NOVARTIS

Clinical Development

AIN457/Secukinumab®

Clinical Trial Protocol CAIN457P12302 / NCT04209205

A randomized, double-blind, placebo-controlled, parallel group, phase III multicenter study of intravenous secukinumab to compare efficacy at 16 weeks with placebo and to assess safety and tolerability up to 52 weeks in subjects with active Psoriatic Arthritis

Statistical Analysis Plan (SAP)

Author:	Statistician, [redacted]

Document type:	SAP Documentation
Document status:	Final
Release date:	03-May-2022
Number of pages:	45

Table of contents

List of tables

List of figures

Document History-Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
03-May-2022	Prior to DB lock	NA	N/A - First version	NA

List of abbreviations

ACR	American College of Rheumatology
AE	adverse event
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
b.i.d.	twice a day
[redacted]	[redacted]
BDR	Bioanalytical Data Report
BMI	Body Mass Index
BSA	Body surface area
BSL	Baseline
CFR	Code of Federal Regulation
C_{max}	Maximum concentration
$C_{max,ss}$	Maximum concentration steady-state
C_{min}	Minimum concentration
$C_{min,ss}$	Minimum concentration steady-state
CMO&PS	Chief Medical Office and Patient Safety
COAs	Clinical Outcome Assessments
CRF	Case Report/Record Form (paper or electronic)
CRO	Contract Research Organization
CRP	C-Reactive Protein
CSR	Clinical study report
CTC	Common Toxicity Criteria
DMARD	Disease Modifying Anti-rheumatic Drug
DSUR	Development Safety Update Report
ECG	Electrocardiogram
EDC	Electronic Data Capture
ELISA	Enzyme-linked immunosorbent assay
EMA/EMEA	European Medical Agency
ESR	Erythrocyte sedimentation rate
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy-Fatigue
FDA	Food and Drug Administration
FSH	Follicle-stimulating hormone
GCP	Good Clinical Practice
GGT	gamma glutamyl transferase
GWA	genome-wide association
HAQ-DI	Health Assessment Questionnaire – Disability Index
hCG	Human chorionic gonadotropin
HDL	High-density lipoprotein
HIV	human immunodeficiency virus

hsCRP	high sensitivity C-Reactive Protein
i.v.	intravenous
IB	investigator brochure
IBD	inflammatory bowel disease
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
IEC	Independent Ethics Committee
IRB	Institutional Review Board
IRT	Interactive Response Technology
IVR	Interactive Voice Response
LDI	Leeds Dactylitis Index
LDL	Low-density lipoprotein
LEI	Leeds Enthesitis Index
LFT	Liver function test
LLN	Lower limit of normal
LLQ	Lower limit of quantification
MCR	Major Clinical Response
MCS	Mental Component Summary
MDA	minimal disease activity
MedDRA	Medical dictionary for regulatory activities
mg	milligram(s)
mL	milliliter(s)
MMP	Matrix Metalloproteinase
MMRM	Mixed effect Repeated Measures model
mNAPSI	Modified Nail Psoriasis Severity Index
MRI	Magnetic Resonance Imaging
MTX	Methotrexate
NSAID	Non-Steroidal Anti-Inflammatory Drug
o.d.	once a day
p.o.	oral
PA	posteroanterior
PASDAS	Psoriatic Arthritis Disease Activity Score
PASI	Psoriasis Area and Severity Index
PCS	Physical Component Summary
[redacted]	[redacted]
[redacted]	[redacted]
PPD	Purified Protein Derivative
PRO	Patient Reported Outcome
PsA	Psoriatic arthritis
QMS	Quality Management System
RBC	red blood cell(s)
RF	Rheumatoid factor

s.c.	subcutaneous
SAE	serious adverse event
SCR	serum creatinine
SD	standard deviation
SGOT	serum glutamic oxaloacetic transaminase
SGPT	serum glutamic pyruvic transaminase
SJC	Swollen Joint Count
SNP	single nucleotide polymorphism
[redacted]	[redacted]
SST	Serum separator tube
SUSAR	Suspected Unexpected Serious Adverse Reactions
TBL	total bilirubin
TFQ	Trial feedback Questionnaire
TJC	Tender Joint Count
TNF/TNFα	Tumor Necrosis Factor
TNF-IR	TNFα Inhibitor Incomplete Responder
ULN	upper limit of normal
UV	ultraviolet
VAS	Visual Analog Scale
WBC	white blood cell(s)
WHO	World Health Organization
[redacted]	[redacted]

1 Introduction

Data will be analyzed by Novartis according to the data analysis section 12 of the clinical study protocol. The statistical methodology is described below and any deviations from the protocol are documented. Additional detailed information regarding the analysis methodology is contained in the Appendix section.

This version of SAP covers the statistical and analytical plans of the CAIN457P12302 trial regarding the final DBL.As this is the Final DBL, analysis will only be done for long-term efficacy and safety. Analyses pertaining to Week 16 will not be repeated. Tables and listing for demographics, baseline disease characteristic, prior medication and medical history will not be reproduced. Any analyses related to the testing hierarchy will not be reproduced.

1.1 Study Design

This multicenter study uses a randomized, double-blind, placebo-controlled, parallel-group design to study the efficacy, safety and tolerability of treatment with intravenous secukinumab in patients with active PsA.

The study population comprises approximately 380 patients with active PsA, despite current or previous NSAID, DMARD and / or TNF inhibitor therapy or intolerance to these therapies.

At baseline, patients will be randomized to one of the two treatment groups in a 1:1 randomization:

- **Group 1:** approximately 190 patients; These patients will receive secukinumab 6 mg/kg i.v. at Baseline (BSL), followed by the administration of secukinumab 3 mg/kg i.v. every four weeks starting at Week 4.
- **Group 2:** approximately 190 patients; These patients will receive i.v. placebo at BSL, Weeks 4, 8, and 12, followed by the administration of secukinumab 3 mg/kg i.v. every four weeks starting at Week 16.

This study will consist of 4 periods: a screening period (up to 10 weeks), treatment period 1 (total duration of 16 weeks) and treatment period 2 (total duration of 36 weeks) followed by a safety follow up period of 8 weeks after the end of treatment visit (i.e., Week 52).

To ensure a balance across both arms, patients previously treated with TNF-inhibitors will be stratified at randomization. No more than 25% of previously treated TNF-inhibitor patients will be enrolled in the study, with this cutoff applied to each group at randomization (no more than 48 patients per group).

Because PsA is considered a chronic disease with no ‘true’ and lasting placebo response, all patients, including those on placebo, will switch to open-label i.v. secukinumab at Week 16. However, all patients and investigators/ site staff will remain blinded to the original randomized treatment group assignment (i.v. secukinumab treatment or placebo).

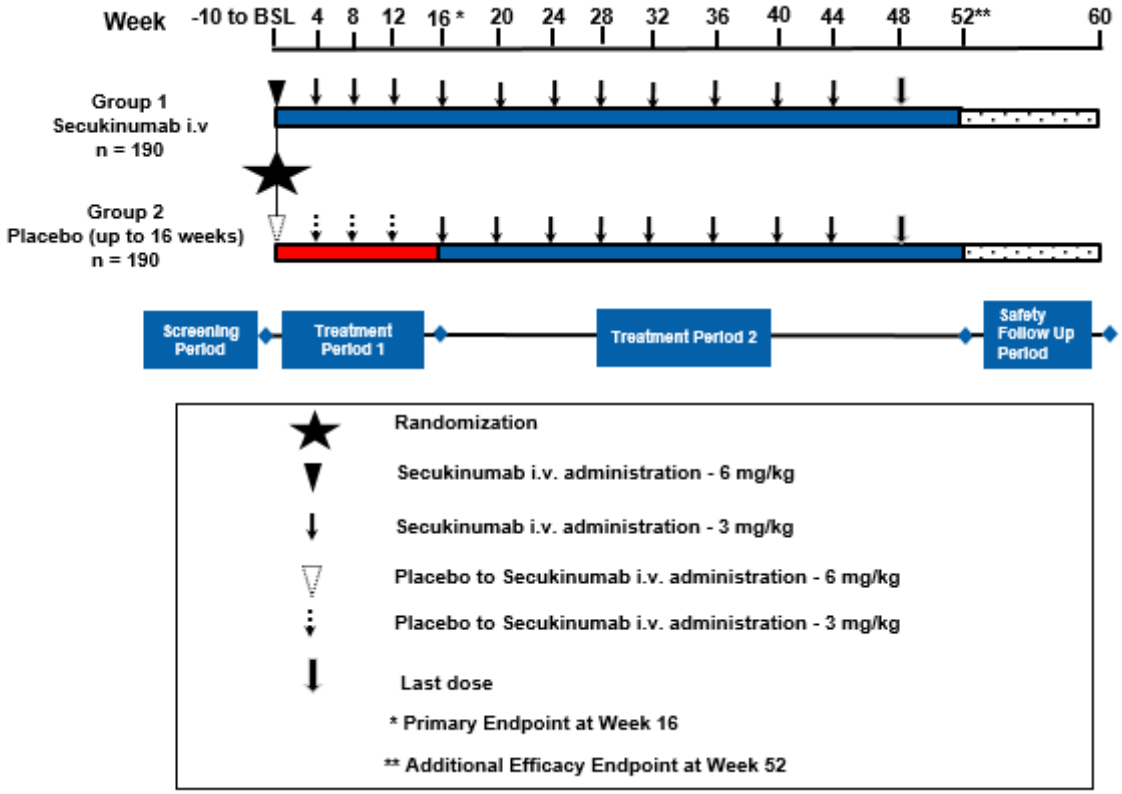
An end of treatment visit (Week 52) will be done 4 weeks after last study treatment administration, and a post treatment follow-up visit (Week 60) is to be done 8 weeks after end of the treatment visit for all subjects (regardless of whether they complete the entire study as planned or discontinue prematurely).

All intravenous infusions will be performed at the study site and site personnel will administer the infusions to subjects.

Rescue medication is not allowed until Week 16. However, subjects who are deemed by the investigator not to be benefiting from the study treatment based on safety and efficacy assessments or for any reason of their own accord will be free to discontinue participation in the study at any time.

The study will have a primary endpoint at Week 16. A Primary endpoint analysis will be performed with Week 16 data (last patient completing Treatment Period 1). Long-term efficacy and safety assessments will be performed up to Week 52.

Figure 1-1 Study Design



1.2 Study Objectives and Endpoints

1.2.1 Primary Objective

To demonstrate that the efficacy of i.v. secukinumab at Week 16 is superior to placebo in subjects with active psoriatic arthritis (PsA) based on the proportion of patients achieving an American College of Rheumatology 50 (ACR50) response.

1.2.2 Secondary Objectives

To evaluate:

- The efficacy of i.v. secukinumab at Week 16 is superior to placebo based on the proportion of subjects achieving an ACR20 response.
- The efficacy of i.v secukinumab at Week 16 is superior to placebo based on the proportion of patients achieving Minimal Disease Activity MDA 5/7.
- The efficacy of i.v. secukinumab at Week 16 is superior to placebo based on the proportion of subjects achieving a PASI90 response in the subgroup of subjects who have ≥3% skin involvement with psoriasis.
- The improvement (change) from baseline on i.v. secukinumab is superior to placebo for the PASDAS at Week 16.
- The improvement (change) from baseline on i.v. secukinumab is superior to placebo for the HAQ-DI at Week 16.
- The improvement (change) from baseline on i.v. secukinumab is superior to placebo for the SF36-PCS at Week 16.
- The improvement (change) from baseline on i.v. secukinumab is superior to placebo for the FACIT-fatigue at Week 16.
- The improvement (change) from baseline on i.v. secukinumab is superior to placebo for the mNAPSI at Week 16 for the subgroup of patients with nail involvement.
- The efficacy of i.v. secukinumab at Week 16 is superior to placebo based on the proportion of subjects with resolution of dactylitis by the Leeds Dactylytis Index in the subset of subjects who have dactylitis at baseline.
- The efficacy of i.v. secukinumab at Week 16 is superior to placebo based on the proportion of subjects with resolution of enthesitis by the Leeds Enthesitis Index in the subset of subjects who have enthesitis at baseline.
- The overall safety and tolerability of i.v. secukinumab compared to placebo as assessed by vital signs, clinical laboratory values, and adverse events monitoring.

2 Statistical Methods

2.1 Data Analysis General Information

Summary statistics for continuous variables include N, mean, standard deviation, minimum, lower quartile, median, upper quartile, and maximum. For binary or discrete variables the absolute number of subjects in each category and relative frequencies will be provided.

Unless otherwise specified, p-values will be presented as 2-sided p-values and the type I error rate (alpha) will be 5%.

Inferential efficacy comparisons with placebo will generally focus on the first 16 weeks of treatment unless otherwise specified.

Efficacy and safety data for the placebo-controlled period (or the entire treatment period as appropriate) will be presented by treatment groups. Subjects may be included in more than one treatment group for some analyses (e.g. exposure-adjusted AEs over the entire treatment period).

Note that the treatment groups for a subject may differ depending on the time period of the analysis and whether one assesses the subject for efficacy or safety.

Data may also be presented by a combination of the ‘original’ and ‘switch’ treatment groups. These treatment groups represent the treatment combinations the subjects experience over the course of the entire trial.

Any data analysis carried out independently by the investigator should be submitted to Novartis before publication or presentation.

Efficacy data following treatment switch

Data will also be presented after Week 16, by a combination of the ‘original’ and ‘switch’ treatment groups and will be referred to as treatment sequence. These treatment sequences represent the treatment combinations the subjects experience over the course of the entire trial

in case of treatment switch (e.g., placebo patients who are reassigned to i.v. secukinumab at Week 16).

All listings will be presented by treatment sequence.

2.2 Analysis Sets and Treatment Groups

The following analysis sets will be used in this trial:

Randomized set: The randomized set will be defined as all subjects who were randomized. Unless otherwise specified, mis-randomized subjects (mis-randomized in Interactive Voice Response (IVR) will be excluded from the randomized set.

Mis-randomized subjects are defined as those subjects who were mistakenly randomized into the IVR prior to the site confirming all eligibility criteria had been met and to whom no study medication was given. Mis-randomized subjects are treated as screen failures.

Full analysis set (FAS): The FAS will be comprised of all subjects from the randomized set to whom study treatment has been assigned. Following the intent-to-treat principle, subjects will be analyzed according to the treatment assigned to at randomization, but with actual anti-TNF status.

Dactylitis subset: The dactylitis subset will include all FAS subjects who have dactylitis at baseline.

Enthesitis subset (LEI): The enthesitis subset will include all FAS subjects who have enthesitis based on LEI at baseline.

[redacted]

Psoriasis subset: The psoriasis subset will include all FAS subject who have >= 3% of the body surface area (BSA) affected by psoriatic skin involvement at baseline.

Nail subset: The nail subset will include all FAS subject who have psoriasis currently in nails at baseline.

Safety set: The safety set includes all subjects who took at least one dose of study treatment during the treatment period. Subjects will be evaluated according to treatment received.

2.2.1 Subgroup of Interest

The primary endpoint(s) and secondary endpoints will be evaluated for TNF-alpha inhibitor status. The primary and some secondary endpoints will be also evaluated for MTX status.

2.2.2 Treatment Groups

For Final DBL analysis, the summaries will be performed by treatment sequence. For some safety summaries (e.g. exposure-adjusted) the 'switch' treatment may be summarized separately.

Randomized treatment:

- AIN457 6mg/kg - 3mg/kg
- Placebo

Treatment sequence:

- AIN457 6mg/kg - 3mg/kg
    - Stay with AIN457 3mg/kg-IV
- Placebo –AIN457 3mg/kg-IV
    - Switch to AIN457 3mg/kg-IV

Additional sequences could be reported such as Any AIN457-IV if applicable.

2.3 Patient Disposition, Demographics and Other Baseline Characteristics

2.3.1 Patient Disposition

The number and percentage of subjects in the randomized set who completed the study periods and who discontinued the study prematurely (including the reason for discontinuation) will be presented at the end of each treatment period (e.g. Week 16, and entire treatment period), if appropriate, for each treatment group and all subjects.

For each protocol deviation (PD), the number and percentage of subjects for whom the PD applies will be tabulated.

2.3.2 Background and Demographic Characteristics

Data on background and demographic characteristics have been summarized in Week 16 Interim Analysis (IA) report. No such tables will be presented at Final DBL analysis.

2.3.3 Medical History

Medical history has been summarized in Week 16 IA report, so no such summary table or listing will be repeated at Final DBL analysis.

2.4 Treatments (Study Treatment, Concomitant Therapies, Compliance)

2.4.1 Study Treatment / Compliance

The analysis of study treatment data will be based on the safety set. The duration of exposure to study treatment will also be summarized by treatment group. In addition, the number and percentage of subjects with cumulative exposure levels (e.g. any exposure, $\geq$ 4 weeks, $\geq$ 8 weeks, etc.) will be presented.

Duration of exposure of a treatment will be defined as the time from first dose of the treatment to the time of treatment switch (for subjects who switch treatment) or minimum of (last dose of the treatment + 84 days) and (last visit date). Patients who switch treatment during the study (e.g. from placebo to active treatment) will have exposure to both medications/doses using the appropriate start and stop dates.

Duration of exposure (years) = duration of exposure (days) / 365.25

Duration of exposure (100 subject years) = duration of exposure (years) / 100

The analyses of duration of exposure described above will be done for the entire study treatment period.

2.4.2 Prior, Concomitant and Post Therapies

Prior medications have been summarized in Week 16 IA report. No summary tables or listings will be repeated at the Final DBL.

Any medication given at least once between the day of first dose of study treatment and the date of within 84 days after last dose will be a concomitant medication, including those which were started pre-baseline and continued into the period where study treatment is administered.

Concomitant medications will be presented by Anatomical Therapeutic Classification (ATC) codes and grouped by anatomical main group. Tables will show the overall number and percentage of subjects receiving at least one treatment of a particular ATC code and at least one treatment in a particular anatomical main group.

Concomitant surgeries and procedures will be summarized by primary system organ class and MedDRA preferred term.

The number and percentage of subjects receiving concomitant psoriatic arthritis therapy will be presented by treatment group as well as the reasons for stopping their therapies (primary lack of efficacy, secondary lack of efficacy, lack of tolerability, other).

2.5 Efficacy Evaluation

2.5.1 Description of Efficacy Variables

ACR 20/50/70

The ACR response will be used to determine efficacy. A subject is defined as an ACR 50 responder if, and only if, the following three conditions hold:

≥ 50% improvement in the number of tender joints (based on 78 joints)

≥ 50% improvement in the number of swollen joints (based on 76 joints)

≥ 50% improvement in three of the following five measures:
Patient's global assessment of disease activity (measured on a VAS scale, 0-100)
Physician's global assessment of disease activity (measured on a VAS scale, 0-100)
Patient's assessment of PsA pain (measured on a VAS scale, 0-100)
Health Assessment Questionnaire- Disability Index (HAQ-DI©) score
Acute phase reactant (hsCRP or ESR)

ACR20 = 20 % improvement in at least 3 of the 5 measures and 20 % improvement in the swollen and tender joint count.

ACR50 = 50 % improvement in at least 3 of the 5 measures and 50 % improvement in the swollen and tender joint count.

[redacted]

ACRn = min (x_1, x_2, x_3), where

x_1 = % improvement from baseline in tender 78-joint count

x_2 = % improvement from baseline in swollen 76-joint count

and x_3 = 3rd largest value of x_4, x_5, x_6, x_7, x_8 where,

x_4 = % improvement from baseline in Patient's assessment of PsA pain (VAS 100 mm)

x_5 = % improvement from baseline in Patient's global assessment of disease activity (VAS 100 mm)

x_6 = % improvement from baseline in Physician's global assessment of PsA disease activity (VAS 100 mm)

x_7 = % improvement from baseline in Patient self-assessed disability (Health Assessment Questionnaire [HAQ©] score)

x_8 = % improvement from baseline in Acute phase reactant (C-reactive protein [hsCRP]) or Erythrocyte sedimentation rate (ESR)

ACRn can be computed even if up to two values of x_4, x_5, x_6, x_7, x_8 are missing. ACRn, theoretically, cannot be computed, if one or both of x_1, x_2 is/are missing OR more than three values of x_4, x_5, x_6, x_7, x_8 are missing.

Health Assessment Questionnaire - Disability Index (HAQ-DI)

The HAQ-DI© (Fries et al 1980) is a validated measure of physical disability and functional status. It has four dimensions: disability, pain, drug side effects and dollar costs, although, the latter three are rarely used in clinical trials. In this trial only the disability dimension will be used. The disability dimension consists of 20 multiple choice items concerning difficulty in performing eight common activities of daily living; dressing and grooming, arising, eating, walking, reaching, personal hygiene, gripping and activities. Subjects choose from four response categories, ranging from 'without any difficulty' to 'unable to do'. The ACR Rheumatology Committee on Outcome Measures in RA recommends the use of this questionnaire in clinical trials.

Scoring of the HAQ-DI©

The HAQ-DI© will be scored in accordance with the recommendation from the developers outlined in the "HAQ PACK" from Stanford University, California.

The following coding is to be used for the 8 categories of the disability outcome dimension:

Without ANY Difficulty	0
With SOME Difficulty	1
With MUCH Difficulty	2
UNABLE to do	3

Within each of the 8 categories only the item indicating the most severe impairment contributes to the category score. If the subject requires the use of aids, devices, or help from another to accomplish any of the activities in an associated category, then the score for that category will be assigned the value 2, unless the score is already 3 (i.e. scores of 0 or 1 are increased to 2).

Associated categories are defined in the "HAQ PACK". From the scores for each category a Standard Disability Index (SDI) is computed by summing the computed scores for each category and dividing by the number of categories answered. The SDI is not computed if the subject does not have scores for at least 6 categories. This SDI is the HAQ-DI© score, which will be used in the statistical analyses of this instrument. The range for this score is (0, 3).

HAQ-DI response is defined by an improvement of at least 0.35 score points compared to baseline.

Minimal disease activity

A subject will be considered a responder of minimal disease activity (MDA, see Coates 2010) if he/she achieves at least 5 of the following 7 items:

≤ 1 tender joint count

≤ 1 swollen joint count

PASI ≤ 1 or body surface area (BSA) ≤3%

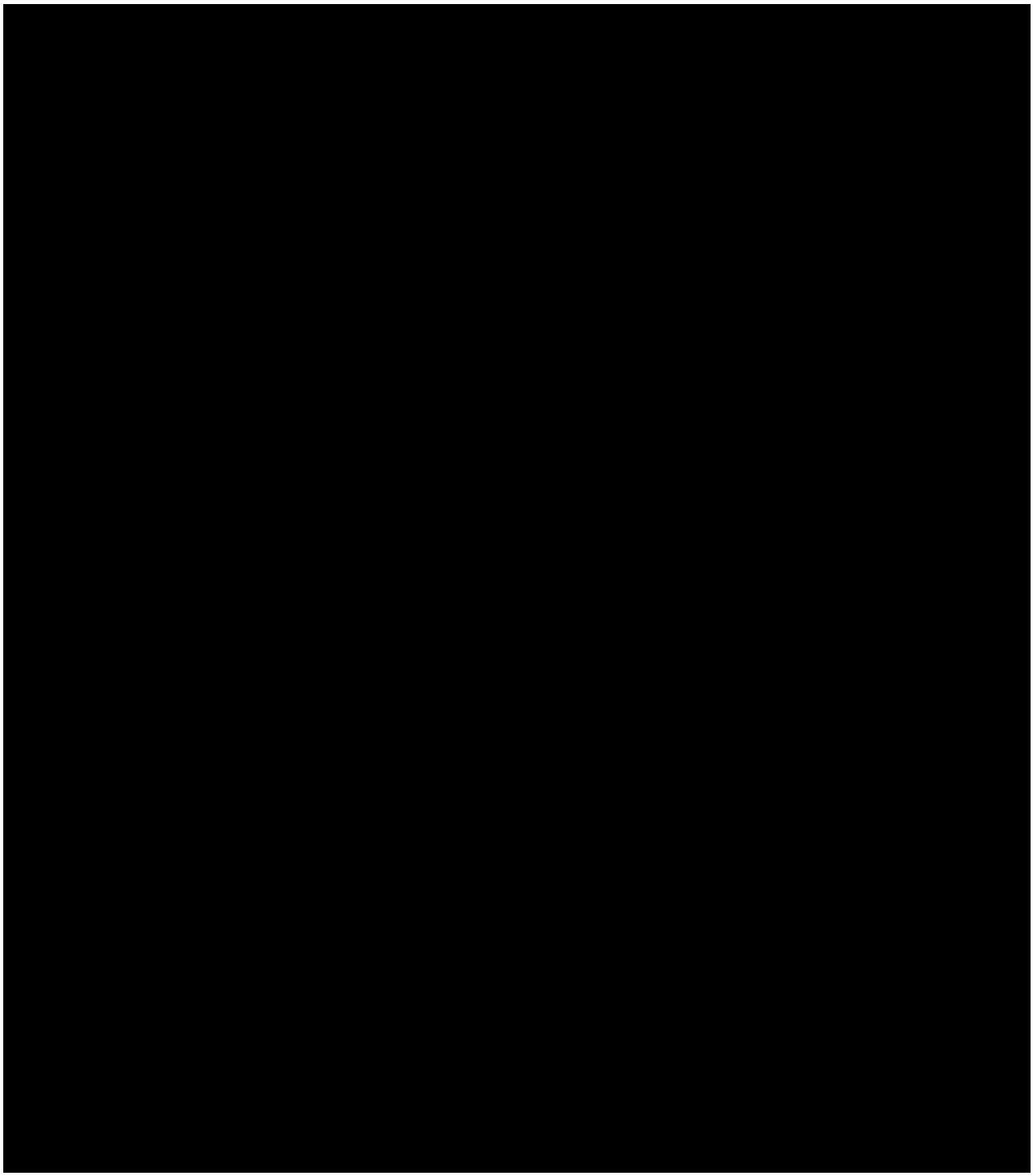
Patient's assessment of PsA pain VAS ≤ 15

Patient's global assessment of disease activity VAS ≤ 20

HAQ-DI ≤ 0.5

tender entheseal points ≤ 1

[redacted]



Leeds Enthesitis Index (LEI)

LEI is a validated enthesitis index that uses 6 sites for evaluation of enthesitis: lateral epicondyle humerus L + R, proximal achilles L + R and medial condyle femur L + R. Tenderness on examination is recorded as either present (1) or absent (0) for each of the 6 sites, for an overall score range of 0–6. Higher count represents greater enthesitis burden.

Resolution of enthesitis

If enthesitis is present with any of the 6 sites (lateral epicondyle humerus L + R, proximal achilles L + R and medial condyle femur L + R), the patient is counted as a patient with enthesitis. The resolution of enthesitis is calculated based on complete resolution of enthesitis as determined by the Leeds Enthesitis Index (LEI) in patients with enthesitis at baseline.

Leeds Dactylitis Index (LDI)

The LDI measures the ratio of the circumference of the affected digit to the circumference of the digit on the opposite hand or foot: a minimum difference of 10% is used to define a dactylitic digit. If both sides are considered involved, a table of normative values is used to provide the comparison. The ratio of circumference is multiplied by a tenderness score, originally based on the Ritchie index (graded 0–3), but a later modification amended this to a binary score (0 for nontender, 1 for tender) — this later modification is referred to as the LDI basic, and is adopted in this study. For each dactylitic digit, the final score is:

[{(A/B) - 1} * 100] * C,

where A is circumference of involved digit, B circumference of opposite (unaffected or from reference) and C is tenderness (0 or 1 in this case). The results from each digit with dactylitis are then summed to produce a final score. Only involved digits are assessed.

Dactylitis count

The dactylitis count is the number of fingers and toes with dactylitis based on the LDI, with a range of 0-20.

Resolution of dactylitis

If dactylitis according to the LDI is present with any finger or toe, the patient is counted as a patient with dactylitis. The resolution of dactylitis is calculated based on complete resolution of dactylitis as determined by the Leeds Dactylitis Index (LDI) in patients with dactylitis at baseline.

Psoriasis Area and Severity Index (PASI)

The PASI assessment will be conducted for subjects in whom at least 3% of the body surface area (BSA) was affected by psoriatic skin involvement at baseline (Psoriasis Subset). The PASI assesses the extent of psoriasis on four body surface areas (head, trunk and upper and lower limbs) and the degree of plaque erythema, scaling and thickness. A PASI score will be derived as indicated in Table 2-1. The head, trunk, upper limbs and lower limbs are assessed separately for erythema, thickening (plaque elevation, induration), and scaling (desquamation). The average degree of severity of each sign in each of the four body regions is assigned a score of 0-4. The area covered by lesions on each body region is estimated as a percentage of the total area of that particular body region. Further practical details help the assessment:

1. The neck is assessed as part of the head.
2. The axillae and groin are assessed as part of the trunk.
3. The buttocks are assessed as part of the lower limbs.
4. When scoring the severity of erythema, scales should not be removed.

Table 2-1 The PASI scoring system

Body region	Erythema (E)	Thickening (plaque elevation, induration, I)	Scaling (desquamation) (D)	Area score (based on true area %, A)*
Head (H)**	0=none 1=slight 2=moderate 3=severe 4=very severe	0=none 1=slight 2=moderate 3=severe 4=very severe	0=none 1=slight 2=moderate 3=severe 4=very severe	0 = no involvement 1 = >0-< 10% 2 = 10-<30% 3 = 30-<50% 4 = 50-<70% 5 = 70-<90% 6 = 90-100%
Trunk, (T)***	0=none 1=slight 2=moderate 3=severe 4=very severe	0=none 1=slight 2=moderate 3=severe 4=very severe	0=none 1=slight 2=moderate 3=severe 4=very severe	0 = no involvement 1 = >0-< 10% 2 = 10-<30% 3 = 30-<50% 4 = 50-<70% 5 = 70-<90% 6 = 90-100%
Upper limbs (U)	0=none 1=slight 2=moderate 3=severe 4=very severe	0=none 1=slight 2=moderate 3=severe 4=very severe	0=none 1=slight 2=moderate 3=severe 4=very severe	0 = no involvement 1 = >0-< 10% 2 = 10-<30% 3 = 30-<50% 4 = 50-<70% 5 = 70-<90% 6 = 90-100%
Lower limbs (L)****	0=none 1=slight 2=moderate 3=severe 4=very severe	0=none 1=slight 2=moderate 3=severe 4=very severe	0=none 1=slight 2=moderate 3=severe 4=very severe	0 = no involvement 1 = >0-< 10% 2 = 10-<30% 3 = 30-<50% 4 = 50-<70% 5 = 70-<90% 6 = 90-100%

Percentage (not score) of body region (not whole body) affected will be entered in the eCRF

**Neck is assessed as part of the Head (H) body region.

***Axillae and groin are assessed as part of the Trunk (T) body region.

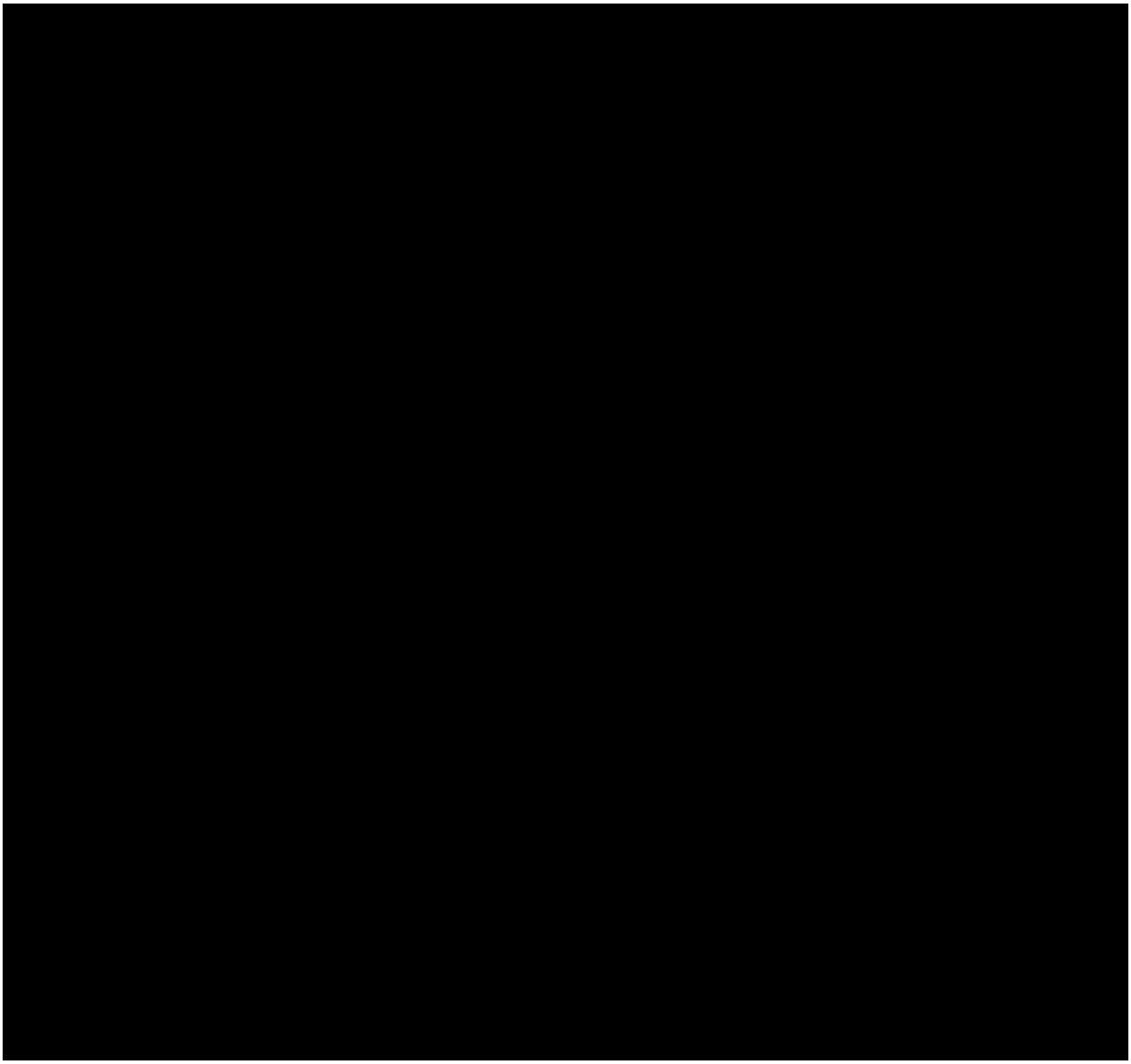
****Buttocks are assessed as part of the Lower limbs (L) body region.

Because the head and neck, upper limbs, trunk and lower limbs correspond to approximately 10%, 20%, 30% and 40% of the body surface area, respectively, the PASI score is calculated using the formula:

$$PASI = 0.1(E_H+I_H+D_H)A_H + 0.2(E_U+I_U+D_U)A_U + 0.3(E_T+I_T+D_T)A_T + 0.4(E_L+I_L+D_L)A_L$$

The keys for the letters are provided in Table 2-1.

PASI scores can range from a lower value of 0, corresponding to no signs of psoriasis, up to a theoretic maximum of 72.0. The total score comes from eCRF.



Modified Nail Psoriasis Severity Index (mNAPSI)

The mNAPSI is an instrument to assess psoriatic nail involvement in subjects with PsA and nail psoriasis. It will be collected only in subjects with psoriatic nail involvement. The modifications on the original NAPSI to create the mNAPSI were made by rheumatologists, with dermatologists' input, as a tool for clinical trials. The creators' goal was to develop a tool to assess disease severity and response to treatment in clinical trials, keeping in mind that the assessor in a clinical trial most likely would not be a trained dermatologist. The mNAPSI scores range from 0-130 for all finger nails. The total mNAPSI score will be calculated as the sum of all the scores from available nails.

Psoriatic Arthritis Disease Activity Score (PASDAS)

PASDAS is a new composite measure developed to access disease activity in Psoriasis (GRACE Project) (Helliwell 2012). It is calculated by utilizing seven measures; the seven components are: Patient reported measures (excluding mental component summary score (MCS) of the medical outcomes survey Short Form-36 (SF-36-PCS)), skin, peripheral joint counts (Tender and Swollen joint counts), Dactylitis (LDI), Enthesitis (LEI), acute phase response (CRP) and Patient & Physician global VAS scores.

PASDAS = (((0.18 x √Physician global VAS)

+ (0.159 x √Patient global VAS)

- (0.253 x √¯SF36-PCS)

+ (0.101 x LN (Swollen joint count + 1))

+ (0.048 x LN (Tender joint count + 1))

+ (0.23 x LN (Leeds Enthesitis Count + 1))

+ (0.377x LN (Dactylitis count + 1))

+ (0.102 x LN (CRP + 1)) +2)*1.5.

If any component measurement is missing, PASDAS will be missing. The patient global VAS defined in the PASDAS score is "VAS for PASDAS assessment" which measures patient's global assessment of PSORIASIS and ARTHRITIS disease activity. The Physician global VAS is the Physician's global assessment of PsA disease activity. Tender Dactylitis count is from "Leeds Dactylitis Instrument" CRF page with "Was dactylitis present?" = "Yes" and "Tender/non-tender" = "Tender".

66-swollen / 68-tender joint counts for PASDAS assessment:

There are only 66 swollen and 68 tender joint counts which will be used for PASDAS assessment. That is the 10 joints in CRF: CMC1, DIP2 (FOOT), DIP3 (FOOT), DIP4 (FOOT), and (DIP5 (FOOT) both left and right, should not be counted in the swollen/tender joint count in PASDAS formula.

If the number of joints for which data were available (e.g., T) is less than 66/68 for the swollen/tender joint assessment, the number of swollen/tender joints (e.g., t) will be scaled up proportionately (i.e., 66*t/T or 68*t/T for swollen or tender joint count).

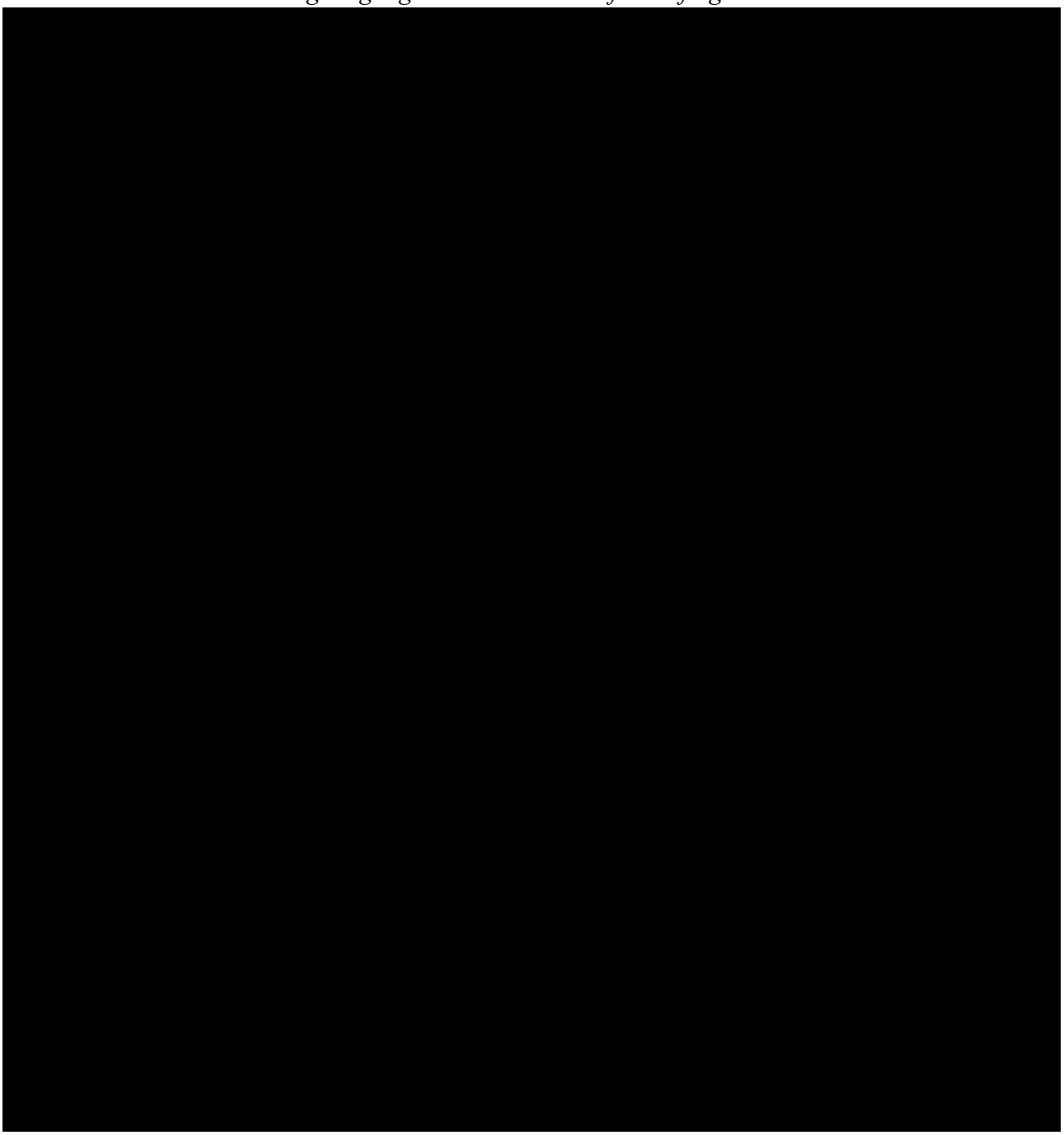
VAS for PASDAS assessment:

Global Disease Activity: The patient's assessment of PSORIASIS and ARTHRITIS will be performed using 100 mm VAS ranging from "Excellent" to "Poor" after the question *"Considering all the ways PSORIASIS and ARTHRITIS affects you, please indicate with a vertical mark (|) through the horizontal line how well you are doing over the past week".*

Physician's global assessment of fingernail disease severity (VAS)

The physician's assessment of nail disease activity will be performed using 100 mm VAS ranging from no disease activity to maximal disease activity, after the question *"After you have*

viewed all the fingernails of a subject, consider all aspects of the subject's fingernails and place a vertical line on the scale giving a global assessment of their fingernails".



2.5.2 Description of Health-related Quality of Life Endpoints

SF-36

The Short Form Health Survey (SF-36) is a widely used and extensively studied instrument to measure health-related quality of life among healthy subjects and patients with acute and chronic conditions. It consists of eight subscales that can be scored individually: Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health. Two overall summary scores, the Physical Component Summary (PCS) and the Mental Component Summary (MCS) also can be computed. The SF-36 has proven useful in monitoring general and specific populations, comparing the relative burden of different disease, differentiating the health benefits produced by different treatments, and in screening individual patients. The eight domains are based on a scale from 0-100 while PCS and MCS are norm-based scores with a mean of 50 and a standard deviation of 10.

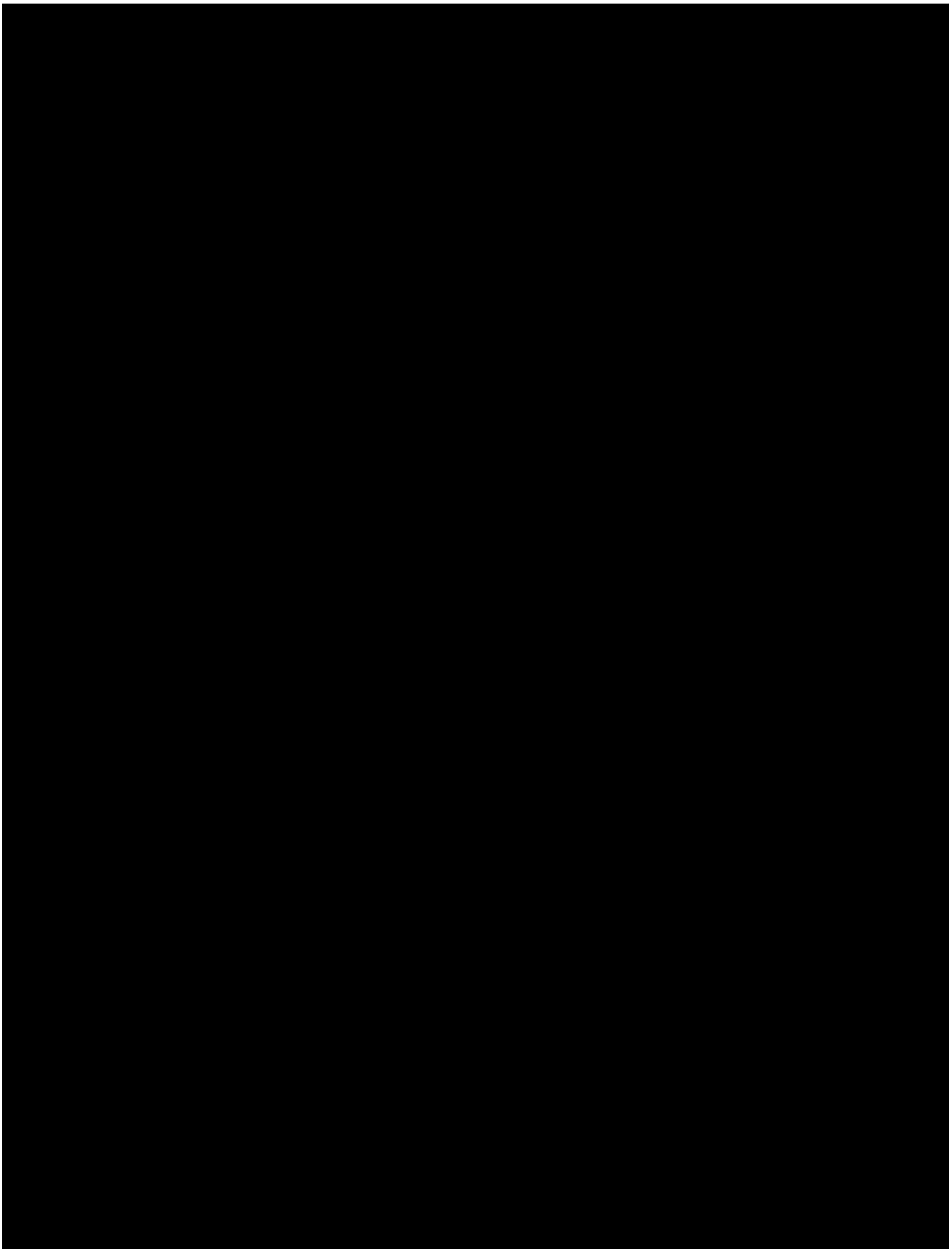
FACIT - Fatigue

The Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-Fatigue©) is a 13-item questionnaire that assesses self-reported fatigue and its impact upon daily activities and function.

Subjects respond to each item on a 5-point Likert-type scale (0 = not at all; 1 = a little bit; 2 = somewhat; 3 = quite a bit; 4 = very much) based on their experience of fatigue during the past 2 weeks. The scale score is computed by summing the item scores, after reversing those items that are worded in the negative direction. Numbering the questions from 1 to 13, it is evident that questions 7 and 8 are worded in the positive direction (4 indicates a desirable response) and all other questions are worded in the negative directions (4 indicates an undesirable response). Thus, it is necessary to reverse the responses for all but questions 7 and 8 (i.e. original response of 0 gets mapped to 4, 1=3, 2=2, 3=1, and 4=0) for scoring purposes.

When there are missing item scores, the subscale score was computed by summing the non-missing item scores, multiplying by 13 (the total number of items in the scale) and dividing by the number of non-missing items (i.e. normalizing the results). The latter rule applied only when at least half of the items (seven or more) are non-missing.

FACIT Fatigue subscale scores range from 0 to 52, where higher scores represent less fatigue (Cella D et al., 2004).



[redacted]

2.5.3 Handling of missing data

Missing data

Missing data for primary and key secondary binary efficacy variables for data up to Week 52 will be handled as follows:

- Subjects who drop out of the trial for any reason will be considered non-responders from the time they drop out through Week 52 for active groups or through Week 16 for placebo group.
- Subjects who do not have the required data at baseline and at the specific time point to compute response will be classified as non-responders.

[redacted]

Data collected after Week 16 will generally be presented as 'observed case'; i.e. all available data for each time point will be included in the analyses. In addition, multiple imputation will be used for ACR50. These will be presented for AIN457 6mg/kg - 3mg/kg treatment group only. Observed data summary will be presented by treatment sequence (Section 2.2.2).

[redacted]

2.7 Safety analyses

Summaries will be performed for the entire treatment period. Safety analyses will be performed on treatment received or actual treatment as described below:

The actual treatment or treatment received for summaries of safety data will differ to the treatment assigned at randomization only if a subject received the wrong treatment during the entire study.

Subjects who switch treatment during the study (e.g. from placebo to active treatment) will be counted to both groups using the appropriate start and stop exposure date.

2.7.1 Adverse events (AEs)

Treatment emergent AEs (events that started after the first dose of study treatment or events present prior to the first dose of study treatment but increased in severity based on preferred term) will be summarized up to 12 weeks (84 days) after the last dose.

AEs will be summarized by presenting, for each treatment group, the number and percentage of subjects having any AE, having an AE in each primary system organ class and having each individual AE (preferred term). Summaries will also be presented for AEs by severity and for study treatment related AEs. If a subject reported more than one AE with the same preferred term, the AE with the greatest severity will be presented. If a subject reported more than one AE within the same primary system organ class, the subject will be counted only once with the

greatest severity at the system organ class level, where applicable. SAEs will also be summarized.

As appropriate, the incidence of AEs will be presented per 100 subject years of exposure.

Separate summaries will be provided for death, SAE, other significant AEs leading to discontinuation and AEs leading to dose adjustment (including study treatment discontinuation).

Adverse events reported will be presented in descending frequency according to its incidence in i.v. secukinumab group starting from the most common event. Summaries (crude incidences only) will also be presented for AEs by severity and for study treatment related AEs. If a particular AE 'severity' is missing, this variable will be listed as missing and treated as missing in summaries. If a subject reported more than one adverse event with the same preferred term, the adverse event with the greatest severity will be presented. If a subject reported more than one adverse event within the same primary system organ class, the subject will be counted only once with the greatest severity at the system organ class level, where applicable.

Separate summaries will be provided for adverse events suspected to be related to study drug, deaths, serious adverse events, and adverse events leading to discontinuation and adverse events leading to temporary dose interruption.

Adverse events will also be reported separately by SMQ according to MedDRA. The MedDRA version used for reporting the study will be described in a footnote.

Non-treatment emergent adverse events will be listed.

Algorithms for date imputations will be provided in Programming Datasets Specifications.

For serious adverse events (SAEs) occurred during screening a listing will be prepared for all subjects screened including screening failures.

The safety analyses that will be performed for treatment emergent AEs and on treatment labs and vital signs for each analysis period is described in Table 2-4.

Table 2-4 Overview of analyses on some safety endpoints

Analysis period	AEs & SAEs	AEs by severity	Study drug related AEs	AEs-SMQ	Risk
Day 1 – Week 16	crude incidence	crude incidence	crude incidence	crude incidence	crude incidence
Entire Treatment	crude incidence exposure time adjusted incidence*	crude incidence	crude incidence	exposure time adjusted incidence	crude incidence exposure time adjusted incidence

Analysis period	AEs & SAEs	AEs by severity	Study drug related AEs	AEs-SMQ	Risk

*Exposure-adjusted incidence rates will be done for the following :
at the PSOC for AE and SAE and Level 1 for Risks and SMQ analyses
at the PT level for common AEs, which is defined as at least 2% of the patients in the i.v. secukinumab during the initial treatment period or events that had an incidence rate of at least 5.0 cases per 100 subject-years in the combined i.v. secukinumab groups during the entire treatment period

2.7.2 Laboratory data

The summary of lab data will only include on treatment data, which are defined as those lab assessments after the first dose of study treatment and on or before last dose + 84 days.

The summary of laboratory evaluations will be presented for three groups of laboratory tests (hematology, chemistry and urinalysis). In addition to the individual laboratory parameters the ratios “total cholesterol / HDL” and “apolipoprotein B / apolipoprotein A1” will be derived and summarized.

For urinalysis, frequency tables will be presented.

Descriptive summary statistics for the change from baseline to each study visit will be presented. These descriptive summaries will be presented by laboratory test and treatment group. Change from baseline will only be summarized for subjects with both baseline and post baseline values and will be calculated as:

change from baseline = post baseline value – baseline value

For each parameter, the maximum change (maximum decrease and maximum increase) from baseline, will be analyzed analogously.

In addition, shift tables may be provided for parameters to compare a subject’s baseline laboratory evaluation relative to the visit’s observed value. For the shift tables, the normal laboratory ranges will be used to evaluate whether a particular laboratory test value was normal, low, or high for each visit value relative to whether or not the baseline value was normal, low, or high. If appropriate, the shifts to the most extreme laboratory test value within a treatment phase will be presented as well (including category “high and low”). These summaries will be presented by laboratory test and treatment group.

The following laboratory parameters will be analyzed with respect to numerical Common Terminology Criteria for Adverse Events (CTCAE) grades, given in Table 2-5: hemoglobin, platelets, white blood cell count, neutrophils, lymphocytes, creatinine, total bilirubin (TBL), gamma-glutamyl transferase (GGT), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), glucose, cholesterol, triglycerides (TG).

These summaries will be split into hematology and chemistry.

Table 2-5 CTCAE grades for laboratory parameters to be analyzed

CTCAE v4.0 Term	Grade 1	Grade 2	Grade 3	Grade 4
HGB decreased (Anemia)	<LLN – 100 g/L	<100 – 80 g/L	<80 g/L	
Platelet count decreased	<LLN – 75.0 x10e9 /L	<75.0 - 50.0 x10e9 /L	<50.0 – 25.0 x10e9 /L	<25.0 x 10e9 /L
White blood cell decreased	<LLN - 3.0 x 10e9 /L	<3.0 - 2.0 x 10e9 /L	<2.0 - 1.0 x 10e9 /L	<1.0 x 10e9 /L
Neutrophil count decreased	<LLN - 1.5 x 10e9 /L	<1.5 - 1.0 x 10e9 /L	<1.0 - 0.5 x 10e9 /L	<0.5 x 10e9 /L
Lymphocyte count decreased	<LLN - 0.8 x 10e9/L	<0.8 - 0.5 x 10e9 /L	<0.5 - 0.2 x 10e9 /L	<0.2 x 10e9 /L
Creatinine increased*	>1 - 1.5 x baseline; >ULN - 1.5 x ULN	>1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	>3.0 baseline; >3.0 - 6.0 x ULN	>6.0 x ULN
TBL increased	>ULN - 1.5 x ULN	>1.5 - 3.0 x ULN	>3.0 - 10.0 x ULN	>10.0 x ULN
GGT increased	>ULN - 2.5 x ULN	>2.5 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
ALT increased	>ULN - 3.0 x ULN	>3.0 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
AST increased	>ULN - 3.0 x ULN	>3.0 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
ALP increased	>ULN - 2.5 x ULN	>2.5 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
Glucose increased (Hyperglycemia)	>ULN - 8.9 mmol/L	>8.9 - 13.9 mmol/L	>13.9 - 27.8 mmol/L	>27.8 mmol/L
Glucose decreased (Hypoglycemia)	<LLN - 3.0 mmol/L	<3.0 - 2.2 mmol/L	<2.2 - 1.7 mmol/L	<1.7 mmol/L
Cholesterol high	>ULN - 7.75 mmol/L	>7.75 - 10.34 mmol/L	>10.34 - 12.92 mmol/L	>12.92 mmol/L
Hypertriglyceridemia	1.71 - 3.42 mmol/L	>3.42 - 5.7mmol/L	>5.7 - 11.4 mmol/L	>11.4 mmol/L

*Note: for “creatinine increased” the baseline criteria do not apply

Shift tables will be presented comparing baseline laboratory result (CTCAE grade) with the worst results (expressed in CTCAE grade) during the entire study analyzed. Of note, baseline will be defined as last assessment prior to first dosing in initial treatment phase.

Summaries for newly occurring or worsening clinically notable lipid abnormalities will also be provided cumulatively for each of the following parameters and categories:

- HDL:
  - <=LLN

  - <0.8 x LLN
- LDL, cholesterol, triglycerides:
  - >=ULN
  - >1.5 x ULN
  - >2.5 x ULN

Newly occurring or worsening liver enzyme abnormalities will also be summarized based on the event criteria given in Table 2-6:

Table 2-6 Liver-related events

Parameter	Criterion
ALT	>3xULN; >5xULN; >8xULN;>10xULN, >20xULN
AST	>3xULN; >5xULN; >8xULN >10xULN; >20xULN
ALT or AST	>3xULN; >5xULN; >8xULN >10xULN; >20xULN
TBL	>1.5xULN, >2xULN, >3xULN,
ALP	>2xULN, >3xULN. >5xULN
ALT or AST & TBL	ALT or AST>3xULN & TBL >2xULN; ALT or AST >5xULN & TBL >2xULN; ALT or AST >8xULN & TBL >2xULN; ALT or AST >10xULN & TBL >2xULN
ALP & TBL	ALP >3xULN & TBL >2xULN ALP >5xULN & TBL >2xULN
ALT or AST & TBL & ALP	ALT or AST>3xULN & TBL >2xULN & ALP <2xULN **(Hy's Law**) Note: elevated ALP may suggest obstruction as a consequence of gall bladder or bile duct disease; ALP may also be increased in malignancy. FDA therefore terms Hy's Law cases as indicators of *pure hepatocellular injury*. This does not mean that cases of ALT or AST >3xULN & TBL >2xULN & ALP ≥2xULN may not result in severe DILI.

Notes:
1) In studies which enroll subjects with pre-existing liver disease, baseline LFT may be increased above ULN; in such a case it is meaningful to add the condition "and worse than baseline" to the abnormality criteria

For a combined criterion to be fulfilled, all conditions have to be fulfilled on the same visit. The criteria are not mutually exclusive, e.g. a subject with ALT = 6.42xULN is counted for ALT >3xULN and ALT>5x ULN.

Individual subject data listings may be provided for subjects with abnormal laboratory data. Data of subjects with newly occurring or worsening liver enzyme abnormalities may be listed in an additional listing.

The laboratory values below Lower Level of Quantification (LLQ) or above Upper Level of Quantification (ULQ) will be imputed as LLQ or ULQ, respectively. The numerical part of the reported result will be treated as the actual LLQ or ULQ.

2.7.3 Vital signs

The summary of vital signs will only include on treatment data, which are defined as those vital sign measurements after the first dose of study treatment and on or before last dose + 84 days.

Analysis in vital sign measurement using descriptive summary statistics for the change from baseline for each post-baseline visit will be performed. These descriptive summaries will be presented by vital sign and treatment group. Change from baseline will only be summarized for subjects with both baseline and post-baseline values and will be calculated as:

change from baseline = post-baseline value – baseline value

The number and percentage of subjects with newly occurring notable vital signs will be presented. Criteria for notable vital sign abnormalities are provided in Table 2-7 below.

Table 2-7 Criteria for notable vital sign abnormalities

Vital sign (unit)	Notable abnormalities
Systolic blood pressure (mmHg)	>= 140 mmHg or < 90 mmHg
Diastolic blood pressure (mmHg)	>=90 mmHg or <60 mmHg
Pulse (bpm)	> 100 bpm or <60 bpm

2.11 Interim analysis

The 16-weeks has already been performed after all patients completed the assessment for the primary and key secondary variables.

3 Sample size calculation

An overall type I error (2-sided) 5% will be used to control type I error. Since one i.v. regimen will be tested versus placebo with respect to the primary endpoint (ACR50 response at Week 16), the type-I-error is 5% two-sided for comparison. The total sample size of 190 subjects per each group is deemed appropriate to achieve adequate power for the primary and secondary endpoints for this study.

3.1.1 Primary endpoint(s)

Analysis of a phase III study (FUTURE 5) showed a placebo response rate of 8.1%, secukinumab 150 mg with SC loading response rate of 35.9%, and secukinumab 300 mg with SC loading response rate of 39.6% at week 16 for ACR50. Assuming that the response rate of i.v. regimen is between the response rates of secukinumab 150 mg with SC loading and 300 mg with SC loading, 190 subjects per group would yield approximately 99% power to detect a

treatment difference between 27.8% (based on the difference between secukinumab 150 mg with SC loading and Placebo) and 31.5% (based on the difference between secukinumab 300 mg with SC loading and Placebo) for the primary endpoint of ACR50 in the FAS population (two-sample Chi-Squared Test, Nquery 7.0).

Table 3-1 Summary of power for binary primary endpoint

Endpoint (Week 16)	Placebo Response Rate	Secukinumab with 150 mg SC loading		Secukinumab with 300 mg SC loading	
ACR50	8.1%	Response Rate	Power N=190/arm	Response Rate	Power N=190/arm
		35.9%	99%	39.6%	99%

3.1.2 Secondary endpoint(s)

A summary of the assumptions and power for the primary and secondary efficacy parameters using the FUTURE5 study week 16 data is shown in the Table 3-2 for binary endpoints and Table 3-3 for continuous endpoints.

Table 3-2 Summary of power for binary secondary endpoints

Endpoint (Week 16)	Placebo Response Rate	Secukinumab with 150 mg SC Loading		Secukinumab with 300 mg SC Loading	
		Response Rate	Power N=190/arm	Response Rate	Power N=190/arm
ACR50	8.1%	35.9%	99%	39.6%	99%
MDA	7.8%	27.7%	99%	32.4%	99%
PASI90	9.3%	36.8%	99%	53.6%	99%
ACR20	27.4%	55.5%	99%	62.6%	99%
Resolution of Enthesitis	35.4%	54.6%	80%	55.7%	84%
Resolution of Dactylitis	32.3%	57.5%	83%	65.9%	98%

Table 3-3 Summary of power for continuous secondary endpoints

Endpoint (Week 16)	Placebo Mean change from baseline	Secukinumab with 150 mg SC Loading			Secukinumab with 300 mg SC Loading		
		Mean change from baseline	Common standard deviation	Power N=190/ arm	Mean change from baseline	Common standard deviation	Power N=190/ arm
HAQ-DI	-0.21	-0.44	0.507	99%	-0.55	0.508	99%
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

mNAPSI	-2.49	-9.31	9.503	99%	-9.18	9.503	99%
FACIT	1.98	6.18	9.303	99%	6.8	9.303	99%
SF36-PCS	1.83	6.3	7.304	99%	7.62	7.33	99%
PASDAS	-0.91	-2.02	1.359	99%	-2.37	1.361	99%

4 Appendix

4.1 Visit Windows

4.1.1 Visit windows

When visit windows are used, all visits will be re-aligned, i.e., they will be mapped into one of the visit windows. E.g., if the *Week 4* visit of a subject is delayed and occurs on Day 46 instead of on Day 29, say, it will be re-aligned to visit window *Week 8*. In the case of major deviations from the visit schedule, or due to unscheduled visits, several assessments of a subject may fall in a particular visit window (either scheduled or unscheduled). Statistical approaches to handle multiple assessments in a given visit window are specified in Table 4-2.

For lab/vital signs, follow-up (F/U) visit is excluded from analysis visit mapping window. Only assessments that come as F/U nominal visit will be directly assigned as analysis F/U visit. Other assessments that are beyond the last on-treatment visit window or after nominal F/U visit date won't be mapped to any analysis visit. F/U visit will not be included in the summary tables by visit.

Of note, subjects are allowed to have gaps in visits. All data collected will be displayed in listings.

Table 4-1 Analysis visit windows

Analysis Visit	Target Day	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
Baseline	1	≤1*	[redacted]	≤1*	≤1*	≤1*	≤1*	[redacted]	≤1*	≤1*	[redacted]
Week 4	29	2 – 43		2 – 43	2 – 43	2 – 43	2 – 43		2 – 43		
Week 8	57	44 – 71		44 – 71	44 – 71	44 – 71	44 – 71		44 – 71	2 – 71	
Week 12	85	72 – 99		72 – 99	72 – 99	72 – 99	72 – 99		72 – 99		
Week 16	113	100 – 127		100 – 141	100 – 141	100 – 141	100 – 127		100 – 141	72– 141	
Week 20	141	128 – 155					128 – 155				
Week 24	169	156 – 183		142 – 183	142 – 183	142 – 197	156 – 183		142 – 197	142 – 291	
Week 25	176										

Analysis Visit	Target Day	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
Week 28	197	184 – 211		184 – 239	184 – 281		184 – 211				
Week 32	225	212 – 239				198 – 253	212 – 239		198 – 253		
Week 36	253	240 – 267					240 – 267				
Week 40	281	268 – 295		240 – 323		254 – 309	268 – 295		254 – 347		
Week 44	309	296 – 323					296 – 323				
Week 48	337	324 – 351				310 – 351	324 – 351				
Week 52	365	352 – 439		324 – 439	282 – 439	352 – 439	352 – 410		347– 439	292 – 439	
Week 60	421						411 - 439				

* The first administration of randomized study treatment (first dose) is defined as Day 1.

Group1: [redacted] ESR, High sensitivity C-Reactive protein, [redacted] Vital sign

Group2: [redacted]

Group3: Hematology, blood chemistry, [redacted]

Group 4: [redacted] SF-36 v2, FACIT, [redacted]

Group 5: LEI, LDI, PASI, [redacted] BSA assessment, Patient's global assessment of psoriasis and arthritis disease activity (VAS), IGA

Group 6: Urine pregnancy test

Group 7: [redacted]

Group 8: mNAPSI, Physician's global assessment of fingernail disease severity (VAS)

Group 9: urinalysis, Lipids

Group 10: [redacted]

The following rules are used to determine the window for an applicable visit post baseline: "Lower limit" = "upper limit of prior applicable visit" + 1. "Upper limit" = "target day of current visit" + integer part of ("target day of next applicable visit" – "target day of current visit")/2. Lower limit of the first applicable visit is always Day 2. Day 1 is the date of the first dose of randomized study treatment

The mapping described above applies to all visits (not just scheduled visits). Repeat and/or unscheduled visits (which will be numbered in the database according to new NCDS standards) will be mapped for analysis purposes in the same way as scheduled visits. This leaves the possibility, then, for multiple measurements within an analysis window. The following

conventions will be used to determine the appropriate measurement to be summarized in the event of multiple measurements within a visit window.

Table 4-2 Rules for flagging variables

Timing of measurement	Type of data	Rule
Baseline	All data	The last non-missing measurement made prior to or on the date of administration of the first dose of study treatment (the reference start date / Day 1). If a patient did not receive any dose of study treatment then the randomization date will be used. Only date part is considered if just one assessment on Day 1. If there are multiple assessments on Day 1, following rules will apply: If assessment time exists, select the last available measurement prior to reference start date/time considering time; if no measurement prior to reference start date/time considering time, select the earliest measurement post reference start date/time considering time . If assessment time does not exist, select the available measurement from the lowest CRF visit number. For X-ray, a baseline value is the last measurement prior to dosing if available. Otherwise, take the first value within 30 days post dosing. For MRI, a baseline value is the last measurement prior to dosing if available. Otherwise, take the first value within 7 days post dosing.
Post-baseline efficacy	All data	For visits without switch of treatment in the window, the measurement closest to the target will be used. In the event two measurements are taken equally apart (e.g. 1 before target date and 1 after) the first one will be used. Cases where the same parameter is recorded more than once on the same date will be handled as follows: If time of completion exists the earliest measurement will be used; If time does not exist the measurement from the lowest CRF visit number will be used.

Timing of measurement	Type of data	Rule
Post-baseline safety	Summary visit information (e.g. lab, etc.)	For visits without switch of treatment in the window, the measurement closest to the target will be used. In the event two measurements are taken equally apart (e.g. 1 before target date and 1 after) the first one will be used. Cases where the same parameter is recorded more than once on the same date will be handled as follows: If time of completion exists the earliest measurement will be used; If time does not exist the measurement from the lowest CRF visit number will be used.
Post-baseline safety	Notable abnormalities (e.g. lab)	The most extreme measurement in the window will be used. Note this means a patient can have a notably high and notably low measurement within a window.

4.2 Detailed on implementation of statistical methodology and assumptions

4.2.1 Analysis of continuous data

4.2.1.1 Summary statistics for continuous data

Summary statistics (including N, mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum) will be provided for continuous data by visit and treatment group.

4.2.1.2 Mixed-effects repeated measures model

Endpoints with continuous data type expected to be normally distributed (e.g, [redacted]) will be analyzed using a mixed-effects repeated measures model (MMRM) with stratification factor and analysis visit as factors; and weight, baseline value, baseline by visit interaction as covariates. An unstructured correlation matrix will be used thus allowing adjustment for correlations between time points within subjects.

SAS code for mixed model:

```
proc mixed data=aaa;
class USUBJID AVISITN STRATA;
model CHG= STRATA AVISITN WEIGHT BASE BASE*AVISITN
      / s ddfm=kr;
lsmeans AVISITN / diff cl;
```

repeated AVISITN / type=un subject=USUBJID;

Run;

In case the MMRM model does not converge the following sequential steps will be used:

1. change ddfm=kr to ddfm=bw. If still no convergence, perform step 2.

2. change type=un to type=cs. If still no convergence, perform step 3.

3. remove covariates in the following order until convergence: WEIGHT, BASE*AVISITN, STRATA.

4.2.2 Analysis of binary (and categorical) data

4.2.2.1 Analyses of Frequencies

Summary statistics for discrete variables will be presented in contingency tables and will include count and frequency in each category. For n subjects, each at risk to experience a certain event with probability π, the crude incidence is estimated as p=x/n, where x is the number of subjects with the event.

If applicable, 95% confidence intervals for the relative frequences will be derived as well based on the score method including continuity correction Newcombe 1998:

With z as (1-alpha/2)-quantile of the standard normal distribution (SAS: z=PROBIT(1-alpha/2), n as total number of subjects (i.e. number of subjects in the denominator), and p as estimated crude incidence (number of subjects with event / n) it is $q = 1 - p$

Then the lower limit is

$$L = \max\left(0, \frac{2np + z^2 - 1 - z\sqrt{z^2 - 2 - 1/n + 4p(nq+1)}}{2(n+z^2)}\right)$$

and the upper limit is

$$U = \min\left(1, \frac{2np + z^2 + 1 + z\sqrt{z^2 + 2 - 1/n + 4p(nq-1)}}{2(n+z^2)}\right).$$

In addition, if L > p then L = p and if U < p then U = p.

For binary response variables (e.g. for ACR20/50/70, HAQ-DI responder, PASI 75, IGA response) the placebo-adjusted response rates including 95% confidence interval will be derived.

SAS code for risk difference:

Proc freq data=acr order=formatted;

Tables response*trt/ riskdiff;

Run;

(Note the response value should be sorted with '1' ahead of '0'.)

Fisher's exact test will be applied to rare events (e.g., MCR), pairwise treatment group comparisons to placebo or active controls.

SAS code for Fisher's exact test:

Proc freq data=mcr order=formmatted;

Tables response*trt/Fisher;

Run;

If appropriate, an exact 100*(1-α)% confidence interval (Clopper-Pearson 1934) will be obtained by using the SAS procedure PROC FREQ with the EXACT BINOMIAL statement. However, the confidence interval derived via the score method including continuity correction will be the default in safety analyses.

Figures will be provided for primary and secondary variables, with means displayed across time for all the treatment groups.

4.2.3 Multiple Imputation

A multiple imputation will be performed based on MAR for AIN457 6mg/kg - 3mg/kg for baseline weight, baseline and post-baseline of each parameter for visits up to the primary time point (e.g Week 52) using Markov Chain Monte Carlo (MCMC) method with EM algorithm.

Impute the missing values 100 times (NIMPUTE) with a seed=457<studycode> as shown below:

```
proc mi data= out=imp minmaxiter=10000000 nimpute=100 seed=4572302;
        by trt;
        var weight_base strata var1_base var1_week1-var1_week52;
        mcmc chain=multiple initial=em;
run;
```

If needed repeat for each component necessary to calculate the final score, e.g. as follows:

```
proc mi data=imp out=imp2 minmaxiter=10000000 nimpute=100 seed=4572302;
        by trt _imputation_;
        var weight_base strata var2_base var2_week1-var2_week52;
        mcmc chain=multiple initial=em;
run;
```

Post-processing of out of range imputed values will be processed with the min and max values of each variable.

The score and ACR response can now be calculated based on the complete data. The response rate will be calculated for each imputation and then combined using Rubin's rules.

In order to calculate the response rate for each imputation, PROC FREQ will be used as follows.

Calculate binomial proportion and standard error for each imputation.

```
proc freq data=<ACR20>;
  by treat visit _imputation_ ;
  tables <response> / binomial (level=2 cl=wilson correct) ;
  ods output BinomialProp=imp_bpr;
run;
```

Transpose the dataset for subsequent use with PROC MIANALYZE.

```
proc transpose data=imp_bpr out=imp_trs(drop=_name_) ;
  by treat visit _imputation_ ;
  var nvalue1; id name1; idlabel label1;
run;
```

Apply LOGIT transformation: y=log(p/(1-p)) and std. err. transformation: <new se> = se/(p*(1-p))

```
data logit;
  set imp_trs(rename=(_bin_=p e_bin=se));
  by treat visit _imputation_ ;
  lmean=log(p/(1-p));
  lse=se/(p*(1-p));
run;
```

The transformed binomial proportion estimates and standard errors are combined by applying Rubin's rules for multiple imputed data sets.

```
proc mianalyze data=logit;
  by treat visit ;
  modeleffects lmean;
  stderr lse;
  ods output ParameterEstimates=logitres;
run;
```

The combined data should be transformed back using the following formula: p=1/(1+exp(-y))

```
data miexpres;
  set logitres;
  by treat visit ;
  resti = 1/(1+exp(-estimate));
```

```
  rlow = 1/(1+exp(-lclmean));
  rupp = 1/(1+exp(-uclmean));
run;
```

Of note, sometimes all responses may be imputed to 0 or 1 at a given combination of response variable, treatment group and visit. Such cases should be considered separately. The combined final response rate would be the same as the original response but the 95% CI will be undefined.

4.2.4 Crude incidence and relativ risk estimates

4.2.4.1 Odds ratio and 100*(1-α)% confidence interval

For an investigational drug group with n_1 subjects at risk, independent from the control group (e.g. placebo or comparator) with n_0 subjects at risk, of whom x_1 and x_0 experience a certain event with probability π_1 and π_0 respectively, the odds ratio is estimated as

$\frac{p_1/(1-p_1)}{p_0/(1-p_0)}$ with $p_1 = x_1/n_1$ and $p_0 = x_0/n_0$. A conditional exact 100*(1-α)% confidence interval will be obtained by using the SAS procedure PROC FREQ with statement EXACT OR.

4.2.5 Exposure adjusted incidence rate and related risk estimates

4.2.5.1 Exposure adjusted incidence rate and 100*(1-α)% confidence interval

It will be assumed that for each of n subjects in a clinical trial the time t_j (j=1,...,n) to the first occurrence of a certain event is observed, or if the event was not experienced, the (censored) time to the end of the observation period. The sequence of first occurrences of an event will be modeled to follow approximately a Poisson process with constant intensity θ. The rate parameter θ will be estimated as λ=D/T, where $T = \sum_{j=1}^{n} t_j$ and D is the number of subjects with at least one event. Conditionally on T, an exact 100*(1-α)% confidence interval for a Poisson variable with parameter θT and observed value D can be obtained based on (Garwood, 1936), from which an exact 100*(1-α)% confidence interval for D/T will be derived as follows (Sahai, 1993; Ulm, 1990):

Lower confidence limit $L = \frac{0.5 c_{\alpha/2,2D}}{T}$ for D>0, 0 otherwise,

Upper confidence limit $U = \frac{0.5 c_{1-\alpha/2,2D+2}}{T}$

where $c_{\alpha,k}$ is the αth quantile of the Chi-square distribution with k degrees of freedom.

The example below shows how this should be handled for cases where subjects switch treatment. In particular for summarizing ‘Any secukinumab’ as a group, one should take into consideration the sequence of treatments while calculating exposure time for subjects.

Table 4-3 Examples for calculating exposure time for incidence rates (IR)

1st treatment	1st exposure	2nd treatment	2nd exposure	Event days (in terms of study day)	Exposure for IR
Placebo	100 days	150 mg	200 days	50 (1st trt) 110 (10 days into 2nd trt)	Placebo: 50 days (event) 150 mg: 10 days (event) Any AIN: 10 days (event)

5 Reference

[SOP-7012382] Analyzing and Reporting Data (PSP 012)

[SOP-0015116] Developing and Completing the Clinical Study Report

[SOP-0018880] Defining, Processing, and Reporting Protocol Deviations

AIN457A efficacy MAP M3, available in Cabinets//CREDI Projects/A/AIN457A /Administrative files/CIS (Clinical Information Sciences)/Biostatistics

Alberti KGMM, Eckel RH, Grundy SM, et al (2009) Harmonizing the metabolic syndrome: a joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; american heart association; world heart federation; international atherosclerosis society; and international association for the study of obesity. Circulation; 120: 1640-1645.

Bishop YMM, Fienberg SE, Holland PW (1991) Discrete multivariate analysis. Cambridge MA, 11th ed. The MIT Press

Bretz F, Maurer W, Brannath W, Posch M (2009) A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine; 28: 586-604.

Clopper CJ, Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 26; 404–413.

Coates LC, Fransen J, Helliwell PS (2010) Defining minimal disease activity in psoriatic arthritis: a proposed objective target for treatment. Ann Rheum Dis.69(1):48-53.

Committee for medicinal products for human use (CHMP), European Medicines Agency for the Evaluation of Medicines for Human Use. (2004) Guideline on clinical investigation of medicinal products indicated for the treatment of psoriasis. CHMP/EWP/2454/02 corr document. London, UK.

CSR template available in the CREDI template area: Cabinets/CREDI Templates /CTD

Fredriksson T, Pettersson U (1978) Severe psoriasis—oral therapy with a new retinoid. Dermatologica; 157:238–44.

Garwood, F (1936). Fiducial limits for the Poisson distribution. Biometrika, 46; 441–453.

Gottlieb A, Griffiths CEM, Ho VC, et al (2005) Efficacy and tolerability of oral pimecrolimus in the treatment of moderate to severe chronic plaque type psoriasis: a double-blind, multicentre, randomized dose-finding trial. Br J Dermatol; 152:1219-27.

Helliwell PS, Fitzgerald O, Fransen J et al (2012) The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project). Ann Rheum Dis 2012 ;0:1–6. doi:10.1136/annrheumdis-2012-201341

Kenward, M. and Roger, J. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics; 53: 983-997.

Koch GG, Tangen CM, Jung JW, et al. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statistics in Medicine; 17:1863-92.

Lachin JM (2000) The assessment of relative risks . New York: Wiley

Leonardi CL, Powers JL, Matheson RT, et al (2003) Etanercept as monotherapy in patients with psoriasis. N Engl J Med 349:2014-22.

Leonardi CL, Kimball AB, Papp KA, et al. (2008) Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 76-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 1). Lancet; 371:1665-74.

Little, R.J.A and Rubin, D.B. (2002). Statistical Analysis with Missing Data. Wiley Series in Probability and Statistics, Chapter 10.

Lubeck, DP. (2004) Patient-reported outcomes and their role in the assessment of rheumatoid arthritis. Pharmacoeconomics 22(1): 27-38.

McInnes, I.B. et al. (2013) Efficacy and safety of ustekinumab in patients with active psoriatic arthritis: 1 year results of the phase 3, multicentre, double-blind, placebo-controlled PSUMMIT 1 trial.Lancet . ISSN 1474-547X (doi:10.1016/S0140-6736(13)60594-2)

Menter A, Gottlieb A, Feldman S, et al (2008) Guidelines of care for the management of psoriasis and psoriatic arthritis – Section 1. Overview of psoriasis and guidelines of care for the treatment of psoriasis with biologics. J Am Acad Dermatol; 58:826-850.

Newcombe RG (1998) Two-sided confidence intervals for the single proportion: comparison of seven methods. Statistics in Medicine; 17: 857-872.

Ostergaard, Mikkel et al (2009). The OMERACT Psoriatic Arthritis Magnetic Resonance Imaging Scoring System (PsAMRIS): Definitions of Key Pathologies, Suggested MRI Sequences, and Preliminary Scoring System for PsA Hands. The Journal of Rheumatology 36:8 1816-1824.

Papp KA, Tyring S, Lahfa M et al (2005) A global phase II randomized controlled trial of etanercept in psoriasis: safety, efficacy, and effect of dose reduction. Br J Dermatol; 152:1304-12.

Papp KA, Langley RG, Lebwohl M, et al. (2008) Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 52-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 2). Lancet; 371:1675-84

Compound Case Retrieval Strategy available in Cabinets/CREDI Projects/A/AIN457A/Integrated Medical Safety

Ritchlin C, Rahman P, Kavanaugh A, et al. (2013). Efficacy and safety of the anti-IL-12/23 p40 monoclonal antibody, ustekinumab, in patients with active psoriatic arthritis despite conventional non-biological and biological anti-tumour necrosis factor therapy: 6-month and 1-year results of the phase 3, multicentre, double-blind, placebo controlled, randomised PSUMMIT 2 trial. Ann Rheum Dis. 2014 Jan 30. doi: 10.1136/annrheumdis-2013-204655. [Epub ahead of print]

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

Safety Profiling Plan stored in CREDI (Cabinets/CREDI Projects/A/AIN457A/Integrated Medical Safety)

Sahai H, Khurshid Anwer (1993). Confidence intervals for the mean of a poisson distribution: a review. Biom J, 35 (7); 857-867

Sampson HA, Muñoz-Furlong A, Bock SA, et al (2005) Symposium on the Definition and Management of Anaphylaxis: Summary report. J Allergy Clin Immunol 2005;115:584-91.

Sampson HA, Muñoz-Furlong A, Campbell RL, et al (2006) Second Symposium on the Definition and Management of Anaphylaxis: Summary Report - Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network Symposium. Ann Emerg Med;47(4):373-80.

Schaefer, J.L. (1997). Analysis of Incomplete Multivariate Data, Chapman&Hall.

Tyring S, Gottlieb A, Papp KA, et al (2006) Etanercept and clinical outcomes, fatigue, and depression in psoriasis: double-blind placebo-controlled randomised phase III trial. Lancet 367:29-35.

Ulm K (1990). A simple method to calculate the confidence interval of a standard mortality ratio. American Journal of Epidemiology, 131(2); 373-375

Van der Heijde DM (1999). How to Read Radiographs According to Sharp/van der Heijde Method. J Rheumatol. 26(3):743-745.

Van der Heijde D, Sharp J, Wassenberg S, Gladman DD (2005) Psoriatic arthritis imaging: a review of scoring methods. Ann Rheum Dis; 64: Suppl 2:ii 61-64.

Weisman S, Pollack CR, Gottschalk RW (2003) Psoriasis disease severity measures: comparing efficacy of treatments for severe psoriasis. J Dermatolog Treat; 14: 158-165.

Zink RC and Koch GG (2012). NParCov3: A SAS/IML Macro for Nonparametric Randomization-Based Analysis of Covariance. Journal of Statistical Software; 50(3): 1-17.

Predictive margins and average marginal effects:http://support.sas.com/kb/63/038.html