STATISTICAL ANALYSIS PLAN

A Multicenter Double-blind, Randomized Controlled Study of Etanercept and Methotrexate in Combination or as Monotherapy in Subjects With Psoriatic Arthritis

Protocol Number: **20130207**

Version: **Version 2.0**

Date: **06 June 2018**

Authors: PPD

NCT Number: 02376790
This NCT number has been applied to the document
for purposes of posting on clinicaltrials.gov

Table of Contents

List of Figures

List of Appendices

Table of Abbreviations

Abbreviation or Term	Definition/Explanation
ACR	American College of Rheumatology
ANCOVA	Analysis of covariance
BMI	body mass index
BSA	body surface area
CDAI	Clinical Disease Activity Index
CMH	Cochran Mantel Haenszel
CRP	C-reactive protein
DAS28-CRP	Disease Activity Score (28 Joints) using CRP
DAPSA	Disease Activity index for Psoriatic Arthritis
DMARD	Disease Modifying Anti-rheumatic Drug
EOS	End of Study
ES	Erosion Score
ET	Early Termination
HAQ-DI	Health Assessment Questionnaire Disability Index
ICH	International Conference on Harmonisation
IP	Investigational product
IPD	Important protocol deviation
ITT	Intent-to-treat
JSN	Joint Space Narrowing
LOCF	Last observation carried forward
LDI	Leeds Dactylitis Index
MDA	minimal disease activity
mNAPSI	Modified Nail Psoriasis Severity Index
mTSS	van der Heijde modified Total Sharp Score
MI	Multiple imputation
NRI	Non-responder imputation
PASDAS	Psoriatic Arthritis Disease Activity Score
PRO	patient reported outcome
PsA	psoriatic arthritis
QW	Weekly
SAP	statistical analysis plan
SDAI	Simplified Disease Activity Index
SF-36 v2	Medical Outcomes Health Survey Short Form 36 items version 2

Abbreviation or Term	Definition/Explanation
SPARCC	Spondyloarthritis Research Consortium of Canada
sPGA	Static Physician Global Assessment
TNF	tumor necrosis factor
VAS	visual analog scale
VLDA	Very Low Disease Activity

1. Introduction

The purpose of this statistical analysis plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol amendment 3 for etanercept Study 20130207 dated 31 August 2016. The scope of this plan includes the final analysis that is planned and will be executed by the Biostatistics department unless otherwise specified.

2. Objectives

2.1 Primary Objective

To evaluate the efficacy of etanercept plus methotrexate therapy and etanercept monotherapy compared to methotrexate monotherapy, in subjects with psoriatic arthritis (PsA) as measured by the proportion of subjects achieving an American College of Rheumatology (ACR) 20 response at week 24.

2.2 Key Secondary Objective

To evaluate the efficacy of etanercept plus methotrexate therapy and etanercept monotherapy compared to methotrexate monotherapy as measured by the proportion of subjects achieving minimal disease activity (MDA) at week 24.

2.3 Other Secondary Objectives

To evaluate the efficacy of etanercept plus methotrexate therapy and etanercept monotherapy compared to methotrexate monotherapy on the following:

- Other measures of arthritis activity
- Measures of non-arthritic PsA disease activity
- Key patient reported outcomes (PRO) related to physical function and quality of life

2.4 Safety Objectives

To evaluate the safety of etanercept and methotrexate.

2.5 Exploratory Objectives

To evaluate the efficacy of etanercept plus methotrexate therapy and etanercept monotherapy compared to methotrexate monotherapy on the following:

- Other measures of disease activity including the skin
- Other PROs
- Inhibition of radiographic progression as measured by the van der Heijde modified Total Sharp score (mTSS) for PsA
- To investigate potential biomarkers of disease activity and response to etanercept
- To investigate the effects of genetic variation in disease genes and drug target genes on PsA and/or subject response to etanercept

3. Study Overview

3.1 Study Design

This is a phase 3, multicenter, randomized, double-blind controlled study in subjects with active PsA naïve to biologics and with no prior use of methotrexate for the treatment of PsA. The study will consist of a 30-day screening period, a 48-week randomized double-blind treatment period, and a 30-day safety follow-up period.

Approximately 840 subjects will be randomly assigned in a 1:1:1 ratio to one of three treatment groups (280 per arm):

- Etanercept 50 mg weekly by subcutaneous injection plus oral methotrexate 20 mg weekly
- Etanercept 50 mg weekly by subcutaneous injection plus oral placebo for methotrexate
- Oral methotrexate 20 mg weekly plus placebo for etanercept.

At or after week 24, subjects who have an inadequate response will receive rescue therapy with etanercept plus methotrexate until the end of the treatment period. Inadequate response is defined by having < 20% improvement in their tender joint counts and < 20% improvement in their swollen joint counts from baseline at or after week 24. Inadequate response may be assessed at a regularly scheduled visit or at a disease assessment visit. All subjects that initiate rescue treatment must complete a disease assessment follow-up visit 4 weeks later to monitor for labs.

3.2 Sample Size

A total of 840 subjects will be enrolled in the study and randomized in a 1:1:1 ratio to etanercept plus methotrexate therapy, etanercept monotherapy, and methotrexate monotherapy.

To preserve the family-wise 2-sided type one error rate at 0.05 for the multiple comparisons of the etanercept plus methotrexate therapy and etanercept monotherapy groups with methotrexate monotherapy, a Bonferroni-based gatekeeping chain procedure (Burman et al 2009 and Millen et al 2011) will be used to determine statistical significance for the primary and key secondary endpoints as described in Section 10.5.1. This procedure will split alpha of 0.05 equally to test in parallel the primary and key secondary endpoints (ACR 20, MDA) for etanercept plus methotrexate therapy vs. methotrexate monotherapy and etanercept monotherapy vs. methotrexate monotherapy sequentially within each parallel path. If one of the parallel paths rejects both hypotheses sequentially, and the other parallel path has at least one hypothesis not

rejected, then the unspent alpha of 0.025 from the successful path will be propagated to the hypotheses in the other path to re-test them sequentially at a level of 0.05.

The sample size is based on the adequacy to evaluate the efficacy of etanercept monotherapy compared with methotrexate monotherapy as measured by the primary endpoint of ACR 20 response at week 24 at a significance level of 0.025 based on the Bonferroni-based gatekeeping chain procedure.

There are no published studies of etanercept monotherapy versus methotrexate monotherapy for the treatment of PsA. Therefore for the purpose of this study the assumed response rates for ACR 20 for the etanercept monotherapy and methotrexate monotherapy arms are derived from separate studies. In the etanercept PsA pivotal phase 3 study (20021630 Study) the ACR 20 response rate at week 24 was 50% for the etanercept arm. However, subjects were allowed to be on a stable dose of methotrexate at study entry. Assuming a methotrexate naïve population would respond better, the response rate for the etanercept monotherapy arm in this planned phase 3 study is assumed to be 60%.

In a recent randomized placebo-controlled trial of methotrexate monotherapy in PsA (Kingsley et al, 2012) the ACR 20 response rate for the methotrexate monotherapy arm was 34%. However, the dose of methotrexate was 15 mg QW. Assuming a higher dose of methotrexate (ie, 20 mg QW) would yield better responses, the response rate for the methotrexate monotherapy arm in this planned phase 3 study is assumed to be 44%.

Therefore, in order to detect a difference of 16% between the etanercept monotherapy and methotrexate monotherapy arms, using a two-sided Chi-square test at a significance level of 0.025 and 90% marginal power, each treatment arm will have approximately 280 subjects adjusting for an anticipated 10% dropout.

This sample size will provide marginal power > 90% to detect a treatment difference in ACR 20 response at week 24 between the etanercept plus methotrexate therapy arm and the methotrexate monotherapy arm at a two-sided 0.025 significance level. As previously mentioned, the ACR response rate at week 24 is assumed to be 44% for the methotrexate monotherapy arm. The ACR response rate at week 24 for the etanercept plus methotrexate therapy arm is assumed to be 5% higher than the etanercept monotherapy arm (ie, 65%). This is based on a longitudinal observational study assessing the role of TNF-inhibitor plus methotrexate combination therapy versus TNF-inhibitor monotherapy (Fagerli et al, 2014).

In addition, this sample size will provide marginal power > 90% to detect treatment differences in MDA response at week 24 between both etanercept arms and the methotrexate arm at a two-sided 0.025 significance level, assuming response rates of 15%, 30%, and 35% for the methotrexate monotherapy, etanercept monotherapy, and etanercept plus methotrexate therapy arms, respectively. In recent PsA trials, the MDA 24-week response rates for the placebo arm range from 3 to 7% (Coates et al, 2010a; Kavanaugh et al, 2013). Therefore, for the methotrexate monotherapy arm, the response rate is assumed to be higher at 15%. In the pivotal phase 3 randomized placebo-controlled trial of etanercept in PsA, the MDA response rate for the etanercept monotherapy arm is estimated to be approximately 20% at week 24 (data on file). For a methotrexate naïve population, the response rate is assumed to be higher at 30%. In addition, the etanercept plus methotrexate therapy arm is assumed to have a 5% improvement over the etanercept monotherapy arm.

4. Study Endpoints and Covariates

4.1 Study Endpoints

4.1.1 Primary Endpoint

- ACR 20 response at week 24

4.1.2 Key Secondary Endpoint

- MDA response at week 24

4.1.3 Other Secondary Endpoints

- ACR 20 response at all measured time points other than week 24
- MDA response at all measured time points other than week 24
- Psoriatic Arthritis Disease Activity Score (PASDAS) at all measured time points and change from baseline at all measured time points
- ACR 50, ACR 70, and components of ACR at all measured time points
- Simplified Disease Activity Index (SDAI) and change from baseline at all measured time points
- Clinical Disease Activity Index (CDAI) and change from baseline at all measured time points
- Disease activity score (28 joint) using the C-reactive protein formula (DAS28-CRP) and change from baseline at all measured time points
- Health Assessment Questionnaire Disability Index (HAQ-DI) score and change from baseline at week 24
- Short Form (36) Health Survey Version 2 (SF-36v2) (MCS and PCS) score and change from baseline at week 24

For subjects with non-zero mNAPSI at baseline,

- mNAPSI and change from baseline at week 24
- achievement of mNAPSI = 0 at week 24

For subjects with non-zero Leeds Dactylitis Index (LDI) at baseline,

- LDI and change from baseline at week 24
- achievement of LDI = 0 at week 24

For subjects with non-zero Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis at baseline,

- SPARCC enthesitis score and change from baseline at week 24
- achievement of SPARCC score = 0 at week 24

For subjects with involved body surface area (BSA) ≥ 3% at baseline, < 3% at baseline, and ≥ 3% but < 10% at baseline, ≥ 10% at baseline:

- Static Physician Global Assessment (sPGA) categorical summary by 0,1,2,3,4,5 at week 24
- sPGA continuous summary at week 24
- sPGA of 0 or 1 (yes/no) at week 24
- sPGA one grade improvement from baseline at week 24
- sPGA two grade improvement from baseline at week 24
- BSA, change and percent improvement from baseline at week 24

4.1.4 Safety Endpoints

- Adverse events
- Serious adverse events
- Laboratory parameters and vital signs

4.1.5 Exploratory

- mTSS, Erosion Score (ES), and Joint Space Narrowing (JSN) at baseline, week 24, and week 48
- Change from baseline in mTSS, ES and JSN at week 24 and week 48
- Percent change from baseline in mTSS, ES and JSN at week 24 and week 48
- Non-progression in mTSS defined as change in 3 levels: change ≤ 0 , change ≤ 0.5, change ≤ 3, from baseline at week 24 and week 48
- Very Low Disease Activity (VLDA) response at week 24 and all other measured time points
- Disease Activity index for Psoriatic Arthritis (DAPSA) and change from baseline at week 24 and all other measured time points

- Health Assessment Questionnaire Disability Index (HAQ-DI) score and change from baseline at all measured time points other than week 24
- Short Form (36) Health Survey Version 2 (SF-36v2) (MCS and PCS) score and change from baseline at all measured time points other than week 24

For subjects with non-zero mNAPSI at baseline,

- mNAPSI and change from baseline at all measured time points other than week 24
- achievement of mNAPSI = 0 at all measured time points other than week 24

For subjects with non-zero Leeds Dactylitis Index (LDI) at baseline,

- LDI and change from baseline at all measured time points other than week 24
- achievement of LDI = 0 at all measured time points other than week 24

For subjects with non-zero Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis at baseline,

- SPARCC enthesitis score and change from baseline at all other measured time points other than week 24
- achievement of SPARCC score = 0 at all measured time points other than week 24

For subjects with involved body surface area (BSA) ≥ 3% at baseline, < 3% at baseline, and ≥ 3% but < 10% at baseline, ≥ 10% at baseline:

- Static Physician Global Assessment (sPGA) categorical summary by 0,1,2,3,4,5 at all measured time points other than week 24
- sPGA continuous summary at all measured time points other than week 24
- sPGA of 0 or 1 (yes/no) at all measured time points other than week 24
- sPGA one grade improvement from baseline at all measured time points other than week 24
- sPGA two grade improvement from baseline at all measured time points other than week 24
- BSA, change and percent improvement from baseline at all measured time points other than week 24

4.2 Planned Covariates

The following covariates will be considered to assess their influence on the primary and key secondary endpoints:

- Age (≤ 65, > 65)
- Sex (male, female)
- Race (Caucasian, non-Caucasian)
- Ethnicity (Hispanic or Latino, non-Hispanic or Latino)

- Baseline body mass index (BMI) (≤ 30, > 30)
- Baseline weight (≤ 100 kg, > 100 kg)
- Baseline BSA (≤ 3%, > 3%, 3-10%, ≥ 10%)
- Prior use of non-biologic DMARD (yes, no)
- Disease duration (≤ median, > median, < = mean, > mean)

Subgroup analyses based on these covariates will be performed for the primary and key secondary endpoints to examine any differences in drug responses.

5. Hypotheses and/or Estimations

The primary hypothesis of this study is that etanercept plus methotrexate therapy and etanercept monotherapy are more efficacious than methotrexate monotherapy as measured by the proportion of subjects with PsA achieving ACR 20 response at week 24.

The secondary hypothesis of this study is that etanercept plus methotrexate therapy and etanercept monotherapy are more efficacious than methotrexate monotherapy as measured by the proportion of subjects with PsA achieving MDA response at week 24.

6. Definitions

6.1 Basic Definitions

Actual treatment received

Actual treatment for a subject is defined as the treatment the subject is randomized. An exception would be when a subject receives a different treatment other than the randomized treatment throughout the whole course of the study period, and then the actual treatment will be defined as the actual treatment received. In this context, treatment may refer to one Investigational product (IP), or a combination of IPs.

The following scenarios are examples of how actual treatment is defined in double blinded period:

Planned Treatment	Actual Treatment
Etanercept + Methotrexate	If subject receives all placebo doses of Etanercept and all placebo doses of Methotrexate then Actual arm = **Etanercept+ Methotrexate** Else if subject receives at least one dose Etanercept and all placebo doses for Methotrexate then Actual arm = **Etanercept + Placebo Methotrexate** Else If subject receives at least one dose Methotrexate and all placebo doses for Etanercept then Actual arm = **Methotrexate + Placebo Etanercept** Else Actual arm = **Etanercept + Methotrexate**
Etanercept + Placebo Methotrexate	If subject receives all doses of placebo Etanercept and at least one dose of placebo Methotrexate then Actual arm = **Etanercept +Placebo Methotrexate ,** Else if subject receives at least one dose Etanercept and all doses of Methotrexate then Actual arm = **Etanercept + Methotrexate** Else Actual arm = **Etanercept + Placebo Methotrexate**
Methotrexate + Placebo Etanercept	If subject receives at least one dose placebo Etanercept and all placebo doses for Methotrexate then Actual arm =**Placebo Etanercept + Methotrexate** Else if subject receives at least one dose Methotrexate and all doses of Etanercept then Actual arm = **Etanercept + Methotrexate** Else Actual arm = **Methotrexate + Placebo Etanercept**

Actual treatment received in rescue period is defined as etanercept + methotrexate.

Age

Number of years at time of enrollment as recorded on the electronic case report form.

Imputed Value

A value substituted for a missing value.

Investigational Product (IP)

Etanercept (or placebo for etanercept) and methotrexate (or placebo for methotrexate).

Last Observation Carried Forward (LOCF) Imputation

A method of imputation where missing post-baseline efficacy data will be carried forward from the last non-missing post-baseline value for that endpoint.

Rescue Therapy

It is defined as subjects whose IP admin CRF forms indicate whether a subject receives rescue IPs or not.

6.2 Study Points of Reference

Study Baseline

The last non-missing measurement for the endpoint of interest taken on or before the first dose of investigational product.

For swollen joint count and tender joint count, the baseline values are the values taken on or before the day of the first dose of investigational product, and on or after randomization date.

Study Day

The number of days from Study Day 1, inclusive, given by the following formula:

Study Day = (Date of Interest – Date of Study Day 1) + 1, for Date of Interest is on/after Study Day 1

Study Day = Date of Interest - Date of Study Day 1, for Date of Interest is before Study Day 1

Study Day 1

The first day of IP administration after randomization, if a subject is never dosed but randomized, then set study day 1 to randomization date.

6.3 Study Dates

End of Study Date

Date of the End of Study (EOS) safety follow-up phone call (for subjects who complete the study) or Early Termination (ET) visit (for subjects who withdraw from the study prematurely).

First Dose Date

The date of the first dose of IP.

Last Dose Date

The date of the last dose of IP.

Randomization Date

The date when a randomization number is assigned through the Interactive Voice/Web Response System (IVRS/IWRS).

6.4 Study Time Intervals

Double-blind Treatment Period

The period of time beginning when a subject receives their first dose of IP through the first dose of rescue treatment, or end of study if they are never rescued. In case where a

subject is randomized but never dosed, the double blind period will start on the randomization date.

Rescue Therapy Period

The period of time between the first dose of open-label rescue therapy at or after week 24 through end of study.

6.5 Calculations

Body Mass Index (BMI)

A measure utilizing both height and weight, as given by the following formula: weight (kg) / squared height (m^2).

Duration of Psoriatic Arthritis

The number of years between the date of diagnosis (DXDT) and study Day 1, rounded to one decimal place, is given by formula below to calculate the duration:

Observed portion	Missing portion	Formula to Calculate Duration
Year, Month, Day		(DAY 1 – DXDT + 1)/365.25
Year, Month	Day	[Year(DAY 1)-Year(DXDT)]+ [Month(DAY 1)-Month(DXDT)]/12 *if duration equals 0, add 1 month or 1/12 years (this is to avoid a disease duration of 0)
Year	Month, Day	[Year(DAY 1)-Year(DXDT)] *if duration equals 0, add 1 month or 1/12 years (this is to avoid a disease duration of 0)

Improvement from Baseline

The arithmetic difference between the baseline value and a post-baseline value, as given by the following formulas:

For endpoints with higher scores as better clinical results,

Improvement from baseline = (Post-baseline Value – Baseline Value)

For endpoints with lower scores as better clinical results,

Improvement from baseline = (Baseline Value – Post-baseline Value)

Percent Improvement from Baseline

The improvement from baseline divided by baseline and multiplied by 100, as given by the following formula:

(Improvement from Baseline / Baseline) * 100

Note: If improvement from baseline is 0, the percent improvement from baseline will be set to 0 regardless of the baseline value. If the baseline value is 0, substitute 1 in the denominator for the percent improvement from baseline calculation.

6.6 Study Endpoints

ACR 20

A positive ACR 20 response is defined as at least 20% improvement from baseline in both tender/painful (68 joints) and swollen joint counts (66 joints), and at least 20% improvement in at least 3 of the following criteria: physician global assessment of arthritis disease activity (PhGA, 0-100 visual analog scale), patient global assessment of arthritis disease activity (PtGA, 0-100 visual analog scale), patient global assessment of joint pain (0-100 visual analog scale), functional disability (HAQ-DI), and acute phase reactant [C-reactive protein (CRP)].

Details of the scoring algorithm are described in Appendix A.

ACR 50

A positive ACR 50 response is defined by using the definition of ACR 20 response described above but requiring at least 50% improvement.

ACR 70

A positive ACR 70 response is defined by using the definition of ACR 20 response described above but requiring at least 70% improvement.

Adverse Event

In addition to the International Conference on Harmonisation (ICH) Guidance definition, an adverse event also includes any occurrence or worsening of a pre-existing medical condition.

Body Surface Area (BSA)

A numerical score (0% to 100%) that measures psoriasis involvement by the proportion of the subject's total body surface area involved with psoriasis.

Clinical Disease Activity Index (CDAI)

A composite score that is based on the number of tender and swollen joints using a 28-joint count, PhGA (0-10 visual analog scale), and PtGA (0-10 visual analog scale). A higher score represents worse disease.

Details of the scoring algorithm are described in Appendix A.

C-Reactive Protein (CRP)

A specific measure of inflammatory activity measured in mg/dL.

Disease Activity Score (DAS28-CRP)

A composite index that was designed to measure disease activity in subjects using the number of tender and swollen joints based upon a 28-joint count, CRP in mg/L, and a 100 mm visual analog scale measuring the subject's general health. A higher score represents worse disease.

Details of the scoring algorithm are described in Appendix A.

Disease Activity index for Psoriatic Arthritis (DAPSA)

This index is a composite measure, which comprises swollen and tender joint counts, patient global and pain assessment and an acute phase reactant. Details of the scoring algorithm are described in Appendix A.

Disability Index of the Health Assessment Questionnaire (HAQ-DI)

The HAQ-DI assesses the subject's physical function or disability. The questionnaire asks about the degree of difficulty a person has in accomplishing tasks in 8 functional areas (dressing, arising, eating, walking, hygiene, reaching, gripping, and usual activities). Responses in each functional area are scored from 0 indicating no difficulty to 3 indicating inability to perform a task in that area.

Details of the scoring algorithm are described in Appendix A.

Leeds Dactylitis Index (LDI)

The Leeds Dactylitis Index (LDI) quantitatively measures dactylitis using the circumference of involved digits and circumference of control digits (either contralateral digit, ie, digits on opposite hand or foot, or digit in standard reference table), and tenderness of involved digits. The ratio of affected digit circumference to control digit circumference is multiplied by a tenderness score if ratio difference is greater than 10%. The results from each involved digit are then summed to provide the final LDI. A higher LDI is associated with worse dactylitis.

Details of the scoring algorithm are described in Appendix A.

Leeds Enthesitis Index

The Leeds enthesitis index assesses enthesitis at 6 sites for palpitation with a resultant score of 0 to 6. Tenderness at each site is quantified on a dichotomous basis (0 = non-tender, 1 = tender). Entheses assessed are lateral epicondyles (left and right), medial femoral condyles (left and right), and achilles tendon insertions (left and right). A higher count represents greater enthesitis burden.

Details of the scoring algorithm are described in Appendix A.

Medical Outcomes Health Survey Short Form 36 Items Version 2 (SF-36 v2)

The SF-36 v2 is a 36-item instrument that measures general health status and will be scored using the Quality Metric Health Outcomes Scoring Software. It includes 8 multi-item scales, each of which assesses one of the following 8 health concepts over the previous month: 1) limitations in physical activities because of health problems; 2) limitations in social activities because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4) bodily pain; 5) general mental health (psychological distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions. SF-36 v2 includes the Physical Component Summary and Mental Component Summary measures. A higher score represents less disability.

Minimal Disease Activity (MDA)

A measure of low disease activity specific for PsA that incorporates measures of joint and entheseal inflammation, skin disease, PROs and functional disability to assess the subject's disease activity. Subjects are classified as achieving MDA if they fulfill 5 of the following 7 outcome measures:

- Tender joint count (0-68) ≤ 1
- Swollen joint count (0-66) ≤ 1
- Body Surface Area (BSA) ≤ 3
- Patient global assessment of joint pain (0-100) ≤ 15
- Patient global assessment of disease activity VAS (0-100) ≤ 20
- HAQ-DI (0-3) ≤ 0.5
- SPARCC enthesitis index ≤ 1

Details of the scoring algorithm are described in Appendix A.

Modified Nail Psoriasis Severity Index (mNAPSI)

The modified NAPSI scale is an objective, numeric, and reproducible grading system for nail psoriasis that incorporates the following 7 clinical features:

- pitting (scores 0-3, depending on the number of pits)
- nail plate crumbling (scores 0-3, depending on the % of nail involvement)
- onycholysis and oil drop dyschromia (scores 0-3, depending on the % of nail involvement)
- leukonychia (0 = absent, 1 = present)
- red spots in lunula (0 = absent, 1 = present)

- nail bed hyperkeratosis (0 = absent, 1 = present)
- splinter hemorrhages (0 = absent, 1 = present)

In randomized subjects with fingernails involved with psoriasis, each fingernail will be scored at baseline to determine the worst fingernail (ie, the fingernail with the highest mNAPSI score). This fingernail will be followed for the remainder of the study. If multiple fingernails have the same worst score, only one target fingernail will be followed. Scores will range from 0-13 where higher scores represent worse nail disease.

Details of the scoring algorithm re described in Appendix A

Patient Global Assessment of Joint Pain (PtGAJP)

An assessment of the severity of the subject's joint pain (0=no pain at all, 100 = worst pain imaginable) completed by the patient using a visual analog scale (VAS).

Patient Global Assessment of Disease Activity (PtGA)

A global assessment of the subject's arthritis (0 = no arthritis activity at all, 100 = worst arthritis activity imaginable) completed by the patient using a visual analog scale (VAS).

Physician Global Assessment of Disease Activity (PhGA)

A global assessment of the subject's arthritis (0 = no activity at all, 100 = worst activity imaginable) completed by the physician using a visual analog scale (VAS).

Psoriatic Arthritis Disease Activity Score (PASDAS)

A score of disease activity consisting of the following domains: physician and patient global assessment of disease activity (0-100 visual analog scale), peripheral joint counts based upon 68/66 joints, tender dactylitis count, Leeds enthesitis index, CRP in mg/L and SF-36 v2 physical component summary.

Details of the scoring algorithm are described in Appendix A.

Simplified Disease Activity Index (SDAI)

A composite score that is based on the number of tender and swollen joints using a 28-joint count, PhGA (0-10 visual analog scale), PtGA (0-10 visual analog scale), and CRP in mg/dL. A higher score represents worse disease.

Details of the scoring algorithm are described in Appendix A.

Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index

The SPARCC enthesitis index assesses enthesitis at 18 sites for palpitation with a resultant score of 0 to 16. Tenderness at each site is quantified on a dichotomous basis (0 = non-tender, 1 = tender). Entheses assessed are medial epicondyle (left and right),

lateral epicondyle (left and right), supraspinatus insertion into greater tuberosity of humerus (left and right), greater trochanter (left and right), quadriceps insertion into superior border of patella (left and right), patellar ligament insertion into inferior pole of patella or tibial tubercle (left and right), Achilles tendon insertion into calcaneum (left and right), plantar fascia insertion into calcaneum (left and right). A higher count represents greater enthesitis burden.

Details of the scoring algorithm are described in Appendix A.

Static Physician Global Assessment of Psoriasis (sPGA)

The static Physician Global Assessment of psoriasis (sPGA) evaluates the physician's global assessment of the subject's psoriasis based on severity of induration, scaling, and erythema. The sPGA is assessed on a scale of 0 (clear) to 5 (severe).

Swollen Joint Count

A score that assesses the presence or absence of swelling in 66 joints. Details of the scoring algorithm are described in Appendix A.

Tender Joint Count

A score that assesses the presence or absence of tenderness in 68 joints. Details of the scoring algorithm are described in Appendix A.

Treatment Emergent AE

A treatment emergent AE is defined as an event that occurs on or after the initiation of the IP and prior to the EOS date (ie, 30 days after last dose of IP). It is determined by the pre-treatment flag (ie, AE.AEPRETRT NE 'Y') on the Events CRF with a start date up to 30 days after the end of investigational product or the End of Study date, whichever is earlier.

Van der Heijde Modified Total Sharp Score (mTSS)

A method for assessing erosions and joint space narrowing of the hands and feet based on the Sharp-van der Heijde method in rheumatoid arthritis. The method is adapted for evaluation of PsA by including the distal interhpalengeal joints of the hand and scoring not only for erosions and joint space narrowing but also for subluxation, ankylosis, gross osteolysis, and pencil in cup phenomena. The maximum possible total score is 528.
A higher score represents worse destruction. Details of the scoring algorithm are described in Appendix A.

Very Low Disease Activity (VLDA)

A measure of very low disease activity specific for PsA that incorporates measures of joint and entheseal inflammation, skin disease, PROs and functional disability to assess the subject's disease activity. Subjects are classified as achieving VLDA if they fulfill all the following 7 outcome measures:

- Tender joint count (0-68) ≤ 1
- Swollen joint count (0-66) ≤ 1
- Body Surface Area (BSA) ≤ 3
- Patient pain VAS (0-100) ≤ 15
- Patient global disease activity VAS (0-100) ≤ 20
- HAQ (0-3) ≤ 0.5
- Tender entheseal points ≤ 1

Details of the scoring algorithm are described in Appendix A.

7. Analysis Subsets

7.1 Full Analysis Set

The full analysis set will include all randomized subjects. Subjects will be analyzed according to their randomized treatment group. Demographics, baseline disease characteristics, and non-psoriasis efficacy analyses will be based on the full analysis set.

7.2 Psoriasis Efficacy Analysis Set

The psoriasis efficacy analysis set will include all randomized subjects with baseline BSA ≥ 3%. Subjects will be analyzed according to their randomized treatment group. Analyses of psoriasis efficacy endpoints (eg, sPGA, BSA) will be based on the psoriasis efficacy analysis set.

7.3 Safety Analysis Set

The safety analysis set will consist of all randomized subjects who receive at least one dose of investigational product. Subjects will be analyzed according to the actual treatment received as defined in Section 6.1. Analyses of safety endpoints will be based on the safety analysis set.

7.4 Adequate Methotrexate Analysis Set

The adequate methotrexate analysis set will consist of all randomized subjects who receive weekly dose of at least 15 mg of methotrexate for at least 15 weeks during the course of week 4 (inclusive) to week 24 (inclusive) on etanercept plus methotrexate arm or methotrexate mono arm, and all subjects randomized to etanercept mono arm.

7.5 Leeds Dactylitis Index (LDI) Analysis Set

The leeds dactylitis index analysis set will consist of all the randomized subjects with non-zero LDI at baseline. Subjects will be analyzed according to their randomized treatment group. Analyses of LDI efficacy endpoints will be based on the leeds dactylitis index analysis set.

7.6 Modified Nail Psoriasis Severity Index (mNAPSI) Analysis Set

The mNAPSI analysis set will consist of all the randomized subjects with non-zero mNAPSI at baseline. Subjects will be analyzed according to their randomized treatment group. Analyses of mNAPSI efficacy endpoints will be based on this analysis set.

7.7 SPARCC Enthesitis Analysis Set

The SPARCC Enthesitis Analysis Set will consist of the randomized subjects with non-zero SPARCC enthesitis at baseline. Subjects will be analyzed according to their randomized treatment group. Analyses of SPARCC efficacy endpoints will be based on the SPARCC Enthesitis analysis set.

7.8 Subgroup Analysis

Subgroup analyses based on age (≤ 65, > 65), sex (male, female), race (Caucasian, non-Caucasian), ethnicity (Hispanic or Latino, non-Hispanic or Latino), baseline BMI (≤ 30, > 30), baseline weight (≤ 100 kg, > 100 kg), baseline BSA (< 3%, > = 3%, > = 3% and < 10%, > = 10%), prior use of non-biologic DMARD (yes, no), and disease duration (≤ median, > median, median is based on all randomized subjects.) will be performed for the primary and key secondary endpoints to examine any differences in drug responses.

Also subgroup analyses based on baseline BSA (< 3%, ≥ 3% but < 10%, and > = 10%) will be performed for sPGA and BSA in exploratory analysis.

Subgroup analysis containing a subgroup of less than 5% of the whole study population will not be conducted.

8. Interim Analysis and Early Stopping Guidelines

No interim analysis is planned for this study.

9. Data Screening and Acceptance

9.1 General Principles

The objective of the data screening is to assess the quantity, quality and statistical characteristics of the data relative to the requirements of the planned analyses.

9.2 Data Handling and Electronic Transfer of Data

Amgen's Clinical Data Management department will provide all data to be used in the planned analyses. This study will use the RAVE database for all data except patient reported outcomes (PROs) and tender and swollen joint counts, which will be collected electronically on a tablet, and radiographic and laboratory data. All PROs, tender and swollen joint counts, radiographic data, and laboratory data will be transferred to Amgen's Clinical Data Management and stored outside the RAVE database as described in the Data Management Plan and Data Transfer Plan.

9.3 Handling of Missing and Incomplete Data

Subjects may be missing specific data points for a variety of reasons. In general, data may be missing due to a subject's early withdrawal from study, a missed visit, or inability to evaluate an endpoint at a particular point in time. The general procedures outlined below describe what will be done when a data point is missing.

9.3.1 Missing Baseline Evaluation

Missing baseline evaluations will not be imputed.

9.3.2 Missing Post-baseline Evaluation

Subjects with missing post-baseline efficacy data will be imputed as non-responder, ie, non-responder imputation (NRI) for the primary analysis of primary and key secondary efficacy endpoints (ACR 20 and MDA). Data will also be analyzed using multiple imputation (MI) as sensitivity analyses for primary and key secondary endpoints (ACR20 and MDA).

For analyses of other efficacy endpoints after week 24, data from subjects who remain in the double-blind period will be analyzed as observed with no imputation for missing data. Data from subjects who enter open-label rescue therapy will be imputed using last observation carried forward (LOCF).

For non-radiograph exploratory endpoints, data will be analyzed as observed without imputation for missing data.

For radiographic exploratory endpoints, data will be analyzed as observed with no imputation for missing data or for subjects who enter rescue treatment. In addition, data will be analyzed using LOCF imputation for missing radiographs and for subjects who enter rescue treatment as sensitivity analysis.

Baseline data will not be carried forward to any missing post-baseline value unless otherwise specified.

Missing post-baseline safety data will not be imputed.

9.3.3 Missing Components of Composite Endpoints

The rules for handling missing individual joints and missing components of composite endpoints are described in Appendix A.

9.3.4 Missing and Incomplete Dates

For any listings, missing or incomplete dates will be listed as is. No imputation will be done on incomplete stop date of an adverse event unless specified otherwise.

For tables, in the case where the start or stop date of an adverse event or concomitant medication is missing or incomplete, the following rule will be applied:

	Missing	Imputation	Exception
Start date (AE, concomitant medication)	Day	01	Default to Study Day 1 if an event starts the same year and month as Study Day 1 and Stop date is after Study Day 1
	Day/Month, or Month	01JAN	Default to Study Day 1 if an event started the same year as Day1 and Stop date is after Study Day 1
	Day/Month/Year	No imputation	

If the imputed start date is after the end date of concomitant medication or AE, then imputed start date will be default to end date of concomitant medication or AE.

9.4 Detection of Bias

The study has been designed to minimize potential bias by selecting subjects, allocating treatment groups, assessing endpoints, and handling withdrawals without knowledge of the treatment. Other factors that may introduce bias include:

- Important protocol deviations likely to impact the analysis or interpretation of results
- Blind breaking before database lock and formal unblinding
- IP dosing non-compliance
- Reasons for early withdrawal from treatment or from study

The incidence of these factors will be assessed. Important protocol deviations likely to impact the analysis and interpretation of results will be listed and/or tabulated in the Clinical Study Report (CSR). If a significant number of major protocol deviations are observed, an additional sensitivity analysis may be performed on the efficacy data excluding subjects with major protocol deviations.

Any breaking of the blind for individual subjects prior to formal unblinding of the study will be documented in the CSR, including the timing and reason for unblinding. The impact of such unblinding on the results will be assessed.

Tabulations of important protocol deviations (IPDs) related to IP dosing non-compliance by treatment group will be provided. A sensitivity analysis including subjects who tolerated at least 15 mg weekly of methotrexate (or placebo for methotrexate) after the initial 4-week titration through week 24 will be performed to assess the potential bias of methotrexate intolerance, as previously described in Section 7.4.

Reasons for early withdrawals from treatment and from study will be summarized. If a significant pattern is observed, additional sensitivity analyses may be performed on efficacy data adjusting for reasons for early withdrawal.

9.5 Outliers

Descriptive summaries will be examined to identify unexpected values.

Outliers due to data entry errors will be corrected by the study team before data lock. Outliers that are not due to data entry will be included in the analysis. If it is deemed necessary after the team reviews the output from the planned analyses after data lock, a post-hoc sensitivity analysis excluding subjects with outliers may be performed.

9.6 Distributional Characteristics

All continuous endpoints that will be analyzed with an ANCOVA model will be evaluated with regards to their distributional characteristics. If they deviate from normality, data will be first transformed to their joint ranks and then further transformed by the van der Waerden method to normalize them before applying the ANCOVA model.

9.7 Validation of Statistical Analyses

Programs will be developed and maintained, and output will be verified in accordance with current risk-based quality control procedures.

Tables, figures and listings will be produced with validated standard macro programs where standard macros can produce the specified outputs.

The production environment for statistical analyses consists of Amgen-supported versions of statistical analysis software, for example the SAS System version 9.2 or later.

10. Statistical Methods of Analysis

10.1 General Principles

The primary analysis will be performed after all subjects have completed the week 48 assessments and the 30-day safety follow-up period and all data have been finalized.

Subject disposition, demographics, and baseline disease characteristics will be summarized descriptively by randomized treatment group.

For categorical endpoints, the descriptive statistics will contain the frequency and percentage. For continuous endpoints, the descriptive statistics will include the number of observations, mean, standard error, standard deviation, median, minimum, and maximum.

Treatment effect will be tested for the primary and key secondary endpoints using a Bonferroni-based gatekeeping chain procedure to control the family-wise 2-sided type one error rate at 0.05. The significance level for the analyses of other secondary endpoints and exploratory endpoints will be 0.05 without adjusting for multiplicity and the p-values for these endpoints will therefore be treated as a descriptive statistic only. Safety endpoints will be tabulated by actual treatment received during the study. No formal statistical testing will be performed for safety analyses.

10.2 Subject Accountability

Subject disposition will be summarized descriptively for all randomized subjects by randomized treatment group.

The disposition for the double-blind treatment period will include the number of subjects who are randomized, who are dosed with investigational product, who complete 24 weeks, who complete 48 weeks, who complete the study, and who withdraw prematurely including their reasons for withdrawal.

The disposition for the rescue therapy subjects will include the number of subjects who enter rescue therapy, who complete 48 weeks, who complete the study, and who withdraw prematurely including their reasons for withdrawal.

10.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject's visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, sub-category codes and descriptions will be used during the course of the study.

10.4 Demographic and Baseline Characteristics

Subject demographic and baseline disease characteristics will be summarized descriptively for all randomized subjects by randomized treatment group. Demographic and baseline disease characteristics may be re-evaluated for the safety evaluable subset if the subset is materially different from the full analysis set.

In addition, subject demographic and baseline disease characteristics will be summarized descriptively for all subjects who receive rescue therapy by randomized treatment group.

10.5 Efficacy Analyses

All efficacy analyses of non-psoriasis efficacy endpoints will be conducted using the full analysis set (ie, intent-to-treat set). Analyses of psoriasis efficacy endpoints (eg, sPGA, BSA) will be based on the psoriasis efficacy analysis set and other subgroups using different cutoffs of baseline BSA. Some PsA efficacy endpoints will be analyzed using analysis sets corresponding to those PsA indications. Subjects will be analyzed according to their original randomized treatment regardless of the actual treatment received during the study.

For the primary and key secondary endpoints, missing post-baseline data will be imputed using non-responder imputation (NRI) for the primary analysis. In addition, the primary and key secondary endpoints will be analyzed using multiple imputation (MI) as a sensitivity analysis. Another sensitivity analysis for primary and key secondary endpoints will also be performed using Adequate Methotrexate Analysis Set. For analyses of other efficacy endpoints after week 24, data from subjects who remain in the double-blind period will be analyzed as observed with no imputation for missing data. Data from subjects who enter rescue period will be imputed using LOCF.

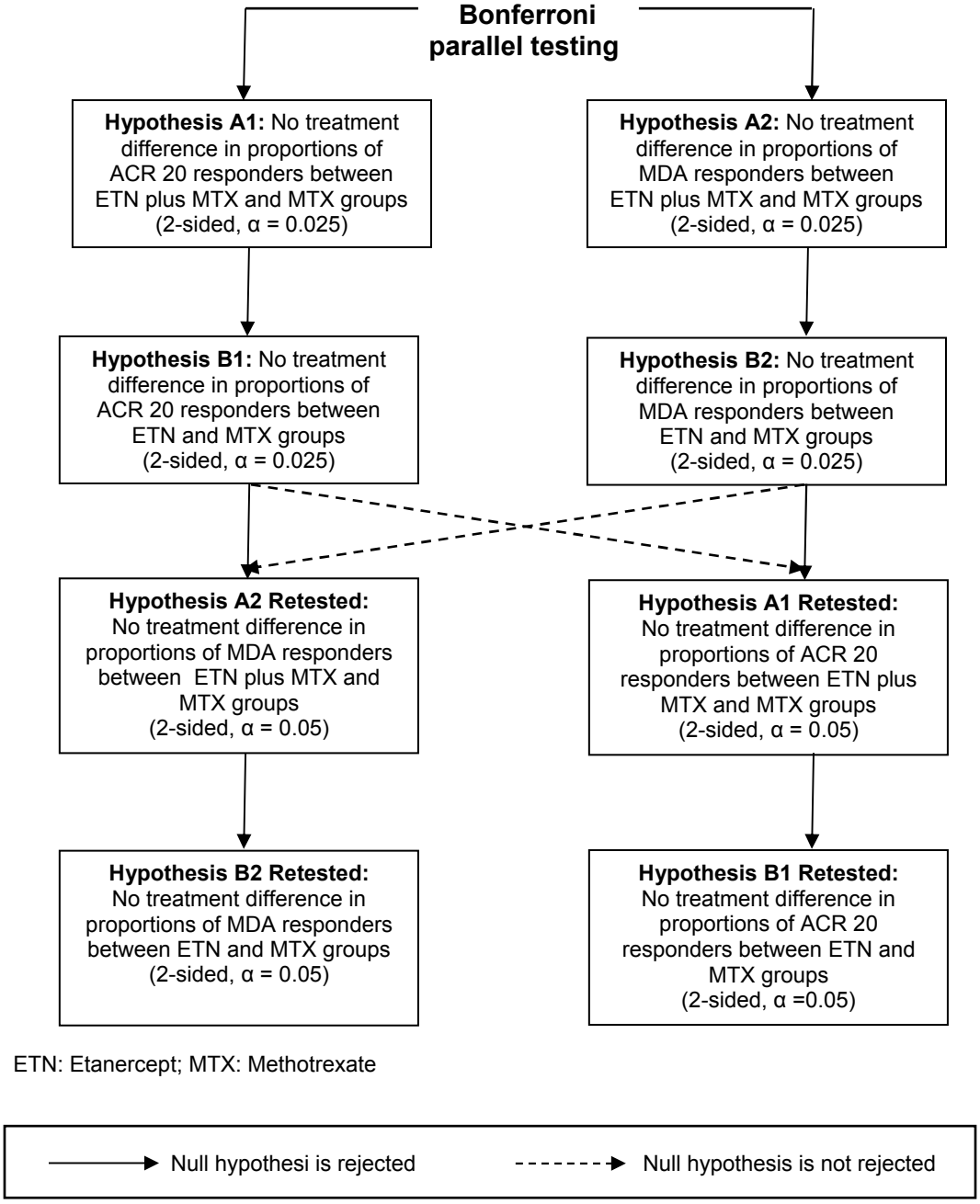
10.5.1 Analyses of Primary and Key Secondary Efficacy Endpoints

The primary analysis of the primary and key secondary endpoints, ACR 20 response and MDA response at week 24, will be performed using the full analysis set. Missing post-baseline data will be imputed using non-responder imputation (NRI) for the primary analysis. Treatment effect will be tested using the stratified Cochran-Mantel-Haenszel (CMH) test with baseline BMI and prior non-biologic DMARD use as the stratification factors.

To control the family-wise 2-sided type one error rate at 0.05 for the multiple comparisons of the etanercept plus methotrexate therapy and etanercept monotherapy groups with methotrexate monotherapy across the primary and key secondary endpoints, the hypotheses will be tested using a Bonferroni-based gatekeeping chain procedure (Burman et al 2009 and Millen et al 2011) as shown in Figure 1. This procedure will split alpha of 0.05 equally to test in parallel the primary and key secondary endpoints (ACR 20, MDA) for etanercept plus methotrexate therapy vs. methotrexate monotherapy and etanercept monotherapy vs. methotrexate monotherapy sequentially

within each parallel path. If one of the parallel paths rejects both hypotheses sequentially, and the other parallel path has at least one hypothesis not rejected, then the unspent alpha of 0.025 from the successful path will be propagated to the hypotheses in the other path to re-test them sequentially at a level of 0.05. For example, suppose the nominal p-values for hypotheses A1, B1, A2, and B2 (Figure 1) are 0.01, 0.02, 0.03, and 0.055, respectively. In the first round of testing, hypotheses A1 and B1 are rejected, whereas hypotheses A2 and B2 are not rejected. However, since both hypotheses A1 and B1 are rejected, the procedure gives a second chance to re-test hypotheses A2 and B2 (according to the left arm of the figure) at a lower threshold of 0.05 rather than the original threshold of 0.025. Note that in this example A2 is rejected and hypothesis B2 is not rejected.

Figure 1. Bonferroni-based Gatekeeping Chain Procedure



ETN: Etanercept; MTX: Methotrexate

In addition, the primary and key secondary endpoints will be analyzed using multiple imputation (MI) as a sensitivity analysis. Another sensitivity analysis for primary and key secondary endpoints will also be performed using Adequate Methotrexate Analysis Set. Sensitivity analyses of primary and key secondary endpoints are descriptively without adjusting for multiplicity.

10.5.2 Analyses of Other Secondary Efficacy and Selected Exploratory Endpoints

The analysis of the following secondary and exploratory endpoints at week 24 will be performed using the LDI analysis set for LDI endpoints, SPARCC Enthesitis analysis set for SPARCC endpoints, mNAPSI analysis set for mNAPSI endpoints, and the full analysis set for the other efficacy endpoints unless otherwise specified:

- LDI change from baseline at week 24 (LDI analysis set)
- Achievement of LDI=0 at week 24 (LDI analysis set).
- SPARCC change from baseline at week 24 (SPARCC analysis set)
- Achievement of SPARCC=0 at week 24 (SPARCC analysis set)
- mNAPSI change from baseline at week 24 (mNAPSI analysis set)
- Achievement of mNAPSI=0 at week 24 (mNAPSI analysis set)
- PASDAS change from baseline at week 24
- DAS28-CRP change from baseline at week 24
- SDAI change from baseline at week 24
- CDAI change from baseline at week 24
- VLDA at week 24
- DAPSA change from baseline at week 24

For binary endpoints, treatment effect will be tested using the stratified CMH test with BMI and prior non-biologic DMARD use as the stratification factors. For continuous endpoints that exhibit normal distribution, ANCOVA model will be used. The model will include treatment groups and baseline BMI and prior non-biologic DMARD as covariates. For continuous endpoints that are not normally distributed, endpoint data will be first transformed to their joint ranks and then further transformed by the Van der Waerden transformation to normalize them before applying the ANCOVA model. The significance level for other secondary endpoints will be 0.05 without adjusting for multiplicity and p-values will therefore be treated as a descriptive statistic only. All the testing will be performed for endpoints at week 24 and methotrexate monotherapy group will be the reference group.

Analyses to other secondary endpoints will be summarized descriptively only. sPGA and BSA summary analyses will be performed using psoriasis efficacy analysis set and subgroups of subjects with the baseline BSA < 3%, > = 3% but < 10%, and > = 10%.

10.5.3 Analyses of Exploratory Endpoints

The analysis of the exploratory endpoints will be performed using the LDI analysis set for LDI endpoints, SPARCC Enthesitis analysis set for SPARCC endpoints, mNAPSI analysis set for mNAPSI endpoints, psoriasis efficacy analysis set and subgroups of subjects with the baseline BSA < 3%, > = 3% but < 10%, and > = 10% for sPGA and BSA summary analyses, and the full analysis set for the other efficacy endpoints unless otherwise specified.

For radiographic exploratory endpoints, data will be analyzed as observed with no imputation for missing data or for subjects who enter rescue treatment. In addition, data will be analyzed using LOCF imputation for missing radiographs and for subjects who enter rescue treatment as sensitivity analysis. Probability plots and mean change plots will also be produced based on observed data.

Analyses to other exploratory endpoints will be summarized descriptively only. For VLDA and DAPSA, additional analyses at week 24 will also be performed as described in Section 10.5.2.

10.5.4 Biomarker Endpoints

Data collected on biomarkers may be analyzed to help characterize subjects who are most likely to respond positively or negatively to etanercept or methotrexate. If the study team decides to perform such an analysis, it will be considered exploratory and will be described in a separate analysis plan.

10.6 Safety Analyses

All safety analyses will be conducted using the safety analysis set. Subjects will be analyzed according to the actual treatment received. Subjects who meet the inadequate response criteria at or after week 24 have the opportunity to enter rescue therapy and begin receiving etanercept plus methotrexate. In order to accurately reflect the safety experiences in these subjects with respect to the treatment received during rescue therapy, all events experienced during rescue therapy will be summarized separately.

10.6.1 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 18.0 or later will be used to code all adverse events (AE) to a system organ class and a preferred term. All adverse event tables will be summarized by treatment group for the double-blind treatment period and the rescue therapy period. The subject incidence of AEs will be

summarized for all treatment-emergent AEs, serious AEs, AEs leading to withdrawal of investigational product, and fatal AEs.

Subject incidence of all treatment-emergent AEs, serious AEs, AEs leading to withdrawal of investigational product, and fatal AEs will be tabulated by system organ class and preferred term in descending order of frequency.

Summaries of treatment-emergent and serious AEs occurring in at least 1% of the subjects by preferred term in any treatment arm will be provided in descending order of frequency.

10.6.2 Laboratory Test Results

Laboratory parameters with grade 3 toxicity or above may be descriptively summarized by study visit and actual treatment received during the study. Shift tables of the worst on-study laboratory toxicity based on the Common Toxicity Criteria (CTC) grade relative to baseline may be tabulated for analytes of interest (ie, creatinine, total bilirubin, AST (SGOT), ALT (SGPT), hemoglobin, platelets, WBC, absolute neutrophil count, lymphocytes) by treatment group. Modifications to CTC version 4.0 used for this analysis are described in Appendix C. Subject listings of grades 3 and 4 laboratory values will be provided.

10.6.3 Vital Signs

Vital signs may be descriptively summarized in baseline disease summaries, if there is unexpected safety finding.

10.6.4 Exposure to Investigational Product

Descriptive statistics will be produced to describe the exposure to investigational product by treatment group, including number of doses, number of missed doses, and duration of dosing.

10.6.5 Exposure to Concomitant Medication

A summary of all concomitant medications reported may be provided by preferred term as coded by the World Health Organization Drug (WHODRUG) dictionary.

11. Changes From Protocol-specified Analyses

Changes are made only for secondary or exploratory endpoints. The changes are instituted to adopt the standardized ways to report the efficacy endpoints in PsA trials. The changes to LDI, SPARCC, and mNAPSI are made to focus on relevant sub-populations for analysis. Modeling analyses described in Section 10.5.2 will be

limited to selected secondary endpoints at week 24 only, for the inclusion in CSR. Further analyses may be performed later on an ad hoc basis.

Specifically, the changes are as follows:

1. These analysis sets are added to the SAP because they are more relevant to the purpose of the analyses focusing on the subpopulations with disease progression.

Leeds Dactylitis Index (LDI) Analysis Set

The leeds dactylitis index analysis set will consist of all the randomized subjects with non-zero LDI at baseline. Subjects will be analyzed according to their randomized treatment group. Analyses of LDI efficacy endpoints will be based on the leeds dactylitis index analysis set.

Modified Nail Psoriasis Severity Index (mNAPSI) Analysis Set

The mNAPSI analysis set will consist of all the randomized subjects with non-zero mNAPSI at baseline. Subjects will be analyzed according to their randomized treatment group. Analyses of mNAPSI efficacy endpoints will be based on this analysis set.

SPARCC Enthesitis Analysis Set

The SPARCC Enthesitis Analysis Set will consist of the randomized subjects with non-zero SPARCC enthesitis at baseline. Subjects will be analyzed according to their randomized treatment group. Analyses of SPARCC efficacy endpoints will be based on the SPARCC Enthesitis analysis set.

2. Modification of the other secondary endpoints limited to specific subpopulation in in Section 4.1.3 as follows:

For subjects with non-zero mNAPSI at baseline,

- mNAPSI and change from baseline at week 24
- achievement of mNAPSI = 0 at week 24

For subjects with non-zero Leeds Dactylitis Index (LDI) at baseline,

- LDI and change from baseline at week 24
- achievement of LDI = 0 at week 24

For subjects with non-zero Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis at baseline,

- SPARCC enthesitis score and change from baseline at week 24
- achievement of SPARCC score = 0 at week 24

For subjects with involved body surface area (BSA) ≥ 3% at baseline, < 3% at baseline, and ≥ 3% but < 10% at baseline, ≥ 10% at baseline:

- Static Physician Global Assessment (sPGA) categorical summary by 0,1,2,3,4,5 at week 24
- sPGA continuous summary at week 24
- sPGA of 0 or 1 (yes/no) at week 24
- sPGA one grade improvement from baseline at week 24
- sPGA two grade improvement from baseline at week 24
- BSA, change and percent improvement from baseline at week 24

The changes are in Section 4.1.5 Exploratory Endpoints:

- mTSS, Erosion Score (ES), and Joint Space Narrowing (JSN) at baseline, week 24, and week 48
- Change from baseline in mTSS, ES and JSN at week 24 and week 48
- Percent change from baseline in mTSS, ES and JSN at week 24 and week 48
- Non-progression in mTSS defined as change in 3 levels: change ≤ 0 , change ≤ 0.5, change ≤ 3, from baseline at week 24 and week 48
- Very Low Disease Activity (VLDA) response at week 24 and all other measured time points
- Disease Activity index for Psoriatic Arthritis (DAPSA) and change from baseline at week 24 and all other measured time points

For subjects with non-zero mNAPSI at baseline,

- mNAPSI and change from baseline at all other measured time points other than week 24
- achievement of mNAPSI = 0 at all other measured time points other than week 24

For subjects with non-zero Leeds Dactylitis Index (LDI) at baseline,

- LDI and change from baseline at all other measured time points other than week 24
- achievement of LDI = 0 at all other measured time points other than week 24

For subjects with non-zero Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis at baseline,

- SPARCC enthesitis score and change from baseline at all other measured time points other than week 24
- achievement of SPARCC score = 0 at all other measured time points other than week 24

For subjects with involved body surface area (BSA) ≥ 3% at baseline, < 3% at baseline, ≥ 3% but < 10% at baseline, and ≥ 10% at baseline:

- Static Physician Global Assessment (sPGA) categorical summary by 0,1,2,3,4,5 at all other measured time points other than week 24
- sPGA continuous summary at all other measured time points other than week 24
- sPGA of 0 or 1 (yes/no) at all other measured time points other than week 24
- sPGA one grade improvement from baseline at all other measured time points other than week 24
- sPGA two grade improvement from baseline at all other measured time points other than week 24
- BSA, change and percent improvement from baseline at all other measured time points other than week 24

3. Changes to Section 10.5.2 are made to focus on selected other secondary and exploratory efficacy endpoints at week 24 only, which is the primary time point of interest, as follows:

The analysis of the following secondary endpoints at week 24 will be performed using the LDI analysis set for LDI endpoints, SPARCC Enthesitis analysis set for SPARCC endpoints, mNAPSI analysis set for mNAPSI endpoints, and the full analysis set for the other efficacy endpoints unless otherwise specified:

- LDI change from baseline at week 24 (LDI analysis set)
- Achievement of LDI = 0 at week 24 (LDI analysis set).
- SPARCC change from baseline at week 24 (SPARCC analysis set)
- Achievement of SPARCC = 0 at week 24 (SPARCC analysis set)
- mNAPSI change from baseline at week 24 (mNAPSI analysis set)
- Achievement of mNAPSI = 0 at week 24 (mNAPSI analysis set)
- PASDAS change from baseline at week 24
- DAS28-CRP change from baseline at week 24
- SDAI change from baseline at week 24
- CDAI change from baseline at week 24
- VLDA at week 24
- DAPSA change from baseline at week 24

12. Literature Citations / References

Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. *Stat. Med*. 2009; 28:739–761

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

Fagerli KM, Lie E, van der Heidje D, et al. The role of methotrexate co-medication in TNF-inhibitor treatment in patients with psoriatic arthritis: results from 440 patients included in the NOR-DMARD study. *Ann Rheum Dis*. 2014 73:132–137.

Kavanaugh A, McInnes IB, Mease PJ, et al. Impact of persistent minimal disease activity on long-term outcomes in psoriatic arthritis: results from 5 years of the long term extension of a randomized, placebo-controlled, Study. Abstract #341. Presented at the ACR/ARHP Annual Meeting, 27 Oct 2013.

Kingsley GH, Kowalcyzk A, Taylor H, et al. A randomized placebo-controlled trial of methotrexate in psoriatic arthritis. *Rheumatology*. 2012;51:1368-1377

Millen B, Dmitrienko A. Chain procedures: a class of flexible closed testing procedures with clinical trial applications. *Statistics in Biopharmaceutical Research* 2011; 3:14–30.

13. Appendices

Appendix A. Technical Detail and Supplemental Information Regarding Statistical Procedures and Programs

Analysis Windows for Evaluations

Per protocol, visits are to be performed within 7 days of the protocol-specified study day. To allow for variations in scheduling, the following visit windows will be applied to selected efficacy and safety evaluations (ie, vital signs, laboratory evaluations) to assign a most appropriate nominal visit for analysis. If more than one assigned visit falls within the same defined window, the closest visit to the target day (ie, scheduled visit week x 7 + 1) will be considered for analysis. If two assessment dates are the same distance from the target day, then the latest visit will be considered for analysis. If more than one evaluation on the same date and time, the average of the results will be used. Only laboratory results collected from the central laboratory will be averaged in the case of duplicate results.

Any visit that falls in an unscheduled visit for that assessment (eg, an enthesitis assessment done at week 16) will be excluded from the analysis.

Visit Windows for Radiographic Endpoints

Visit Week	Target Day	Window Definition	Interval (days)
Baseline	1	Last Evaluation prior to or on Study Day 1	n/a
Week 24	169	Study Day 15 to 252	118
Week 48	337	>=Study Day 253	n/a

[a] If results from baseline radiograph are not available, the results from scan taken on or before Study Day 14 will be considered as baseline values.

Visit window for SPARCC, Dactylitis, mNAPSI, sPGA, BSA

Visit Week	Target Day	Window Definition	Interval (days)
Baseline	1	Last Evaluation prior to or on Study Day 1	n/a
Week 4	29	Study Day 2 to 56	54
Week 12	85	Study Day 57 to126	70
Week 24	169	Study Day 127 to 210	84
Week 36	253	Study Day 211 to 294	84
Week 48	337	>=Study Day 295	n/a

Visit window for SF-36 and PASDAS

Visit Week	Target Day	Window Definition	Interval (days)
Baseline	1	Last Evaluation prior to or on Study Day 1	n/a
Week 12	85	Study Day 2 to 126	124
Week 24	169	Study Day 127 to 210	84
Week 36	253	Study Day 211 to 294	84
Week 48	337	>=Study Day 295	n/a

For the rest of the endpoints of interest, use the following window:

Visit Week	Target Day	Window Definition	Interval (days)
Baseline[a]	1	Last Evaluation prior to or on Study Day 1	n/a
Week 4	29	Study Day 2 to 42	41
Week 8	57	Study Day 43 to 70	28
Week 12	85	Study Day 71 to 98	28
Week 16	113	Study Day 99 to 140	42
Week 24	169	Study Day 141 to 210	70
Week 36	253	Study Day 211 to 294	84
Week 48	337	>=Study Day 295	n/a

[a] For swollen joint count and tender joint count, the baseline values are the values taken on the day of randomization before the first dose of investigational product.

Scoring Algorithms for Efficacy Endpoints

ACR Responder

The ACR response is a composite endpoint that classifies a patient as a responder or non-responder based upon improvement from baseline as defined by 7 endpoints. To be considered an ACR 20/50/70 responder, a subject must experience the following:

- At least 20/50/70% improvement in tender joint count (68 joints), AND
- At least 20/50/70% improvement in swollen joint count (66 joints), AND
- At least 20/50/70% improvement in 3 of the 5 following endpoints:
    - Physician global assessment (PhGA)
    - Patient global assessment (PtGA)
    - Patient global assessment of joint pain
    - Functional disability (HAQ-DI)
    - Acute phase reactant (CRP)

If joint data are partially missing, the joint counts will be prorated prior to deriving ACR responses. There is no imputation for missing secondary individual components (ie, HAQ-DI, CRP or ESR, pain VAS, PtGA, and PhGA).

In the case that some ACR components are missing, the ACR scores will be based on the non-missing components. If a subject's non-missing components are not sufficient to determine ACR response, then that subject will be considered as missing ACR response. The corresponding algorithm is listed below (using ACR 20 as an example):

1. If either swollen or tender joint counts is not improved at least 20%, the subject is an ACR non-responder
2. If both tender and swollen joint counts are improved at least 20%, and
   a. If 3 or more of the other 5 ACR components are improved at least 20%, the subject is an ACR 20 responder;
   b. If 3 or more of the other 5 ACR components are not improved at least 20%, the subject is an ACR 20 non-responder;
   c. If none of the above, the ACR response for the subject cannot be determined due to missing data and is therefore set to missing.
3. If the percent improvement of either swollen or tender joint counts is missing, and
   a. If 3 or more of the other 5 ACR components are not improved at least 20%, the subject is an ACR 20 non-responder;
   b. Otherwise, the ACR response for the subject cannot be determined due to missing data and is therefore set to missing.

Clinical Disease Activity (CDAI)

CDAI will be calculated based on the number of tender (TEN28) and swollen (SW28) joint using a 28-joint count, physician global assessment on a 0-10 scale (PhGA), and patient global assessment on a 0-10 scale (PtGA). Since PhGA and PtGA are collected on a 0-100 scale, the formula has been modified to convert these scores to a 0-10 scale as such:

$$CDAI = SW28 + TEN28 + PhGA/10 + PtGA/10$$

If one or more components are missing in the formula, then CDAI is set to be missing.

Leeds Dactylitis Index (LDI)

The Leeds Dactylitis Index (LDI) will be calculated based on the tenderness and circumference of dactylitic digits. First, identify involved digit (A) as a digit with tenderness score greater than 0. Second, identify control digit (B) as in the following scenarios:

1. contralateral digit (the digit on opposite hand and foot) if tenderness score associated with contralateral digit is 0.
2. If the tenderness score for contralateral digit is also greater than 0, then control digit is the digit listed in standard reference table below.

Note in scenario 2), the contralateral digit is also an involved digit (A), look up its control digit from standard reference table below.

For any digit with a missing tenderness score, it will be removed from LDI calculation.

For any involved digit with contralateral digit circumference missing, only calculate LDI when both involved digit and its contralateral are tender (tenderness score > 0). In latter case, refer to standard reference table in the same fashion as in scenario 2 above.

To calculate LDI for (A), take the ratio of circumference in mm of the involved digit (A) and control digit (B) . If the ratio difference [(A/B)-1] is greater than 10%, multiply 100 and the ratio difference [(A/B)-1] by the tenderness score corresponding to involved digit (A): :

LDI = (A/B-1) x100 x tenderness score of A, if [(A/B)-1] > 10%

The scores from each involved digit are summed by subject by visit to give a grand total
Standard reference table for LDI:

Digit	Men	Women
thumb	70	58
index	63	54
middle finger	63	54
ring	59	50
little finger	52	44
Great Toe	82	72
2nd toe	52	46
middle toe	50	44
4th toe	50	44
toe	52	45

Example:

Raw data for subject 1:

Folder	BODYCD	SEX	circum_right	circum_left	tender_right	tender_left
D1	Toe #2	Male	70	61	1	0
D1	Toe #3	Male	71	61	1	2
D1	Finger #4	Male	85	90	0	1

There are 4 involved digits: Toe# 2 right, Toe#3 left, Toe#3 right, Finger#4 left.

To calculate LDI for Toe#2 right, A = 70, B = 61, ratio difference = A/B-1 = 14.75%, therefore LDI = (A/B-1) x 100 x tenderness score= 14.75*1 = 14.75

Digit	Circumference of Involved Digit (A)	Contralateral Digit (B)	Ratio [(A/B)-1] (C)	Tenderness 0, 1, 2, 3 (D)	Final Score if C≥10% C*D*100
Finger #4	90	85	5.88%	1	n/a
Toe #2	70	61	14.75%	1	14.75
Toe #3	71	50	42%	1	42
Total LDI at D1					14.75+42+44=100.75

Tender Dactylitis Count

Tender dactylitis count will be calculated based on the tenderness of each dactylitis digits. If the tenderness for a digit (finger or toe) is greater than 0, then the count will be coded as 1 for that digit. The total score for tender dactylitis count will be the sum of the tender dactylitis count for each digit, ranging from 0 to 20.

Disease Activity Score (DAS28-CRP)

DAS28-CRP will be calculated based on the number of tender (TEN28) and swollen (SW28) joints using a 28-joint count, CRP in mg/L, and a 100 mm visual analog scale measuring the subject's general health (PtGA). Since CRP is collected in mg/dL, the formula has been modified to convert this measure to mg/L as such:

$$DAS28\text{-}CRP = 0.56*\sqrt{TEN28} + 0.28*\sqrt{SW28} + 0.36*ln[(CRP*10)+1] + 0.014*PtGA + 0.96$$

If one or more components are missing in the formula, then DAS28_CRP is set to be missing.

Disease Activity index for Psoriatic Arthritis (DAPSA) (range 0-164)

DAPSA will be calculated based on the number of tender (TEN68) and swollen (SW66) joints, patient global assessment on a 0-10 scale (PtGA), patient global assessment of pain on a 0-10 scale (PtGAJP) and CRP in mg/dL. Since PtGA and PtGAJP are collected on a 0-100 scale, the formula has been modified to convert these scores to a 0-10 scale as such:

$$DAPSA = SW66 + TEN68 + PtGA/10 + PtGAJP/10 + CRP$$

If one or more components are missing in the formula, then DAPSA is set to be missing.

Leeds Enthesitis Index

The Leeds enthesitis index will be used for the PASDAS derivation only and will be calculated based on the following entheses:

Entheses	Tenderness (0=Not Tender, 1= Tender)
Medial femoral condyle (left)	0 or 1
Medial femoral condyle (right)	0 or 1
Lateral epicondyle (left)	0 or 1
Lateral epicondyle (right)	0 or 1
Achilles tendon into calcaneum (left)	0 or 1
Achilles tendon into calcaneum (right)	0 or 1
Total Leeds Enthesitis Index	sum (range 0-6)

If one or more components are missing, then enthesitis index is set to be missing.

SPARCC enthesitis index

SPARCC enthesitis index will be calculated based on the following entheses:

Entheses	Tenderness (0=Not Tender, 1= Tender)
Medial epicondyle (left)	0 or 1
Medial epicondyle (right)	0 or 1
Lateral epicondyle (left)	0 or 1
Lateral epicondyle (right)	0 or 1
Supraspinatus (left)	0 or 1
Supraspinatus (right)	0 or 1
Greater trochanter (left)	0 or 1
Greater trochanter (right)	0 or 1
Quads into superior patella (left)	0 or 1
Quads into superior patella (right)	0 or 1
Max {Patella ligament into inferior patella (left), Patella ligament into tibial tubercle (left)}	0 or 1
Max{Patella ligament into inferior patella (right), Patella ligament into tibial tubercle (right)}	0 or 1
Achilles tendon into calcaneum (left)	0 or 1
Achilles tendon into calcaneum (right)	0 or 1
Plantar fascia into calcaneum (left)	0 or 1
Plantar fascia into calcaneum (right)	0 or 1
Total SPARCC Enthesitis Index	sum (range 0-16)

If one or more components are missing, then SPARCC is set to be missing.

HAQ-DI

The HAQ-DI is the disability assessment component of the HAQ that assesses a patient's level of functional ability. It is calculated based on the mean score from the following eight categories: dressing & grooming, rising, eating, walking, hygiene, reach, grip, and usual activities. Each category is comprised of two or three questions that are scored as 0 (without any difficulty), 1 (with some difficulty), 2 (with much difficulty), or 3 (unable to do). The highest score for questions in each category (range 0 to 3)

determines the score for the category, unless aids or devices are required. Dependence on equipment or physical assistance increases a lower score (ie, scores of 0 or 1) to the level of 2 to more accurately represent underlying disability. If the subject's highest score is a 3 then it stays a 3. Devices and aids are assigned to the categories as follows:

HAQ-DI Category	Companion AIDS OR DEVICES item
DRESSING & GROOMING	Devices used for dressing (button hook, zipper pull, long-handled shoe horn, etc.)
ARISING	Built up or special chair
EATING	Built up or special utensils
WALKING	Cane, walker, crutches, wheelchair
HYGIENE	Raised toilet seat, bathtub seat, bathtub bar, long-handled appliances in bathroom
REACH	Long-handled appliances for reach
GRIP	Jar opener (for jars previously opened)

The eight category scores are averaged into an overall HAQ-DI score on a scale from 0 (no disability) to 3 (completely disabled) when 6 or more categories are non-missing. If more than 2 categories are missing, the HAQ-DI score is set to missing.

Minimal Disease Activity (MDA)

MDA response will be derived based on the following criteria:

Outcome	Criteria
Tender joint count (0-68)	≤ 1
Swollen joint count (0-66)	≤ 1
Body surface area (0-100)	≤ 3
Patient global assessment of joint pain – PTGAJP (0-100) *(used for patient pain VAS)*	≤ 15
Patient global assessment of disease activity – PTGA (0-100) *(used for patient global disease activity VAS)*	≤ 20
HAQ-DI (0-3) *(used for HAQ)*	≤ 0.5
SPARCC enthesitis index (0-16) *(used for tender entheseal points)*	≤ 1

Use prorated scores for Tender and Swollen joint count. If 5 of 7 criteria are met then the subject has achieved MDA response.
If all components are missing, then MDA is set to be missing. If one or more but not all components are missing in the formula, then MDA is set to be zero.

Very Low Disease Activity (VLDA)

VLDA response will be derived based on the following criteria:

Outcome	Criteria
Tender joint count (0-68)	≤ 1
Swollen joint count (0-66)	≤ 1
Body surface area (0-100)	≤ 3
Patient global assessment of joint pain (0-100) *(used for patient pain VAS)*	≤ 15
Patient global assessment of disease activity (0-100) *(used for patient global disease activity VAS)*	≤ 20
HAQ-DI (0-3) *(used for HAQ)*	≤ 0.5
SPARCC enthesitis index (0-16) *(used for tender entheseal points)*	≤ 1

Use prorated scores for Tender and Swollen joint count. If all 7 criteria are met then the subject has achieved VLDA response. If all components are missing, then VLDA is set to be missing. If one or more but not all components are missing in the formula, then VLDA is set to be zero.

mNAPSI

The modified NAPSI will be calculated for all fingernails at baseline. For remaining visits, the modified NAPSI will only be calculated for the worst fingernail (ie, the fingernail with the highest score) based on the baseline assessment. The total score will range from 0-13 as follows:

Clinical feature	Score
Pitting (0-3 based on severity)*	0, 1, 2 or 3
Nail plate crumbling (0-3 based on severity)*	0, 1, 2 or 3
Onycholysis and oil drop dyschromia (0-3 based on severity)*	0, 1, 2 or 3
Leukonychia (0=absent, 1=present)	0 or 1
Red spots in lunula (0=absent, 1=present)	0 or 1
Nail bed hyperkeratosis (0=absent, 1=present)	0 or 1
Splinter hemorrhages (0=absent, 1=present)	0 or 1
Total mNAPSI Score	sum (range 0-13)

Severity Codes				
Clinical feature	Score=0	Score=1	Score=2	Score=3
Pitting	no pits	1-10	11-49	≥50
Nail plate crumbling	no crumbling	1-25%	26-50%	>50%
Onycholysis and oil drop dyschromia	no onycholysis	1-10%	11-30%	>30%

If one or more components are missing, then mNAPSI is set to be missing.

PASDAS

PASDAS will be calculated based on the physician global assessment (PhGA), patient global assessment (PtGA), SF-36 physical component scale (SF-36PCS), number of swollen joints based on a 66 joint count (SW66), number of tender joints based on a 68 joint count (TEN68), Leeds enthesitis index (0-6), tender dactylitis count (0-20), and CRP in mg/L. Since CRP is collected in mg/dL, the formula has been modified to convert this measure to mg/L as such:

$$\text{PASDAS} = (0.18\sqrt{\text{PhGA}}) + (0.159\sqrt{\text{PtGA}}) - (0.253\sqrt{\text{SF-36PCS}}) + (0.101*\ln(\text{SW66}+1)) + (0.048*\ln(\text{TEN68}+1)) + (0.23*\ln(\text{Leeds enthesitis index} +1)) + (0.37*\ln(\text{tender dactylitis count} +1)) + (0.102*\ln((\text{CRP}*10)+1) +2) * 1.5$$

If one or more components are missing, then PASDAS is set to be missing.

Simplified Disease Activity Index (SDAI)

SDAI will be calculated based on the number of tender (TEN28) and swollen (SW28) joint using a 28-joint count, physician global assessment on a 0-10 scale (PhGA), patient global assessment on a 0-10 scale (PtGA) and CRP in mg/dL. Since PhGA and PtGA are collected on a 0-100 scale, the formula has been modified to convert these scores to a 0-10 scale as such:

$$SDAI = SW28 + TEN28 + PhGA/10 + PtGA/10 + CRP$$

If one or more components are missing, then SDAI is set to be missing.

Sharp-van der Heijde Modified Scoring (mTSS) of Radiographs

Erosions are scored on a 0 to 5 scale for the hands and on a 0 to 10 scale for the feet. The following 20 locations in each hand will be scored for hand erosions .

- 4 Distal inter-phalangeal joints (2-5)
- 4 Proximal Inter-Phalangeal Joints (2-5)
- 5 Metacarpo-Phalangeal Joints (1-5)

- Inter-Phalangeal Joint of the thumb
- Proximal first Metacarpal Bone
- Radius Bone
- Ulnar Bone
- Trapezium and Trapezoid (as one unit; multangular)
- Navicular Bone
- Lunate Bone

The following 6 locations in each foot will be scored for foot erosions :

- 5 Metatarso-phalangeal joints (1-5)
- Inter-phalangeal joint of the great toe

The maximum score per joint is 10, with a maximum of 5 at each side of the joint. There is a maximum of two large, middle line passing erosions per joint in the foot (in the hand, technically only one is allowed as a score of six (6) is not permitted). Likewise, a maximum of 4 large erosions not passing the middle line is permitted (five '2s' to achieve a score of 10 is not feasible). Thus, the maximum erosion score (ES) in the hands/wrist is 200 and in the feet 120.

Joint space narrowing (JSN) is scored on a scale of 0 to 4 for hands and feet resulting in a maximum JSN score of 160 and 48, respectively. The following 20 joints in each hand and 6 joints in each foot will be scored for JSN:

Hand Joints:

- 4 Distal inter-phalangeal joints (2-5)
- 5 Metacarpo-phalangeal joints (1-5)
- 4 Proximal inter-phalangeal joints (2-5)
- 3 Carpo-metacarpal joints (3-5)
- Interphalangeal Joint of thumb (IP)
- Radio-carpal joint
- Multangular-navicular joint
- Capitate-navicular-lunate joint

Foot Joints:

- 5 Metatarso-phalangeal joints
- Inter-phalangeal joint of great toe

To obtain the total mTSS to identify cases requiring adjudication, scores for erosions and Joint Space Narrowing (JSN) in both the hands and feet will be added together. Any "P" or "G" will be considered the maximal score for the feature (erosions and JSN) per location in the calculation of the total mTSS. Any "N" or "S" will be considered null in the calculation of the total mTSS. The range of scores is summarized below.

Parameter	**hands/wrists**	**feet**	**Total (hands + feet)**
ES range	40 joints * 5 score = 0-200	12 joints * 10 score = 0- 120	0-320
JSN range	40 joints * 4 score =0- 160	12 joints * 4 score = 0-48	0-208
mTSS range	0-360	0-168	**0-528**

In addition, since the radiographs will be read at a central location by two independent readers and possibly an adjudicator, calculate the score using the following rules:

- Use the scores from adjudicator where available. If adjudicator results are present for one subject at any time point, then use adjudicator results for the same subject at all time points and ignore the results from either reader 1 or reader 2.
- Use the mean of the scores from 2 readers if both are non-missing when adjudicator scores are unavailable.
- Use the non-missing score from the reader if the score from the other reader is missing, and adjudicator scores are unavailable
- Set to missing if both scores from 2 readers are missing and no adjudicator scores are available

Swollen Joint Count

The swollen joint count is an assessment of the swelling of joints using 0-1 point scale (0 = none, 1 = present). The total swollen joint count is calculated by summing the number of joints with present swelling. If a joint has been replaced, fused, fractured or receives an intra-articular steroid injection, it will be counted as swollen at all visits subsequent to the replacement or injection.

Swollen Joint Count (66)	
Temporomandibular joints (n=2)	Metacarpophalangeal joints (MCP) (n=10)
Sternoclavicular joints (n=2)	Fingers distal interphalangeal joints (DIP) (n=8)
Acromioclavicular joints (n=2)	Knees (n=2)
Shoulders (n=2)	Ankles (n=2)
Elbows (n=2)	Tarsi (n=2)
Wrists (n=2)	Metatarsal phalangeal joints (MTP) (n=10)
Fingers proximal interphalangeal joints (PIP) (n=10)	Toes proximal interphalangeal joints (PIP) (n=10)

Swollen Joint Count (28)	
Shoulders (n=2)	Metacarpophalangeal joints (MCP) (n=10)
Elbows (n=2)	Proximal interphalangeal joints (Fingers PIP) (n=10)
Wrists (n=2)	Knees (n=2)

If at least half but not all joints are evaluable (33 for the 66 joint count or 14 for the 28 joint count), then the prorated swollen joint count will be calculated. The prorated score will be adjusted based upon the number of evaluable joints: the counted score will be multiplied by 66 or 28 as applicable and divided by the number of joints evaluated. Otherwise, the number of swollen joints is missing.

Tender Joint Count

The tender joint count is an assessment of the pain and/or tenderness of joints using 0-1 point scale (0 = none, 1 = present). The total tender joint count is calculated by summing the number of joints with present tenderness. If a joint has been replaced, fused, fractured or receives an intra-articular steroid injection, it will be counted as tender at all visits subsequent to the replacement or injection.

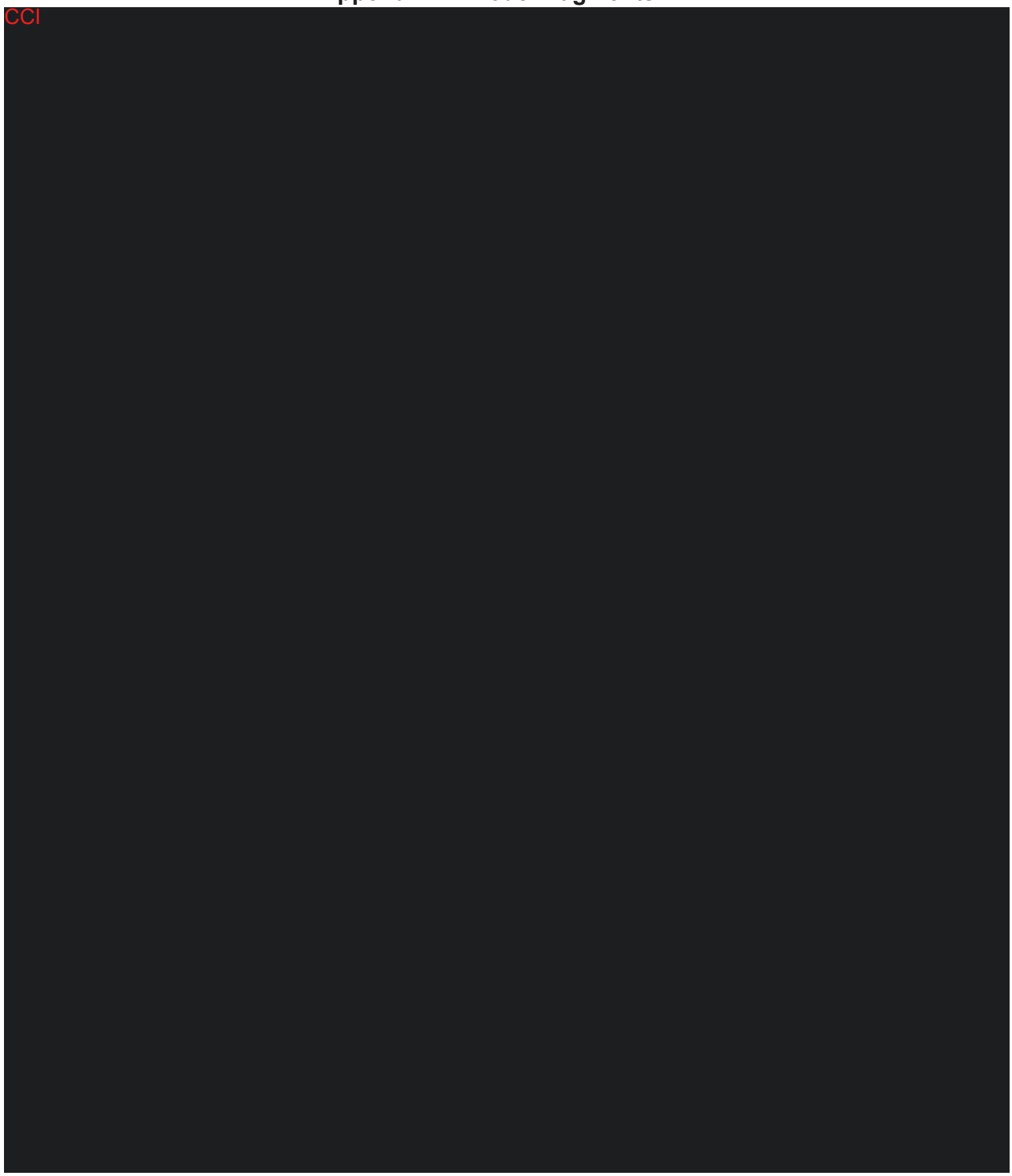
Tender Joint Count (68)	
Temporomandibular joints (n=2)	Metacarpophalangeal joints (MCP) (n=10)
Sternoclavicular joints (n=2)	Metatarsal phalangeal joints (MTP) (n=10)
Acromioclavicular joints (n=2)	Toes proximal interphalangeal joints (PIP) (n=10)
Shoulders (n=2)	Knees (n=2)
Elbows (n=2)	Ankles (n=2)
Wrists (n=2)	Tarsi (n=2)
Fingers proximal interphalangeal joints (PIP) (n=10)	Hip (n=2)
Fingers distal interphalangeal joints (DIP) (n=8)	

Tender Joint Count (28)	
Shoulders (n=2)	Metacarpophalangeal joints (MCP) (n=10)
Elbows (n=2)	Proximal interphalangeal joints (Fingers PIP) (n=10)
Wrists (n=2)	Knees (n=2)

If at least half but not all joints are evaluable (34 for the 68 joint count or 14 for the 28 joint count), then the prorated tender joint count will be calculated. The prorated score will be adjusted based upon the number of evaluable joints: the counted score will be multiplied by 68 or 28 as applicable and divided by the number of joints evaluated. Otherwise, the number of tender joints is missing.

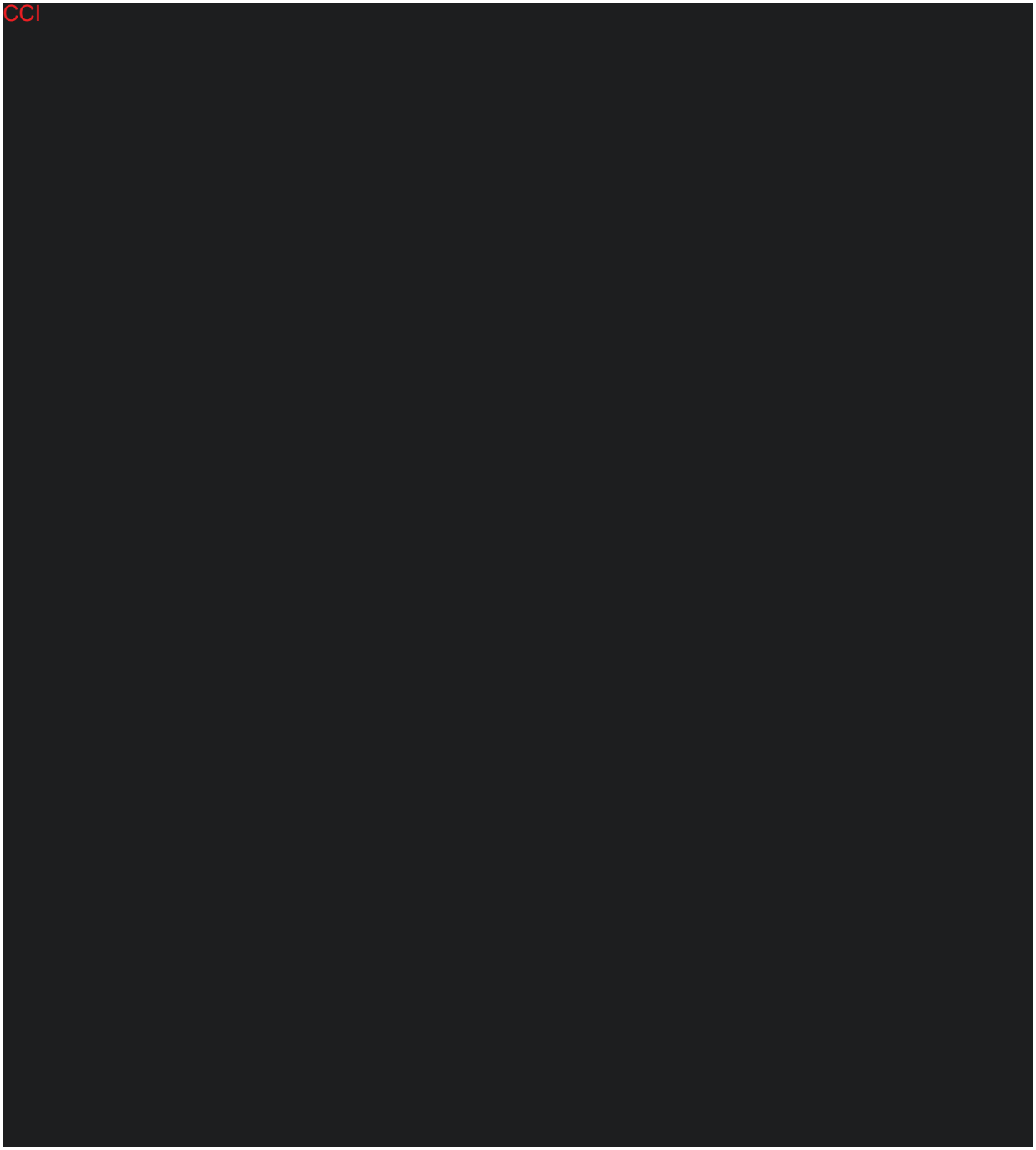
Appendix B. Code Fragments

CCI



CCI

[redacted]


CCI

Appendix C. Reference Values/Toxicity Grades

Adverse event severity is graded based on the National Cancer Institute Common Toxicity Criteria for Adverse Events (CTCAE) version 4.0.

Laboratory toxicity is graded based on the National Cancer Institute Common Toxicity Criteria (CTC) version 4.0 with the following additions:

	Grade				
	0	**1**	**2**	**3**	**4**
BUN	WNL	>1.5 – 2.0 x ULN	>2.0 – 3.0 x ULN	>3.0 – 4.0 x ULN	>4.0 x ULN
HGB	WNL	>ULN	>(ULN+1) – (ULN+2)	>(ULN+2)	-
Urine Protein	NEG, NEGATIVE, TRACE, NORMAL, NOT DETECTED	+, +1, 1+	++, +2, 2+, +++, +3, 3+	++++, +4, 4+	-
Platelet Count	WNL	> ULN – 1.0×10^{12}/L	> 1.0×10^{12}/L	-	-
Leukocyte Count	WNL	> ULN – 2.0×10^{10}/L	> 2.0×10^{10}/L – 4.0×10^{10}/L	> 4.0×10^{10}/L	-