1.0 Title Page

Statistical Analysis Plan

Study M16-244

Phase 2 Single-Arm Open-Label Extension Study to Investigate Safety with Risankizumab in Psoriatic Arthritis Subjects Who Have Completed Week 24 Visit of Study 1311.5

Date: 16 Jul 2018

Version 1.0

2.0 Table of Contents

List of Tables

List of Figures

3.0 Introduction

This statistical analysis plan (SAP) describes the statistical analyses to be completed by the Data and Statistical Science Department for ABBV-066 Study M16-244. It provides details to further elaborate statistical methods as outlined in the protocol and describes analysis conventions to guide the statistical programming work.

Unless noted otherwise, all analyses will be performed using SAS version 9.4 (SAS Institute Inc., Cary, NC 27513) under the UNIX operating system.

4.0 Study Objectives, Design and Procedures

4.1 Primary Objectives

The primary objective of Study M16-244 is to assess the safety and tolerability of risankizumab in PsA subjects who have completed all doses of study drug and the Week 24 visit in Study 1311.5.

4.2 Secondary Objectives

The secondary objectives include the assessment of symptomatic and structural efficacy of risankizumab in PsA subjects who completed all doses of study drug and the Week 24 visit in Study 1311.5.

4.3 Overall Study Design and Plan

This is a 52-week multicenter, single-arm, OLE study to assess safety, tolerability and efficacy of risankizumab in PsA subjects who have completed all doses of study drug and the Week 24 visit of Study 1311.5 (Phase 2 randomized clinical trial in PsA subjects, sponsored by Boehringer Ingelheim).

Subjects who complete all doses of study drug and the Week 24 visit in Study 1311.5, meet all inclusion criteria (Section 5.2.1 of the protocol) and none of the exclusion criteria (Section 5.2.2 of the protocol) have the option to sign the informed consent form and enter

the OLE study, receiving risankizumab 150 mg administered every 12 weeks with an additional dose of risankizumab made available at the Week 4 of the OLE study visit for those subjects that have not achieved a protocol defined response (defined as an improvement in tender and swollen joint count of $\geq$ 20% compared with the baseline in Study 1311.5), provided the investigator believes this may be beneficial to the subject.

If the Study M16-244 Week 0 visit occurs on the same day or in the 4 days following the Study 1311.5 Week 24 visit, only selected activities are required as some activities are common to both studies and are not required twice. See Study Activity Table in Appendix C of the protocol. If the OLE Week 0 visit occurs > 4 days and $\leq$ 8 weeks after the Study 1311.5 Week 24 visit, all procedures specified in Appendix C of the protocol at Week 0 of the OLE study are required.

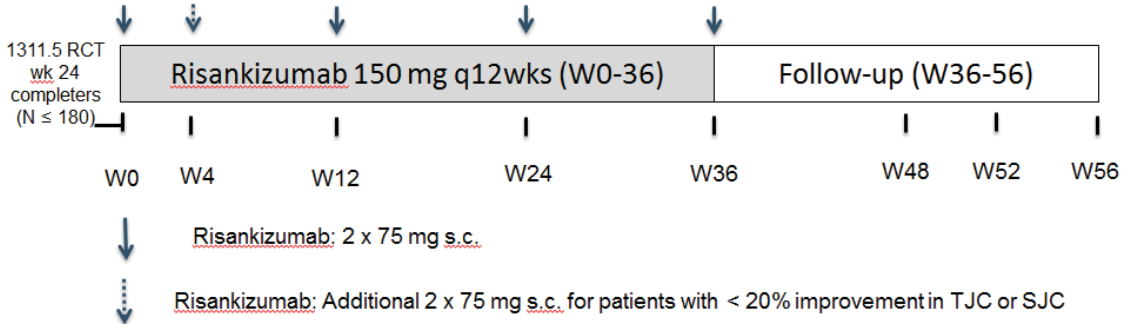
Subjects that enroll in the OLE study are no longer enrolled in Study 1311.5 and will not complete the Week 28 or Week 32 visits in Study 1311.5.

Subjects who are not enrolled in Study M16-244 within 8 weeks of the Study 1311.5 Week 24 visit are not eligible to participate in Study M16-244.

This study is anticipated to enroll up to 180 subjects. Study visits will occur at Week 0/Enrollment, Week 4, 12, 24, 36, 48 and 52. A ±4 day window is permitted around scheduled study visits after enrollment.

A schematic of the study design is shown below in Figure 1.

Figure 1. Study M16-244 Study Design



4.4 Sample Size

Study 1311.5 will enroll approximately 180 subjects. The sample size of Study M16-244 will depend on how many subjects completed all doses of study medication and the Week 24 visit of Study 1311.5 and enroll in this OLE.

4.5 Interim Analysis

No interim efficacy analysis is planned for this study.

4.6 Data Monitoring Committee

An independent DMC operating under a charter is responsible for reviewing safety data periodically to ensure subject safety and to monitor the conduct of the trial and the integrity of the data. An independent analytic team external to both BI and AbbVie will provide the unblinded results to the DMC.

5.0 Analysis Sets

5.1 Definition for Analysis Sets

Safety Analysis Set

The Safety Analysis Set consists of all subjects who have received at least one dose of study medication in Study M16-244. All efficacy and safety analyses will be based on the Safety Analysis Set.

All efficacy and safety analyses will be summarized for "All ABBV-066 150 mg q12wks with optional wk4 dose" group, which includes all subjects who received at least one dose of ABBV-066 150 mg every 12 weeks in Study M16-244 (with an optional additional dose at Week 4 for subjects with < 20% improvement in tender joint count (TJC) or swollen joint count (SJC)), regardless of what treatment group they were randomized to in Study 1311.5.

5.2 Variables Used for Stratification of Randomization

No stratified randomization is planned. This is an open-label study.

6.0 Analysis Conventions

6.1 Definition of Baseline

The last non-missing measure collected on or before the day of the first dose of study drug injection in Study 1311.5 for each randomized subject will be used as the Baseline for summary of demographics, disease characteristics and safety analyses. For efficacy endpoints except modified total Sharp score (mTSS), the baseline is defined as the last non missing pre-treatment observation prior to first dose in Study 1311.5. The Baselines for mTSS are defined as the last non-missing observation prior to Day 13 and closest to Day 1 in Study 1311.5.

For mTSS, data for Week 24 in Study 1311.5 and Week 24 in Study M16-244 are considered in analyses. In the context of visits corresponding to mTSS data collection

they will be referred to as Week 24 (Study 1311.5) and Week 24 (Study M16-244) respectively.

If multiple measurements are recorded on the same day, the worst of the multiple measurements will be used.

6.2 Definition of Final Observation (Applicable to Safety Analyses Other than Laboratory and Vital Signs)

Final observation is defined as the last non-missing observation collected within 140 days (20 weeks) following the last dose of study drug.

6.3 Definition of Rx Days (Days Relative to the First Dose of Study Drug)

Rx days are calculated for each time point relative to the date of first dose of study drug. They are defined as the number of days between the date of the first dose of study drug and the specific time point. Rx days are negative values when the time point of interest is prior to the first study drug dose day. Rx days are positive values when the time point of interest is on or after the first study drug dose day. The day of the first dose of study drug is defined as Rx Day 1, while the day prior to the first study drug dose is defined as Rx Day –1 (there is no Rx Day 0).

6.4 Definition of Analysis Windows

All time points and corresponding time windows are defined based on Rx Days.

For efficacy analyses, laboratory parameters, and vital sign variables, analysis windows are constructed using the following algorithm:

- Determine the nominal Rx day for each visit (e.g., Week 4 [4 weeks after Baseline visit] equals Rx Day 29).
- In order to include all post baseline data, the first post-baseline interval starts on the first day after the first dose of study drug (Rx Day 2).

- Determine the window around a specific nominal Rx day by adding or subtracting half of the interval between adjacent visits (e.g., days between Week 2 and Week 4 is 14). The threshold between adjacent visits is determined by splitting the interval evenly between the visits. If the resulting split is between Rx days, then the threshold is determined as the midpoint between the adjacent visits. If the resulting split is on a Rx day, then the threshold is determined as being between that Rx day and the Rx day prior to it (e.g., the split between Week 2 and Week 4 would be between Rx Days 22 and 23).
- If more than one assessment is included in a time window the assessment closest to the nominal day will be used. If there are two observations equidistant to the nominal day, the one after the nominal day will be used in analyses. If more than one assessment is included on the same day, then the worst assessment on that day will be used in analyses. For laboratory values, if there are multiple measurements on the same day, average value will be used.

The protocol specified visits and corresponding time windows used in the various efficacy analyses, laboratory parameters, and vital sign variables, are presented in the following tables.

Table 1. Analysis Windows for Efficacy Variables (PtGA, Patient's Assessment of PsA Pain Intensity, HAQ-DI, SF-36v2, TJC68, SJC66, PASI, BSA, sPGA, PhGA, hsCRP for DAS)

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
0	1	1	1
4	2	29	57
12	58	85	127
24	128	169	211
36	212	253	295
48	296	337	351
52	352	365	378

Table 2. Analysis Windows for BASDAI

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
0	1	1	1
24	2	169	211
36	212	253	295
48	296	337	351
52	352	365	378

Table 3. Analysis Windows for FACIT-FATIGUE

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
0	1	1	1
24	2	169	211
36	212	253	295
48	296	337	378

Table 4. Analysis Windows for LDI, LEI, SPARCC, mNAPSI During the Open Label Treatment Period

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
0	1	1	1
12	2	85	127
24	128	169	211
36	212	253	295
48	296	337	351
52	352	365	378

Table 5. Analysis Windows for mTSS

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
0 (Study 1311.5)	–999 (Study 1311.5)	1 (Study 1311.5)	13 (Study 1311.5)
24 (Study 1311.5)	14 (Study 1311.5)	169 (Study 1311.5)	Min (306, Last Day) (Study 1311.5)
24 (Study M16-244)	2 (Study M16-244)	169 (Study M16-244)	253 (Study M16-244)
48 (Study M16-244)	254 (Study M16-244)	337 (Study M16-244)	420 (Study M16-244)

Table 6. Analysis Windows for Analysis of Safety Laboratory Tests and Vital Signs

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
0	1	1	1
4	2	29	57
12	58	85	127
24	128	169	211
36	212	253	295
48	296	337	351
52	352	365	378

Table 7. Analysis Windows for Local Tolerability During the Open Label Treatment Period

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
0	1	1	1
4	2	29	57
12	58	85	127
24	128	169	211
36	212	253	295

6.5 Missing Data Handling

All data will be summarized as observed, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the summary for that visit.

7.0 Demographics, Baseline Characteristics, Medical History, and Previous/Concomitant Medications

7.1 Demographic and Baseline Characteristics

Demographic data and baseline characteristics are reported for the last non-missing observation before the first dose of study medication in the preceding RCT Study 1311.5. Demographic data and baseline characteristics are not re-assessed before the start of Study M16-244.

Demographic and baseline characteristics will be summarized for the Safety Analysis Set of Study M16-244 for the "All ABBV-066 150 mg q12wks with optional wk4 dose" treatment group. Summaries of baseline characteristics will be presented for the Baseline of Study 1311.5.

The number of observations, mean, standard deviation, median, minimum, Q1, Q3 and maximum will be summarized for continuous variables. Categorical or discrete variables will be summarized via frequencies and percentages. The summary statistic will be computed for the analysis treatment group defined in Section 5.1. No statistical test will be performed.

Main Demographic and Baseline Characteristics

- Sex (male/female)
- Ethnicity (Hispanic or Latino, Other)
- Age (years), defined as the number of years from date of birth to date of first drug
- Age Categories (< 65, ≥ 65)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White)
- Ethnicity (Hispanic or Latino, Other)
- Body Weight (kg)
- Body Weight Categories (< 100 kg, ≥ 100 kg)

- Height (cm)
- Body Mass Index (BMI) (kg/m^2)
- Body Mass Index (BMI) Category (kg/m2) (BMI < 30 vs BMI ≥ 30)

PsA Medical History and Disease Characteristics at Baseline

- Categorical:
  - MDA (Minimal Disease Activity) for PsA
  - sPGA (Static Physician Global Assessment) assessed in subjects with ≥ 3% BSA of psoriatic plaques
  - Presence of Dactylitis (yes/no)
  - Presence of Enthesitis based on LEI (yes/no)
  - Presence of Enthesitis based on SPARCC (yes/no)
  - Presence of nail psoriasis (yes/no)
  - Body Surface Area (BSA) of psoriatic plaques: ≥ 3% vs < 3%
  - Prior exposure to TNF antagonists (TNFi's; Experienced versus Naive)
  - Concurrent exposure to Methotrexate (Concurrent use versus Non-concurrent use)
  - Presence of inflammatory spondylitis (yes/no)
  - Duration of PsA Categories (≤ 5, [5, 10], > 10)
  - Rheumatoid Factor (RF) status: Positive (≥ 15) or Negative (< 15)
- Continuous
  - PASI (Psoriasis Area and Severity Index) assessed in subjects with ≥ 3% BSA of psoriatic plaques
  - DAS28 (Disease Activity Score in 28 joints) – hsCRP
  - Dactylitis count in subjects with dactylitis at baseline 0 – 20
  - Leeds Dactylitis Index (LDI)in subjects with dactylitis at baseline
  - Leeds Enthesitis Index (LEI) in subjects with enthesitis (based on LEI) at baseline (yes/no) 0 – 6
  - Duration since diagnosis of PsA in years

- CASPAR classification criteria total score
- SPARCC (Spondyloarthritis Research Consortium of Canada) Enthesitis Index in subjects with enthesitis (based on SPARCC) at baseline (yes/no) 0 – 16
- mNAPSI (Modified Nail Psoriasis Severity Index) assessed in subjects with nail psoriasis at baseline 0 – 130

ACR and/or DAS Components at Baseline

- SJC (Swollen Joint Count(s) 66) 0 – 66
- TJC (Tender Joint Count(s) 68) 0 – 68
- SJC (Swollen Joint Count(s) 28) 0 – 28
- TJC (Tender Joint Count(s) 28) 0 – 28
- Patient's assessment of pain on VAS 0 – 100
- Patient's global assessment of the disease on VAS 0 – 100
- Physician's global assessment of the disease on VAS 0 – 100
- HAQ-DI (Health Assessment Questionnaire-Disability Index) 0 – 3
- High sensitivity C-Reactive Protein (hsCRP) mg/L

Patient Report Outcomes at Baseline

- SF-36v2 (Short Form-36 Health Survey)
- FACIT-FATIGUE (Functional Assessment of Chronic Illness Therapy-Fatigue) 0 to 52
- BASDAI (Bath AS Disease Activity Index) (in subjects with baseline inflammatory spondylitis) 0 to 10

Imaging Characteristics at Baseline

- mTSS (Modified Total Sharp Score) 0 – 528
  - Total Erosion Score 0 – 320
  - Joint Space Narrowing Score 0 – 208

Clinical Tests at Screening

- Tuberculin PPD skin test (Negative/Positive/Undetermined), QuantiFERON TB Gold test,
- Hepatitis Testing
- Pregnancy test (Negative/Positive/Not Done/NA)

General Use

- Smoking status (Never-smoked, Ex-smoker, Currently smokes)
- Alcohol status (Non-drinker, drinks – no interference, drinks – possible interference)

7.2 Medical History

Medical history other than psoriatic arthritis or cardiovascular diseases will be summarized using body systems and condition/diagnosis as captured on the eCRF. The body systems will be presented in alphabetical order and the conditions/diagnoses will be presented in alphabetical order within each body system. The number and percentage of subjects with a particular condition/diagnosis will be summarized for each treatment arm. Subjects reporting more than one condition/diagnosis within a body system will be counted only once for that body system.

7.3 Prior Treatment and Concomitant Medications

Prior and concomitant medications will be summarized. A prior medication is defined as any medication taken prior to the first dose of study drug in Study 1311.5. A concomitant medication is defined as any medication that started prior to the first dose of study drug in Study 1311.5 and continued to be taken after the first dose of study drug or any medication that started after the first dose of study drug, until 140 days (20 weeks) after the last dose of study drug. The number and percentage of subjects who had taken medications will be summarized by generic drug name assigned by the World Health

Organization (WHO) Drug Dictionary for both prior and concomitant medications. The following medication will be summarized:

- Prior systemic corticosteroid
- Prior MTX
- Prior csDMARDS other than MTX
- Prior biologic DMARDS
- Prior NSAIDS
- Concomitant systemic corticosteroid
- Concomitant MTX
- Concomitant csDMARDS other than MTX
- Concomitant NSAIDS

Concomitant biologic DMARDS and csDMARDS other than MTX and on or before Week 16 in Study 1311.5 are protocol violations.

7.4 Protocol Deviation

Number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though she/he did not satisfy entry criteria
- Subject who received wrong treatment or incorrect dose
- Subject who received excluded or prohibited concomitant treatment
- Subject who developed withdrawal criteria during the study and was not withdrawn

8.0 Subject Disposition

The number of subjects will be tabulated by country, investigator site and overall for the Safety Analysis Set. The number of subjects for each of the following categories will be summarized, for overall and for each treatment group in the Safety Analysis Set.

- Number of subjects that took at least one dose of study drug
- Number of subjects who completed Week 36 dose
- Number of subjects who prematurely discontinued from the study before Week 36 dose due to other reasons.

In addition, the reasons for premature discontinuation (primary reason and all reasons) from the trial and/or from the medication collected from CRF by the following categories will be summarized with frequencies and percentages. Subjects with multiple reasons for premature discontinuation will be counted once in the calculation of the number and percentage of total discontinuations ("Premature Discontinuation").

- Adverse event
- Protocol violation
- Lost to follow-up,
- Withdrew by subject,
- Others.

9.0 Study Drug Exposure and Compliance

A summary of study drug duration will be provided for the Safety Analysis Set.

Study drug duration (days) will be summarized with the number of subjects, mean, standard deviation, minimum, median and maximum for each treatment arm. It will also be summarized in 4-week intervals with frequencies and percentages for the number of subjects receiving study drug doses in each interval. In addition cumulative duration of risankizumab (including total subject years) will be summarized.

The study drug duration will be calculated as follows:

Duration = date of last injection – date of first injection + 84 day

Total drug duration across Study 1311.5 (feeder) and Study M16-244 (OLE) will be calculated as follows:

Duration over both studies = date of last injection in Study M16-244 – date of first injection in Study 1311.5 + 84 day – MAX(0, date of first injection in Study M16-244 – date of last injection in Study 1311.5 – 84)

Compliance

Each subject will receive five risankizumab injections during the OLE study. Treatment compliance (TC) will be summarized based on Safety Analysis Set. The treatment compliance is defined as the number of visits that the subject received injections divided by the number of visits a subject is supposed to receive injections during the treatment period (i.e., from the date of the subject's first injection through the date of the last injection). Subjects with missing data for study drug administration will be excluded from the summary. Specifically, TC will be calculated using the following formula:

TC (%) = (number of visits where injections are received/number of visits where injections are supposed to be received) * 100

The number of visits where injections are supposed to be received for a subject = Max (Round ([Last Study Drug Dose Date – First Study Drug Dose Date – 1]/84 + 1 + 1*I(Does the Subject Receive the Optional Week 4 dose)),1), where round (x) rounds x to the nearest integer.

The maximum number of visits where injections are supposed to be received will be limited to 5.

The actual drug and dose delivered to the subject will be listed.

10.0 Efficacy Analysis

10.1 General Considerations

The table below provides the overview of the efficacy analyses to be performed on different endpoints.

There will be no statistical testing and only descriptive statistics will be provided. These include the number of observations, mean, standard deviation, median, minimum and maximum for continuous variables; and frequencies and percentages for categorical or discrete variables.

Table 8. Summary of Efficacy Variables and Corresponding Analyses

Efficacy Variables	Summary Method
Categorical Efficacy Endpoints:	
Secondary Binary Endpoint: ACR20 responses at all measured time points	• Point estimates and 95% CIs of the response rates • Imputation: OC
Other Categorical Endpoints: **At all measured time points:** • ACR50 • ACR70 • Minimal disease activity (MDA) • PASI75 response in subjects with a ≥ 3% baseline PsO BSA • PASI90 response in subjects with a ≥ 3% baseline PsO BSA • PASI100 response in subjects with a ≥ 3% baseline PsO BSA • sPGA clear and almost clear in subjects with a ≥ 3 baseline PsO BSA • Psoriatic Arthritis Response Criteria (PsARC) • European League Against Rheumatism (EULAR) Response • Presence of dactylitis (yes/no) • No Progression of mTSS	• Point estimates and 95% CIs of the response rates • Imputation: OC
Continuous Efficacy Endpoints	
Secondary Continuous Endpoint: Change from baseline in Study 1311.5 at Week 24 (Study 1311.5), Week 24 (Study M16-244) and Week 48: • mTSS Change from baseline in Study M16-244 at all measured timepoints: • HAQ-DI • SF-36	• Point estimates, 95% CIs of mean change from baseline together with SD, Min, Q1, Median, Q3 and Max • Imputation: OC

Table 8. Summary of Efficacy Variables and Corresponding Analyses (Continued)

Other Continuous Endpoints: **At all measure time points as compared to baseline in Study M16-244:** • Change in Tender Joint Count (TJC) • Change in Swollen Joint Count (SJC) • Change in Dactylitis Count • Change in SPARCC Enthesitis Index • Change in mNAPSI • Change in PhGA(VAS) • Change in Patient's Assessments of PsA Pain Intensity(VAS) • Change in PtGA assessments • Change in High sensitivity C-Reactive Protein (hsCRP) • Change in DAS28-hsCRP • Change in LDI • Change in LEI • Change in FACIT-FATIGUE Change in BASDAI in subjects with baseline inflammatory spondylitis, based on investigator judgment, in Study 1311.5.	• Point estimates, 95% CIs of mean change from baseline together with SD, Min, Q1, Median, Q3 and Max • For SF-36, spyderdiagram will be provided • Imputation: OC

For Binary Endpoints:

Estimates of the response rate and the associated 95% confidence interval will be calculated using the exact method.

For Continuous Endpoints:

The number of observations, mean, standard deviation, median, minimum, Q1, Q3 and maximum will be summarized. Point estimates and 95% CIs of mean change from baseline will be provided.

10.2 Primary Efficacy Analysis

There is no primary endpoint for this study.

10.3 Secondary Efficacy Analyses

The following secondary endpoints will be summarized:

- ACR 20 response at all measured time points in Study M16-244
- Change in mTSS at Week 24 (Study 1311.5), Week 24 (Study M16-244), and Week 48 as compared to baseline in Study 1311.5
- Change in HAQ-DI at all measured time points in Study M16-244 compared to baseline in Study 1311.5
- Change SF-36 at all measured time points in Study M16-244 compared to baseline in Study 1311.5

For the change from baseline in continuous endpoints, the number of observations, mean, standard deviation, median, minimum, Q1, Q3 and maximum will be summarized. Point estimates and 95% CIs of mean change from baseline will be provided.

For the summary of binary endpoints by visit, point estimate and 95% CI will be calculated and presented.

The primary analysis for the above secondary efficacy endpoints will be performed based on the Safety Analysis Set with observed cases.

10.4 Further Efficacy Analyses

Further efficacy endpoints will include, but are not limited to, the following:

- ACR 50 response at all measured time points in Study M16-244
- ACR 70 at all measured time points in Study M16-244
- Change in Tender Joint Count at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Change in Swollen Joint Count at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Change in HAQ-DI at all measured time points in Study M16-244 as compared to baseline in Study 1311.5

- Change in SF-36 at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Change in Dactylitis count at all measured time points in Study M16-244 as compared to baseline in Study 1311.5 in subjects with dactylitis at baseline
- Change in SPARCC Enthesitis Index at all measured time points in Study M16-244 as compared to baseline in Study 1311.5 in subjects with enthesitis based on SPARCC sites at baseline
- Change in mNAPSI at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Change in PhGA (VAS) at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Change in Patient's Assessment of PsA Pain Intensity (VAS) at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Change in PtGA assessments at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Change in High sensitivity C-Reactive Protein (hsCRP) at all measured time points in Study M16-244 as compared to baseline in Study 1311.5.
- Minimal disease activity (MDA) at all measured time points in Study M16-244
- Change in DAS28-hsCRP at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- PsO endpoints assessed at all measured time points in Study M16-244 in subjects with a ≥ 3% baseline PsO BSA in Study 1311.5:
  - PASI 75 and PASI 90 response
  - Change in sPGA clear and almost clear
- EULAR (European League Against Rheumatism) response at all measured time points in Study M16-244
- Change in PsARC (Psoriatic Arthritis Response Criteria) at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Presence of dactylitis (yes/no) at all measured time points in Study M16-244 in subjects with dactylitis at baseline

- Change in LDI at all measured time points in Study M16-244 as compared to baseline in Study 1311.5 in subjects with dactylitis at baseline
- Change in LEI at all measured time points in Study M16-244 as compared to baseline in Study 1311.5 in subjects with enthesitis based on LEI sites at baseline
- Change in FACIT-FATIGUE at all measured time points in Study M16-244 as compared to baseline in Study 1311.5
- Change in BASDAI at all measured time points in Study M16-244 (in subjects with baseline inflammatory spondylitis, based on investigator judgment, in Study 1311.5) as compared to baseline in Study 1311.5

Further efficacy endpoints will be summarized descriptively. Continuous endpoints will be summarized with the use of box plots, while proportions will be displayed by histograms as appropriate.

10.5 Handling of Multiplicity

There is no multiplicity adjustment needed.

10.6 Efficacy Subgroup Analysis

A subgroup analyses based on prior randomization and weight will be performed to evaluate the consistency of efficacy in the secondary endpoints.

10.7 Efficacy Variables Definitions and Conventions

10.7.1 ACR Criteria

ACR criteria is a commonly used standard criteria mentioned in the guidance of American College of Rheumatology to evaluate the effectiveness of investigational drugs in PsA clinical trials. It is a composite measurement calculated based on the improvement over a set of core measurements.

ACR20 is defined as at least 20% improvement (compared to baseline values) in tender and swollen joint counts and at least 20% improvement in 3 of the remaining 5 core set measures (patient's assessment of pain [VAS], PtGA, PhGA, HAQ-DI and CRP).

ACR50 and ACR70 are similarly defined with at least 50% and 70% improvement, respectively.

Calculation Rules for ACR Criteria

A subject will be classified as an ACR20 (ACR50, ACR70) responder, if the following conditions are met:

1. ≥ 20% (50%, 70%) improvement from baseline in tender joint count (TJC)* and
2. ≥ 20% (50%, 70%) improvement from baseline in swollen joint count (SJC)* and
3. ≥ 20% (50%, 70%) improvement from baseline in at least 3 of the following 5:
    - Patient's assessment of pain
    - Patient's Global Assessment of Disease Activity for Arthritis (PtGA)
    - Physician's Global Assessment of Disease Activity for Arthritis (PhGA)
    - Patient's self-assessment of physical function (i.e., measured by Health Assessment Questionnaire (HAQ-DI score)
    - Acute-phase reactant value CRP

There are seven components to be evaluated to define an ACR response. Missing values for each component can occur due to missing assessment, a missed visit or due to dropout from the study. Depending on the pattern of the missing components, ACR responses may be or may not be determined using observed values only.

To maximize the utilization of observed information at certain visits and to be scientifically robust as possible, the principle to calculate ACR response is to minimize imputation whenever possible. "As Observed" ACR responses will be calculated

first based on a derived visit window instead of the nominal visit identifier collected from the CRF.

To calculate "as observed" ACR responses:

- Identify the observed component xx% improvement indicator (0/1/missing), 1 means achieving ≥ xx% improvement from baseline and 0 means < xx% improvement from baseline (e.g., xx% representing 20%/50%/70%).
- ACRxx = 0 if TJC indicator = 0 OR SJC indicator = 0 OR at least 3 out of 5 components improvement indicators = 0;
- ACRxx = 1 if TJC indicator = 1 AND SJC indicator = 1 AND at least 3 out of 5 components improvement indicators = 1
- For all other cases, "as observed" ACRxx = missing since ACRxx cannot be determined.

The following table illustrates examples for as-observed ACR calculations.

Example	TJC 68	SJC 66	Component 1	Component 2	Component 3	Component 4	Component 5	ACR20-Response?
A	1	1	1	1	1	.	.	Yes
B	1	0	1	1	1	1	1	No
C	.	0	.	.	.	.	.	No
D	1	.	1	1	1	1	1	.
E	1	1	0	0	0	1	1	No
F	.	.	0	0	0	.	.	No
G	1	1	1	1	0	0	.	.

Legend: 1 = ≥ 20% improved compared to baseline; 0 = < 20% improved compared to baseline; "." missing

Derived Visit Windowing Rule for ACR Response Calculation

To identify the component value in a visit window:

- ACR component values will first be determined at each date within a visit window.

- ACR component values at each date will be combined to determine the "as observed" ACR composite score at each date in each window.
- After this calculation, if multiple non-missing ACR composite scores are available within a given visit window, the non-missing ACR composite score closest to the target day will be used. If two composite scores have the same distance from the target day, the later one will be used. The corresponding date will be used as the "as observed" ACR response date in the derived efficacy dataset.
- If a non-missing ACR composite score is not available for any day within a given visit window, the windowed component values for that visit will be used to calculate the ACR composite score for that visit window (component value windowing follow the same rules as in steps described above). The date of observed ACR composite score will be determined by the first available ACR component date, in the order of TJC, SJC, Pain, PGA, PhGA, HAQ-DI, CRP/ESR, in the derived efficacy dataset.

10.7.2 Joint Evaluation

Anatomical joints are evaluated for swelling and tenderness at every study visit. The 34 anatomical joints in Table 9 are assessed in this study for both the left and right side of the body.

Table 9. Anatomical Joints Assessed for Calculation of Tender and Swollen Joint Counts (TJC68 and SJC66)

Temporomandibular	Sternoclavicular	Acromio-clavicular	Shoulder
Elbow	Wrist	Metacarpophalangeal I	Metacarpophalangeal II
Metacarpophalangeal III	Metacarpophalangeal IV	Metacarpophalangeal V	Interphalangeal Thumb Joint
Proximal Interphalangeal Joint II of the Hand	Proximal Interphalangeal Joint III of the Hand	Proximal Interphalangeal Joint IV of the Hand	Proximal Interphalangeal Joint V of the Hand
Distal Interphalangeal Joint II of the Hand	Distal Interphalangeal Joint III of the Hand	Distal Interphalangeal Joint IV of the Hand	Distal Interphalangeal Joint V of the Hand
Hip[a]	Knee	Ankle	Transverse Tarsus
Metatarsophalangeal I	Metatarsophalangeal II	Metatarsophalangeal III	Metatarsophalangeal IV
Metatarsophalangeal V	Interphalangeal Joint I of the Foot	Proximal Interphalangeal Joint II of the Foot	Proximal Interphalangeal Joint III of the Foot
Proximal Interphalangeal Joint IV of the Foot	Proximal Interphalangeal Joint V of the Foot		

a. Hip joints are not assessed for swelling.

At each study visit, a joint evaluator assessed whether a particular joint was "tender or painful" where presence of tenderness was scored as "1" and the absence of tenderness was scored as "0." The total tender joint count (TJC68), which is based on 68 joints, will be derived as the sum of all "1"s and proportional extrapolation will be used to compute joint counts for the joints that are replaced or not assessed. A similar method will be followed for the derivation of total swollen joint count (SJC66), which is based on 66 joints as the hip joints are excluded. Thus, the range for TJC68 will be 0 to 68 and 0 to 66 for SJC66. Joints with surgery (e.g., joint replacement) will not be assessed.

10.7.3 Patient's Global Assessment of Disease Activity Visual Analog Scale (VAS)

The patient's global assessment of disease activity will be performed using a horizontal 100 mm VAS, ranging from 0 (very well) to 100 (very poor) after the question:

"Considering all the ways psoriatic arthritis affects you, please indicate with a vertical mark (|) through the horizontal line how well you are doing today."

10.7.4 Physician's Global Assessment of Disease Activity Visual Analog Scale (VAS)

The physician will assess Patient's disease activity at the time of visit using a physician's Global Assessment of Disease VAS. The range is 0 to 100 mm with no activity being indicated by 0 and severe activity by 100.

10.7.5 Patient's Global Assessment of Pain

The patient's assessment of pain will be performed using a horizontal 100 mm visual analog scale (VAS), ranging from 0 (no pain) to 100 (severe pain) after the question:

"Please indicate with a vertical mark (|) through the horizontal line the most pain you had from your psoriatic arthritis today."

10.7.6 Disease Activity Score Based on DAS28 (hsCRP)

The Disease Activity Score (DAS) is a combined index used to measure the disease activity in patients with PsA. The DAS provides a score between 0 and 10, indicating how active the joint disease in Psoriatic Arthritis is at the time of measurement. DAS28 (hsCRP) score will be determined based on a continuous scale of combined measures of TJC, SJC, Patient Global Assessment of Disease Activity (PtGA) (in mm), and hsCRP (in mg/L).

DAS28 (hsCRP) = 0.56 $\sqrt{}$ ("TJC28") + 0.28 $\sqrt{}$ ("SJC28") + 0.36 ln (hsCRP + 1) + 0.014 PtGA + 0.96

where $\sqrt{}$ is square root and ln is natural log. A score > 5.1 generally means high disease activity; 2.6 ≤ DAS28 (hsCRP) < 3.2 generally means low disease activity (LDA); DAS28 (hsCRP) < 2.6 is generally considered clinical remission (CR).

Table 10. Anatomical Joints for DAS28 (hsCRP) Calculation

Shoulder	Elbow	Wrist	Thumb Interphalangeal
Metacarpophalangeal I	Metacarpophalangeal II	Metacarpophalangeal III	Metacarpophalangeal IV
Metacarpophalangeal V	Proximal Interphalangeal II	Proximal Interphalangeal III	Proximal Interphalangeal IV
Proximal Interphalangeal V	Knee		

10.7.7 Health Assessment Questionnaire – Disability Index (HAQ-DI)

HAQ-DI[1] is a self-reported patient outcome measurement tool commonly used in RA and PsA clinical trials to measure physical functioning in RA and PsA patients. The HAQ-DI composite score is calculated as the mean of the scores from the 8 following categories with a range of 0 – 3 (0 = no disability; 3 = worst disability): Dressing and Grooming, Rising, Eating, Walking, Hygiene, Reach, Grip, and Activities. The higher the score, the more likely to be associated with morbidity and mortality for the PsA patient. Under each category there are 2 – 3 items on the amount of difficulty they have in performing specific activities with four response options from 0 (no difficulty) to 3 (unable to do). In addition to these eight categories, there is an aids or devices/help from other person section ("companion items") that is used to record the type of assistance, if any, a subject uses for his/her usual activities in each of the eight categories.

The preferred and traditional scoring method for scoring HAQ-DI is the Standard HAQ-DI approach which takes into account the use of the aids/devices section. There are three steps to compute the HAQ-DI score and a patient must have a score for at least six of the eight categories, otherwise a HAQ-DI score cannot be computed. The first step is to compute each of the category score. The maximum score for all the questions in each category is considered as the score for the category. The second step is to adjust the score of each category based on use or no use of aids/devices and/or help from another person when indicated. If aids or devices and/or assistance from another person are

checked for a disability category, and the score for the category is 0 (no difficulty) or 1 (some difficulty), increase it to 2 (much difficulty). If the score for the category is a 2, it remains a 2, and if it is a 3, it remains a 3. The third step is to sum the adjusted categories scores and divide by the number of categories answered (minimum 6) to obtain a HAQ-DI score of 0 to 3.

10.7.8 SF-36v2 (Acute Form)

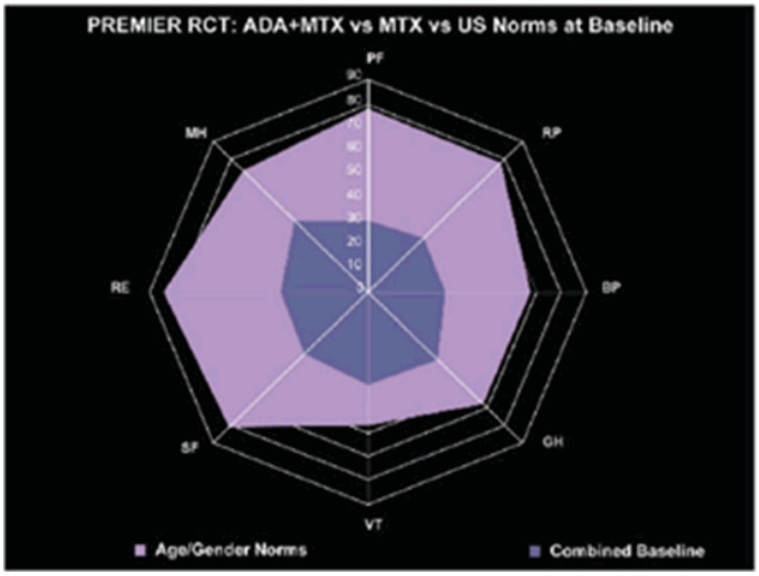
The 36-Item Short Form, Version 2 (SF-36v2) Questionnaire with 1 week recall consists of 36 general health questions and this study is using the form for 4 weeks recall period (standard form). It has 2 components: physical and mental. For each component, a transformed summary score is calculated using 8 subdomains: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health.[2,3]

The coding and scoring for the SF-36 will use the software provided by Optimum.

Changes from baseline for each component and each of the 8 subdomains will be summarized by visit and treatment group. Analysis will be based on Observed Case only.

Spydergrams will be plotted for the 8 sub-domains of SF-36v28. In a spydergram, physical function (PF) is at the top, 12 o'clock, followed clockwise by role physical (RP), bodily pain (BP) and general health perceptions (GH), and vitality (VT) at the 6 o'clock position, followed by social functioning (SF), role emotional (RE) and mental health index (MH) clockwise (Figure 2). Domain scores are plotted from 0 (worst) at the center to 100 (best) at the outside; demarcations along axes of the domains present changes of 10 points, representing one to two times minimally clinical important differences (MCID). An example of a spydergram is shown below. The spydergram is a radar chart that can be plotted using PROC GRADAR in SAS version 9.4.

Figure 2. An Example of Spydergram of SF-36



Note: Strand 2009.[4]

10.7.9 Psoriasis Area and Severity Index (PASI)

Four anatomic sites – head, upper extremities, trunk, and lower extremities – are assessed for erythema, induration and desquamation using a 5-point scale:

- 0 = no symptoms
- 1 = slight
- 2 = moderate
- 3 = marked
- 4 = very marked

Based on the extent of lesions in a given anatomic site, the area affected is assigned a numerical value:

- 1 = < 10%
- 2 = 10% – 29%

- 3 = 30% – 49%
- 4 = 50% – 69%
- 5 = 70% – 89%
- 6 = 90% – 100%

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

$PASI = 0.1(E_h + I_h + D_h)A_h + 0.2(E_u + I_u + D_u)A_u + 0.3(E_t + I_t + D_t)A_t + 0.4(E_l + I_l + D_l)A_l$

where *E*, *I*, *D*, and *A* denote erythema, induration, desquamation, and area, respectively, and *h*, *u*, *t*, and *l* denote head, upper extremities, trunk, and lower extremities, respectively. PASI scores range from 0.0 to 72.0 with the highest score representing complete erythroderma of the severest possible degree. Typically scores of 3 or less represent mild disease, scores over 3 and up and including 15 represent moderate disease and scores over 15 are considered to be associated with severe disease.

The frequency and percentage of PASI/75/90/100 (defined as 75%, 90% and 100% improvement in PASI score compared to baseline) will be summarized by treatment group in subjects with ≥ 3% BSA (Body Surface Area) psoriasis involvement at baseline.

BSA will be approximated by the following formula using the affected area information (lower bound of each degree) collected for PASI.

BSA = 0.1*Lower(Area_h) + 0.2*Lower(Area_u) + 0.3*Lower(Area_t) + 0.4*Lower(Area_l)

where Lower() takes the lower bound percentage corresponding to the affected area score. For example, if Area_h = 1, then Lower(Area_h) = 0%. If Area_h = 5, then Lower (Area_h) = 70%.

10.7.10 Modified Nail Psoriasis Severity Index (mNAPSI)

mNAPSI is the tool to assess each abnormality for each of a subject's fingernails. Three features or groups of features (pitting, onycholysis and oil-drop dyschromia, and crumbling) of each fingernail will be graded on a scale from 0 to 3. Four features (leukonychia, splinter hemorrhages, hyperkeratosis, and red spots in the lunula) will be graded as either present or absent for each fingernail. After a subject has assessed all of their nails consider all aspects of all of the subject's fingernails and place a mark on the visual analog scale giving a global assessment of their fingernails. For detailed instruction, please see protocol Appendix K. The overall mNAPSI is calculated as the sum of all the components for all of a subject's nails. In the case of missing data for any component of mNAPSI sum up all available component values and divide it by the total weight (if a score ranges 0 – 1, the weight is 1. If a score ranges 0 – 3, the weight is 3), then multiply by 130 (the upper range of the total score). If missing more than half of the total weight then the total score will be set as missing.

10.7.11 European League Against Rheumatism (EULAR) Response Criteria

EULAR response will be evaluated according to the following table:

	Improvement in DAS28 from baseline:		
DAS28 at endpoint	**> 1.2**	**> 0.6 and ≤ 1.2**	**≤ 0.6**
≤ 3.2	Good	Moderate	None
> 3.2 and ≤ 5.1	Moderate	Moderate	None
> 5.1	Moderate	None	None

EULAR response will be analyzed as a binary variable with Good being defined as response and Moderate & None being defined as non-response.

10.7.12 Minimal Disease Activity (MDA) in PsA

Table 11. Minimal Disease Activity (MDA) Criteria in Psoriatic Arthritis

A patient is classified as in MDA when 5 of the following 7 criteria are met:
• TJC68 ≤ 1 • SJC66 ≤ 1 • PASI ≤ 1 or BSA ≤ 3 • Patient assessment of pain ≤ 15 • PtGA ≤ 20 • HAQ-S ≤ 0.5 • Tender entheseal points ≤ 1

Derivation rules for MDA follow the same logic as ACR. MDA response can be determined if at least 5 of the 7 criteria are met (responder), or if at least 3 of the 7 criteria are not met (non-responder). Selection of multiple MDA responses within one visit window follows the same rules as ACR.

10.7.13 Static Physician's Global Assessment (sPGA)

This sPGA is a 5 point score ranging from 0 to 4, based on the physician's assessment of the average thickness, erythema, and scaling of all psoriatic lesions (Table 12).

The assessment is considered "static" which refers to the subjects disease state at the time of the assessments, without comparison to any of the subject's previous disease states, whether at Baseline or at a previous visit.

A lower score indicates less body coverage, with 0 being clear and 1 being almost clear.

The investigator (or qualified site personnel) scores the erythema, induration and scaling of all psoriatic lesions from 0 – 4 based on the following descriptors:

Scoring

A composite score is generated from the above data and the final sPGA is determined from this composite score as follows:

Clear 0 = 0 for all three

Almost clear 1 = mean > 0, < 1.5

Mild 2 = mean > = 1.5, < 2.5

Moderate 3 = mean > = 2.5, < 3.5

Severe 4 = mean > = 3.5

Table 12. sPGA Rating Scale for Overall Psoriatic Disease

Score	Short description	Detailed description
0	**clear**	No signs of psoriasis. Post-inflammatory hyperpigmentation may be present.
1	**almost clear**	Normal to pink coloration; Just detectable (possible slight elevation above normal skin) No to minimal focal scaling
2	**mild**	Pink to light red coloration Mild thickening (slight but definite elevation, typically edges are indistinct or sloped) Predominantly fine scaling
3	**moderate**	Dull to bright red coloration Clearly distinguishable to moderate thickening Moderate scaling
4	**severe**	Bright to deep dark red coloration; Severe thickening with hard edges Severe coarse scaling covering almost all or all lesions

10.7.14 Psoriatic Arthritis Response Criteria (PsARC) Response of Psoriasis

A subject is defined as a PsARC responder if, and only if, they have an improvement in two of the following four factors (with at least one factor being a joint count) and no worsening in the remaining factors:

- Patient global assessment of disease activity (0 – 100 mm VAS scale, improvement defined as decrease of ≥ 20 mm)
- Physician global assessment of disease activity (0 – 100 mm VAS scale, improvement defined as decrease ≥ 20 mm)
- Tender 68-joint count (improvement defined as decrease of ≥ 30%)
- Swollen 66-joint count (improvement defined as decrease of ≥ 30%)

Derivation rules for PsARC follow the same logic as ACR.

10.7.15 Leeds Dactylitis Index (LDI)

The LDI measures the ratio of the circumference of the affected digit to the circumference of the digit on the opposite hand or foot. The ratio of circumference is multiplied by a tenderness score, using a modification of LDI which is a binary score (1 for tender, 0 for non-tender), using a minimum difference of 10% to define a dactylitic digit. When the difference is less than 10% the LDI for that digit is 0. If both sides are considered involved, or the circumference of the contralateral digit cannot be obtained, the number will be compared to data provided in the standard reference tables The reference tables for standard digit circumference are listed in Appendix A1. This modification is referred to as LDI basic and will be applied in this study. The LDI requires a finger circumference gauge or a tape measure to measure digital circumference. Overall LDI for a patient will be calculated as the sum of their individual LDIs from the digits where the assessment was done. The LDI for a single digit will be calculated using the formula (A/B-1)*100*C, where A is the circumference of the affected digit, B is the circumference of the contralateral digit (or digit from the reference table if it is also affected), and C is the binary tenderness score for the digit.

Dactylitis Count:

The dactylitis count is the number of fingers and toes with dactylitis that are affected, tender, and with a minimum circumference difference of 10%, with a range of 0 – 20. If a site is not assessed assign it a value of "0."

Presence of Dactylitis:

If dactylitis is present with any finger or toe, the subject is counted as a subject with dactylitis.

10.7.16 Leeds Enthesitis Index (LEI)

LEI is a validated enthesitis index that uses 6 sites for evaluation of enthesitis: lateral epicondyle humerus left and right, Achilles tendon insertion left and right and medial condyle femur left and right. The LEI demonstrated substantial to excellent agreement with other scores in the indication of psoriatic arthritis. The LEI score ranges from 0 to 6.

Enthesitis Count:

Tenderness on examination is recorded as either present (1) or absent (0) for each of the 6 sites (If a site is not assessed assign it a value of "0"), for an overall score range of 0 – 6.

Presence of Enthesitis Based on LEI:

If enthesitis is present with any of the 6 sites for LEI, the subject is counted as a subject with enthesitis based on LEI.

10.7.17 Enthesitis (SPARCC)

The SPARCC Enthesitis index is an outcome measure for enthesitis in SpA created by the Spondyloarthritis Research Consortium of Canada.[5] Sixteen sites are evaluated. Tenderness at each site is quantified on a dichotomous basis: 0 means non-tender and 1 means tender. The SPARCC Enthesitis index is calculated by taking the sum of the scores from the 16 sites. If a site is not assessed assign it a value of "0." The SPARCC score ranges from 0 to 16.

Table 13. Enthesial Sites Examined for SPARCC Calculation (Left and Right)

Medial epicondyle	Lateral epicondyle	Supraspinatus insertion into greater tuberosity of humerus	Greater trochanter
Quadriceps insertion into superior border of patella	Patellar ligament insertion into inferior pole of patella or tibial tubercle	Achilles tendon insertion into calcaneum	Plantar fascia insertion into calcaneum

Presence of Enthesitis Based on SPARCC:

If enthesitis is present with any of the 16 sites for SPARCC, the subject is counted as a subject with enthesitis based on SPARCC.

10.7.18 Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue)

The FACIT-Fatigue is a 13-item questionnaire[6] that assesses self-reported fatigue and its impact upon daily activities and function.

The purpose of FACIT-Fatigue in this study is to assess the impact of fatigue on subjects with PsA.

Number of items: 13 items: I feel fatigued (-), I feel weak all over (-), I feel listless (washed out) (-), I feel tired (-), I have trouble starting things because I am tired (-), I have trouble finishing things because I am tired (-), I have energy (+), I am able to do my usual activities (+), I need to sleep during the day (-), I am too tired to eat (-), I need help doing my usual activities (-), I am frustrated by being too tired to do the things I want to do (-), I have to limit my social activity because I am tired (-).

Response options/scale: Answers are based on a 5-point Likert scale. Responses of "not at all," "a little," "somewhat," "quite a bit," and "very much" are available for each question, and correspond to scores of 0, 1, 2, 3, and 4, respectively (4 = not at all fatigued

to 0 = very much fatigued). The FACIT Fatigue Scale is ranged from 0 to 52 and the higher the score, the better the quality of life.

Score for each item is calculated by either subtracted from 4 or adding 0 depending on whether it is a reversal item or not. FACIT Fatigue Scale is then calculated by adding up all item scores, multiplied by 13 and divided by the number of items answered. It is essentially a prorated subscale if there are missing values for some items. If less than or equal to 50% of the items are answered (e.g., 6 out of 13), the proration is not acceptable and the scale will not be computed.

Recall period for items: 7 days.

10.7.19 Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

The BASDAI is composed of 6 questions investigating 5 domains (fatigue, spinal pain, joint pain/swelling, areas of localized tenderness, morning stiffness), with 1 item for each of the first four domains and 2 items for the last domain (morning stiffness). Each item is scored on a one through 10 scale. A lower score indicates less disease activity.

Scoring of the BASDAI is as follows:

1. Measure each item of the BASDAI in centimeters (out of a total of 10)
2. BASDAI Score = 0.2*(Question 1 + Question 2 + Question 3 + Question 4 + 0.5* Question 5 + 0.5* Question 6)

The BASDAI Score ranges from 0 – 10. If one or more items are unanswered, take the average of the non-missing items. If there is more than one mark for an item, take the average.

Below are the 6 questions.

3. How would you describe the overall level of fatigue/tiredness you have experienced?
4. How would you describe the overall level of AS neck, back or hip pain you have had?
5. How would you describe the overall level of pain/swelling in joints other than neck, back or hips you have had?
6. How would you describe the overall level of discomfort you have had from any areas tender to touch or pressure?
7. How would you describe the overall level of morning stiffness you have had from the time you wake up?
8. How long does your morning stiffness last from the time you wake up?

10.7.20 Modified Total Sharp Score (mTSS)

The radiographic outcome will be assessed and scored according to Sharp's method (Van der Heijde modification) centrally by two qualified physicians/radiologists who will be blinded to the site number, subject number, treatment allocation, time sequence and clinical response.

Calculation of the Modified Total Sharp Score

To obtain the total mTSS score, scores for erosions and JSN in both the hands and feet will be added together.

The range of scores is summarized below.

	Hands	Feet	Total (Hands and Feet)
Erosion Score Range	0 - 200	0 - 120	0 - 320
Joint Space Narrowing Range	0 - 160	0 - 48	0 - 208
mTSS Range for Erosion and JSN	0 - 360	0 - 168	0 - 528

The following joints will be examined for assessing Erosions:

Foot[a]	Hand[b]			
1st IP	1st IP	3rd MCP	5th MCP	Ulnar
1st MTP	1st MCP	3rd DIP	5th DIP	MC1
2nd MTP	2nd PIP	4th PIP	Multangular[c]	
3rd MTP	2nd MCP	4th MCP	Navicular	
4th MTP	2nd DIP	4th DIP	Lunate	
5th MTP	3rd PIP	5th PIP	Radius	

a. IP: Inter-Phalangeal, MTP: Metatarso-Phalangeal.

b. IP: Inter-Phalangeal, PIP: Proximal Inter-Phalangeal, MCP: Metacarpophalangeal, DIP: Distal Inter-Phalangeal.

c. Trapezium/Trapezoid as read as one unit-Multangular.

The following joints will be examined for assessing Joint Space Narrowing:

Foot[a]	**Hand**[b]			
1st IP	1st MCP	3rd DIP	5th DIP	RC
1st MTP	2nd PIP	4th PIP	3rd CMC	1st IP
2nd MTP	2nd MCP	4th MCP	4th CMC	
3rd MTP	2nd DIP	4th DIP	5th CMC	
4th MTP	3rd PIP	5th PIP	MN	
5th MTP	3rd MCP	5th MCP	CNL	

a. IP: Inter-Phalangeal, MTP: Metatarso-Phalangeal.

b. PIP: Proximal Inter-Phalangeal, MCP: Metacarpo-Phalangeal, CMC: Carpo-Metcarpal, MN: Multangular-Navicular, CNL: Capitate-Navicular Lunate, RC: Radio-Carpal, DIP: Distal Inter-Phalangeal.

For each Joint and Bone assessed scores range as follows:

- Erosions: 0 – 5 (hands/wrists) or 0 – 10 (feet) to characterize the extent of erosions (where 0 denotes no erosion).
- Joint Space Narrowing: 0 – 4 to characterize the extent of Joint Space Narrowing (JSN) (where 0 denotes no narrowing).
- The categorical scores G and P could be possible for the entry for the following situation:
  - Osteolysis in the form of pencil-in-cup: Osteolysis of proximal phalanx and the base of the distal phalanx resulting in a pencil like proximal phalanx covered by cup like base of the distal phalanx. Pencil-in-cup will be scored as "P" where applicable.
  - Gross Osteolysis: Osteolysis of the phalanx resulting a loss of the normal joint structure, ususlly accompanied by shorting of the length of the phalanx. Gross osteolysis will be scored as "G" where applicable.

Erosion and JSN scores for each reader are calculated by taking the sum of the left and right joints as shown below.

$$\text{Erosion}_{\text{Reader i}} = \text{Erosion}_{\text{Left}} + \text{Erosion}_{\text{Right}}$$

$$JSN_{Reader\,i} = JSN_{Left} + JSN_{Right} \text{ for } i = 1, 2.$$

Thus, the maximum erosion score for all 40 joints in hands/wrists is 200. The maximum erosion score for all 12 joints in feet is 120. Thus, the total erosion score for hands/wrists and feet is 320 assessed on a total of 52 erosion joints.

The maximum score for JSN in all 40 hand/wrist joints is 160. The maximum score for JSN in all 12 feet joints is 48. Thus, the total JSN score for hand/wrist and feet is 208 based on total 52 JSN joints.

Since two independent readers evaluate each film, the mean score will be calculated for the two readers from the individual erosion and JSN scores as shown below:

$$\text{Erosion} = \frac{\text{Erosion}_{Reader1} + \text{Erosion}_{Reader2}}{2}$$

$$\text{JSN} = \frac{\text{JSN}_{Reader1} + \text{JSN}_{Reader2}}{2}$$

The mTSS for each reader is defined as the sum of the erosion and JSN scores:

$$TSS_{Reader\,i} = \text{Erosion}_{Reader\,i} + JSN_{Reader\,i} \text{ for } i = 1, 2.$$

The mTSS from Erosion score plus JSN score will be used for all x-ray endpoint calculations.

$$TSS = Erosion + JSN.$$

Handling of missing joints in mTSS score calculation

For categorical score G and P, the maximum score per location should be assigned before any imputation.

Missing joint score imputation will be performed for Erosion and JSN respectively. If a score at a location/joint is missing then the methods described below will be used to calculate Erosion and JSN total score for that visit. However, if the missing score was present in a previous reading of the image (Study M16-002) then that previous reading's score will be used for missing score imputation.

Method 1 (used in Study 1311.5):

- If the scores of more than 50% of JSN scoring locations are available (e.g., 27 or more JSN scoring locations out of 52 total JSN scoring locations), the total JSN score would be calculated as Total JSN Score = Avg of all available scores*52
- If 50% or less (e.g., 26 or fewer JSN scoring locations out of 52 total JSN scoring locations) of the JSN scoring locations are readable, the Total JSN Score will not be calculated
- If the scores of more than 50% of Erosion scoring locations are available (e.g., 27 or more Erosion scoring locations out of 52 total Erosion scoring locations), the total Erosion score would be calculated as Total Erosion Score = average of all available hand scores*40 + average of all available foot scores*12.
- If 50% or less (e.g., 26 or fewer erosion scoring locations out of 52 total erosion scoring locations) of the erosion scoring locations are readable, the Total erosion Score will not be calculated

Immunology Convention Method:

If a score at any location/joint is missing, the method described below will be used for deriving mTSS.

- If the score for a location/joint is missing at Baseline, this joint will not contribute to the calculation of mTSS for this subject at any visit within the reading session (even if the score for this location/joint is available at post-baseline visits).

- If the score for a location/joint is missing at all post-baseline visits within a reading session, this joint will not contribute to the calculation of mTSS for this subject at any visit within the reading session (even if the score for this location/joint is available at Baseline).
- If the score for a location/joint is available at Baseline and at least one post-baseline visit, missing scores for this joint at any other post-baseline visit will be imputed assuming no progression from the previous time point with available score.

Sensitivity analysis may be performed as needed.

Adjudication process

Two reviewers will independently review the images. Adjudication will occur for all subjects with pre-specified criteria for that study/indication between the two reviewers' mTSS change scores, in which case another reviewer, different from the reviewers who performed primary assessments, will make a third, independent assessment.

For the calculation of erosion and JSN total score, individual visit score of the 2 closest of the 3 readings (2 primary readers and adjudicator) will be used to determine the final score. If one score is in the middle (equally close), take the middle score. If one of 3 readings is missing, take the average of two non-missing readings. If two of 3 readings are missing, take the non-missing reading. If all 3 readings are missing, erosion/JSN = missing and hence mTSS = missing.

No Progression

No progression of mTSS is defined as a change in mTSS score less than or equal to 0.

11.0 Safety Analysis

11.1 General Considerations

Safety analyses will include reporting of adverse events, laboratory, and vital signs measurements. Safety analyses will be carried out using Safety Analysis Set, which includes all subjects who received at least one dose of study medication in Study M16-244.

Mean changes from baseline in all continuous laboratory parameters and vital signs variables at each visit will be summarized by treatment group. The baseline, minimum, maximum, and final value means will be presented for subjects who have both baseline and post-baseline values. Categorical data will be summarized using frequencies and percentages. Additionally, the number of non-missing values will be given.

11.2 Analysis of Adverse Events

Treatment-emergent adverse events (TEAEs) will be summarized and reported. Frequency of TEAEs will be summarized using subject data from Study M16-244 and a combination of data from Study 1311.5 and Study M16-244 that includes any subject receiving a dose in Study M16-244. Only the combination of data from Study 1311.5 and Study M16-244 that includes any subject receiving a dose in Study M16-244 will be used for per 100 patient years of exposure analyses.

11.2.1 Treatment-Emergent Adverse Events

All adverse events occurring between start of treatment and end of the residual effect period (REP) will be considered 'treatment emergent.' The REP is defined as 20 weeks (140 days) after the last trial medication application and will include adverse events reported through EOS visit. Adverse events that start before first drug intake and deteriorate under treatment will also be considered as 'treatment-emergent.'

Events where the onset date is the same as the study drug start date are assumed to be treatment emergent, unless the study drug start time and the adverse event start time are

collected and the adverse event start time is prior to the study drug start time. If an incomplete onset date is collected for an adverse event, the event will be assumed to be treatment-emergent unless there is other evidence that confirms that the event is not treatment-emergent (e.g., the event end date is prior to the study drug start date).

Adverse events that occur outside of 140 days from last dose in Study 1311.5 but before first dose in Study M16-244 will not be considered treatment emergent.

The number and percent of subjects experiencing treatment-emergent TEAEs will be tabulated using the Medical Dictionary for Drug Regulatory Activities (MedDRA®) version 20.0 by system organ class (SOC) and preferred term (PT).

Summary tables of TEAEs will be presented as follows:

1. **Overview of TEAEs**

   The number and percentage of subjects experiencing TEAEs will be summarized for the following adverse event categories:

   - Any TEAEs
   - Any severe TEAEs
   - Any serious TEAEs
   - Any related TEAEs
   - Any TEAEs leading to discontinuation of study drug
   - Any TEAEs leading to death
   - Any deaths
   - Any TEAEs of area of safety interest

2. **Adverse Events by System Organ Class and Preferred Term**

   The number and percentage of subjects experiencing TEAEs will be tabulated according to the primary MedDRA SOC and PT for each prior and current treatment group. Subjects reporting more than one adverse event for a given MedDRA PT

will be counted only once for that term (most severe incident for the severity tables and most related incident for the relationship tables). Subjects reporting more than one type of adverse event within a SOC will be counted only once for that SOC. Subjects reporting more than one type of adverse event will be counted only once in the overall total.

The system organ classes will be presented in alphabetical order and the preferred terms will be presented in alphabetical order within each system organ class.

3. **Adverse Events by Maximum Severity/Toxicity**

The severity grading of AEs follows Rheumatology Common Toxicity Criteria (RCTC).

- Grade 1 – mild
- Grade 2 – moderate
- Grade 3 – severe
- Grade 4 – life threatening

The TEAEs will also be summarized by maximum severity. If a subject has an adverse event with unknown severity, then the subject will be counted in the severity category of "unknown," even if the subject has another occurrence of the same event with a severity/toxicity present. The only exception is if the subject has another occurrence of the same adverse event with the most extreme severity (e.g., "Life threatening"). In this case, the subject will be counted under the "Life threatening" category.

4. **Adverse Events by Maximum Relationship**

Adverse events will be summarized by maximum relationship to study drug, as assessed by the investigator. Relationship of an AE to study drug is assessed by the investigator and collected in the CRF as 'Yes' or 'No.' If a subject has an adverse event with unknown relationship, the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the same event with a relationship of "No." If the subject has another occurrence of the same

adverse event with a relationship assessment of "Yes," the subject will be counted under the "yes" category.

5. **TEAEs by Preferred Term in Descending Frequency**

   TEAEs, SAEs, and TEAEs leading to discontinuation of study drug will be summarized by prior and current treatment group in decreasing order of frequency of MedDRA PT in the total risankizumab doses, including AEs occurring after the first dose in Study M16-244 for subjects who received Placebo in Study 1311.5, as well as the highest dose arm from Study 1311.5. The most frequent TEAEs, SAEs and TEAEs leading to discontinuation of study drug can be identified from this summary.

6. **Serious Adverse Events (Including Deaths) and TEAEs Leading to Study Drug Discontinuation**

   All serious adverse events (SAEs), deaths, and TEAEs leading to discontinuation of study drug will be listed. The number and percentage of subjects experiencing SAEs (including deaths) and TEAEs leading to discontinuation of study drug will be tabulated by SOC and PT for each prior and current treatment group.

7. **Areas of safety interest**

   The list of Areas of safety interest will be based on the most updated version of the risankizumab Product Safety Statistical Analysis Plan, which is consistent with the most updated risankizumab Product Safety Plan.

   Grouped terms

   Grouped terms will be summarized by grouped term or SMQ, including sub-SMQs and preferred term for any adverse event, adverse events leading to discontinuation, serious adverse events, moderate or severe adverse events and related adverse events. The groupings are provided in below Table 14.

Table 14. Areas of Safety Interest Listing

ASI Grouping	Categories (ASI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Adjudicated CV Events	MACE	Adjudicated terms will be identified using CECAT and CETERM from the CE SDTM dataset.	Display underlined terms defined by the following adjudicated terms: • CV Death which includes adjudicated results of: Sudden Cardiac death, Death due to Acute MI, Death due to Heart Failure, Death due to CV Procedures, Death due to CV Hemorrhage, Death due to Other CV Causes (specify), Fatal PE, Fatal Non-Cardiac/Non-Neuro Arterial Thrombosis/Thromboembolism, Undetermined Death, Not assessable death, fatal stroke (ischemic, hemorrhagic, undetermined) • Myocardial infarction which includes adjudicated results of Type 1 Myocardial Infarction, Type 2 Myocardial Infarction, Type 3 Myocardial Infarction, Type 4 Myocardial Infarction, Type 5 Myocardial Infarction • Stroke: Ischemic stroke, Hemorrhagic stroke, Undetermined stroke	Y
	Extended MACE	Adjudicated terms will be identified (for MACE +) using CECAT and CETERM from the CE SDTM dataset.	Display underlined terms from MACE and underlined terms below: • Hospitalization for Unstable Angina • Coronary Revascularization Procedures	N

Table 14. Areas of Safety Interest Listing (Continued)

ASI Grouping	Categories (ASI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Adjudicated CV Events (continued)	Other CV events	Adjudicated terms will be identified using CECAT and CETERM from the CE SDTM dataset.	Display underlined terms defined by the following adjudicated terms: • Thrombotic events which includes adjudicated results of: Stent Thrombosis, DVT, TIA, PE, Non-fatal Non-Cardiac/Non-Neurologic Arterial Thrombosis/Thromboembolism, Other Venous Thrombosis, specified (non-fatal), Carotid revascularization • Cardiac arrhythmia which includes adjudicated results of: Supraventricular Arrhythmia, Ventricular Arrhythmia, Heart Block, Other Clinically Significant Arrhythmia (no evidence of ischemia) • Congestive heart failure which includes adjudicated results of Heart Failure – Requiring hospitalization, Heart Failure – Urgent heart failure visit • Hypertensive emergency	N
Serious infections, TB, fungal and opportunistic infections (including herpes zoster)	Serious infections	Serious PTs of the SOC Infections and Infestations	PTs	Y
	TB	Tuberculosis (including Investigations) CMQ (code 80000033)	PTs	Y
	Opportunistic infections	Opportunistic infections CMQ (code 80000073)	PTs	Y
	Fungal infections	Fungal infections CMQ (code 80000063)	PTs	N
	Herpes Zoster	Herpes zoster CMQ (code 80000175)	PTs	N

Table 14. Areas of Safety Interest Listing (Continued)

ASI Grouping	Categories (ASI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Malignancies	All possible malignancies	Narrow Malignancies (SMQ 20000090)	PTs	N
	Malignant Tumours	Narrow Malignant tumours (SMQ 20000194)	PTs	Y
	Non-melanoma skin cancer (NMSC)	Broad Skin malignant tumours (SMQ 20000204) excluding terms identified by the Melanoma CMQ (code 80000119)	PTs	N
	Malignancies excluding NMSC	'Malignancies excluding NMSC' is identified by the 'Malignant Tumours' search excluding terms identified by the 'Non-melanoma skin cancer (NMSC) search.	PTs	Y
Hypersensitivity Reaction (Hypersensitivity Serious Event only OR Anaphylactic Reaction = "Y")	Hypersensitivity	Narrow Hypersensitivity (SMQ 20000214)	PTs	Y
	Anaphylactic Reaction	Narrow Anaphylactic reaction (SMQ 20000021)	PTs	Y
Depression, Suicidal ideation and behavior (SIB)	Suicidal ideation and behavior (SIB)	Suicide/self-injury (SMQ 20000037)	PTs	N

Table 14. Areas of Safety Interest Listing (Continued)

ASI Grouping	Categories (ASI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Hepatic Events	Hepatic Events	Broad Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ 20000013) Broad Hepatitis, non-infectious (SMQ 20000010) Broad Cholestasis and jaundice of hepatic origin (SMQ 20000009) Broad Liver related investigations, signs and symptoms (SMQ 20000008) Narrow Liver-related coagulation and bleeding disturbances (SMQ 20000015)	PTs	N

8. **Adjudicated events including CCVT**

Table A4 (See Appendix) illustrates the events and sub-events that will be adjudicated along with which events will comprise the definition of a MACE event.

9. **Adverse Event per 100 Patient Years of Exposure in Study 1311.5 and Study M16-244**

The treatment-emergent adverse events occurring during Study 1311.5 and Study M16-244 will be presented by event rate per 100 patient year. These will be

presented for any TEAEs, serious adverse events, AE of special interest in addition to MACE events.

AEs per 100 patient years of exposure is defined as the number of AEs divided by the total exposure in 100 patient years. Note that one event per preferred term per day per subject will be counted in the calculation of the number of AEs (i.e., a preferred term will not be counted twice on the same day for the same subject). See the calculation method below.

$$100 \times \frac{\text{Number of TEAEs}}{\text{Total Patient Years}}$$

where total patient years is defined as the sum of the study drug exposure (defined in Section 9.0) of all subjects normalized by 365.25, and rounded to one decimal place.

10. Safety Subgroup Analysis

The following subgroup analyses will be performed to evaluate the safety in the primary endpoint across demographic data, baseline disease characteristics for PsA.

Subgroup Factor	Categories
Age	< 55 years, ≥ 55 years
Sex	Male or Female
Race	White or Others
Prior TNFi use	Naïve or experienced
Concurrent MTX use	Yes or no

11. Listing of Adverse Events

The following additional summaries of AEs will be prepared.

- Listing of Subjects with TEAE of Area of Safety Interests
- Listing of Subjects with Pretreatment Serious Adverse Evemts
- Listing of Subjects with Treatment-Emergent Serious Adverse Events
- Listing of all adverse events that led to discontinuation of study drug

- Listing of all deaths

11.3 Analysis of Laboratory Data

Changes from Baseline in Study 1311.5 in continuous laboratory parameters will be summarized by n, mean, standard deviation, minimum value, median, and maximum value for the group of subjects that received Placebo in Study 1311.5 and for the group of subjects that received Risankizumab in Study 1311.5.

Shift tables from Baseline to the final value (the last assessment during each treatment period) according to the normal range will be provided for each hematology, clinical chemistry parameter and urinalysis parameter. The laboratory data will be categorized as low, normal, or high based on the normal ranges of the laboratory used in this study. The shift tables will tabulate the number and percentage of subjects with Baseline in Study 1311.5 values below/within/above the normal range versus final values below/within/above the normal range.

Frequencies and percentages of subjects with post Baseline lab values meeting the following criteria in Table 15 and Table 16 will be summarized.

Table 15. Criteria for Potentially Clinically Important Chemistry Values

		Definition of Potentially Clinically Important Current (Version 3) NCI CTCAE Grade 3	
Chemistry Variables	**Units**	**Very Low**	**Very High**
TBL	mcmol/L		> 3.0 × ULN
ALP	U/L		> 5.0 × ULN
SGOT/AST	U/L		> 5.0 × ULN
SGPT/ALT	U/L		> 5.0 × ULN
Albumin	g/L	< 20	
Glucose	mmol/L	< 2.2	> 13.9
Triglycerides	mmol/L		> 5.7
Creatinine	mcmol/L		> 3.0 × ULN
Potassium	mmol/L	< 3.0	> 6.0
Calcium	mmol/L	< 1.75	> 3.1
CK	U/L		> 5.0 × ULN
Total Cholesterol	mmol/L		> 10.34
GGT			> 5.0 × ULN

Table 16. Criteria for Potentially Clinically Important Hematology Values

		Definition of Potentially Clinically Important Current (Version 4) CTCAE Grade 3 or greater
Hematology Variables	**Units**	**Very Low**
Hemoglobin	g/dL	< 8.0
Platelets count	10^9/L	< 50.0
WBC count	10^9/L	< 2.0
Neutrophils	10^9/L	< 1.0
Lymphocytes	10^9/L	< 0.5

Though the protocol indicates utilizing the Rheumatology Common Toxicity Criteria (RCTC) scale for grading laboratory values, given that the National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI CTCAE) scale includes a more comprehensive list of laboratory values; the lab analyses based on the NCI CTCAE scale will be presented.

For selected laboratory parameter with CTCAE a listing of all subjects with any laboratory determinations meeting CTC Version 4.0 (or later) of Grade ≥ 3 will be provided. For each of these subjects, the whole course of the parameter will be listed. For subjects with laboratory values with CTC ≥ 3 in addition to being a higher grade than the baseline CTC grade for that lab parameter, all of the laboratory parameters for those subjects will be listed.

11.3.1 Variables and Criteria Defining Abnormality

Clinical laboratory tests conducted in the study are listed in Table 17.

Table 17. Laboratory Tests

Category	Test name
Haematology	Hematocrit (Hct) Hemoglobin (Hb) White Blood Cells/Leukocytes Platelet Count/Thrombocytes
Enzymes	AST (GOT) ALT (GPT) Alkaline Phosphatase (AP) Creatine Kinase (CK) Gamma-Glutamyl Transferase (GGT/γ-GT) Lactic Dehydrogenase (LDH) Amylase Lipase
Electrolytes	Calcium Sodium Potassium Chloride Bicarbonate
Substrates	Glucose BUN (blood urea nitrogen) Creatinine Bilirubin Total Albumin Cholesterol, total Triglycerides LDL-Cholesterol HDL-Cholesterol C Reactive Protein

11.3.2 Statistical Methods

1. **Analysis of Quantitative Laboratory Parameters (Hematology, Chemistry and Urinalysis)**

   Changes from Baseline in Study 1311.5 to each scheduled visit and to the final value in continuous laboratory parameters will be summarized with mean, standard

deviation, median, minimum, Q1, Q3 and maximum. The Baseline and visit/final value means will also be presented for subjects who have both the Baseline and visit/final values (see Section 6.0 for the definition of Baseline and final values).

If there are multiple measurements on the same day, average value will be used.

Additional summaries will be presented for liver function tests including ALT or serum glutamic-pyruvic transaminase (SGPT), AST or serum glutamic-oxaloacetic transaminase (SGOT), alkaline phosphatase, and total bilirubin. Each laboratory value will be categorized as follows:

- < 1.5 × ULN
- ≥ 1.5 × ULN – < 3.0 × ULN
- ≥ 3.0 × ULN – < 5.0 × ULN
- ≥ 5.0 × ULN – < 10.0 × ULN
- ≥ 10.0 × ULN – < 20.0 × ULN
- ≥ 20.0 × ULN

where ULN is the upper normal limit.

Shift tables of Baseline in Study 1311.5 to the maximum (relative to the normal range, i.e., the largest multiple relative to the upper limit of normal) values, and from Baseline to final value will be presented using these five categories. A listing of potentially clinically significant liver function laboratory values will be provided. The listing will include all subjects who met any of the following four criteria:

- ALT ≥ 3 × ULN, or
- AST ≥ 3 × ULN, or
- Alkaline phosphatase ≥ 1.5 × ULN, or
- Total bilirubin ≥ 2 × ULN.

A listing of possible Hy's Law cases, defined as those who meet all of the following conditions will be provided.

- ALT of > 3 × ULN or AST of > 3 × ULN,
- associated with an increase in bilirubin ≥ 2 × ULN,

- Alkaline phosphatase < 2 × ULN

In addition, a graphical summary highlighting potential cases of Hy's Law within each treatment group will be presented. The maximum on-treatment values of total bilirubin and ALT will be plotted each on a scale as multiples of the upper limit of normal. The figure will show areas that meet the criteria of cholestasis (total bilirubin > 2 × ULN), Temple's corollary (ALT > 3 × ULN) and Hy's Law as the combination of these two factors.

Figure 3. An Example of eDISH Plot

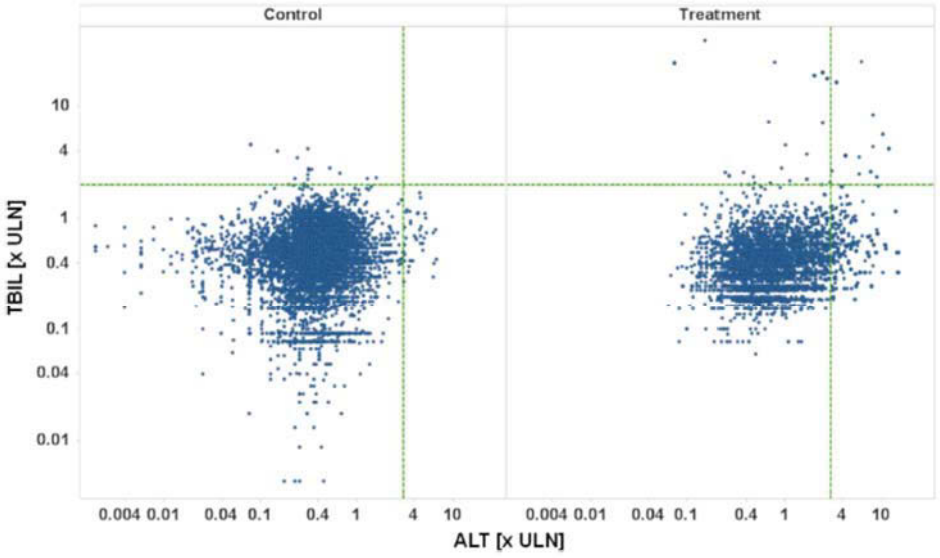


Fig. 1 eDISH plot, TBIL [× ULN] vs. ALT [× ULN] on a log/log scale, treatment by panel, pooled active versus control. *ULN* upper limit of normal, *ALT* alanine aminotransferase, *TBIL* total bilirubin

Note: Merz 2014.[7]

11.4 Analysis of Vital Signs

All analyses will be conducted in the Safety Analysis Set. The analyses of vital sign data will be descriptive.

11.4.1 Variables and Criteria Defining Abnormality

Table 18 presents the Criteria for Potentially Clinically Significant Vital Sign Findings.

Table 18. Criteria for Potentially Clinically Significant Vital Sign Findings

Vital Signs Variables	Criterion	Definition of Potentially Clinically Important
Systolic Blood Pressure (mmHg)	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 180 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic Blood Pressure (mmHg)	Low	Value ≤ 50 mmHg and decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 105 mmHg and increase ≥ 15 mmHg from Baseline
Heart Rate (bpm)	Low	Value ≤ 50 bpm and decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm and increase ≥ 15 bpm from Baseline

11.4.2 Statistical Methods

Changes from Baseline in Study 1311.5 to each visit and to the final value in vital sign parameters will be summarized with the mean, standard deviation, median, minimum, Q1, Q3 and maximum. The Baseline and final value means will also be presented for subjects who have both the Baseline and final values (see Section 6.0 for the definition of Baseline and final values).

For systolic blood pressure, diastolic blood pressure and pulse, a listing of all subjects with any vital sign value meeting criteria for potentially clinically significant values will be provided. For each of these subjects, the whole course of the respective parameter will be listed. The number and percentage of subjects who have at least one value meeting criteria for potentially clinically significant values will be provided for each selected vital sign parameter.

11.5 Analysis of ECG Parameters

ECG abnormalities will be captured as AE. They will not be captured if they are normal. Hence no ECG analyses (categorical or outlier will be performed).

11.6 Local Tolerability

Local tolerability at the administration site of the subcutaneous injection will be assessed by the investigator according to "swelling," "induration," "heat," "redness," "pain," or "other findings" at the specified visits post dosing.

A frequency table for the local tolerability will be provided at each visit by treatment group based on Safety Analysis Set.

12.0 Biomarkers Analysis

No biomarkers are to be collected so no analysis will be performed.

13.0 Pharmacokinetic Analysis

Descriptive statistics of risankizumab trough concentration measurements and risankizumab anti-drug antibody and neutralizing antibody titers by treatment group and visit will be summarized.

14.0 Summary of Changes

This is the first version of this document.

15.0 Appendix

A1. Standard Digit Reference Tables

Table A1.1. Hands (in cm)

Digit	Men	Women
Thumb	7.0	5.8
Index	6.3	5.4
Middle	6.3	5.4
Ring	5.9	5.0
Little	5.2	4.4

Table A1.2. Feet (in cm)

Digit	Men	Women
Central toe	8.2	7.2
Second	5.2	4.6
Middle	5.0	4.4
Fourth	5.0	4.4
Little	5.2	4.5

A2. Modified Nail Psoriasis Severity Index (mNAPSI)

Modified NAPSI Instructions

This tool will ask you to assess each abnormality for each of a subject's fingernails. If you question which grade to give, your answer should be the lower of the grades.
Three features or groups of features (pitting, onycholysis and oil-drop dyschromia, and crumbling) of each fingernail will be graded on a scale from 0 to 3, according to the directions below. Four features (leukonychia, splinter hemorrhages, hyperkeratosis, and red spots in the lunula) will be graded as either present or absent for each fingernail.

1. Onycholysis: Separation of the nail plate from the nail bed. The separated part of the nail is opaque and can have white, yellow, or greenish tinge. If there is a piece

of nail missing, estimate where the nail normally would have ended at the end of the nail bed, and count that missing part as involved in onycholysis.

Oil-drop (salmon patch) dyschromia: Reddish-brown discoloration under the nail plate.

Onycholysis and oil-drop dyschromia are considered together. When looking at the nail, combine the total percentage area of the nail that is affected by either and use that combined total to score the nail.

Score	Percent of Nail with Onycholysis or Oil-Drop Dyschromia Present
0	No onycholysis or oil drop dyschromia present
1	1 – 10% of the nail has onycholysis or oil-drop dyschromia
2	11 – 30% of the nail has onycholysis or oil-drop dyschromia
3	> 30% of the nail has onycholysis or oil-drop dyschromia

2. Pitting: Small, sharply defined depressions in the nail surface. Pits are discrete abnormalities ("ice-pick-like"). If there is nail plate crumbling that is confluent with pits, do not score for pits. If the pits are separate from crumbling, they may be scored regardless of whether crumbling is present or not.

Score	Number of Pits
0	0
1	1 – 10
2	11 – 49
3	> 50

3. Nail plate crumbling: Crumbling or fragmentation of friable nail plate which may be associated with confluent pitting. Crumbling involves alteration of the nail plate surface. Horizontal ridging of the nail, "wave-like" appearance, and horizontal lines are all features of crumbling.

Score	Percent of Nail with Crumbling Present
0	No crumbling
1	1 – 25% of the nail has crumbling
2	26 – 50% of the nail has crumbling
3	> 50% of the nail has crumbling

The next 4 abnormalities are scored only by their presence or absence. A score of 1 indicates present and a score of zero indicates not present.

9. Leukonychia: White spots in the nail plate due to psoriasis in the mid matrix. Leukonychia are just color changes. If it appears that there is depression or irregularity to the nail surface, this may be pitting or crumbling, not leukonychia. If the leukonychia is adjacent to, or confluent with crumbling or pits, it is counted as part of the crumbling or pitting and not as a separate abnormality.
10. Splinter hemorrhages: Small, longitudinal, linear, dark brown hemorrhage under the fingernail.
11. Nail bed hyperkeratosis: Thickened keratin in the nail bed.
12. Red spots in the lunula: Small pink or red macules in the lunula.

Table A3. Adjudicated Events and Composition of MACE Event Definition

Decision Form (CECAT from SDTM.CE – Select Records Using this Variable)	Event (for NON-FATAL Events, CETERM from SDTM.CE – to Be Produced in Table; for FATAL EVENTS, CETERM = 'DEATH' and Reason for Death in DDORRES (Note that DDORRES = Null Indicates 'NOT ASSESSABLE')) Take AESTDT from CEREFID < AELLTCD: AETERM: AESTDTC >	Charter Event	MACE	MACE+	Other CV
NON-FATAL CARDIOVASCULAR DECISION	MI	MI Type 1	X	X	
		MI Type 2	X	X	
		MI Type 4	X	X	
		MI Type 5	X	X	
NON-FATAL NEUROLOGICAL DECISION	Stroke	Ischemic Stroke	X	X	
		Hemorrhagic Stroke	X	X	
		Undetermined Stroke	X	X	
FATAL CARDIOVASCULAR DECISION	Fatal CV	Sudden Cardiac Death	X	X	
		Due to Acute MI	X	X	
		Death due to Heart Failure	X	X	
		Death due to CV Procedures	X	X	
		Death due to CV Hemorrhage	X	X	
		Death due to Other CV Causes	X	X	
FATAL NEUROLOGICAL DECISION	Fatal Stroke	Ischemic Stroke	X	X	
		Hemorrhagic Stroke	X	X	
		Undetermined Stroke	X	X	

Table A4. Adjudicated Events and Composition of MACE Event Definition (Continued)

Decision Form (CECAT from SDTM.CE – Select Records Using this Variable)	Event (for NON-FATAL Events, CETERM from SDTM.CE – to Be Produced in Table; for FATAL EVENTS, CETERM = 'DEATH' and Reason for Death in DDORRES (Note that DDORRES = Null Indicates 'NOT ASSESSABLE')) Take AESTDT from CEREFID < AELLTCD: AETERM: AESTDTC >	Charter Event	MACE	MACE+	Other CV
FATAL THROMBOTIC DECISION	Fatal PE	Fatal PE	X	X	
FATAL THROMBOTIC DECISION	Fatal Non-Cardiac/Non-Neuro Arterial Thrombosis/Thromboembolism	Fatal Non-Cardiac/Non-Neuro Arterial Thrombosis/Thromboembolism	X	X	
FATAL CARDIOVASCULAR*	Non-CV Death	Non-cardiac Death			
FATAL CARDIOVASCULAR	Undetermined Death	Undetermined Death	X	X	
FATAL CARDIOVASCULAR/ NEUROLOGICAL/ THROMBOTIC DECISION	Not assessable deaths (cardiac/neuro/thrombotic)	Not assessable Death	X	X	
NON-FATAL CARDIOVASCULAR DECISION	Hospitalization for Unstable Angina	Hospitalization for Unstable Angina		X	
NON-FATAL CARDIOVASCULAR DECISION	Coronary Revascularization Procedures	PCI		X	
		CABG		X	
NON-FATAL CARDIOVASCULAR DECISION	Hypertensive emergency	Hypertensive Emergency			X
NON-FATAL NEUROLOGICAL DECISION	TIA	TIA			X

Table A4. Adjudicated Events and Composition of MACE Event Definition (Continued)

Decision Form (CECAT from SDTM.CE – Select Records Using this Variable)	Event (for NON-FATAL Events, CETERM from SDTM.CE – to Be Produced in Table; for FATAL EVENTS, CETERM = 'DEATH' and Reason for Death in DDORRES (Note that DDORRES = Null Indicates 'NOT ASSESSABLE')) Take AESTDT from CEREFID < AELLTCD: AETERM: AESTDTC >	Charter Event	MACE	MACE+	Other CV
NON-FATAL THROMBOTIC DECISION	Deep Vein Thrombosis	DVT			X
NON-FATAL THROMBOTIC DECISION	Pulmonary Embolism	PE			X
NON-FATAL THROMBOTIC DECISION	Non-fatal Non-Cardiac/Non-Neurological Arterial Thrombosis/Thromboembolism	Non-fatal Non-Cardiac/Non-Neurological Arterial Thrombosis/Thromboembolism			X
NON-FATAL THROMBOTIC DECISION	Other Venous Thrombosis, specified (non-fatal)	Other Venous Thrombosis, specified (non-fatal)			X
NON-FATAL CARDIOVASCULAR	Heart Failure	Heart Failure – Requiring hospitalization			X
NON-FATAL CARDIOVASCULAR DECISION		Heart Failure – Urgent heart failure visit			X
NON-FATAL CARDIOVASCULAR DECISION	Clinically Significant Arrhythmia (no evidence of ischemia)	Supraventricular Arrhythmia			X
		Ventricular Arrhythmia			X
		Heart Block			X
		Other			X

16.0 References

1. Daltroy LH, Larson MG, Roberts NW, et al. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol. 1990;17(7):946-50.
2. Ware JE, Jr, Kosinski M, Keller SD. SF 36 physical and mental health summary scales: a user's manual. MA: The Health Institute, New England Medical Center. 1994.
3. Ware JE. Measuring patients' views: the optimum outcome measure. SF 36: valid, reliable assessment of health from the patient's point of view. BMJ. 1993;306:1429-30.
4. Strand V, Crawford B, Singh J, et al. Use of "spydergrams" to present and interpret SF-36 health-related quality of life data across rheumatic diseases. Ann Rheum Dis. 2009;68(12):1800-04. doi:10.1136/ard.2009.115550.
5. Maksymowych WP, Mallon C, Morrow S, et al. Development and validation of the Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index. Ann Rheum Dis. 2009;68(6):948-53.
6. Minnock P, Kirwan J, Veale D, et al. Fatigue is an independent outcome measure and is sensitive to change in patients with psoriatic arthritis. Clin Exp Rheumatol. 2010;28(3):401-4.
7. Merz M, Lee KR, Kullak-Ublick GA, et al. Methodology to assess clinical liver safety data. Drug Saf. 2014;37 (Suppl 1) S33-45. doi: 10.1007/s40264-014-0184-5.

Document Approval

Study M16244 - Statistical Analysis Plan Version 1 - 16Jul2018 (E3 16.1.9)

Version: 1.0 **Date:** 17-Jul-2018 11:00:00 AM **Company ID:** 07172018-00F9F683DE8CDF-00001-en

Signed by:	Date:	Meaning Of Signature:
[redacted]	16-Jul-2018 10:49:41 PM	Approver
	17-Jul-2018 10:59:59 AM	Approver