STATISTICAL ANALYSIS PLAN

A PHASE 4, MULTICENTER, SINGLE-ARM, OPEN-LABEL STUDY TO EVALUATE THE IMPACT OF APREMILAST (CC-10004) ON MRI OUTCOMES IN SUBJECTS WITH PSORIATIC ARTHRITIS

INVESTIGATIONAL PRODUCT (IP):	**Apremilast**
PROTOCOL NUMBER:	**CC-10004-PsA-014**
DATE FINAL:	12-MAY-2022
SAP Version Status:	**1.0**

Prepared by:

Amgen

One Amgen Center Drive

Thousand Oaks, CA 91320

NCT Number: NCT03783026
This NCT number has been applied to the document for purposes of posting on Clinicaltrials.gov

HISTORY OF AMENDMENT(S)

Amendment number	Date of amendment	Significant amendments and rationales	Impact to statistical analysis and descriptions

TABLE OF CONTENTS

1. LIST OF ABBREVIATIONS

AE	Adverse event
ALT	Alanine aminotransferase (SGPT)
APR	Apremilast
AST	Aspartate aminotransferase (SGOT)
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BID	Twice per day
BME	Bone marrow edema
BSA	Body surface area
CASPAR	Classification Criteria for Psoriatic Arthritis
CBC	Complete blood count
c-DAPSA	Clinical Disease Activity Index for Psoriatic Arthritis
CI	Confidence interval
CRP	Clinical Research Physician
DAS	Disease Activity Score
DCE-MRI	Dynamic Contrast-enhanced MRI
[redacted]	
eCRF	Electronic case report form
EGA	Evaluator's Global Assessment of Disease activity
ET	Early Termination
EULAR	European League Against Rheumatism
[redacted]	
FAS	Full Analysis Set
FCBP	Female of childbearing potential
FDA	Food and Drug Administration
GCP	Good Clinical Practice
GEI	Gladman Enthesitis Index

GRAPPA	Group for Research and Assessment of Psoriasis and Psoriatic Arthritis
HbA1c	Glycated hemoglobin
HAQ-DI	Health Assessment Questionnaire Disability Index
hsCRP	High sensitivity C-reactive protein
IP	Investigational product
LDI	Leeds Dactylitis Index.
LEI	Leeds Enthesitis Index
L/R	Left/right
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified intent to treat
MRI	Magnetic resonance imaging
MTX	Methotrexate
NR	Non-responder
NRS	Numeric rating scale
NSAID	Nonsteroidal anti-inflammatory drug
PASDAS	Psoriatic Arthritis Disease Activity Score
[redacted]	
PGA	Patient's Global Assessment of Disease activity
PP	Per Protocol
PsA	Psoriatic arthritis
PsAID-12	Psoriatic Arthritis Impact of Disease 12-domain questionnaire
PsAMRIS	Psoriatic Arthritis Magnetic Resonance Imaging Score
QoL	Quality of life
RBC	Red blood cell
SAE	Serious adverse event
SAP	Statistical analysis plan
[redacted]	
SJC	Swollen joint count

SOP	Standard operating procedure
SPARCC	Spondyloarthritis Research Consortium of Canada Enthesitis
TEAE	Treatment-emergent adverse event
TJC	Tender joint account
WBC	White blood cell
WB-MRI	Whole-body magnetic resonance imaging
WHO	World Health Organization
WHODD	World Health Organization Drug Dictionary
Wk	Week
[redacted]	

2. INTRODUCTION

This statistical analysis plan (SAP) describes the analyses and data presentations for Amgen's protocol CC-10004-PSA-014 "A Phase 4, Multicenter, Single-Arm, Open-Label Study to Evaluate the Impact of Apremilast (CC-10004) On MRI Outcomes in Subjects with Psoriatic Arthritis". This SAP contains definitions of analysis populations, derived variables and statistical methods for the efficacy and safety analyses. The efficacy and safety analyses will be based on the final database locked at the completion of the study.

In this SAP, the term "treatment group" refers to the apremilast 30 mg BID group. The term "APR" refers to the apremilast.

This SAP provides a more technical and detailed elaboration of the statistical analyses, as outlined and/or specified in the CC-10004-PSA-014 study protocol amendment 4 dated May 4, 2020. All statistical analyses detailed in this SAP will be conducted using SAS® Version 9.3 or higher.

3. STUDY OBJECTIVES

3.1. Primary Objective

The primary objective of the study is to evaluate the efficacy of apremilast 30 mg twice a day (BID) on inflammation indices, assessed by magnetic resonance imaging (MRI) of the hand.

3.2. Secondary Objectives

The secondary objectives of the study are to evaluate:

- The efficacy of apremilast 30 mg BID on imaging outcomes associated with structural progression, assessed by MRI of the hand
- The efficacy of apremilast 30 mg BID on disease activity and functionality, assessed by clinical outcomes
- The efficacy of apremilast 30 mg BID on inflammation indices of peripheral arthritis and enthesitis, total (sum of arthritis and enthesitis) and separately, assessed by whole-body MRI (WB-MRI)
- The efficacy of apremilast 30 mg BID on The European League Against Rheumatism Psoriatic Arthritis Impact of Disease 12 domains (PsAID-12)
- The safety and tolerability of apremilast 30 mg BID

3.3. Exploratory Objectives

The exploratory objectives of the study are:

[redacted]

[redacted]

4. INVESTIGATIONAL PLAN

4.1. Overall Study Design and Plan

This is a Phase 4, single-arm, multicenter, open-label study to evaluate the impact of apremilast (CC-10004), either in monotherapy or with stable methotrexate (MTX) on MRI outcomes in subjects with active PsA with up to 5 years of disease duration (since diagnosis).

Approximately 120 subjects will receive apremilast 30 mg BID, after a 5-day titration period, with or without stable MTX. All subjects will be permitted to take nonsteroidal anti-inflammatory drugs (NSAIDs) and/or low-dose oral glucocorticoids (prednisone ≤ 10 mg/day or equivalent) throughout the study. The NSAIDs and low-dose oral glucocorticoids must be on a stable regimen for at least 4 weeks prior to baseline. MTX (≤ 25 mg/week) will be permitted if duration of treatment is ≥ 6 months and on a stable regimen for at least 3 months prior to baseline. In addition, the stable doses of MTX, NSAIDs, and low-dose glucocorticoids must be continued from Day 1 through the Week 24 Visit. Change in doses, increase or decrease, and/or discontinuation will not be allowed except for safety reasons or for lack of availability. After the Week 24 Visit, the doses of MTX, NSAIDs, or glucocorticoids may be adjusted as clinically required.

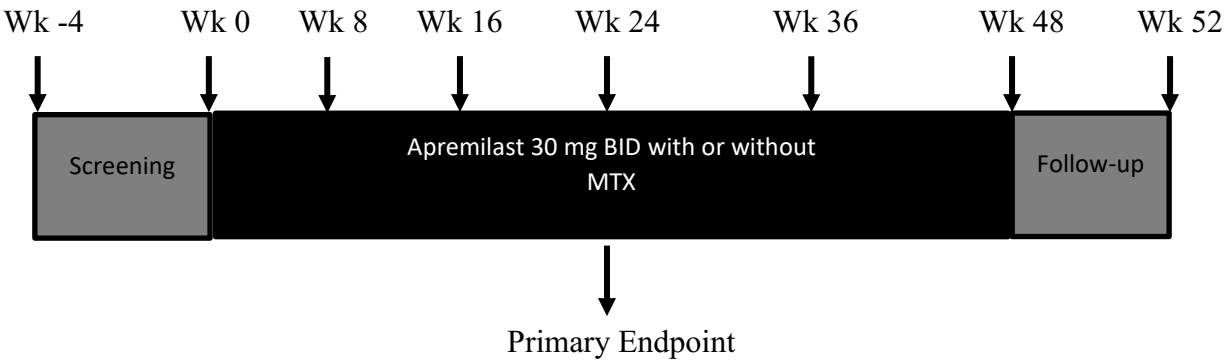
This is a 56-week study comprised of 3 phases:

- Screening Phase – up to 4 weeks
- Single-arm, Open-label Treatment Phase – Weeks 0 to 48
  - 120 subjects will receive apremilast 30 mg BID (after a 5-day titration period) for the entire duration of this phase.

- Observational Follow-up Phase – 4 Weeks
  - All subjects who complete the study or discontinue early will participate in the 4-week Post-Treatment Observational Follow-up Phase.

The study schematic is presented in Figure 1.

Figure 1: Overall Study Design



BID = twice daily; MTX = methotrexate; Wk = Week.

4.2. Study Endpoints

Efficacy and safety endpoints are listed in Sections 4.2.1 through 4.2.2, respectively. Derivations of select efficacy and safety endpoints are described in Sections 4.2.4 and 4.2.5, respectively.

4.2.1. Efficacy Endpoints

The following sections provide the efficacy endpoints as specified in the protocol.

4.2.1.1. Primary Efficacy Endpoint

The primary endpoint of the study is the change from baseline in the composite score of bone marrow edema (BME), synovitis, and tenosynovitis assessed by Psoriatic Arthritis Magnetic Resonance Imaging Score (PsAMRIS) at Week 24.

4.2.1.2. Secondary Efficacy Endpoints

- Change from baseline in the composite score of BME, synovitis, and tenosynovitis assessed by PsAMRIS at Week 48.
- Change from baseline in the composite score of BME and synovitis assessed by PsAMRIS at Weeks 24 and 48.
- Change from baseline in the PsAMRIS total inflammation score (BME, synovitis, tenosynovitis and periarticular inflammation) at Weeks 24 and 48.
- Change from baseline in BME assessed by PsAMRIS at Weeks 24 and 48.
- Change from baseline in synovitis assessed by PsAMRIS at Weeks 24 and 48.
- Change from baseline in tenosynovitis assessed by PsAMRIS at Weeks 24 and 48.
- Change from baseline in periarticular inflammation assessed by PsAMRIS at Weeks 24 and 48.

- Change from baseline in the PsAMRIS total damage score (erosion + bone proliferation) at Weeks 24 and 48.
- Change from baseline in bone erosion assessed by PsAMRIS at Weeks 24 and 48.
- Change from baseline in bone proliferation/enthesophytes assessed by PsAMRIS at Weeks 24 and 48.
- Change from baseline in the Swollen Joint Count (SJC) at Weeks 24 and 48.
- Change from baseline in the Tender Joint Count (TJC) at Weeks 24 and 48.
- Change from baseline in the Clinical Disease Activity Index for Psoriatic Arthritis (c-DAPSA) Score at Weeks 24 and 48.
- Change from baseline in Spondyloarthritis Research Consortium of Canada (SPARCC), in subjects with preexisting enthesopathy at Weeks 24 and 48.
- Change from baseline in the Leeds Enthesitis Index (LEI), in subjects with preexisting enthesopathy at Weeks 24 and 48.
- Proportion of subjects with baseline enthesitis whose enthesitis improves to 0 at Weeks 24 and 48.
- Change from baseline in the Leeds Dactylitis Index (LDI), in subjects with preexisting dactylitis at Weeks 24 and 48.
- Proportion of subjects with baseline dactylitis whose tender dactylitis count improves to 0 at Week 24 and 48.
- Change from baseline in the Psoriatic Arthritis Disease Activity Score (PASDAS) at Weeks 24 and 48.
- Change from baseline in the evaluator's global assessment of disease activity at Weeks 24 and 48.
- Change from baseline in the subject's global assessment of disease activity at Weeks 24 and 48.
- Change from baseline in the subject's assessment of pain at Weeks 24 and 48.
- Change from baseline in Health Assessment Questionnaire-Disability Index (HAQ-DI score) at Weeks 24 and 48.
- Change from baseline in the peripheral enthesitis inflammation index assessed by WB-MRI at Weeks 24 and 48.
- Change from baseline in peripheral joints inflammation index assessed by WB-MRI at Weeks 24 and 48.
- Change from baseline in the total peripheral (joints and enthesitis) inflammation index assessed by WB-MRI at Weeks 24 and 48.

- Change from baseline in the Bath Ankylosing Spondylitis Disease Activity Index BASDAI (for subjects deemed to have PsA spondylitis by the investigator and with BASDAI item 2 ≥ 4 at baseline) at Weeks 24 and 48.
- Change from baseline in the PsAID-12 at Weeks 24 and 48.

4.2.1.3. Exploratory Efficacy Endpoints

[redacted]

4.2.2. Safety Endpoints

The follow safety parameters will be evaluated throughout the duration of the study:

- Adverse events (AEs)
    - Type, frequency, severity and relationship of adverse events to study medication
    - Number of subjects who discontinue study medication due to any adverse event
- Deaths
- Laboratory evaluations
    - Laboratory marked abnormalities (refer to Appendix A1.2)
    - Observed value and change from baseline over time in continuous laboratory parameters
    - Shifts from baseline to post-baseline time points and to the worst post-baseline value in terms of low/normal/high/both low and high in laboratory parameters, as appropriate
- Vital signs and weight

- Observed value and change from baseline over time in vital signs (weight, temperature, pulse, and blood pressure)
- Shifts from baseline to post-baseline time points, and to the worst post-baseline value in terms of normal/abnormal in pulse and blood pressure (normal ranges are defined as: 60-100 beats/minute for pulse, 90-140 mmHg for systolic blood pressure, 35.5-37.9 C for temperature, and 60-90 mmHg for diastolic blood pressure)
- Observed value, change and percent change from baseline over time in weight

4.2.3. Health-related Quality of Life Endpoints

All endpoints related to the health-related quality of life (shown below) are included in the efficacy endpoints above.

Table 1: Health-related Quality of Life Endpoints

Secondary	The European League Against Rheumatism Psoriatic Arthritis Impact of Disease 12 domains (PsAID-12)	Change from baseline in the PsAID-12	Week 24 Week 48
Exploratory	[redacted]		
Exploratory			
Exploratory			
Exploratory			

4.2.4. Derivations of Efficacy Endpoints

The definition of baseline, applicable to all continuous efficacy endpoints, is given in Section 5.4. Change from baseline is calculated as on-treatment value minus the baseline value. Handling of time points is described in Section 5.5.

4.2.4.1. Magnetic Resonance Imaging (MRI)

In this study, the [redacted]

[redacted]

The MRI assessments will be conducted following the procedures outlined in the MRI manual, and must be performed on the same day of or within 7 days after the clinical visit. Therefore, the baseline MRI could be performed 7 days after the first administration of IP. The subsequent MRIs should be performed at Week 24 + 7 days and Week 48 + 7 days. All MRI visits should be scheduled well in advance to allow for proper planning. In addition, it is recommended to schedule 2 MRI visits (an initial and a repeat visit) for each MRI time point to ensure a high-quality MRI scan at each protocol-specified time point. The initial and repeat MRI visits should be scheduled approximately 7 to 14 days apart. If the initial MRI scans are of acceptable quality, then the scheduled repeat MRI session can be canceled. Maximum two repeated MRI scans can be scheduled. All images will be read regardless of deviations in acquisition window.

4.2.4.2. MRI of the Most Affected Hand

The conventional MRI assessment will be performed at baseline, Week 24, Week 48 and ET visit on the hand with the greater inflammatory burden of swollen joints and/or dactylitis. If both hands are equally affected, the dominant hand will be designated as the index hand.

4.2.4.3. PsAMRIS

The Outcome Measures in Rheumatology Clinical Trial (OMERACT) Psoriatic Arthritis Magnetic Resonance Imaging Score (PsAMRIS) is a validated scoring system used to assess the images from the MRI of the most affected hand.

OMERACT PsAMRIS assesses metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints of fingers 2 to 5, and evaluates synovitis (score 0-3), flexor tenosynovitis (score 0-3), periarticular inflammation (absent or present), bone edema (score 0-3), bone erosion (score 0-10) and bone proliferation/enthesophytes (absent or present).

All MRI readings are performed by two radiologists blinded to visit sequence and the patients. The adjudication of the results, when two readers are involved, will be done through the consensus procedure between the two readers. The readers will be able to share the screen and discuss the data, then amend the final scoring. The adjudication might involve the original reading and repeat reading. The adjudication score is considered as final assessment score and used for analysis. The intent of adjudication is to verify those cases that will have the greatest impact on results.

For more detailed information on the specific procedures and the consistent acquisition of MRI images, please refer to the Imaging Charter. The handling of missing data in PsAMRIS will be discussed in Section 4.2.4.4.

4.2.4.3.1. Synovitis

Synovitis is to be scored 0-3 at MCP, PIP and DIP joints of fingers 2 to 5. Grading scale: score 0 is normal, while score 1–3 is mild, moderate, severe, respectively. A total of 12 joints in the most affected hand are evaluated, the maximum possible score is 36.

4.2.4.3.2. Flexor Tenosynovitis

Flexor tenosynovitis is to be scored 0-3 at the same 12 joint regions as evaluated for synovitis. The maximal thickness of enhancing/bright signal tenosynovium is to be assessed as follows: 0: none; 1: < 1/2 tendon thickness; 2: ≥ 1/2 and < 1 tendon thickness; 3: ≥ 1 tendon thickness. Like the synovitis score in Section 4.2.4.3.1, the maximum possible score is 36.

4.2.4.3.3. Periarticular Inflammation

Periarticular inflammation is to be scored 0–1 separately at volar and dorsal aspects of the same 12 joint regions as evaluated for synovitis and flexor tenosynovitis. Grading scale: 0: absent; 1: present. Thus, the maximum possible score is 24.

4.2.4.3.4. BME score assessed by PsAMRIS

Bone marrow edema score, the bone edema score from PsAMRIS, is assessed on a scale of 0-3, based on the proportion of bone with edema, compare to the "assessed bone volume", judged on all available images; where 0: no edema; 1: 1–33% of bone oedema; 2: 34–66% of bone oedema; 3: 67–100% of bone oedema (as per RAMRIS[1]). A total of 12 joint regions are evaluated, and as each joint has proximal and distal components, thus the maximum possible score is 72.

4.2.4.3.5. Bone Erosion

Bone erosion is assessed on a scale of 0-10, based on the proportion of eroded bone compared to the "assessed bone volume," judged on all available images: 0: no erosion; 1: 1–10% of bone eroded; 2: 11–20%, etc. The "assessed bone volume" is from the articular surface (or its best estimated position if absent) to a depth of 1 cm (as per RAMRIS[1]). A total of 12 joints in the most affected hand are evaluated and each joint has proximal and distal components; the maximum possible score is 240.

4.2.4.3.6. Bone Proliferation

Bone proliferation is to be scored 0–1 at MCP, PIP, and DIP joints of fingers 2 to 5. Grading scale: 0: absent; 1: present. Thus, the maximum possible score is 12.

4.2.4.3.7. Inflammation Score of BME, Synovitis and Tenosynovitis

To give each of the three symptoms equal weighting, the inflammation score of BME, synovitis and tenosynovitis is the [redacted] The maximum possible score is 216.

4.2.4.3.8. Composite Score of BME and Synovitis

The composite score of BME and synovitis is the [redacted] like the inflammation score estimation in Section 4.2.4.3.7. The maximum possible score is 144.

4.2.4.3.9. Total Inflammation Score

The total inflammation score consists of BME, synovitis, tenosynovitis and periarticular inflammation. The total inflammation score is computed as [redacted]

[redacted]. The maximum possible score is 288.

4.2.4.3.10. Total Damage Score

The total damage score consists of erosion score and bone proliferation score. To give these 2 symptoms equal weights, the total damage score is [redacted] The maximum possible score is 480.

4.2.4.3.11. Smallest Detectable Change (SDC)

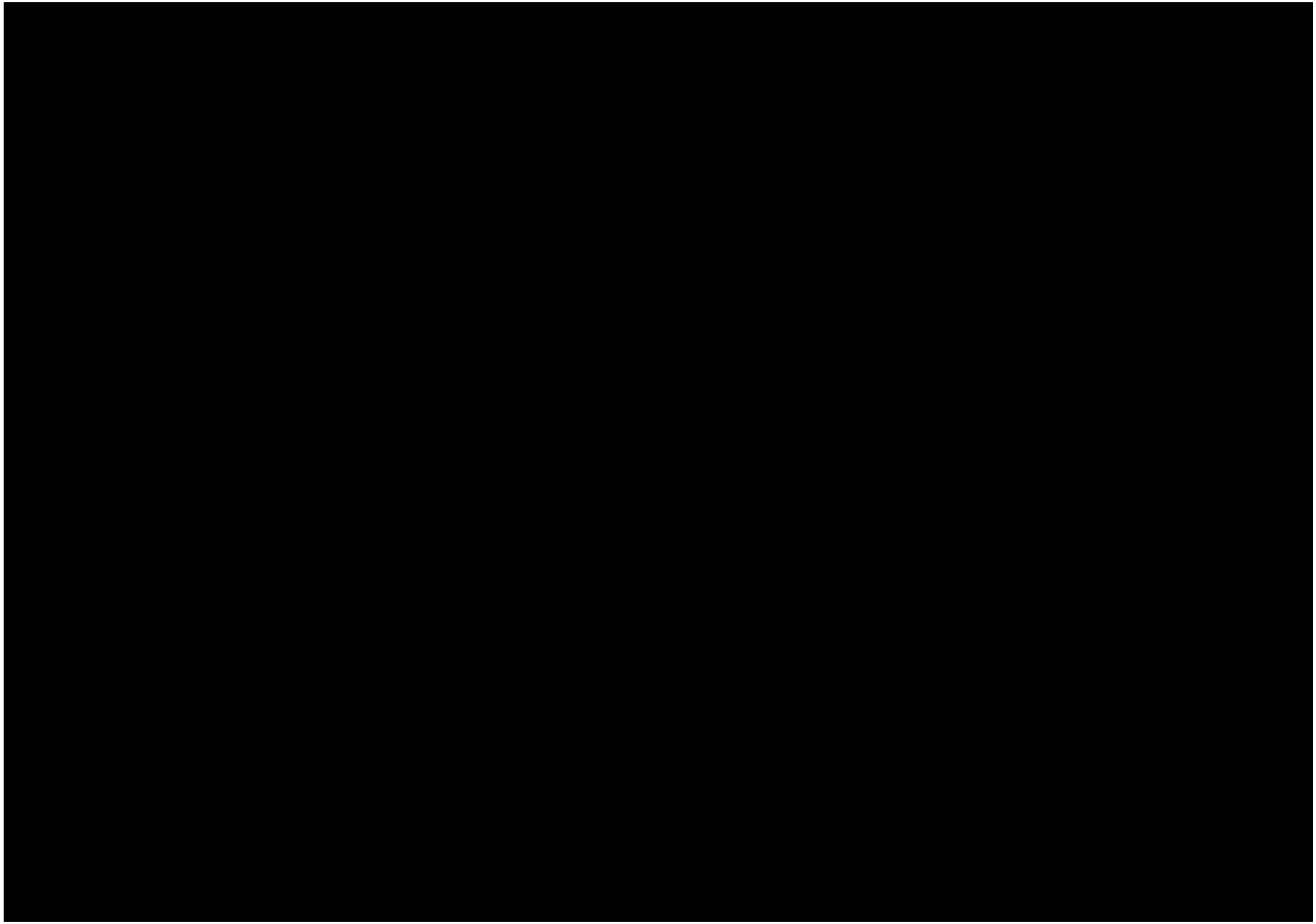
To maintain objectivity in the evaluation of imaging, two different readers will evaluate all time points for a given subject. All time points for a subject will be presented in arbitrary order. The reviewers will have no knowledge of the time point chronological sequence, subject identity.

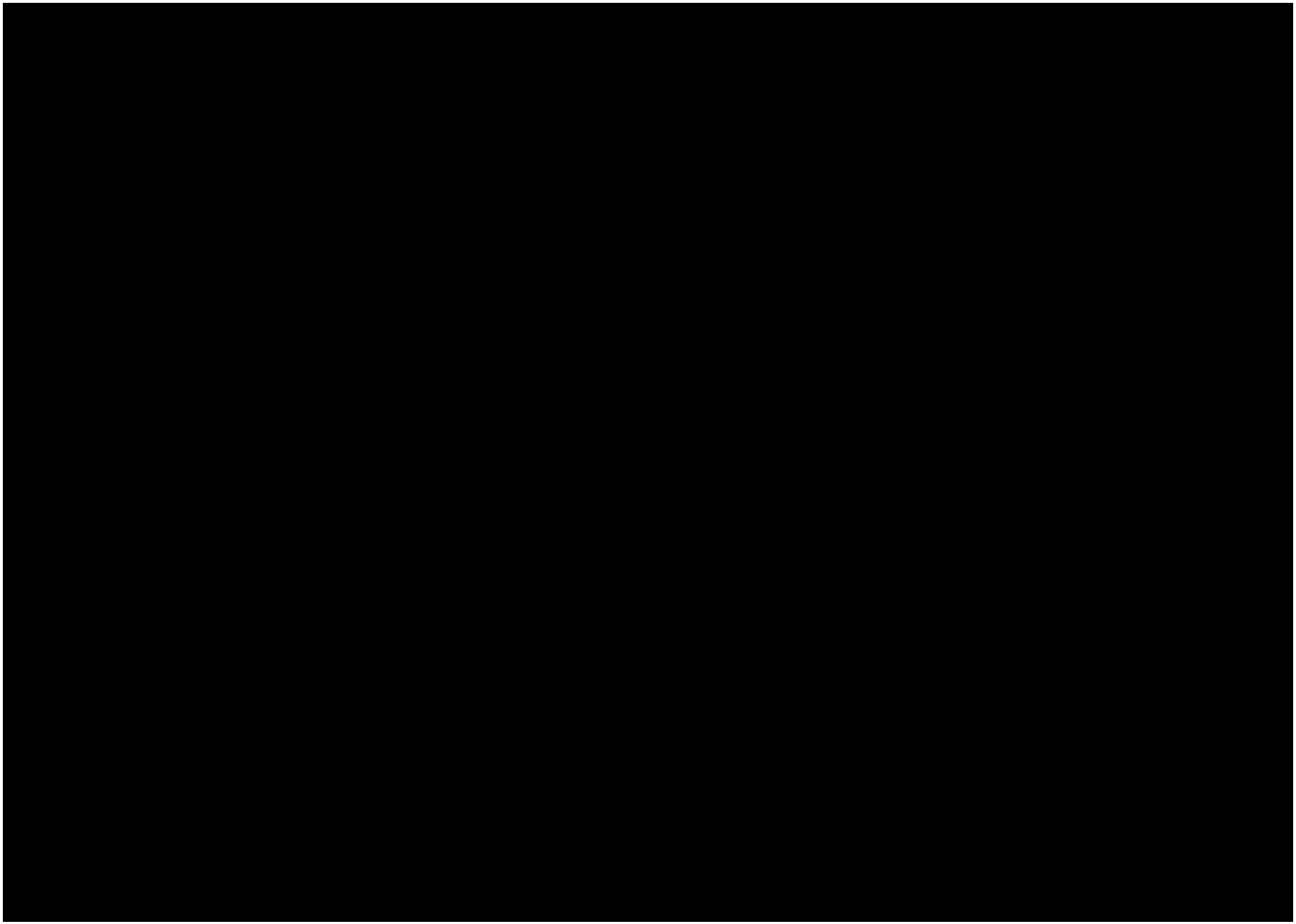
[redacted]

4.2.4.4. Handling of Missing Data in PsAMRIS

4.2.4.4.1. Handling of Missing Individual Joint Evaluation in PsAMRIS

PsAMRIS readings between the two readers will be adjudicated by the consensus method. For more detailed information on the adjudication of MRI images, please refer to the Imaging Charter. Missing or unclassifiable readings for individual joint/location evaluations will be imputed by the following guideline:





4.2.4.4.2. Handling of Missing PsAMRIS Total Score Assessment

If a subject does not have any results within the visit window (referred to Section 4.2.4.1), the PsAMRIS results for this subject will be considered as missing at the time point, no imputation is needed and the detailed statistical analysis will be described in section 6.1

4.2.4.5. Whole Body-MRI (WB-MRI)

Selected peripheral joints and entheses, [redacted] will be assessed by WB-MRI.

[redacted]

For completed details on the reading, scoring, and review process, please refer details in the Imaging Charter.

The parameters for WB-MRI are listed below:

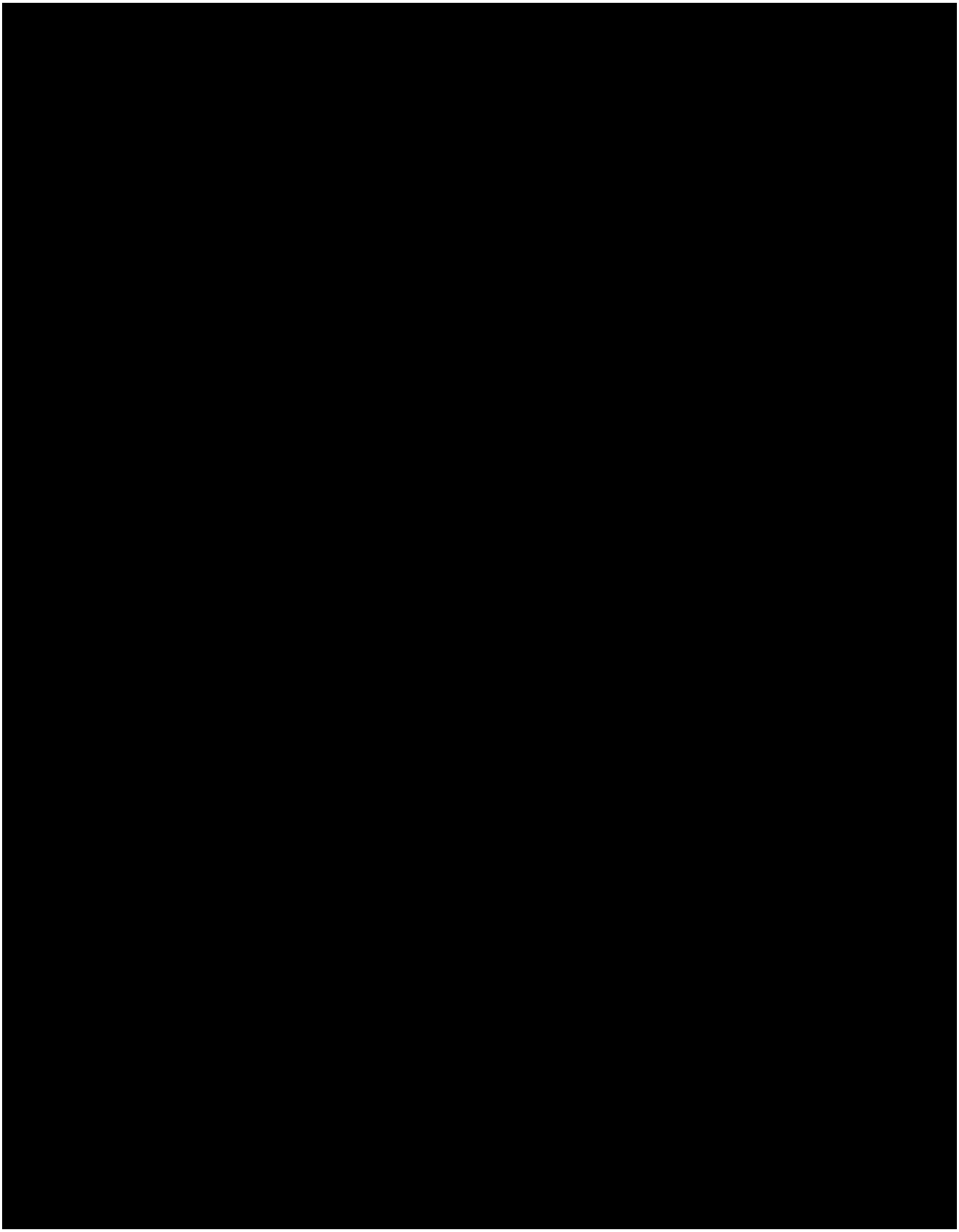
- SPARCC spine and sacroiliac joint indices

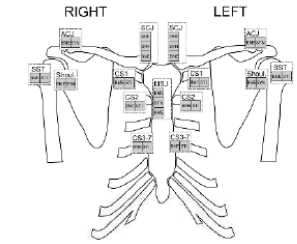
[redacted]

- MRI inflammation indices of peripheral arthritis and enthesitis

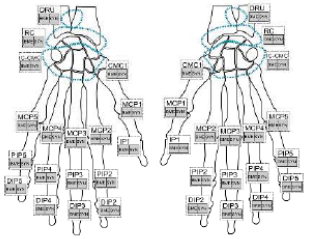
[redacted]

[redacted]

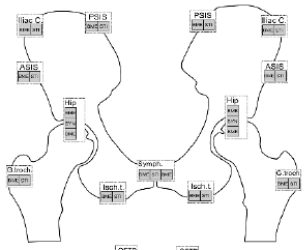




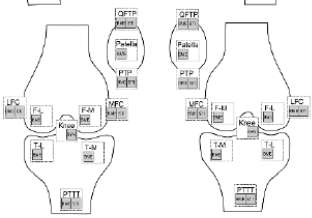
Shoulder/ACW	No. areas	Score range
Joints		
Synovitis	7	0-21
Osteitis	10	0-30
Entheses		
Soft tissue infl.	8	0-24
Osteitis	8	0-24



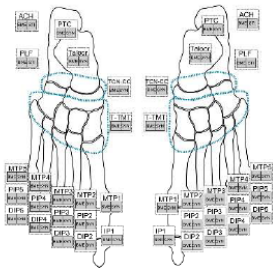
Hands	No. areas	Score range
Joints		
Synovitis	36	0-108
Osteitis	36	0-108
Entheses		
Soft tissue infl.	0	0
Osteitis	0	0



Pelvis	No. areas	Score range
Joints		
Synovitis	2	0-6
Osteitis	4	0-12
Entheses		
Soft tissue infl.	11	0-33
Osteitis	12	0-36



Knees	No. areas	Score range
Joints		
Synovitis	2	0-6
Osteitis	10	0-30
Entheses		
Soft tissue infl.	10	0-30
Osteitis	10	0-30



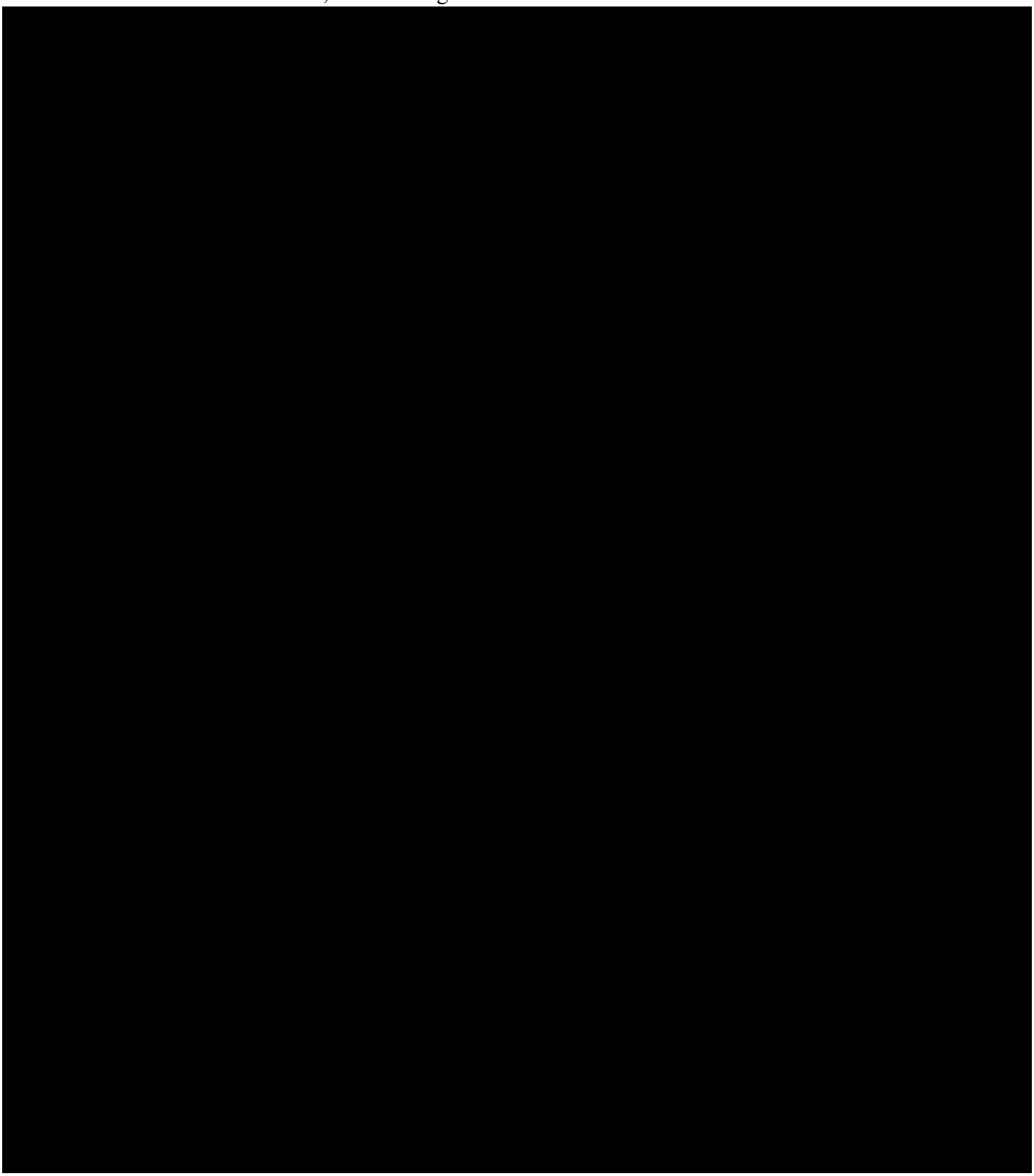
Feet	No. areas	Score range
Joints		
Synovitis	36	0-108
Osteitis	36	0-108
Entheses		
Soft tissue infl.	4	0-12
Osteitis	4	0-12

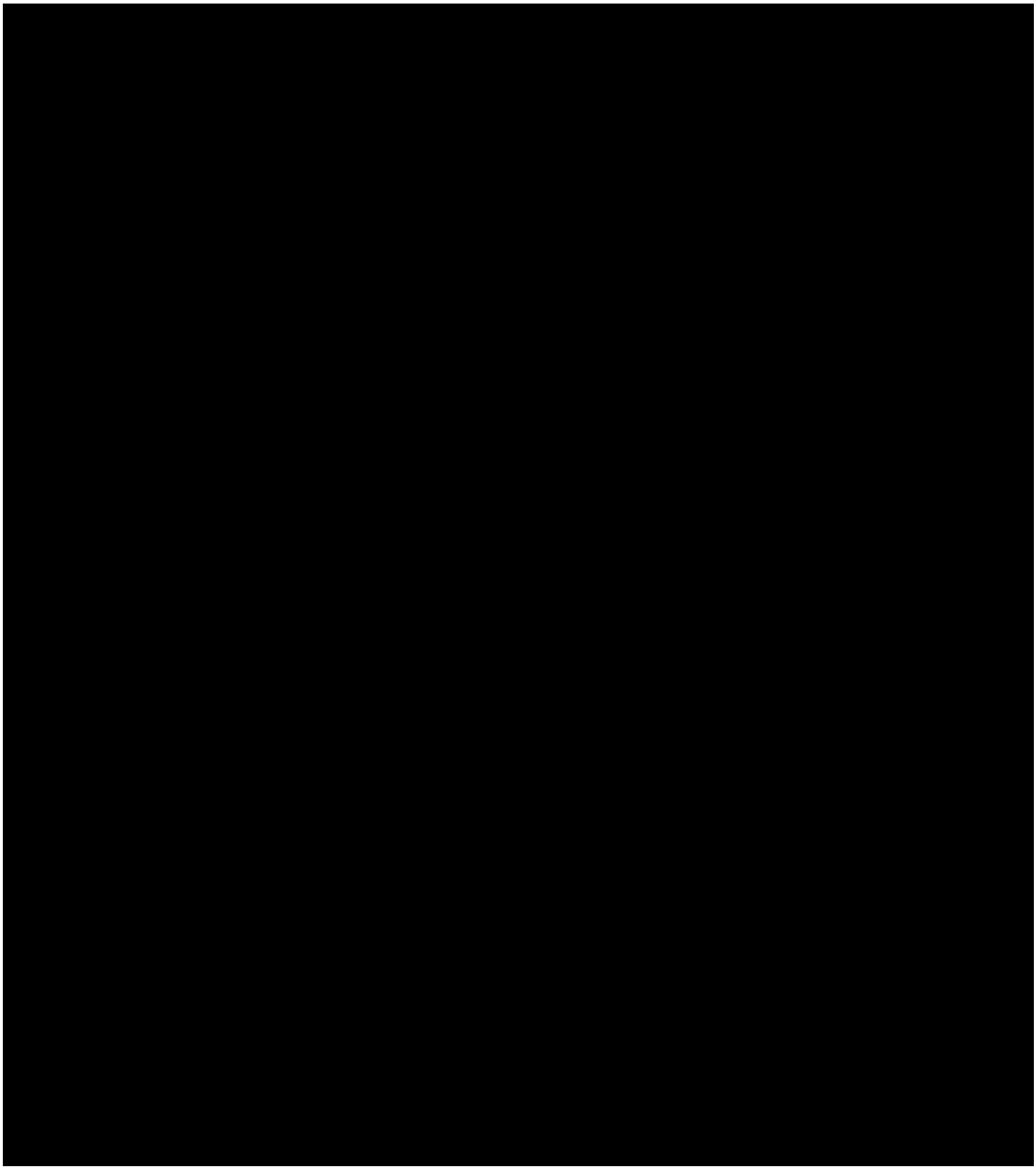
Total

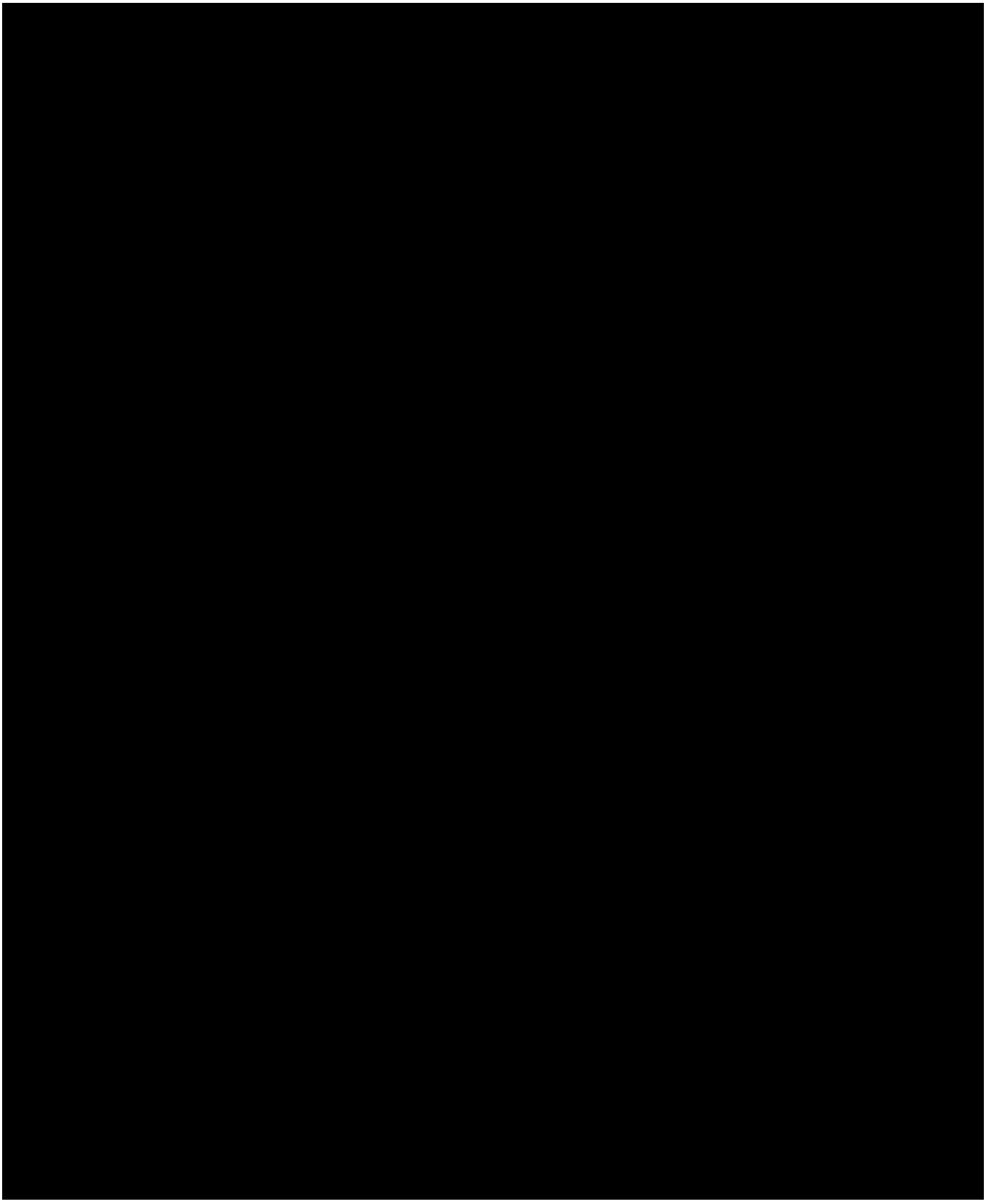
Total	No. areas	Score range
Joints		
Synovitis	83	0-249
Osteitis	96	0-288
Entheses		
Soft tissue infl.	33	0-99
Osteitis	34	0-102

Figure 4: Areas to be scored for the Peripheral Arthritis and Enthesitis.

The range of total synovitis joints is 0-249, the range of total osteitis joints is 0-288, the range of total osteitis entheses is 0-102, and the range of total STI is 0-99.







[redacted]

4.2.4.8. Handling of Missing [redacted]

If a subject does not have any results within the visit window (referred to Section 4.2.4.1), the [redacted] results for this subject will be considered as missing at the time point, no imputation is needed and the detailed statistical analysis will be described in section 6.1.

4.2.4.9. Tender and Swollen Joint Counts

For tender and swollen joint counts, a joint assessment will be recorded as "Temporarily Not Assessable Due to IA Injection" (TNA-IA), "Permanently Not Assessable" (PNA), "Temporarily Not Assessable – Other" (TNA-O), "Tender Only", "Swollen Only", or "Tender and Swollen". In this section, the term "assessed" refers to the joints that are "Tender Only", "Swollen Only", or "Tender and Swollen".

The approaches below used to handle joints at screening, baseline and post-baseline visits will be applied, unless otherwise specified.

Table 3: TJC and SJC Handling method

Visit	Scenarios	Handling method
Screening Visit	The joint is assessable	Count it into the number of tender/swollen joints
	The joint is PNA	Exclude it from joint count throughout the study
	The joint is TNA-IA or TNA-O	Exclude it from joint count at screening visit
Baseline Visit	The joint is assessable	Count it into the number of tender/swollen joints
	The joint is TNA-IA or PNA or TNA-O	Exclude it from joint count throughout the study (both baseline and post-baseline visits)
Post-Baseline Visit	The joint is assessable at both baseline and post-baseline visit	Count it into the number of tender/swollen joints
	The joint is TNA-IA or PNA or TNA-O	The last observed joint assessment (including baseline and post-baseline assessment) will be carried forward to this visit

After applying these data handling rules, TJC will be calculated by adding the number of tender joints with outcomes of 'Tender Only' or 'Tender and Swollen'. SJC will be calculated by adding

the number of joints with outcomes of 'Swollen Only' or 'Tender and Swollen'. TJC and SJC will be derived based on all the joints (76 joints for swellingness and 78 joints for tenderness).

Data resource: Joint assessment is entered by evaluator via site pad provided by eRT.

Swollen Joint Count/Tender Joint Count over 48 weeks based on all joints will be presented graphically. Proportion of Subjects with Swollen Joint Count/Tender Joint Count <=1 Over 48 Weeks will be plotted.

4.2.4.10. Subject's Pain Numeric Rating Scale

The pain Numeric Rating Scale (NRS) is the subject's assessment of how much pain he/she had, on average, in the previous week in his/her joints due to Psoriatic Arthritis (PsA). The subject will be asked to mark the box with an X on a 0- to 10-unit NRS on which the left-hand box of 0 represents "No Pain," and the right-hand box of 10 represents "Pain As Bad As You Can Imagine". This value will be used as a numerical result.

Data resource: Subject's pain score is entered by patient via site pad provided by ERT.

Patient's assessment of pain over 48 weeks will be presented graphically.

4.2.4.11. Patient's Global Assessments of Disease Activity (PGA)

The Patient's (Subject's) Global Assessments of Disease Activity is an assessment of how active a subject's PsA was on average during the previous week. Subject will select a number from a 0- to 10-unit NRS on which the 0 represents "Very Well" and 10 represents "Very Poor".

Data resource: PGA is entered by patient via site pad provided by ERT.

4.2.4.12. Evaluator's Global Assessments of Disease Activity (EGA)

The Evaluator's Global Assessment (EGA) is an assessment of how active a subject's arthritis today. Evaluator will select a number from a 0- to 10-unit NRS on which the 0 represents "No Arthritis Activity" and 10 represents "Extreme Active Arthritis".

Data resource: EGA is entered by evaluator via site pad provided by ERT.

4.2.4.13. Clinical Disease Activity Index for Psoriatic Arthritis (c-DAPSA) Score

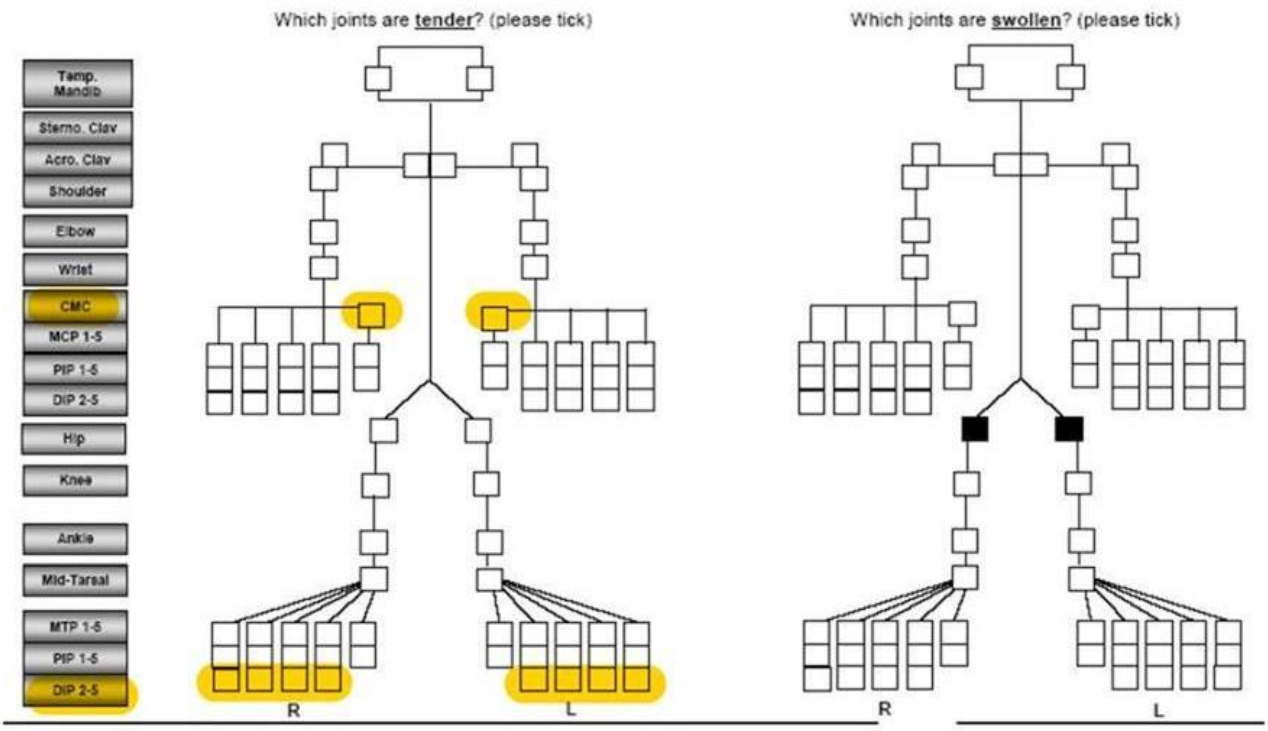
The c-DAPSA is a measure of PsA disease activity, associated with functional and structural outcomes (Aletaha, 2017). The c-DAPSA score is computed using the following formula:

c-DAPSA = TJC68 + SJC66 + PGA + Pain,

where the TJC68 is the 68-tender joint count (0-68), the SJC66 is the 66-swollen joint count (0-66), PGA is the patient (subject) global assessment of disease activity (0-10 NRS), and Pain is the subject's pain numeric rating scale (0-10 NRS).

10 joints (CMC and DIP 2-5 joints on left and right side) will be excluded from TJC78 and SJC76. See Figure 5 for the schematic diagram.

Figure 5: Schematic diagram of 68 tenders and 66 swollen joint assessment



The c-DAPSA score ranges from 0 to 154, and it will be set to missing if any of the 4 components is missing. A higher score indicates greater disease activity. Proposed cut off for c-DAPSA are a score ≤ 4 for remission, a score > 4 and ≤ 13 for low disease activity, > 13 and ≤ 27 for moderate disease activity and >27 for high disease activity. The proportion of subjects who achieve PSA remission, low disease activity, moderate disease activity and high disease activity will be summarized.

c-DAPSA score over 48 weeks will be tabulated and presented graphically.

4.2.4.14. Enthesitis

Enthesitis or the inflammation of the sites where tendons or ligaments insert into the bone, is a prominent clinical manifestation in subjects with PsA. In this study, Spondyloarthritis Research Consortium of Canada (SPARCC) and Leeds Enthesitis Index (LEI) will be used to evaluate entheseal inflammation. The details about SPARCC and LEI are mentioned in Sections 4.2.4.14.1 and 4.2.4.14.2.

Pre-existing SPARCC enthesopathy is defined as subjects who have a baseline SPARCC score greater than 0. Pre-existing LEI enthesopathy is defined as subjects who have a baseline LEI score greater than 0. The subjects with baseline SPARCC enthesitis is defined as the subjects whose baseline SPARCC scores greater than 0. The subjects with baseline LEI enthesitis is defined as the subjects whose baseline LEI scores greater than 0Spondyloarthritis Research Consortium of Canada (SPARCC)

The SPARCC Enthesitis Index will be assessed through physical examination of the following sites:

- Medial epicondyle (left/right [L/R])
- Lateral epicondyle (L/R)
- Supraspinatus insertion into greater tuberosity of humerus (L/R)
- Greater trochanter (L/R)
- Quadriceps insertion into superior border of patella (L/R)
- Patellar ligament insertion into inferior pole of patella or tibial tubercle (L/R)
- Achilles tendon insertion into calcaneum (L/R)
- Plantar fascia insertion into calcaneum (L/R)

Tenderness on examination is recorded as either present (1) or absent (0) for each of the 16 sites, for an overall score range of 0–16. Higher count represents greater enthesitis burden.

Similar to the tender and swollen joint assessments, the SPARCC assessment may be not done or the joints may be recorded as "Temporarily Not Assessable" or "Permanently Not Assessable", thus the approaches of dealing with not done assessment and handling joints that are not assessable will be employed (see Section 4.2.4.9 for details).

Data resource: SPARCC questionnaire is entered by evaluator via site pad provided by ERT.

4.2.4.14.1. Leeds Enthesitis Index (LEI)

The Leeds Enthesitis Index will be assessed through physical examination of the following sites. Tenderness (yes/no) will be assessed at 6 sites of tendon insertions (0-6) as depicted below:

- Lateral epicondyle, left and right
- Medial femoral condyle, left and right
- Achilles tendon insertion, left and right

Tenderness on physical examination is recorded as either present (1) or absent (0) for each of the 6 sites, for an overall score range of 0 to 6. Higher count represents a greater enthesitis burden.

Similar to the tender and swollen joint assessments, the LEI assessment may be not done, or the joints may be recorded as "Temporarily Not Assessable" or "Permanently Not Assessable", thus the approaches of dealing with assessments that are not completed and handling of joints that are not assessable will be employed (see Section 4.2.4.9 for details).

Data resource: LEI is entered by evaluator via site pad provided by ERT.

LEI score over 48 weeks will be presented graphically.

4.2.4.15. Dactylitis

Dactylitis ("sausage digit") is characterized by the swelling of the entire finger or toe. It will be assessed using the Leeds Dactylitis Index (LDI) (Helliwell, 2005).

4.2.4.15.1. Leeds Dactylitis Index (LDI)

In this study, the dactylitis will be assessed by LDI, which is a function of finger circumference and tenderness (0 = No Tenderness, 1 = Tender, 2 = Tender and wince, 3 = Tender and withdraw).

In addition, as part of the LDI, number of affected digits will be evaluated. Each digit on the hands and feet will be rated as zero for no dactylitis or 1 for dactylitis present. The dactylitis score is the sum of the individual scores for each digit.

The dactylitis score for individual affected finger or toe is defined as below.

[(Circumference of involved digit/circumference of contralateral digit – 1) *100] *Tenderness Score.

Preexisting dactylitis will be defined as subjects who have a baseline LDI score greater than 0 . A higher score indicates greater disease activity.

Data resource: LDI is entered by evaluator via site pad provided by ERT.

4.2.4.16. Psoriatic Arthritis Disease Activity Score (PASDAS)

The PASDAS is a weighted index comprised of assessments of joints, function, acute phase response, QoL, and patient and physician visual analogue scores (VAS) (Coates, 2014).

Note: This study does not utilize VASs, but rather NRSs. The patient's and physician's global assessment of disease activity captured on a NRS will be multiplied by a factor of 10 to derive the VAS, used to calculate PASDAS.

The calculation of the PASDAS will be performed programmatically using the formula below:

$$\text{PASDAS} = [(0.18\times\sqrt{\text{physician global VAS}})+(0.159\times\sqrt{\text{patient global VAS}})-(0.253\times\sqrt{\text{SF36}-\text{PCS}})+(0.101\times\text{LN (SJC66 +1)})+(0.048\times\text{LN (TJC68 +1)})+(0.23\times\text{LN (Leeds enthesitis count+1)})+(0.377\ \text{LN (dactylitis count+1)})+(0.102\times\ \text{LN (CRP+1)}+2)]*1.5.$$

CRP = C-reactive protein; LN = natural logarithm; PCS = physical component summary scale of SF36; SF36 = Medical Outcomes Study Short Form-36; SJC = swollen joint count; TJC = tender joint count; VAS = visual analogue scale. All VAS scores are 0–100 mm. Swollen joint count is 66 joints, and tender joint count 68.

Note: SJC66 and TJC68 will be used in the derivation of PASDAS, refer section 4.2.4.13. If one or more components are missing, then PASDAS is set to be missing. The score range of the PASDAS is 0–10, with worse disease activity represented by higher scores. Response criteria for the Psoriatic Arthritis Disease Activity Score (PASDAS) is detailed in Table 4. PASDAS Score Over 48 Weeks will be presented graphically.

Table 4: Response criteria for PASDAS

	Improvement		
Final PASDAS Score	≥ 1.6	< 1.6 BUT > 0.8	≤0.8
≤ 3.2	1	2	3
>3.2 but <5.4	2	2	3
≥ 5.4	2	3	3

1: good response; 2: moderate response; 3: poor response

4.2.4.17. Health Assessment Questionnaire-Disability Index (HAQ-DI) Score

The HAQ-DI is composed of 20 items in 8 categories (Dressing and Grooming, Arising, Eating, Walking, Hygiene, Reach, Grip, Activities). Each category has 2 or 3 items. For each item, there are 4 response options: without any difficulty, with some difficulty, with much difficulty, and unable to do (scored 0-3, respectively). The HAQ-DI score will be set to missing if a subject does not have values for at least 6 of the 8 categories.

Computing the HAQ-DI Score will follow the 2 steps below:

1. Find the highest score for each category. If use of aids/devices and/or help from another person are indicated, adjust the highest score for the associated category by increasing a lower score, which is between 0 and 2, to represent underlying disability more accurately. For example, if use of aids/devices and/or help from another person are indicated, adjust the highest score to 2 if the highest score is a 0 or a 1 If the highest score is 2 or 3, retain the value.
2. Sum the highest scores from all categories answered and divide it by the number of categories answered (must be a minimum of 6) to obtain an HAQ-DI score of 0-3 (3=worst functioning).

Please note, only one score will be calculated per HAQ-DI assessment.

If the "other" category is checked on the HAQ-DI assessment, then this is expected to be an aide not associated for any of the existing categories, thus this will be excluded from the HAQ-DI score calculation.

Data resource: HAQ-DI is entered by the patient via site pad provided by ERT.

Change from baseline in HAQ-DI score over 48 weeks will be presented graphically.

4.2.4.18. Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) Score

The BASDAI consists of 6 questions to assess 5 major symptoms (fatigue, spinal pain, peripheral joint pain/swelling, areas of localized tenderness, and morning stiffness), with each answered on a 0- to 10-unit NRS. To give each of the 5 symptoms equal weighting, the mean of the two

scores relating to morning stiffness (questions 5 and 6) is taken. Each of the other 4 questions counts as 1 item by itself. The final BASDAI score is defined by calculating the mean of the 5 items. The BASDAI score ranges from 0 to 10, with higher scores reflecting greater disease activity. In case any of the scores for questions 1-4 is missing, or the scores for both questions 5 and 6 are missing, the BASDAI score will be set to missing. In case of a missing score for either question 5 or 6, the score for the other question answered will be used as the mean of questions 5 and 6.

Data resource: BASDAI is entered by patient via site pad provided by ERT.

BASDAI score over 48 weeks will be presented graphically.

4.2.4.19. Psoriatic Arthritis Impact of Disease 12-domain Questionnaire (PsAID-12)

The PsAID consists of 12 physical and psychological domains and is based on 0- to 10-unit NRS. Each domain is rated as a number between 0 and 10. The final PsAID score is the sum of each domain with different weighting and divided by 20 as below.

Final PsAID value =[(PsAID1 (pain) NRS value (range 0-10) x 3)
+ (PsAID2 (fatigue) NRS value (range 0-10) x 2)
+ (PsAID3 (skin) NRS value (range 0-10) x 2)
+ (PsAID4 (work and/or leisure activities) NRS value (range 0-10) x 2)
+ (PsAID5 (function) NRS value (range 0-10) x 2)
+ (PsAID6 (discomfort) NRS value (range 0-10) x 2)
+ (PsAID7 (sleep) NRS value (range 0-10) x 2)
+ (PsAID8 (coping) NRS value (range 0-10) x 1)
+ (PsAID9 (anxiety) NRS value (range 0-10) x 1)
+ (PsAID10 (embarrassment and/or shame) NRS value (range 0-10) ×1)
+ (PsAID11 (social life) NRS value (range 0-10) x 1)
+ (PsAID12 (depression) NRS value (range 0-10) x 1)]/20

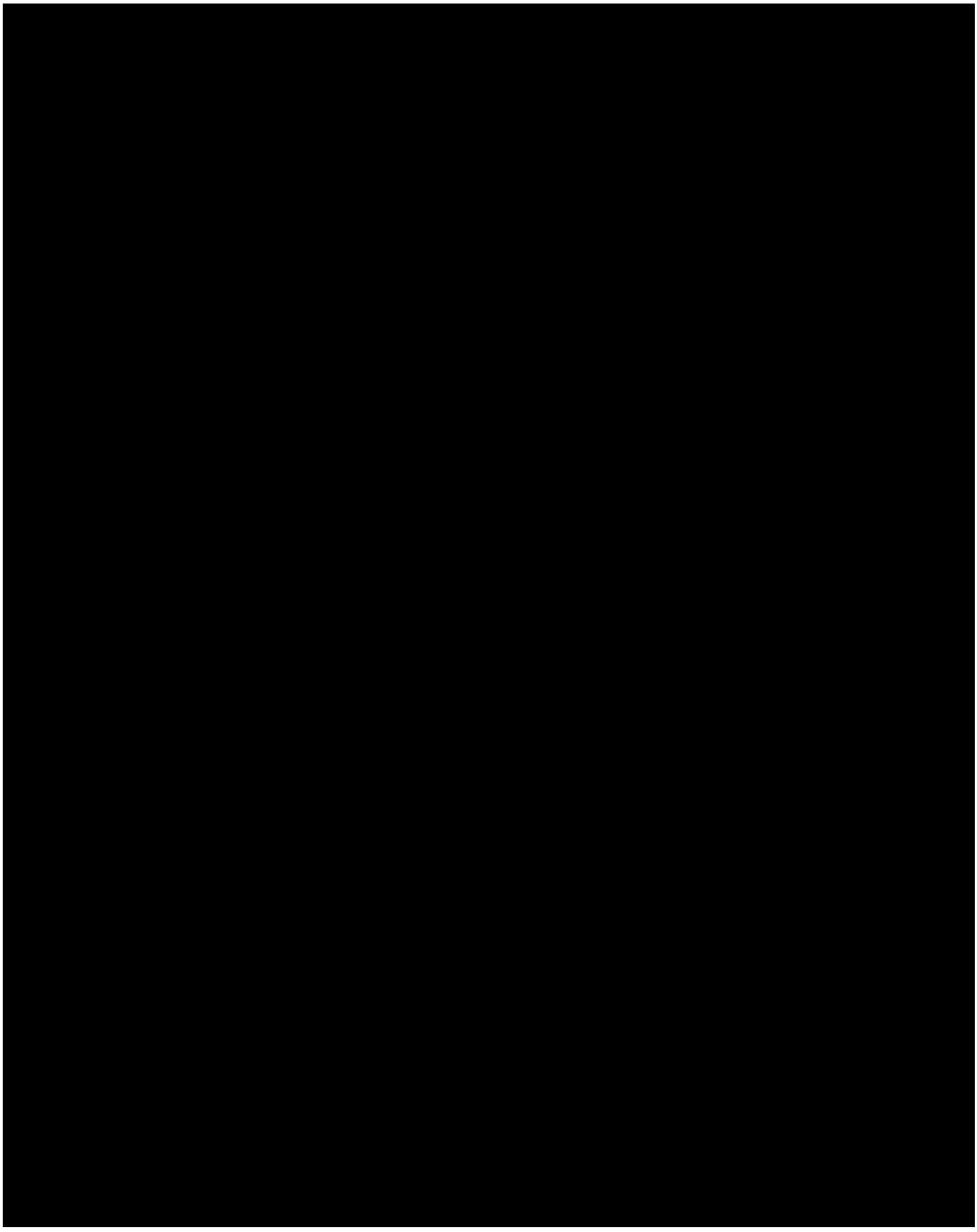
The final score has a range from 0 (best status) to 10 (worst status) with a cut-off of 4 (Gossec, 2014).

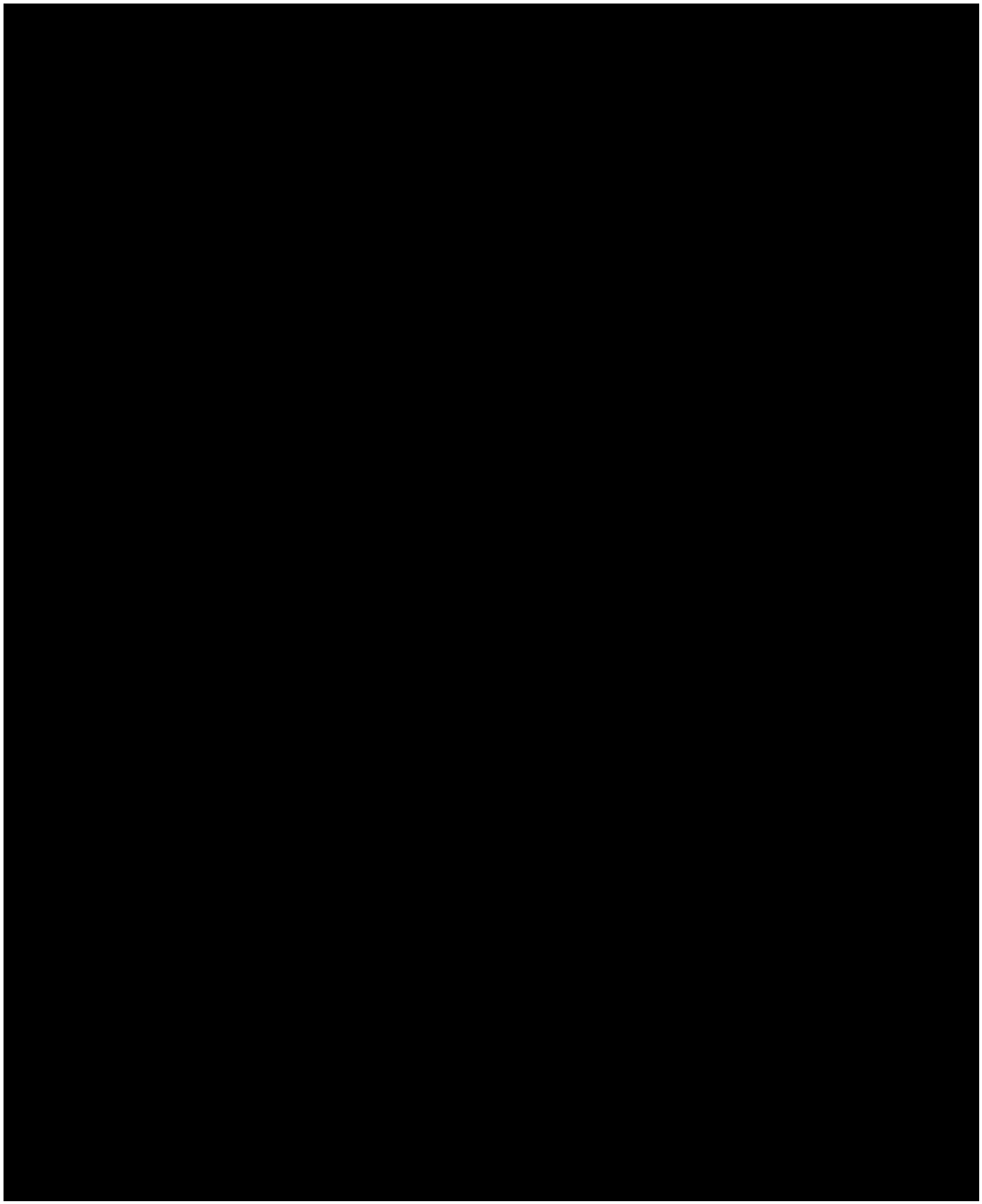
If one of the 12 domains composing the PsAID is missing, it will be imputed by the mean value of the 11 non-missing NRS scores. If two or more of the domains are missing, the final PsAID score will be considered as missing.

Data resource: PsAID-12 is entered by patient via site pad provided by ERT.

PsAID-12 score over 48 weeks will be presented graphically.

[redacted]





4.2.5. Derivations of Safety Endpoints

Baseline definition for all safety endpoints is given in Section 5.4. Change from baseline is calculated as on-treatment value minus the baseline value. Handling of time points is described in Section 5.5.

4.2.5.1. Treatment-Emergent Adverse Event

A treatment-emergent adverse event (TEAE) is defined as any AE that begins or worsens on or after the date of the first dose of investigational product (IP) and no later than 28 days after the

last dose of IP for subjects who have completed the study or have discontinued early by the time of database cut.

If the treatment-emergent status of an AE is unclear due to a missing/incomplete start date, it will always be considered treatment-emergent, unless shown otherwise by data. Date imputation rules for partially missing AE start dates are described in Appendix A1 1.1.

4.2.5.2. Treatment-emergent Adverse Events Leading to Drug Withdrawal, Leading to Drug Interruption, and Leading to Death, and Drug-related Treatment-emergent Adverse Events

A TEAE leading to drug withdrawal is a TEAE for which the investigator indicates that the action taken with respect to IP is withdrawn permanently (action taken on AE eCRF is recorded as "DRUG WITHDRAWN"). A TEAE leading to drug interruption is a TEAE for which the investigator indicates that the action taken with respect to IP is drug interrupted. A TEAE leading to death is a TEAE for which the outcome is fatal. Relationship to IP is based on the investigator's causality judgment; that is, a drug-related AE is an AE indicated by the investigator to have a suspected relationship to IP.

4.3. Sample Size and Power

While no MRI data are available for apremilast, assumptions used for the sample size calculation are primarily derived from the literature (Glinatsi, 2015). With an approximate 20% dropout rate, a sample size of 120 is required to produce a two-sided 95% confidence interval, for the expected change from baseline in the composite score of BME, synovitis, and tenosynovitis, with a distance from the mean to the limit equal to 1.24, assuming the standard deviation is 6.2.

5. GENERAL STATISTISTICAL CONSIDERATIONS

5.1. Reporting and Analysis Conventions

Summary statistics for continuous variables include sample size (n), mean, standard deviation (SD), median, minimum, the 25th (Q1) and 75th (Q3) percentiles, and maximum. All mean and median values will be formatted to one more decimal place than the measured value. SD values will be formatted to two more decimal places than the measured value. Minimum and maximum values will be presented to the same number of decimal places as the measured value.

Frequency summary for categorical variables includes number and percentage. All percentages will be rounded to one decimal place, unless presenting two decimal places is necessary when the percentages are very small. Number and percentage values will be presented as xx (xx.x%).

All analysis and summary tables will have the population sample size for the treatment in the column heading. P-values (2 sided) will be presented with 4 decimal places. All laboratory data except for hs-CRP will be reported using standard international (SI) units; hs-CRP will be reported in mg/dL.

Change from baseline will be calculated as (post-baseline minus baseline value).

Percent (%) change from baseline will be calculated as (post-baseline minus baseline value) divided by (baseline value), then multiply by 100. If a baseline value is zero, the percent change from baseline value will be excluded from the summaries.

Subject data listings will be provided to support the tables and graphs will be provided for selected parameters.

5.2. Definition of Analysis Populations

5.2.1. Full Analysis Set

The Full Analysis Set (FAS) will be the primary population for efficacy analyses. The FAS will consist of all subjects who are enrolled in the study. Subjects who are enrolled but do not receive any dose of IP will be excluded from FAS.

5.2.2. Modified Intent to Treat (mITT)

The mITT will consist of all subjects who are in FAS, have baseline and at least one postbaseline of primary efficacy endpoint PsAMRIS. mITT might be the population for the sensitivity analysis for the primary efficacy endpoint if needed.

5.2.3. Per-protocol Population

The Per Protocol (PP) population will consist of all subjects who receive at least one dose of study medication and have no major protocol violations which might affect primary efficacy endpoint. The final determination of major protocol violation criteria will be made prior to the database lock and will be documented.

5.2.4. Safety Population

The Safety population will consist of all subjects who received at least one dose of study medication.

5.3. Analysis Periods

In this study, the analysis periods are divided into open label treatment period and observational follow-up period below.

- Open label treatment period – Week 0 to Week 48

  This period starts on the day after the first IP is dispensed (Visit 2/Week 0), and stops on either: (1) the completion date of treatment period collected in the eCRF; or (2) the early termination date of treatment period collected in the eCRF; or (3) the last known study date if the subject is lost to follow-up.

- Observational follow-up period – 4 weeks

  Subjects who complete the study or early terminate from the study, will be followed up for 28 days after the Visit 7/Week 48 or Early Termination visit. The period starts at the end date of treatment period + 1 and stops on either (1) the completion date of follow-up period in eCRF; (2) the last known study date if the subject is lost to follow-up.

For efficacy analysis, only the open-label treatment period will be used.

For safety analysis, the open-label treatment period and observational follow-up period will be included into the data summary.

5.4. Baseline Definitions

For all MRI assessments (MRI of the most affected hand, WB-MRI, and [redacted]), the baseline is defined as nominal baseline result from IAG. For other assessments related to efficacy endpoints and baseline disease characteristics, the baseline is defined as the last value measured on or before the day of the first dose of IP.

For safety analyses, the baseline is defined as the last value measured on or before the day of the first dose of IP.

5.5. Time Points

Time points in all analyses are based on the visits/study weeks as recorded in the database. Appropriate dates (e.g., date of measurement or date of specimen collection) will first be used to ensure only data (including data from scheduled, unscheduled and discontinuation) measured or collected within the specific analysis period (Section 5.3) are included, and then the visits/study weeks as recorded in the database will be used to assign one value (possibly missing) to each time point at the subject level according to the following rules:

Scheduled Visits

- If the value at a scheduled visit is missing but a value is available at an associated unscheduled visit, the value from the associated unscheduled visit will be used for the

study week. If values from more than one associated unscheduled visit (e.g., Visits 8.01 and 8.02) are available, then the one from the first associated unscheduled visit (e.g., Visit 8.01) will be used for the study week.

- If the value at a scheduled visit is missing and there is no value available from an associated unscheduled visit, the value at the study week will be missing.
- Please note that unscheduled visits are recorded to one general visit code. The visit numbering for unscheduled visits will be re-numbered based on the unscheduled visit date (and time) comparison to the schedule visit date (and time). If the unscheduled record falls in between two scheduled visits, the re-numbered value will be assigned. For example, if the unscheduled visit date is "on or after visit 1" AND "prior to visit 2", then the re-numbered value will be 1.01. If there are multiple unscheduled visits in that interval, the unscheduled visit date (and time) will be used to ordered the re-number values like 1.01, 1.02, 1.03, so on.

Early Termination Visit

- A value from the early termination visit will be mapped to the next scheduled visit after the last visit the subject has completed.

Unscheduled Visit

- Please note that only vital signs, local lab data, central lab data and procedures MRI will have unscheduled visits and each unscheduled visit will be recorded to one specific visit code.

Analysis Visit Window for Safety Data

For safety analyses, all available values obtained between 2 planned visits including unscheduled measurements will be pooled to the analysis visit window according to the following definition:

Table 2: Analysis visit window for Safety Data

Visit Number	Visit Name	Target Day	Analysis Window for Safety Data
1	Screening (1)	<1	<1
2	Baseline (2)	1	1
3	Week 8 (3)	56	[2, 84]
4	Week 16 (4)	112	[85, 140]
5	Week 24 (5)	168	[141, 210]
6	Week 36 (6)	252	[211, 294]
7	Week 48 (7)	336	[295, 350]

For multiple measurements in the same visit category, the value measured nearest to the target day will be assigned to the visit; if they are at the same distance to the target day, the latest one will be used.

For the efficacy data, no analysis window will be mapped, and the nominal visit will be used.

6. STATISTICAL METHODOLOGY FOR EFFICACY

The analyses of efficacy endpoints will be primarily based on the FAS. In addition, supportive analyses using the Per Protocol population and mITT might be conducted for the primary and the key secondary efficacy endpoints.

6.1. Analysis of Primary Efficacy Endpoint

The primary endpoint of the study is the change from baseline to Week 24 in the composite score of BME, synovitis, and tenosynovitis assessed by PsAMRIS.

Descriptive statistics of the baseline, post-baseline, change from baseline and percent change from baseline values will also be presented by the scanner type and the combination of both scanner types for the FAS.

The boxplots of change from baseline of primary efficacy endpoint over week 48 will be presented graphically.

Sensitivity analyses

The supportive/sensitivity analyses of the primary endpoint will be performed with the same analysis model based on PP population and mITT population.

6.2. Analyses of Secondary Efficacy Endpoints

The secondary efficacy endpoints are divided into binary and continuous endpoints.

6.2.1. Analyses of Binary Secondary Endpoints

The binary secondary endpoints list is presented below:

- Proportion of subjects with baseline Leeds Enthesitis Index (LEI) whose LEI improves to 0 at Week 24 and Week 48

- Proportion of subjects with baseline SPARCC whose SPARCC improves to 0 at Week 24 and Week 48
- Proportion of subjects with baseline dactylitis whose tender dactylitis count improves to 0 at Week 24 and Week 48

For these binary endpoints, the number and percentages of subjects will be summarized by baseline and post-baseline visits, and the two-sided 95% Clopper-Pearson confidence intervals will be provided for the proportion. In addition, the proportion will be presented graphically. The number of subjects with missing values at each visit will be provided in a separate category.

6.2.2. Analyses of Continuous Secondary Endpoints

Due to the impact from MRI scanner type, the statistical method for the continuous secondary endpoints assessed by all MRI assessments (MRI of the most affected hand, WB-MRI, and DCE-MRI) will be different from other continuous secondary endpoints.

6.2.2.1. Analyses of Continuous Secondary Endpoints Assessed by MRI

The analyses of secondary endpoints assessed by all MRI assessments will be performed in a manner similar as the primary analysis for primary endpoint (refer to Section 6.1) due to the impact from scanner type and baseline score and time. [redacted]

6.2.2.2. Analyses of Other Continuous Secondary Endpoints

Descriptive statistics (n, mean, SD, median, Q1, Q3, min, max) for observed values and changes from baseline (including percent changes from baseline, if needed) will be presented at specified time points among subjects who have both values at baseline and the time point of interest. In addition, two-sided 95% CIs based on normal distribution for mean changes (and percent changes from baseline) will be provided at each time point of interest. In addition, data will be reported based on the proportion of subjects above or below the smallest detectable changes (SDC).

6.3. Analyses of Exploratory Efficacy Endpoints

[redacted]

6.4. Subgroup Analysis

To determine whether the treatment effect is consistent across various subgroups, the primary endpoint will be provided, where feasible, within each category of the following classification variables (subcategories under a specific category of a variable, where applicable, are indicated in brackets):

Demographic subgroups

- Sex (male, female)
- Age (< 65, ≥ 65 years)
- Baseline weight (< 70, ≥ 70 to < 85, ≥ 85 to < 100, ≥ 100 kg)
- Baseline BMI (< 25, ≥ 25 to < 30, ≥ 30 to < 35, ≥ 35 to < 40, ≥ 40 kg/m^2)
- Region (North America, Europe, Rest of the World [including Russia])
- Prior glucocorticosteroid use (yes, no)
- Prior use of MTX (Yes/No)
- Concomitant use of MTX (Yes/No)

Baseline disease subgroups

- Duration of PsA (≥3 months to ≤ 2, > 2 to ≤ 5 years)

The consistency of the treatment effect across subgroups will be assessed in the context of the primary methods used for the analyses with the entire FAS.

7. INTERIM ANALYSIS

No interim analysis will be conducted.

At the end of the study, after all subjects have completed the treatment period or early terminated, the final analysis will be performed, and a final clinical study report will be generated.

8. Summary of Subject Disposition

The number of subjects screened, the number and percentage of subjects, who complete screening, who do not complete screening, and who are screen failure among all subjects screened, and the eligibility criteria failed will be summarized. The above percentages will be based on the number of subjects screened.

For this single-arm, open-label study, the number and percentage of subjects who enroll, receive at least one dose of IP, complete, and discontinue, and the number and percentage of primary reasons for discontinuation (except for the observational follow-up) will be summarized. The number and percentage of subjects who enter/complete the observational follow-up, will also be provided. The primary reasons for discontinuation are collected in the study disposition eCRF and will be summarized with the following categories:

- Adverse event(s)
- Lack of efficacy
- Non-compliance with study drug
- Withdrawal by subject
- Study terminated by sponsor
- Lost to follow-up
- Pregnancy
- Death
- Protocol deviation
- Other

The number and percentage of subjects included in the FAS, mITT, PP, and safety populations will be summarized; the percentages will be based on the number of subjects enrolled. A listing of subjects excluded from the analysis populations, with the reasons for exclusions, will be provided.

The number and percentage of subjects by region, country, and study site will be tabulated based on the FAS.

9. Demographics and Baseline Characteristics

Demographics and baseline characteristics will be summarized descriptively for the FAS. Each relevant characteristic will be assessed descriptively in table format; no statistical hypothesis tests will be performed on these characteristics.

For summaries (number and percentage) of categorical variables in demographics, baseline disease characteristics, and alcohol and tobacco use, an "unknown" category may be added (in addition to the categories specified below) to account for missing data as appropriate.

9.1. Demographics

Summary statistics will be provided for the following continuous variables as appropriate:

- Age (years)
- Weight (kg)
- Height (cm)
- Body mass index (BMI; kg/m^2)

Number and percentage will be provided for the following categorical variables as appropriate:

- Age category (< 65, ≥ 65 years; < 40, 40 to < 65, 65 to < 75, 75 to < 85, ≥ 85 years)
- Sex (male, female)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islanders, White, Other)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Region (North America, Europe, Rest of the World [including Russia])
- Weight category (< 70, ≥ 70 to < 85, ≥ 85 to < 100, ≥ 100 kg)
- BMI category (< 25, 25 to < 30, 30 to < 35, 35 to < 40, ≥ 40 kg/m^2)

9.2. Baseline or Disease Characteristics

Summary statistics will be provided for the following continuous variables as appropriate:

- Duration of PsA (from the date of diagnosis to the date of informed consent; years)
- Psoriasis body surface area (BSA; %)
- Inflammation score of BME, synovitis, and tenosynovitis assessed by PsAMRIS
- Composite score of BME and synovitis assessed by PsAMRIS
- BME score assessed by PsAMRIS
- Synovitis score assessed by PsAMRIS

- Tenosynovitis score assessed by PsAMRIS
- Periarticular inflammation score assessed by PsAMRISTotal Inflammation score
- Total damage score (bone erosion + bone proliferation) assessed by PsAMRIS
- Bone erosion score assessed by PsAMRIS
- Bone proliferation score assessed by PsAMRIS
- Tender joint count
- Swollen joint count
- C-DAPSA score
- SPARCC score
- SPARCC score (for subjects with pre-existing enthesopathy)
- LEI score
- LEI score (for subjects with pre-existing enthesopathy)
- LDI score
- LDI score (for subjects with pre-existing dactylitis)
- PASDAS score
- Physician global assessment of disease activity (0-10 NRS)
- Subject global assessment of disease activity (0-10 NRS)
- Subject assessment of pain (0-10 NRS)
- HAQ-DI score
- Peripheral enthesitis inflammation index assessed by WB-MRI
- Peripheral joints inflammation index assessed by WB-MRI
- Total peripheral (joints and enthesitis) inflammation index assessed by WB-MRI
- BASDAI score (for subjects deemed to have PsA spondylitis by the investigator and with BASDAI item 2 $\geq$ 4 at baseline)
- PsAID-12 score

[redacted]

Number and percentage will be provided for the following categorical variables (subcategories under a specific category of a variable, where applicable, are indicated in square brackets) as appropriate:

- Baseline use of methotrexate (MTX) (yes, no)
- Psoriasis BSA ≥ 3% (yes, no)
- Prior use of NSAIDs
- Prior use of low-dose oral glucocorticoids (prednisone ≤ 10 mg/day or equivalent)
- Prior use of narcotic analgesics (yes, no)
- Rheumatoid factor (normal, high)
- Pre-existing SPARCC enthesopathy (yes, no)
- Pre-existing LEI enthesopathy (yes, no)
- Pre-existing dactylitis (yes, no)
- PsA spondylitis assessed by investigator (yes, no)

- BASDAI iterm 2 ≥ 4 (yes, no)
- PsA spondylitis assessed by investigator and BASDAI iterm 2 ≥ 4 (yes, no)
- Prior csDMARD usage (yes, no)
- Prior biologic treatment (yes, no)
- C-DAPSA (<=4, >4 to <=13, >13 to <=27, >27)
- As based on the trial design (Section 4.1) and permitted concomitant medications and procedures from protocol, MTX (≤ 25 mg/week) will be permitted if treatment duration is ≥ 6 months and on a stable regimen for at least 3 months prior to baseline. The NSAIDs, narcotic analgesics and low-dose oral glucocorticoids (prednisone ≤ 10 mg/day or equivalent) must be on a stable regimen for at least 4 weeks prior to baseline. Thus, baseline use of MTX is defined as MTX (≤ 25 mg/week) used as a stable regimen for at least 3 months before the first dose of IP. For the NSAIDs, narcotic analgesics and low-dose oral glucocorticoids medications, the baseline use of these medications are defined as medications used as a stable regimen for at least 4 weeks before the first dose of IP.

9.3. Medical History

Medical history will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA; Version 21.0). A frequency summary (number and percentage) of medical history will be presented by system organ class (SOC), and preferred term (PT).

9.4. Prior Medications

Prior medications will be collected from prior and concomitant medications and history of psoriatic arthritis medications panels in the CRF pages. The prior medications are defined as medications that were taken before the first dose of IP. The Anatomical Therapeutic Chemical (ATC) coding scheme of the World Health Organization Drug Dictionary (WHODD; Version March 2018) will be used to group prior medications and history of psoriatic arthritis medications into relevant categories. Frequency summaries will be provided by ATC 2 level, and standardized medication name for the FAS population.

9.5. Prior Procedures

Prior procedures are defined as the procedures that occurred before the first dose of IP. The prior procedures will be coded using the Medical Dictionary for Regulatory Activities (MedDRA; Version 21.0). Frequency summaries of prior procedures will be provided for the FAS population by SOC, and PT.

10. Study Treatments and Extent of Exposure

Overall treatment duration will be summarized using descriptive statistics. Individual subject listings will be provided to support the tables.

10.1. Treatment Duration

Summary statistics for treatment duration (in weeks), as well as a frequency summary of treatment duration categories (e.g., < 8 weeks, ≥ 8 to < 16 weeks, etc.) will be summarized for FAS and safety population. A subject data listing of study drug exposure will be also provided.

Treatment duration (in weeks) will be calculated as (the date of the last dose of IP - the date of first dose of IP +1)/7 and rounded to one decimal place. The specific definitions of the first and last dose dates are given below:

- First dose date: the date of the first dose of IP at Visit 2/ Week 0.
- Last dose date:
    - The date of last dose of IP captured at Visit 7/Week 48 or Early Termination Visit, whichever occurred first.
    - The date prior to Visit 7/Week 48, if the date of last dose of IP is unknown at Visit 7/Week 48
    - The date prior to Early Termination Visit, if the date of last dose of IP is unknown at Early Termination Visit
    - The last known study date if the subject is lost to follow-up

Imputation rules for partially or completely missing last dose date is specified in Appendix A1 1.

10.2. Treatment Compliance

Treatment compliance (%) will be calculated for each subject based on the number of tablets dispended and the number of tablets returned at each visit (except Visit 1). Treatment compliance will be summarized.

The treatment compliance (%) for each subject is calculated as below:

(100* the total number of tablets taken)/ the intended total number of tablets that should have been taken over the same period

Where the total number of tablets taken = the total number of tablets dispensed – the total number of tablets returned. The intended total number of tablets that should have been taken over the same period is [(last dose date in the period-first dose date in the period)+1]*2 – 1 since the patient only took one dose at study day 1.

Please note that treatment compliance will not be calculated for subjects who have only the dispense record at Visit 2/Week 0 and no other drug accountability records.

Additionally, the number and percentage of subjects will be summarized by compliance categories (e.g., < 75%, ≥ 75% to ≤ 120%, and > 120%). A subject data listing of drug accountability records will be provided.

11. Protocol Deviations/Violations

The protocol deviations or violations will be identified and assessed by clinical research physician (CRP) or designee throughout all study periods. A list of protocol deviations and important protocol deviations for all subjects for final analysis will be defined and finalized prior to the beginning of any data analyses or the final database lock, whichever occurred first. This listing will also identify which subjects are to be excluded from the per-protocol population.

Protocol deviations and important protocol deviations will be summarized for all subjects and FAS population. Summary tables showing the number and percent of subjects with at least one protocol deviation and important protocol deviations and by each category of protocol deviations and important protocol deviations will be provided. Summary of protocol deviations related to COVID-19 control measures will be provided. Listings of subjects with protocol deviations and important protocol deviations will also be provided.

12. SAFETY ANALYSIS

Safety will be evaluated via descriptive statistics and point estimates. Safety summaries will be provided based on all the subjects in the safety population . Individual subject listings will be provided to support the summary tables.

For the analyses of AEs and marked abnormalities, the following point estimates are provided, unless otherwise specified:

- Subject incidence: Subject incidence (i.e., percentage [%] used in a frequency summary) is defined as the number of subjects with the specific event divided by the number of subjects included in the analysis, multiplied by 100. Subjects with multiple occurrences of a specific event in a defined exposure interval will be counted only once in the numerator.
- Exposure-adjusted incidence rate (EAIR) per 100 subject-years: The EAIR per 100 subject-years is defined as 100 times the number of subjects with the specific event divided by the total exposure time (in years) among subjects included in the analysis. Subjects with multiple occurrences of the specific event in a defined exposure interval will be counted only once in the numerator. The exposure time for a subject without the specific event is the treatment duration, whereas the exposure time for a subject with the specific event is the treatment duration up to the start date (inclusive) of the first occurrence of the specific event. The total exposure time in years is calculated by dividing the sum of exposure time in days over all subjects included in the analysis by 364. The EAIR per 100 subject-years is interpreted as the expected number of subjects with at least one occurrence of the specific event per 100 subject-years of exposure to the IP.

In summaries of continuous variables with summary statistics by time point, values for baseline, time point, and change from baseline will be summarized at each time point for subjects who have values at both baseline and the specific post-baseline time point. Frequency summaries of shifts from baseline to post-baseline time points will be provided for subjects who have values at both baseline and each individual specific post-baseline time point. In addition to the scheduled visits, shift summaries will also be provided from baseline to the end of period and to the worst value in a specific post-baseline period.

12.1. Adverse Events

AEs will be coded according to the MedDRA Version 21.0. Unless otherwise specified, AEs will be summarized by SOC and PT, with SOCs presented in the standard international order and PTs within each SOC in descending order of subject incidence.

A subject data listing of all AEs (including TEAEs and non-TEAEs) will be provided.

12.1.1. Overall Summary of Adverse Events

An overall summary of the following AE categories will be provided:

- Any TEAE
- Any drug-related TEAE
- Any severe TEAE
- Any serious TEAE
- Any serious drug-related TEAE
- Any TEAE leading to drug interruption
- Any TEAE leading to drug withdrawal
- Any TEAE leading to death

12.1.2. All TEAEs

Summarization of all TEAEs by the following categories will be provided as appropriate:

- Any TEAEs (by SOC and PT, and by PT only)
- Any TEAEs by exposure interval (≤ 1, > 1 to ≤ 4, > 4 to ≤ 12, > 12)
- Any TEAEs by age category (<65, ≥ 65 years) andgender
- Any TEAEs by prior methotrexate use (Yes/No)
- Any TEAE by concomitant methotrexate use (Yes/No)

For the summary of any TEAEs by exposure interval, the AE data will be handled per the following rules, unless otherwise specified.

- An event with a start date falling into an exposure interval is considered a new event for that interval.
- Each subject is counted once for either subject incidence or EAIR for each applicable specific TEAE in each exposure interval where an event started.
- The denominator of a subject incidence is the number of subjects with treatment duration exceeding the lower bound of the particular exposure interval, while the denominator of an EAIR is the sum of the exposure time during the exposure interval (up to the first event start date for subjects with at least one event starting in the interval) among the same number of subjects as in the denominator of the corresponding subject incidence.

Summarization of the TEAEs with the following specifications will be provided as appropriate:

- Most frequent TEAEs (≥ 5%; by SOC and PT, and by PT only in descending order of subject incidence)
- Most frequent TEAEs by onset interval and event duration (for special interest of PTs: Diarrhoea, Headache, Nasopharyngitis, Nausea, Hypertension, and Upper respiratory tract infection)
- Any drug-related TEAEs (by SOC and PT)
- Any drug-related TEAEs by exposure interval (refer to Section 12.1.2)
- Any serious TEAEs (by SOC and PT)
- Any serious TEAEs by exposure interval (refer to Section 12.1.2)
- Any serious TEAEs by age category (<65, ≥ 65 years) and gender
- Any serious drug-related TEAEs (by SOC and PT)
- Any serious drug-related TEAEs by exposure interval (refer to Section 12.1.2)
- TEAEs by maximum severity (by SOC and PT)
- Any TEAEs leading to study drug withdrawal (by SOC and PT)
- Any TEAEs leading to study drug withdrawal by age category (<65, ≥ 65 years) and gender
- Any TEAEs leading to study drug withdrawal by prior use of psoriatic arthritis medications
- Any TEAE leading to study drug withdrawal by concomitant use of psoriatic arthritis medications
- Any TEAEs leading to study drug interruption (by SOC and PT)
- Any TEAEs leading to death (by SOC and PT)

For drug-related TEAEs, the relationship of an AE to IP is recorded on the Adverse Events - Serious Adverse Events eCRF page as “NOT SUSPECTED” and “SUSPECTED”. Adverse events with missing relationship to IP will be assessed as drug-related.

For TEAEs by maximum severity, if a subject reports multiple occurrence of a specific event within a specific analysis period, the subject will be counted only once by the maximum severity (mild, moderate, severe, or, if needed, missing). If a subject reports multiple TEAEs within an SOC, the subject will be counted only once for that SOC at the maximum severity. If the severity

is missing for one or more of the occurrences/TEAEs, the maximum severity of the remaining occurrences will be used. If the severity is missing for all the occurrences/TEAEs, the subject will be counted only once in the "missing" category of severity.

In addition, subject data listings for serious AEs (both TEAEs and non-TEAEs), TEAEs leading to drug withdrawal, and deaths will be provided.

12.2. Adverse Events of Special Interest

Adverse events of special interest (AESI) include diarrhea.

Protocol defined diarrhea AEs, referring to events with 2 or more watery or liquid stools in a day, will be captured on the site pad by ERT. Duration and frequency of the individual diarrhea will be captured. The diarrheas reported on the site pad will need to be reconciled with the events captured on Medidata's RaveX EDC platform.

Frequencies (Daily, 5 – 6 days per week, 2 – 4 days per week, once every other week to once a week) of these events will be summarized.

A listing of AESIs will be also provided.

12.3. Clinical Laboratory Evaluations

Clinical laboratory evaluations will be available from the central laboratory and include:

- Hematology: complete blood count (CBC) (red blood cell [RBC] count, hemoglobin, hematocrit, white blood cell [WBC] count and differential, absolute WBC counts, platelet count).
- Serum chemistries: total protein, albumin, calcium, phosphorous, glucose, triglycerides, total cholesterol [TC], total and direct bilirubin, alkaline phosphatase, AST, ALT], sodium, potassium, chloride, bicarbonate [carbon dioxide, CO_2], blood urea nitrogen [BUN], creatinine, lactate dehydrogenase [LDH], and magnesium.
- Urinalysis: specific gravity, pH, glucose, ketones, protein, blood, bilirubin, leukocyte esterase, nitrite, and urobilinogen; if urinalysis is abnormal, microscopic urinalysis (epithelial cells, RBC, WBC, and casts) will be performed.

Summary statistics of observed values and changes from baseline in hematology and serum chemistry parameters will be provided over time. Frequency summaries (shift tables) of shifts from baseline to post-baseline time points and to the worst post-baseline value, in terms of low/normal/high/both low and high, will be provided for hematology and serum chemistry parameters. Subject data listings of all laboratory data, including urinalysis, will be provided.

Summary of the categories of laboratory marked abnormalities will be provided based on the marked abnormalities criteria. Subject incidence and EAIR for each abnormality will be calculated based on subjects with a baseline value and at least one post-baseline value for criteria

requiring baseline or subjects with at least one post-baseline value for criteria not requiring baseline.

A separate summary of laboratory marked abnormalities will also be presented by categories of subjects with normal values at baseline and subjects with abnormal values at baseline. To the purpose of these summaries, subjects with abnormal (normal) values at baseline are defined as those whose baseline value is low (not low) for criteria concerning low values and those whose baseline value is high (not high) for criteria concerning high values; both low and high are relative to the laboratory normal range.

A subject data listing of laboratory marked abnormalities will be provided. In the listing, all recordings of a lab analyte will be displayed for the subject if that subject has one or more marked abnormalities for the test.

For comorbid conditions of interest or a subgroup of subjects, summary statistics of relevant laboratory parameters over time may be provided. A subject listing will be also provided for all available pregnancy test results.

12.4. Vital Sign Measurements and Weight

For vital signs and weight, summary statistics (n, mean, standard deviation, median, minimum, the 25th (Q1) and 75th (Q3) percentiles, maximum, percent change from baseline for weight only) of observed and change from baseline values will be presented.

For vital signs (excluding weight), Frequency summaries (shift tables) of shifts from baseline to post-baseline or worst post-baseline values (low/normal/high/both low and high) during the treatment in below, within and above the normal ranges will be displayed in cross–tabulations.

For weight, change and percent change from baseline to the end of the study will be presented, as well as by baseline BMI and baseline weight categories.

12.5. Physical Examination

Physical examination findings will be captured as medical history (screening) or abnormal physical examination findings will be reported as adverse events and will be reported in those summaries.

12.6. Concomitant Medications/Procedures

Concomitant medications and procedures will be summarized for the safety population. Subject data listings for concomitant medications and procedures will be also provided.

12.6.1. Concomitant Medications

Concomitant medications are defined as medications that administered on or after the first dose of the IP or those started prior to the first dose of the IP but stopped on or after the date of the first dose of IP, or recorded as ongoing. The Anatomical Therapeutical Chemical (ATC) coding scheme of the World Health Organization Drug Dictionary (WHODD; Version March 2018) will

be used to group medications into relevant categories. A frequency summary of concomitant medications will be provided by ATC2 level and standardized medication name.

12.6.2. Concomitant Procedures

Concomitant procedures are defined as procedures that were done on or after the first dose of the IP or those started prior to the first dose of the IP but stopped on or after the date of the first dose of IP, or recorded as ongoing. Concomitant procedures will be coded according to MedDRA Version 21.0. A frequency summary of concomitant procedures will be provided by SOC and preferred term.

13. REFERENCES

Aletaha D, Alasti F, Smolen JS. Disease activity states of the DAPSA, a psoriatic arthritis specific instrument, are valid against functional status and structural progression. Ann Rheum Dis. 2017;76(2):418-21.

Glinatsi D, Bird P, Gandjbakhch F, Mease PJ, Boyesen P, Peterfy CG, et al. Validation of the OMERACT Psoriatic Arthritis Magnetic Resonance Imaging Score (PsAMRIS) for the hand and foot in a randomized placebo-controlled trial. J Rheumatol. 2015;42(12):2473-9.

APPENDIX A1

Following are the general conventions for various computations and imputations for references. Users may need to consult with study team for specific study practices or regulatory guidelines.

1. Guideline of Partially Missing Date Imputation

1.1 Adverse Events

The principle of the imputation rules for subjects who were treated with apremilast initially is to treat the AE as treatment-emergent, i.e., occurring on or after the date of the first dose of IP, if possible.

Partially missing AE start date will be imputed in the ADaM dataset for AEs, but partially missing AE end dates will not be imputed in the same dataset. If the AE end date is complete with no missing year, month, or day, and the partially missing start date imputed by the rules below is after the AE end date, then the start date will be imputed by the AE end date.

The following scenarios are considered in the imputation rules:

- If AE start day is missing, and AE start month and year are not missing:
    - If AE start year and month are the same as the year and month of first dose, then impute AE start day using the day of first dose date.
    - If AE start date is less than the first dose date, then impute AE start day using the last day of AE start month.
    - If AE start date is greater than the first dose date, then impute AE start day using the first day of AE start month.
- If both AE start month and day are missing, and AE start year is not missing:
    - If AE start year is less than the year of first dose, then impute AE start date using the last day of the AE start year.
    - If AE start year is equal to the year of first dose, then impute AE start date using the first dose date.
    - If AE start year is greater than the year of first dose, then impute AE start date using the first day of the AE start year.

1.2 Prior/Concomitant Medications and Procedures

Partially missing start/stop dates for prior/concomitant medications and partially missing start dates for prior/concomitant procedures will be imputed in the ADaM dataset for prior/concomitant medications/procedures.

For prior/concomitant medications, if the stop date is complete with no missing year, month, or day, and the partially missing start date imputed by the rule below is after the stop date, then the start date will be imputed by the stop date.

Partially missing prior/concomitant medication/procedure start dates will be imputed by the earliest possible date given the non-missing field(s) of the date.

Partially missing prior/concomitant medication stop dates will be imputed by the latest possible date given the non-missing field(s) of the date.

1.3 Medical History

Partially missing medical history start dates will be imputed in the ADaM dataset for medical history (for the purposes of calculating durations of PsA and psoriasis). The 16_{th} of the month will be used to impute a partially missing start date that has only the day missing, and July 1st will be used to impute a partially missing start date that has both the month and day missing.

1.4 Treatment Duration

Partially or completely missing last dose dates will be imputed in the ADaM dataset for treatment duration. When partially missing last dose date is available, set last dose date to the maximum of [the earliest possible date given the non-missing field(s) of last dose date, the minimum of (the latest possible date given the non-missing field(s) of last dose date, last known date in database, first non-missing Early Termination (ET) visit date)] When last dose date is completely missing, set last dose date to the minimum of (last known date in database, first non-missing Early Termination (ET) visit date) Last known date in database is defined as maximum of (last visit date, lab, vital signs, AE start or end dates, concomitant medications start or end dates, concomitant procedure date, last dose date from 'Disposition- Treatment' page, treatment exposure start date or end date where doses were completely or partially taken, death date).

2. Marked Abnormalities Criteria

Table 8: Marked Abnormalities Criteria

Category / Analyte	SI Units	Criteria
Chemistry/ ALT/AST	U/L	3*ULN
Alanine Aminotransferase (SGPT)	U/L	3*ULN
Albumin	kg/m^3	<25
Alkaline Phosphatase	U/L	> 400
Aspartate Aminotransferase (SGOT)	U/L	3*ULN
Bilirubin	µmol/L	> 1.8xULN
Bilirubin and ALT/AST	umol/L U/L	(ALT or AST > 3xULN) and bilirubin > 2xULN
Blood Urea Nitrogen	mmol/L	>15
Calcium	mmol/L	1. < 1.8 2. > 3.0
Cholesterol	mmol/L	> 7.8
Creatinine	µmol/L	> 1.7*ULN
Glucose	mmol/L	1. < 2.8 2. > 13.9
Lactate Dehydrogenase	U/L	> 3*ULN
Magnesium	mmol/L	> 1.2
Phosphate	mmol/L	1. < 0.64 2. >1.6
Potassium	mmol/L	1. < 3.0 2. > 5.5
Sodium	mmol/L	1. < 130 2. > 150
Urate	umol/L	Male: > 590; Female: > 480
Hematology/ Basophils	10^9/L	> 3
Eosinophils	10^9/L	1. > 0.8 2. Increase ≥ 100% and value > 0.8

Category / Analyte	SI Units	Criteria
Hematocrit	Ratio	1. < 0.27 2. Male: > 0.55; Female: > 0.50
Hemoglobin	g/dL	1. Male: < 10.5 or Female < 8.5 2. Male: > 18.5 or Female > 17
Hemoglobin A1C	%	> 9
Leukocytes	10^9/L	< 1.5
Lymphocytes	10^9/L	< 0.8
Neutrophils	10^9/L	< 1
Platelets	10^9/L	1. < 75 2. > 600

ULN = upper limit of normal