improvement in EASI from baseline to week 24
- An improvement of ≥ 2 in IGA scale from baseline to Week 24
- An IGA of 0 or 1 at Week 24

Appendix 5 Potentially clinically significant values (PCSV) for clinical laboratory tests

Parameter	SI Units	Conventional Units	PCSV
WBC	x10e9/L	x10e3/µL	<3.0 Giga/L & ≥3.0G/L at Baseline (non-Black) <2.0 Giga/L & ≥2.0G/L at Baseline (Black) ≥16.0 G/L & <16G/L at Baseline
Neutrophils ABS	x10e9/L	x10e3/µL	<1.5 Giga/L & ≥1.5G/L at Baseline (non-Black) <1.0 Giga/L & ≥1.0G/L at Baseline (Black)
Eosinophils ABS	x10e9/L	x10e3/µL	(>0.5 Giga/L and >ULN) and (≤0.5 Giga/L or ≤ULN at Baseline)
Basophils ABS			>0.1 Giga/L and ≤0.1 Giga/L at Baseline
Monocytes ABS			>0.7 Giga/L and ≤0.7 Giga/L at Baseline
Lymphocytes ABS	x10e9/L	x10e3/µL	>4.0 Giga/L and ≤4.0 Giga/L at Baseline <0.5 Giga/L and ≥0.5 Giga/L at baseline
Hemoglobin *Female *Male	g/L	g/dL	≤100 g/L and Baseline >100 g/L for Male ; ≤ 90 g/L and Baseline >90 g/L for Female ≥200 g/L and baseline <200 g/L for Male; ≥180 g/L and Baseline <180 g/L for Female
Platelets	x10e9/L	x10e3/µL	<100 Giga/L and ≥100G/L at Baseline ≥700 G/L & <700G/L at Baseline
Sodium	mmol/L	mmol/L	≤130mmol/L and Baseline >130 ≥155 mmol/L and Baseline <155
Potassium	mmol/L	mmol/L	<3mmol/L and Baseline ≥3 >6.0 mmol/L and Baseline ≤6.0
Calcium (Total) Corrected serum calcium	mmol/L	mg/dL	<7 mg/dL and Baseline ≥7 mg/dL >12.5 mg/dL and Baseline ≤12.5 mg/dL
Glucose	mmol/L	mg/dL	<2.2mmol/L and ≥2.2 mmol/L at Baseline

Parameter	SI Units	Conventional Units	PCSV
			>13.9mmol/L and ≤13.9mmol/L at Baseline
Creatinine	µmol/L	mg/dL	≥150µmol/L (Adults) and Baseline <150µmol/L
AST	U/L		>3 ULN and Baseline ≤3 ULN
ALT	U/L		>3 ULN and baseline ≤3 ULN
Alkaline phosphatase	U/L		>2.5 ULN and Baseline ≤2.5 ULN
Bilirubin total	µmol/L	mg/dL	>2.0 ULN and Baseline ≤2.0 ULN
CPK	U/L		>3.0 ULN and Baseline ≤3.0 ULN
Fibrinogen	g/L	mg/dL	Not provided
GGT	U/L		>2.5 ULN and Baseline ≤2.5 ULN
Albumin	g/L	g/dL	≤30 g/L and Baseline >30g/L
Uric acid *Male16-59 *Female>16 *Male >60	mmol/L	mg/dL	>600 µmol/L and ≤600 µmol/L at Baseline for Male; >500 µmol/L and ≤500 µmol/L at Baseline for Female
Cholesterol total	mmol/L	mg/dL	>10.34 mmol/L and ≤10.34 mmol/L at Baseline
Triglycerides	mmol/L	mg/dL	>5.7 mmol/L and ≤5.7 mmol/L at Baseline

PAREXEL International Electronic Signature Page

This page is the manifestation of the electronic signature(s) used in compliance with PAREXEL International's electronic signature policies and procedures and in compliance with applicable regulations.

UserName: [redacted]
Title: [redacted]
Date: [redacted] 08:09 PM GMT Standard Time
Meaning: Document contents approved.
==========