1.0 Title Page

Statistical Analysis Plan Revisions

Boehringer Ingelheim Study 1311.5

(AbbVie Study M16-002)

A Randomized, Double-Blind, Placebo-Controlled Proof-of-Concept Dose-Ranging Study of Risankizumab in Subjects with Active Psoriatic Arthritis

Date: 16 October 2017

Version 1.0

3.0 Introduction

This document is an explanation of differences between the analysis performed after final database lock and those detailed in the SAP.

The changes include:

- **Section 6.4:** Updates to the visit window for mTSS.
  Rationale: Updated window more accurately encompasses observations appropriate to label as baseline.
- **Section 6.5:** mTSS and PsAMRIS removed from 6.5 Missing Data Handling.
  Rationale: No subject level imputation will be performed for continuous imaging data.
- **Section 7.1:** Removed statistical testing between placebo arm and the mixed two arms at baseline.
  Rationale: *No statistical testing for baseline characteristics.*
- **Section 9.0:** Planned number of visits when injection planned has been capped at 5.
  Rationale: *No subject should have more than 5 planned visits.*
- **Section 10.1:** Table 12 has been updated.
  Rationale: Some efficacy endpoints have been added and some analyses have changed.
- **Section 10.6** List of further efficacy endpoints has been updated.
  Rationale: Some efficacy endpoints have been added and some analyses have changed
- **Section 10.9.10:** The addition of a PASI LOCF sensitivity analysis.
  Rationale: Sensitivity analysis added due to the presence of PASI missing data.
- **Section 10.9.12:** Updates to what defines a dactylic digit.
  Rationale: Added a missing component of the definition of dactylitis.
- **Section 10.9.20:** Updates to the scoring and missing data rules for mTSS.
  Rationale: *Two separate methods of analysis will be performed for mTSS.*

- **Section 10.9.21:** Updated PsAMRIS analysis to be descriptive statistics only.
  Rationale: There are far too few observations with PsAMRIS data to provide a more detailed analysis. Additionally hand and foot scores are not comparable.
- **Section 11.2:** Changed "Adverse Events by "Reasonably Possibly Related" Relationship" to "Adverse Events by Maximum Relationship."
  Rationale: *Updates to the PSSAP.*
- **Section 11.2:** Addition of unknown category to Adverse Events by Maximum Relationship.
  Rationale: *Updates to the PSSAP.*
- **Section 11.4:** Change of wording in Criteria for Potentially Clinically Significant Vital Sign Findings.
  Rationale: *Updates to the PSSAP.*
- **Various:** Several small editorial changes.
  Rationale: Various small updates to wording were made for clarity.

6.0 Analysis Conventions

6.4 Definition of Analysis Windows

Table 6 previously read:

Table 1. Visit Windows for Analysis of mTSS

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	[–999, 13]
Week 24	VISIT 9	169	[14, 306]

Table 6 now reads:

Table 1. Visit Windows for Analysis of mTSS

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	[–999, 30]
Week 24	VISIT 9	169	[31, 306]

6.5 Missing Data Handling

This section no longer includes the subject level imputation method below:

Missing Data Handling for mTSS and PsAMRIS

The following imputation method on subject level will be used as the primary approach to impute missing mTSS and PsAMRIS:

- Calculate average change from Baseline to follow-up visit (for example, Week 24 for mTSS and Week 16 for PsAMRIS) computed for all observed non-missing bones/joints data at both time points from a given subject.
- If the Baseline score is available for a bone/joint, then the impute value at the follow-up visit will be the Baseline score plus the average change score from the other scored bone/joint locations.
- If the Baseline visit is unavailable but the follow-up visit score is available, the imputed value at the Baseline will be computed from the follow-up time point score minus the average change, if the average change is smaller than the follow-up time point score at that location. Otherwise, the Baseline visit imputed value will use the most recent follow-up time point score carried backward.
- If both Baseline and follow-up time point visit score values are missing for a bone/joint, no values will be imputed to either visit, and that joint's bone/joint contribution will be zero, which means assigning 0 to both baseline and the follow-up time point for that joint.

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

7.1 Demographic and Baseline Characteristics

The first paragraph in Section 7.1 previously read:

Demographics and Baseline characteristics will be summarized for each arm and for overall of the FAS populations. Continuous variables will be summarized with the number of non-missing observations by mean, standard deviation, first quartile, median, third quartile, minimum and maximum values. Categorical data will be summarized using frequencies and percentages. Statistical tests will be performed to assess the comparability of the placebo arm and the mixed two arms at baseline. Treatment comparison will be made based on non-missing information. Continuous variables will be analyzed using one-way analysis of variance (ANOVA). Categorical variables will be analyzed using Fisher's exact test.

The first paragraph in Section 7.1 now reads:

Demographics and Baseline characteristics will be summarized for each arm and for overall of the FAS populations. Continuous variables will be summarized with the number of non-missing observations by mean, standard deviation, first quartile, median, third quartile, minimum and maximum values. Categorical data will be summarized using frequencies and percentages.

PsA Medical History and Disease Characteristics at Baseline No Longer Includes:

- Categorical:
  - PASI (Psoriasis Area and Severity Index) assessed in subjects with ? 3% BSA of psoriatic plaques
  - ContinousAnti-cyclic citrullinated peptide (Anti-CCP) (units)

PsA Medical History and Disease Characteristics at Baseline Now Includes:

- Continuous:

- PASI (Psoriasis Area and Severity Index) assessed in subjects with ≥ 3% BSA at baseline
- PsA Disease duration based on diagnosis date in years

9.0 Study Drug Duration and Compliance

Compliance now includes:

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

10.0 Efficacy Analysis

Summary of Efficacy Variables and Corresponding Analyses previously read:

Table 2. Summary of Efficacy Variables and Corresponding Analyses

Efficacy Variables	Analysis Method
Binary Efficacy Endpoints:	
• **Primary and Secondary Endpoints:** o ACR20/50/70 responses at Week 16 o PASI90 response at Week 16 in subjects with PsO BSA ≥ 3% at baseline	• Point estimate and 90% CI of the response rate for each treatment group using exact method • Point estimate, 90% CI and p-value of the response rate difference between combined Arm 1 and Arm 2 vs the placebo, as well as the response rate difference between combined Arm 2 and Arm 3 using the stratified Cochran-Mantel-Haenszel test. • Pairwise comparisons of the each risankizumab dose group vs placebo will be conducted using the same stratified Cochran-Mantel-Haenszel methods. • Histogram of response rates by treatment group • MCP-Mod to evaluate dose-response for ACR 20/50/70 • Forest plot • No multiplicity adjustments. • Imputation: NRI (primary), OC
Further Binary Endpoints: At all measured time points except Week 16: • ACR20/50/70 At all measured time points: • PASI75 in subjects with a ≥ 3% baseline PsO BSA • PASI100 in subjects with a ≥ 3% baseline PsO BSA • sPGA Response (0 = clear and 1 = almost clear) in subjects with a ≥ 3% baseline PsO BSA • EULAR response • Presence of Dactylitis • MDA • PsARC	• Point estimate and 90% CI of the response rate for each treatment group using exact method • Point estimate, 90% CI and p-value of the response rate difference between combined Arm 1 and Arm 2 vs the placebo, as well as the response rate difference between combined Arm 2 and Arm 3 using the stratified Cochran-Mantel-Haenszel risk difference estimate. • Pairwise comparisons of the risankizumab dose groups vs placebo will be conducted using the same stratified Cochran-Mantel-Haenszel methods. • Plot of ACR20/50/70 response rates overtime by treatment group • Ordinal Logistic Regression of EULAR response • Forest plot • Imputation: NRI (primary), OC

Table 2. Summary of Efficacy Variables and Corresponding Analyses (Continued)

Efficacy Variables	Analysis Method
Continuous Efficacy Endpoints:	
Secondary Endpoints: Change from baseline at Week 16: • TJC68 • SJC66 • HAQ-DI • SF-36 • Dactylitis Count in subjects with dactylitis at baseline • SPARCC in subjects with enthesitis at baseline • mNAPSI assessed in subjects with nail psoriasis at baseline **Further Continuous Endpoints:** Change from baseline at all measured time points: • Physician's Global (VAS) • Patient's Pain (VAS) • Patient's Global Activity (VAS) • CRP • DAS28(CRP) • LDI in subjects with dactylitis at baseline • LEI in subjects with Enthesitis at baseline • Presence of enthesitis • FACIT-F Change from baseline at Week 16 and 24: • BASDAI in subjects with baseline inflammatory spondylitis	• Point estimate and 90% CI within each treatment group using mixed model repeated measures (MMRM) model; • Point estimate, 90% CI and p-value for the mean change difference between combined Arm 1 and Arm 2 vs placebo, as well as the mean change difference between combined Arm 2 and Arm 3 with mixed model repeated measures (MMRM) model • Pairwise comparisons of the each risankizumab dose group vs placebo will be conducted MMRM model • Box plots and change from baseline over time will be plotted for key continuous variables • For SF-36, spyderdiagram will be provided • Imputation: OC (primary)

Summary of Efficacy Variables and Corresponding Analyses currently reads:

Table 2. Summary of Efficacy Variables and Corresponding Analyses (Continued)

Efficacy Variables	Analysis Method
Continuous Efficacy Endpoints (continued):	
mTSS and PsAMRIS: • Change from baseline (CFB) at Week 24 in mTSS • Change from baseline (CFB) at Week 16 in PsAMRIS	• Mean CFB and 90% CI within each treatment group; • Mean CFB , 90% CI and p-value for the difference between combined Arm 1 and Arm 2 and the placebo, as well as the difference between combined Arm 2 and Arm 3 using ANCOVA model • Pairwise comparisons of the each risankizumab dose group vs placebo will be conducted with ANCOVA model

Table 3. Summary of Efficacy Variables and Corresponding Analyses now reads:

Efficacy Variables	Analysis Method
Binary Efficacy Endpoints:	
• **Primary and Secondary Endpoints:** o ACR20/50/70 responses at Week 16 o PASI90 response at Week 16 in subjects with PsO BSA ≥ 3% at baseline	• Point estimate and 90% CI of the response rate for each treatment group using exact method • Point estimate, 90% CI and p-value of the response rate difference between combined Arm 1 and Arm 2 vs placebo, as well as the response rate difference between combined Arm 2 and Arm 3 vs placebo using the stratified Cochran-Mantel-Haenszel test. • Pairwise comparisons of the each risankizumab dose group vs placebo using the same stratified Cochran-Mantel-Haenszel test. • Bar plot of response rates by treatment group • MCP-Mod to evaluate dose-response for ACR 20/50/70 • Forest plot • No multiplicity adjustments. • Imputation: NRI (primary), OC, LOCF additionally for PASI90
Further Binary Endpoints: At all measured time points except Week 16: • ACR20/50/70 At all measured time points: • PASI75 in subjects with a ≥ 3% baseline PsO BSA • PASI100 in subjects with a ≥ 3% baseline PsO BSA • sPGA Response (0 = clear and 1 = almost clear) in subjects with a ≥ 3% baseline PsO BSA • EULAR response (Multinomial) • EULAR response (Binary) • Presence of Dactylitis • MDA • PsARC • No Progression of mTSS	• Point estimate and 90% CI of the response rate for each treatment group using exact method • Point estimate, 90% CI and p-value of the response rate difference between combined Arm 1 and Arm 2 vs the placebo, as well as the response rate difference between combined Arm 2 and Arm 3 vs placebo using the stratified Cochran-Mantel-Haenszel risk difference estimate. • Pairwise comparisons of the risankizumab dose groups vs placebo will be conducted using the same stratified Cochran-Mantel-Haenszel methods. • Plot of ACR20/50/70 response rates overtime by treatment group • Ordinal Logistic Regression of EULAR multinomial response • Forest plot • Imputation: NRI (primary), OC, LOCF additionally for PASI75/100

Table 3. Summary of Efficacy Variables and Corresponding Analyses (Continued)

Efficacy Variables	Analysis Method
Continuous Efficacy Endpoints:	
Secondary Endpoints: Change from baseline at Week 16: • TJC68 • SJC66 • HAQ-DI • SF-36 • Dactylitis Count in subjects with dactylitis at baseline • SPARCC in subjects with enthesitis at baseline • mNAPSI assessed in subjects with nail psoriasis at baseline **Further Continuous Endpoints:** Change from baseline at all measured time points: • Physician's Global (VAS) • Patient's Pain (VAS) • Patient's Global Activity (VAS) • CRP • DAS28(CRP) • LDI in subjects with dactylitis at baseline • LEI in subjects with Enthesitis at baseline • Presence of enthesitis • FACIT-F Change from baseline at Week 16 and 24: • BASDAI in subjects with baseline inflammatory spondylitis	• Point estimate and 90% CI within each treatment group using mixed model repeated measures (MMRM) model; • Point estimate, 90% CI and p-value for the mean change difference between combined Arm 1 and Arm 2 vs placebo, as well as the mean change difference between combined Arm 2 and Arm 3 vs placebo with mixed model repeated measures (MMRM) model • Pairwise comparisons of the each risankizumab dose group vs placebo will be conducted MMRM model • Change from baseline over time will be plotted for key continuous variables • For SF-36, spyderdiagram will be provided • Imputation: OC (primary)

Table 3. Summary of Efficacy Variables and Corresponding Analyses (Continued)

Efficacy Variables	Analysis Method
Continuous Efficacy Endpoints (continued):	
mTSS: • Change from baseline at Week 24 in mTSS	• Point estimate and 90% CI within each treatment group; • Point estimate, 90% CI and p-value for the difference between combined Arm 1 and Arm 2 and the placebo, as well as the difference between combined Arm 2 and Arm 3 using ANCOVA model • Pairwise comparisons of the each risankizumab dose group vs placebo using an ANCOVA model
PsAMRIS: • Change from baseline at Week 16 in PsAMRIS	• Descriptive statistics at Week 16 • Imputation: OC

10.7 Further Efficacy Analysis

The last paragraph in Section 10.7 previously read:

Analyses at all measured time points will be conducted in a similar fashion as for Week 16 endpoints except for EULAR response, mTSS, and PSAMRIS. EULAR response will be analysed using ordinal logistic regression with treatment group, prior TNFi use, and concurrent MTX use as main factors and baseline DAS value as a covariate. mTSS and PSAMRIS will be analysed using ANCOVA with treatment group, prior TNFi use, and concurrent MTX use as main factors and baseline value as a covariate. 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.

The last paragraph in Section 10.7 now reads:

Analyses at all measured time points will be conducted in a similar fashion as for Week 16 endpoints except for mTSS and PSAMRIS. Multinomial EULAR response will additionally be analysed using ordinal logistic regression with treatment group, prior

TNFi use, and concurrent MTX use as main factors and baseline DAS value as a covariate. mTSS will be analysed using ANCOVA with treatment group, prior TNFi use, and concurrent MTX use as main factors and baseline value as a covariate. PsAMRIS will be described descriptively separately for hand and foot images as well as a descriptive analysis of the PsAMRIS feature scores, also separate for hand and foot images. Further efficacy endpoints will be summarized descriptively. Data will be displayed by graphics as appropriate.

10.9 Efficacy Variables Definition and Conventions

10.9.7 European League Against Rheumatism (EULAR) Response Criteria

This section now additionally includes:

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

For missing data use Non-Responder Imputation using the score of "None" as non-response. For the binary response Good vs (Moderate & None) Non-Responder Imputation use Moderate & None as non-response.

10.9.10 Psoriasis Area and Severity Index (PASI) and Body Surface Area (BSA)

This section now includes:

An additional LOCF analysis will be performed, using last observation carried forward imputation for visits post-baseline with PASI70/90/100 response then being based on the imputed PASI score.

10.9.12 Leeds Dactylitis Index (LDI)

This section previously read:

The LDI basic 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).

This section now reads:

The LDI basic 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.

Dactylitis Count previously read:

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

Dactylitis Count now reads:

The dactylitis count is the number of fingers and toes with dactylitis, with a range of 0 – 20. If a site is not assessed assign it a value of "0."

10.9.20 Modified Total Sharp Score (mTSS)

This section now reads:

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.

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

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

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 total 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:

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

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

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

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

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.

Method 1:

- 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

Method 2:

Missing joints for baseline will be imputed as 0 (denote no erosion or narrowing).

Missing joint score imputation will be performed for Erosion and JSN respectively. If a score at any location/joint is missing for a post-baseline visit, and if there are >= 27 joints (i.e., more than 50%) in Erosion (or JSN) with **non-missing** change from baseline scores, the method described below will be used to impute the missing joint score for that visit. Otherwise, the erosion/JSN score will be missing so TSS is missing.

- Step 1: Calculate average change from Baseline to follow-up visit (for example, Week 24) computed for all observed non-missing change from baseline scores for a given subject.
- Step 2: The imputed value at the follow-up visit will be the Baseline score plus the average change score from step 1. If the imputed score exceeds the upper range of that bone/joint location, take the upper range to be the imputed value.

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.

10.9.21 Psoriatic Arthritis Magnetic Resonance Image Scoring System (PsAMRIS)

Psoriatic Arthritis Magnetic Resonance Image Scoring System (PsAMRIS) previously read:

The PsAMRIS will be evaluated at Baseline and Week 16. Change from baseline in PsAMRIS score at Week 16 will be summarized. PSAMRIS will have six separate scores

for Synovitis (0 – 72), Tenosynovitis (0 – 72), Periaticular Inflammation (0 – 48), Osteitis (0 – 144), Bone Erosion (0 – 480), and Bone Proliferation (0 – 24).

Psoriatic Arthritis Magnetic Resonance Image Scoring System (PsAMRIS) now reads:

The PsAMRIS will be evaluated at Baseline and Week 16. PSAMRIS will have six separate scores. Four scores for Inflammation: Synovitis, Tenosynovitis, Periaticular Inflammation, Osteitis. Two scores for Damage, Bone Erosion, and Bone Proliferation. Descriptive statistics will be used to summarize Inflammation, Damage, and Total Scores for each visit separately for hands and feet.

11.0 Safety Analysis

11.2 Analysis of Adverse Events

Table 20 for DILI previously read:

ASI Grouping	Categories (ASI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Drug induced liver injury (DILI)	Drug induced liver injury (DILI)	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

Table 20 now reads:

ASI Grouping	Categories (ASI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Possible Drug induced liver injury (DILI)	Possible Drug induced liver injury (DILI)	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

4. Adverse Events by "Reasonably Possibly Related" Relationship has changed to read:

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.

11.3 Analysis of Laboratory Data

Table 21 used to read:

Table 21. 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 21 now reads:

Table 21. Criteria for Potentially Clinically Important Chemistry Values

		Definition of Potentially Clinically Important Current (Version 3) NCI CTCAE Grade 3 or greater	
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 22 used to read:

Table 22. Criteria for Potentially Clinically Important Hematology Values

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

Table 22 now reads:

Table 22. 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

11.4 Analysis of Vital Signs and Weight

11.4.1 Variables and Criteria Defining Abnormality

Table 24 used to read:

Table 24. Criteria for Potentially Clinically Significant Vital Sign Findings

Vital Signs	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg or decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 180 mmHg or increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg or decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 105 mmHg or increase ≥ 15 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm or decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm or increase ≥ 15 bpm from Baseline

Table 24 now reads:

Table 24. 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

Document Approval

Study 13115 - Statistical Analysis Plan Version 1 Revisions - 16Oct2017 (E3 16.1.9)

Version: 1.0 **Date:** 17-Oct-2017 03:14:35 PM **Company ID:** 10172017-00F9F683AED334-00001-en

Signed by:	Date:	Meaning Of Signature:
[redacted]	16-Oct-2017 11:09:39 PM	Approver
	17-Oct-2017 03:14:35 PM	Approver

1.0 Title Page

Statistical Analysis Plan

Boehringer Ingelheim Study 1311.5

(AbbVie Study M16-002)

A Randomized, Double-Blind, Placebo-Controlled Proof-of-Concept Dose-Ranging Study of Risankizumab in Subjects with Active Psoriatic Arthritis

Date: 12 May 2017

Version 1.0

2.0 Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACR	American College of Rheumatology
AE	Adverse Event
ASI	Areas of Special Interest
ALT	Alanine Aminotransferase
Anti-CCP	Anti-Cyclic-Citrullinated Peptide
AST	Aspartate Aminotransferase
BASDAI	Bath AS Disease Activity Index
BI	Boehringer Ingelheim
BSA	Body Surface Area
CASPAR	Classification criteria for Psoriatic Arthritis
CRP	C-Reactive Protein
DAS28	Disease Activity Score in 28 joints
DMARDs	Disease-Modifying Antirheumatic Drugs
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EOS	End Of Study
EOT	End Of Treatment
EULAR	European League Against Rheumatism
FACIT-F	Functional Assessment of Chronic Illness Therapy-Fatigue
FAS	Full Analysis Set
FU	Follow-Up
HAQ-DI	Health Assessment Questionnaire-Disability Index
HIV	Human Immunodeficiency Virus
IRT	Interactive Response System
i.v.	intravenous
LDI	Leeds Dactylitis Index
LEI	Leeds Enthesitis Index
MACE	Major Adverse Cardiovascular Event
MAR	Missing At Random
MDA	Minimal Disease Activity
MedDRA	Medical Dictionary for Drug Regulatory Activities

MMRM	Mixed Model Repeated Measures Model
mNAPSI	Modified Nail Psoriasis Severity Index
MRI	Magnetic Resonance Imaging
mTSS	Modified Total Sharp Score
MTX	Methotrexate
NRI	Non-Responder Imputation
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
OLE	Open-Label Extension
OMERACT	Outcome Measures in Rheumatology
PASI	Psoriasis Area and Severity Index
PD	Pharmacodynamic
PK	Pharmacokinetic
PsA	Psoriatic Arthritis
PsAMRIS	Psoriatic Arthritis Magnetic Resonance Image Scoring System
PsARC	Psoriatic Arthritis Response Criteria
PsO	Psoriasis
RCTC	Rheumatology Common Toxicity Criteria
REML	Residual Maximum Likelihood Method
REP	Residual Effect Period
SAE	Serious Adverse Event
SF-36	Short Form-36 Health Survey
SJC	Swollen Joint Count(s)
SPARCC	Spondyloarthritis Research Consortium of Canada
sPGA	Static Physician Global Assessment
PtGA	Patient's Global Assessment of Disease Activity for Arthritis
PhGA	Physician's Global Assessment of Disease Activity for Arthritis
TJC	Tender Joint Count(s)
TNF	Tumor necrosis factor
TNFi	Tumor necrosis factor inhibitor(s)
VAS	Visual Analog Scale
W	Week
WBC	White Blood Count

3.0 Introduction

This statistical analysis plan (SAP) describes the statistical analysis to be completed by the AbbVie Clinical Statistics Department for Boehringer Ingelheim (BI) Study 1311.5 (AbbVie Study M16-002) dated 12 October 2016. This study was sponsored by BI and data was collected using their internal SOPs.

This SAP will provide details to further elaborate statistical methods as outlined in the BI Protocol 1311.5 and will describe analysis conventions to guide the statistical programming work. This SAP will be signed off before the study database is locked.

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 Objectives

The objectives of this study are to provide proof-of-concept and dose-ranging data of risankizumab in subjects with active psoriatic arthritis (PsA) to support advancement and dose selection for the pivotal program in this indication. The proof-of-concept will be achieved through the primary endpoint comparison (ACR 20 rates at Week 16) from the pooling of the two highest exposure arms versus placebo. In addition, clinical efficacy (based on secondary and further endpoints) will be evaluated. Onset of response and assessment of duration of response will be assessed during treatment and through Week 32 follow-up, or until last visit for subjects enrolling into the open-label extension (OLE) Study M16-244.

The safety of risankizumab will be evaluated during treatment and through the Week 32 follow-up, or through last visit for subjects enrolling into the OLE study. The effect of emergence of anti-drug antibodies (ADA) on safety and efficacy will be explored.

In addition, risankizumab pharmacokinetic (PK) exposure will be assessed to provide data for subsequent pharmacokinetic- pharmacodynamic (PK-PD) modelling. The

risankizumab exposure-response profile will be characterized through a relatively wide exposure range achieved by 150 mg every 4 weeks to 75 mg single dosing.

Prevention of structural damage will be explored through X-ray endpoints at Week 24 and magnetic resonance imaging (MRI) endpoints at Week 16. Influence of risankizumab on pathway gene and protein expression levels as well as disease specific protein markers will be explored.

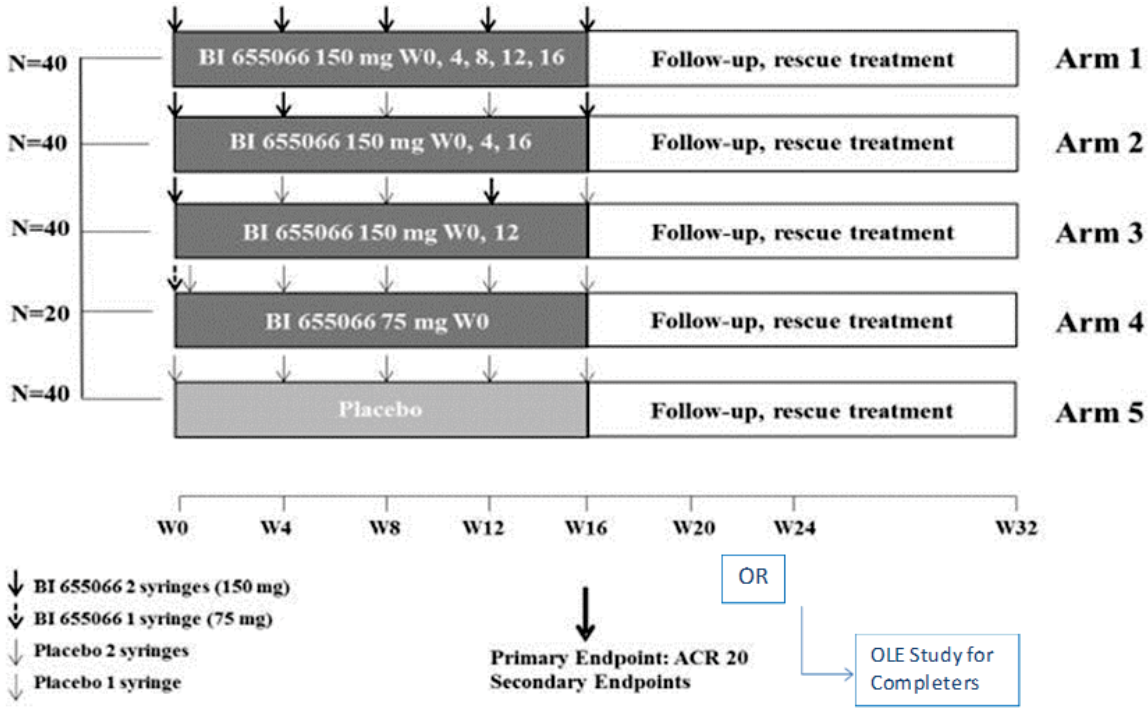
4.2 Overall Trial Design and Plan

This is a multi-national, randomized, parallel-design, dose-ranging, multiple-dose, placebo-controlled, double-blinded Phase 2 study. Approximately 180 eligible subjects with active PsA were randomized at 2:2:2:1:2 ratio, stratified based on prior tumor necrosis factor inhibitor(s) (TNFi) use and concurrent methotrexate (MTX) use into five treatment arms shown in Figure 1. Subjects with prior TNFi experience were to be capped at approximately 70%.

The overall trial treatment duration is 16 weeks with an additional 16 weeks follow-up period, or subjects who have completed all doses and Week 24 visit will potentially have the option to enroll into a separate OLE known as Study M16-244. Subjects electing to participate in the OLE study will not complete any remaining visit in this study (BI Study 1311.5) after signing the OLE ICF and will solely follow the procedures in the OLE study protocol.

Subjects in Study 1311.5 will receive risankizumab/placebo as described in Figure 1. The overall dosing schedule will be the same in all treatment arms in order to keep the blind; this is further described in Section 4.1.4 in the protocol.

Figure 1. Trial Design



If subjects have a reduction in both tender and swollen joint count of < 20% relative to baseline at the Week 16 assessment, they were permitted to alter their concomitant PsA treatment or start an additional treatment, except biologics, according to investigator's judgment and local standard of care, after receiving Week 16 study drug dose. After Week 24, subjects not enrolling in the OLE study are allowed to start rescue treatment with local standard of care, including biologic treatment, if considered appropriate by the investigator, regardless of the joint count improvement. For details, please refer to protocol Section 4.2.1.

In addition to clinical endpoint assessments, subjects will have X-rays performed at Weeks 0 and 24 for assessment of the modified total Sharp score (mTSS). Approximately 90 subjects from selected sites will be included in the MRI sub-study, with MRI assessments performed at Weeks 0 and 16.

The analysis of the primary endpoint (ACR 20) will be performed after the last subject completes the Week 16 visit. The trial will be unblinded to the Sponsor, for analysis, but subjects and investigators will remain blinded until after the completion of the trial.

Individual subject participation is concluded when the subject has completed the last planned visit (Week 32 or early End of Study visit if a patient prematurely discontinues or the final visit prior to enrolling in the OLE study). The "last-subject-last-visit-primary-endpoint" is the last scheduled primary endpoint visit at Week 16 completed by the last subject. The end of the trial is defined as "last subject out," i.e., last scheduled visit completed by last subject.

4.3 Sample Size

The sample size was determined on the basis of a one-sided comparison between the average rate of ACR 20 response at Week 16 of Arm 1 and Arm 2 versus placebo with the assumed Week 16 ACR 20 response rate of 38% in the combined arms (Arm 1 and Arm 2) and of 15% in the placebo arm, 40 participants each for Arm 1, Arm 2 and placebo will provide 85% power to detect a 23% difference in proportion (combination of Arm 1 and Arm 2 versus placebo) using a one-sided test of 0.05 significance (equivalent to two sided test of 0.1 significance). Table 1 shows the estimated power based on a range of possible placebo rates. Power analysis was conducted using the Chi-Square Test; the software used was ADDPLAN version 6.0.4.

Although not included in the hypothesis testing strategy, Arm 3 will have about 40 participants. In addition, we plan to enroll 20 participants into Arm 4 dose response modeling. The total sample size for this study is therefore 180 participants. Figure 1 provides details on the study arms. This study is not powered to detect statistically significant differences between the different risankizumab treatment arms.

Table 1. Power and Sample Size for the Superiority Test of ACR 20 at Week 16

	Risankizumab	Placebo	Difference	1-Sided Type I Error	Total Sample Size of Arm 1 + 2 and Placebo	Drop-Out Rate	Power
ACR 20 Rate at Week 16	38%	15%	23%	0.05	120	0%	85%
	40%	17%					84%
	42%	19%					83%
	44%	21%					82%
	46%	23%					81%
	48%	25%					80%
	38%	17%	21%				78%
	38%	19%	19%				69%

4.4 Interim Analysis

No interim efficacy analysis is planned for this study.

The primary efficacy analysis will be performed once all subjects completed Week 16 visits based on the interim locked database. Unblinded group level results will be shared with the AbbVie internal study team. No type I error adjustment will be done.

4.5 Data Monitoring Committee (DMC)

An independent DMC operating under a charter is responsible for reviewing safety data periodically to ensure subjects 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 Populations

5.1 Definition for Analysis Populations

Full Analysis Set (FAS)

The FAS consists of all randomized subjects who have received at least one dose of study medication. The subjects will be grouped by the treatment group assigned at the time of randomization regardless of the actual treatment they receive during the study.

All baseline and efficacy analyses will be based on FAS.

Safety Analysis Set

The Safety Analysis Set consists of all subjects who have received at least one dose of study medication. For safety parameters, the treatment that was actually used by the subject will be applied in the analysis (an as-treated analysis).

5.2 Variables Used for Stratification of Randomization

Randomization will be stratified with respect to naïve or experienced to TNFi therapy and concurrent MTX use (yes or no) as determined at baseline.

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 for each randomized subject will be used as Baseline for summary of demographics, disease characteristics and safety analyses. For efficacy endpoints except mTSS and PsAMRIS, the baseline is defined as the last non missing pre-treatment observation prior to first dose. The Baselines for the modified total Sharp score (mTSS) and the MRI assessment, the imaging endpoints, are defined as the non-missing observation prior to Day 13 and closest to Day 1.

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 105 days (15 weeks) following the last dose of study drug for subjects who did not enter the OLE.
- The last non-missing observation collected on or before the subjects signed the Informed Consent Form (ICF) for the OLE and entered the OLE.

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 an 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.

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 Table 2, Table 3, Table 4, Table 5, Table 6, Table 7, Table 8, and Table 9.

Table 2. Visit Windows for Analysis of Efficacy Variables (ACR20/50/70, BSA, PSARC, MDA, TJC68 and SJC66, Patient's Assessment of Pain/Patient's Global and Physician's Global Assessment of Disease Activity, HAQ-DI, CRP, LDI, LEI, Dactylitis Count, Dactylitis Presence, SPARCC Enthesitis Index, PASI, sPGA)

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	< 1
Week 2	VISIT 3	15	[2, 22]
Week 4	VISIT 4	29	[23, 43]
Week 8	VISIT 5	57	[44, 71]
Week 12	VISIT 6	85	[72, 99]
Week 16	VISIT 7 (EOT)	113	[100, 127]
Week 20	VISIT 8	141	[128, 155]
Week 24	VISIT 9	169	[156, 183]
Week 28	VISIT 10	197	[184, 211]
Week 32	EOS	225	[212, 238]

Table 3. Visit Windows for Analysis of BASDAI

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	< 1
Week 16	VISIT 7 (EOT)	113	[2, 141]
Week 24	VISIT 9	169	[142, 196]

Table 4. Visit Windows for Analysis of SF-36 v2 and FACIT-F

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	< 1
Week 4	VISIT 4	29	[2, 71]
Week 16	VISIT 7 (EOT)	113	[72, 141]
Week 24	VISIT 9	169	[142, 196]

Table 5. Visit Windows for Analysis of mNAPSI

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	< 1
Week 4	VISIT 4	29	[2, 43]
Week 8	VISIT 5	57	[44, 71]
Week 12	VISIT 6	85	[72, 99]
Week 16	VISIT 7 (EOT)	113	[100, 127]
Week 20	VISIT 8	141	[128, 155]
Week 24	VISIT 9	169	[156, 183]
Week 28	VISIT 10	197	[184, 211]
Week 32	EOS	225	[212, 238]

Table 6. Visit Windows for Analysis of mTSS

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	[–999, 13]
Week 24	VISIT 9	169	[14, 306]

Table 7. Visit Windows for Analysis of PsAMRIS

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	[–999, 13]
Week 16	VISIT 7 (EOT)	113	[14, 224]

Table 8. Visit Windows for Local Tolerability

Window Label	Nominal Visit	Target Day	Interval
Week 2	VISIT 3	15	[1, 22]
Week 4	VISIT 4	29	[23, 43]
Week 8	VISIT 5	57	[44, 71]
Week 12	VISIT 6	85	[72, 99]
Week 16	VISIT 7 (EOT)	113	[100, 127]
Week 20	VISIT 8	141	[128, 155]

Table 9. Visit Windows for Analysis of Safety Laboratory Tests, and Vital Signs

Window Label	Nominal Visit	Target Day	Interval
Baseline	VISIT 2	1	< 1
Week 2	VISIT 3	15	[2, 22]
Week 4	VISIT 4	29	[23, 43]
Week 8	VISIT 5	57	[44, 71]
Week 12	VISIT 6	85	[72, 99]
Week 16	VISIT 7 (EOT)	113	[100, 127]
Week 20	VISIT 8	141	[128, 155]
Week 24	VISIT 9	169	[156, 183]
Week 28	VISIT 10	197	[184, 211]
Week 32	EOS	225	[212, 238[a]]

a. Within 105 days of last double-blind dose for safety analyses.

The same visit windows will be used for safety and laboratory assessments.

The time windows specified in Table 10 will be used for the summary of study drug injections.

Table 10. Visit Windows for Summary of Risankizumab Injections for Treatment Group Risankizumab 150 mg at Week 0, 4, 8, 12 and 16

Window Label	Nominal Visit	Target Day	Interval
Week 0	VISIT 2	1	N/A
Week 4	VISIT 4	29	[2, 43]
Week 8	VISIT 5	57	[44, 71]
Week 12	VISIT 6	85	[72, 99]
Week 16	VISIT 7 (EOT)	113	[100, 126]

6.5 Missing Data Handling

Missing data will be imputed for efficacy variables only for data up to Week 24. Data after Week 24 will not be imputed and only be analyzed as observed. Missing values can occur due to a missed visit or due to dropout from the study. Data will be categorized "as observed" and "imputed" as deemed appropriate. The following imputation approaches will be used to impute missing data:

- For the binary variables, the Non-responder Imputation (NRI) approach will be used for imputing missing data.
- For the continuous variables, Last Observation Carried Forward (LOCF) approach will be used for imputing missing data.

Missing data up to Week 24 will be addressed by considering the participant a non-responder for their binary endpoint regardless of adherence to treatment or early discontinuation.

Modified Non-Responder Imputation (mNRI) Approach:

According to the non-responder imputation principle, any missing binary response at a given visit will be imputed as non-responders for that visit. This imputation approach ensures the most conservative estimate for a response for active treatment arms.

For composite categorical efficacy endpoint such as ACR20/50/70, PsARC, MDA:

- **Step 1:** all missing components will be imputed using Last Observation Carried Forward (LOCF) first, and then calculate the composite score.
- **Step 2:** if the composite score still cannot be determined by Step 1, the composite score will be imputed as 0. In addition, all subjects who prematurely discontinue from the study will be considered as non-responders for all subsequent visits up to Week 24 after the discontinuation date.

For categorical efficacy endpoint derived from a single component such as EULAR response, presence of dactylitis, the missing score will be imputed as non-responder.

For categorical efficacy endpoint derived from a single score such as PASI75 and sPGA Response that will only be measured at visits when BSA ≥ 3%: The missing case will be imputed as a non-responder if the BSA ≥ 3%. The missing case will instead be imputed as a responder if the BSA < 3% at visits after baseline. If BSA is missing the missing case will be imputed as a non-responder. PASI90/100 will follow this rule with the exception that if PASI is missing and BSA < 3% then PASI90/100 will be missing.

Last Observation Carried Forward (LOCF) Approach:

The LOCF approach will be used for imputing missing data for continuous variables.

The LOCF analyses will use the completed evaluation from previous visit for efficacy measures assessed to impute missing data at later visits. Only post-baseline value will be carried forward. Baseline data will not be carried forward to post baseline visit.

For composite efficacy endpoint such as DAS28

- **Step 1:** all missing components will be imputed using LOCF carrying forward the most recent previous non-missing component value (excluding baseline values), and then a composite score will be calculated by following the "as observed" calculation.
- **Step 2:** if a composite score is still missing after Step 1, the composite score of the most recent previous visit with a non-missing composite score will be carried forward.

For composite PROs such as HAQ-DI, BASDAI, SF-36, FACIT-F, scoring rule will be followed when missing component presents. If the composite score still cannot be determined, then the entire composite score from the previous visit will be carried forward.

Observed Cases (OC)

The OC analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a schedules visit will be excluded from the OC analysis for that visit.

Missing Data Handling for mTSS and PsAMRIS

The following imputation method on subject level will be used as the primary approach to impute missing mTSS and PsAMRIS:

- Calculate average change from Baseline to follow-up visit (for example, Week 24 for mTSS and Week 16 for PsAMRIS) computed for all observed non-missing bones/joints data at both time points from a given subject.
- If the Baseline score is available for a bone/joint, then the impute value at the follow-up visit will be the Baseline score plus the average change score from the other scored bone/joint locations.
- If the Baseline visit is unavailable but the follow-up visit score is available, the imputed value at the Baseline will be computed from the follow-up time point score minus the average change, if the average change is smaller than the follow-up time point score at that location. Otherwise, the Baseline visit imputed value will use the most recent follow-up time point score carried backward.
- If both Baseline and follow-up time point visit score values are missing for a bone/joint, no values will be imputed to either visit, and that joint's bone/joint contribution will be zero, which means assigning 0 to both baseline and the follow-up time point for that joint.

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

7.1 Demographic and Baseline Characteristics

Demographics and Baseline characteristics will be summarized for each arm and for overall of the FAS populations. Continuous variables will be summarized with the number of non-missing observations by mean, standard deviation, first quartile, median, third quartile, minimum and maximum values. Categorical data will be summarized using frequencies and percentages. Statistical tests will be performed to assess the comparability of the placebo arm and the mixed two arms at baseline. Treatment comparison will be made based on non-missing information. Continuous variables will be analyzed using one-way analysis of variance (ANOVA). Categorical variables will be analyzed using Fisher's exact test.

The following demographic and baseline parameters will be summarized.

Subject Demographics

- Sex (male, female)
- Ethnicity (Hispanic or Latino, Other)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White)
- Age (years), defined as the number of years from date of birth to date of first drug
- Age Categories (< 65, ≥ 65)
- 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/m^2) (BMI < 30 vs BMI ≥ 30)

PsA Medical History and Disease Characteristics at Baseline

- Categorical:
  - MDA (Minimal Disease Activity) for PsA
  - PASI (Psoriasis Area and Severity Index) assessed in subjects with ≥ 3% BSA of psoriatic plaques
  - 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 LEI and 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)
  - Anti-cyclic citrullinated peptide (Anti-CCP) status: Positive or Negative
  - Rheumatoid Factor (RF) status: Positive (≥ 15) or Negative (< 15)
- Continous
  - 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 and SPARCC) at baseline (yes/no) 0 – 6
  - Duration of PsA in years
  - CASPAR classification criteria total score
  - SPARCC (Spondyloarthritis Research Consortium of Canada) Enthesitis Index in subjects with enthesitis

- mNAPSI (Modified Nail Psoriasis Severity Index) assessed in subjects with nail psoriasis at baseline 0 – 130
- Anti-cyclic citrullinated peptide (Anti-CCP) (units)

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-F (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
- PsAMRIS (Psoriatic Arthritis Magnetic Resonance Image Scoring System)
  - Synovitis 0 – 72
  - Tenosynovitis 0 – 72
  - Periarticular Inflammation 0 – 48

- Osteitis 0 – 144
- Bone Erosion 0 – 480
- Bone Proliferation 0 – 24

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 Previous 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. A concomitant medication is defined as any medication that started prior to the first dose of study drug 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 105 days (15 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 biologic DMARDS
- Concomitant NSAIDS

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

7.4 Protocol Deviations

As BI and AbbVie have different criteria of protocol deviation, a list of important protocol deviations according to BI and AbbVie is given in Appendix Section 15.A4. 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 following sets: Full Analysis Set and 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 FAS population:

- Number of subjects randomized
- Number of subjects that took at least one dose of study drug
- Number of subjects who completed EOT
- Number of subjects who prematurely discontinued from the study before EOT
- Number of subjects who elect to participate in the open label extension Study M16-244

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
  - Worsening of disease under study
  - Worsening of other preexisting condition
  - Other adverse event
- Protocol violation
- Lost to follow-up,
- Withdrew by subject,
- Others.

9.0 Study Drug Duration and Compliance

A summary of study drug duration will be provided for each treatment arm in each treatment part in the corresponding Safety populations.

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 the follows:

Duration = date of last injection – date of first injection + 28 days

Compliance

Each subject will receive five risankizumab (active or placebo) injections Treatment compliance (TC) will be summarized based on Safety Analysis Set for each treatment group. 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 = Round ([Last Study Drug Dose Date – First Study Drug Dose Date – 1]/28 + 1), where round (x) rounds x to the nearest integer.

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

10.0 Efficacy Analysis

10.1 General Considerations

Table 11 provides the overview of all efficacy endpoint assessments during the study.

Table 11. Overview of Efficacy Endpoint Assessments

Trial Periods	Treatment							Follow-Up		
Visit	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10[1]**	**EOS[2]**
Week	**0**	**2**	**4**	**8**	**12**	**16**	**20**	**24**	**28**	**32**
Tender Joint Count (68)	x	x	x	x	x	x	x	x	x	x
Swollen Joint Count (66)	x	x	x	x	x	x	x	x	x	x
Patient's assessment of pain (VAS)	x	x	x	x	x	x	x	x	x	x
Patient's global assessment of disease activity (VAS)	x	x	x	x	x	x	x	x	x	x
Physician's global assessment of disease activity (VAS)	x	x	x	x	x	x	x	x	x	x
HAQ-DI	x	x	x	x	x	x	x	x	x	x
CRP	x	x	x	x	x	x	x	x	x	x
Leeds Dactylitis Index, dactylitis count, dactylitis presence	x	x	x	x	x	x	x	x	x	x
Leeds Enthesitis Index (0 – 6)	x	x	x	x	x	x	x	x	x	x
SPARCC Enthesitis Index (0 – 16)	x	x	x	x	x	x	x	x	x	x
BSA	x	x	x	x	x	x	x	x	x	x
PASI, assessed in patients with ≥ 3% BSA of psoriatic plaques at baseline and/or at current visit	x	x	x	x	x	x	x	x	x	x
sPGA assessed in patients with ≥ 3% BSA of psoriatic plaques at baseline and/or at current visit	x	x	x	x	x	x	x	x	x	x
mNAPSI	x		x	x	x	x	x	x	x	x
SF-36 Version 2	x		x			x		x		
FACIT-F	x		x			x		x		
BASDAI, assessed only in patients with baseline inflammatory spondylitis based on investigator judgement	x					x		x		

Table 11. Overview of Efficacy Endpoint Assessments (Continued)

Trial Periods	Treatment							Follow-Up		
Visit	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10[1]**	**EOS[2]**
Week	**0**	**2**	**4**	**8**	**12**	**16**	**20**	**24**	**28**	**32**
DAS28(CRP) (calculated)	x	x	x	x	x	x	x	x	x	x
EULAR response criteria (calculated)		x	x	x	x	x	x	x	x	x
ACR 20, 50, 70 (calculated)		x	x	x	x	x	x	x	x	x
PsARC (calculated)	x	x	x	x	x	x	x	x	x	x
MDA (calculated)	x	x	x	x	x	x	x	x	x	x
mTSS (X-Ray)	x							x		
PsAMRIS	x					x				

1. Subjects that enrol in the extension study, at Week 24 or Week 28 visit, do not need to attend any remaining visits in Study 1311.5; however, the Trial Completion page and Visit date page at the EOS visit in the eCRF must be completed.
2. End of Study (EOS).

Table 12 provides an overview of the efficacy endpoints to be analyzed and the general analysis methods. The general analysis methods for binary and continuous endpoints will be described in detail following the table. Some endpoints may require additional analyses other than the general ones listed in Table 12 – these additional analyses will be described in detail in later sections.

Table 12. Summary of Efficacy Variables and Corresponding Analyses

Efficacy Variables	Analysis Method
Binary Efficacy Endpoints:	
• **Primary and Secondary Endpoints:** o ACR20/50/70 responses at Week 16 o PASI90 response at Week 16 in subjects with PsO BSA ≥ 3% at baseline	• Point estimate and 90% CI of the response rate for each treatment group using exact method • Point estimate, 90% CI and p-value of the response rate difference between combined Arm 1 and Arm 2 vs the placebo, as well as the response rate difference between combined Arm 2 and Arm 3 using the stratified Cochran-Mantel-Haenszel test. • Pairwise comparisons of the each risankizumab dose group vs placebo will be conducted using the same stratified Cochran-Mantel-Haenszel methods. • Histogram of response rates by treatment group • MCP-Mod to evaluate dose-response for ACR 20/50/70 • Forest plot • No multiplicity adjustments. • Imputation: NRI (primary), OC
Further Binary Endpoints: At all measured time points except Week 16: • ACR20/50/70 At all measured time points: • PASI75 in subjects with a ≥ 3% baseline PsO BSA • PASI100 in subjects with a ≥ 3% baseline PsO BSA • sPGA Response (0 = clear and 1 = almost clear) in subjects with a ≥ 3% baseline PsO BSA • EULAR response • Presence of Dactylitis • MDA • PsARC	• Point estimate and 90% CI of the response rate for each treatment group using exact method • Point estimate, 90% CI and p-value of the response rate difference between combined Arm 1 and Arm 2 vs the placebo, as well as the response rate difference between combined Arm 2 and Arm 3 using the stratified Cochran-Mantel-Haenszel risk difference estimate. • Pairwise comparisons of the risankizumab dose groups vs placebo will be conducted using the same stratified Cochran-Mantel-Haenszel methods. • Plot of ACR20/50/70 response rates overtime by treatment group • Ordinal Logistic Regression of EULAR response • Forest plot • Imputation: NRI (primary), OC

Table 12. Summary of Efficacy Variables and Corresponding Analyses (Continued)

Efficacy Variables	Analysis Method
Continuous Efficacy Endpoints:	
Secondary Endpoints: Change from baseline at Week 16: • TJC68 • SJC66 • HAQ-DI • SF-36 • Dactylitis Count in subjects with dactylitis at baseline • SPARCC in subjects with enthesitis at baseline • mNAPSI assessed in subjects with nail psoriasis at baseline **Further Continuous Endpoints:** Change from baseline at all measured time points: • Physician's Global (VAS) • Patient's Pain (VAS) • Patient's Global Activity (VAS) • CRP • DAS28(CRP) • LDI in subjects with dactylitis at baseline • LEI in subjects with Enthesitis at baseline • Presence of enthesitis • FACIT-F Change from baseline at Week 16 and 24: • BASDAI in subjects with baseline inflammatory spondylitis	• Point estimate and 90% CI within each treatment group using mixed model repeated measures (MMRM) model; • Point estimate, 90% CI and p-value for the mean change difference between combined Arm 1 and Arm 2 vs placebo, as well as the mean change difference between combined Arm 2 and Arm 3 with mixed model repeated measures (MMRM) model • Pairwise comparisons of the each risankizumab dose group vs placebo will be conducted MMRM model • Box plots and change from baseline over time will be plotted for key continuous variables • For SF-36, spyderdiagram will be provided • Imputation: OC (primary)

Table 12. Summary of Efficacy Variables and Corresponding Analyses (Continued)

Efficacy Variables	Analysis Method
Continuous Efficacy Endpoints (continued):	
mTSS and PsAMRIS: • Change from baseline (CFB) at Week 24 in mTSS • Change from baseline (CFB) at Week 16 in PsAMRIS	• Mean CFB and 90% CI within each treatment group; • Mean CFB , 90% CI and p-value for the difference between combined Arm 1 and Arm 2 and the placebo, as well as the difference between combined Arm 2 and Arm 3 using ANCOVA model • Pairwise comparisons of the each risankizumab dose group vs placebo will be conducted with ANCOVA model

The treatment effect will be evaluated based on a two-sided significance level of 0.1 (when rounded to three decimal places). The efficacy analysis will be conducted in the FAS.

For categorical variables, frequencies and percentages will be summarized by treatment group. The treatment groups will be compared using the Cochran-Mantel-Haenszel (CMH) risk difference estimate stratified by the randomization factors with weights proposed by Greenland & Robins.[10] Pairwise comparisons of all doses of treatment received versus placebo will be conducted using the same Cochran-Mantel-Haenszel methods, 90% confidence intervals as well as nominal p-values of the comparison of doses will be provided. There will be no adjustments for multiplicity in these analyses.

For continuous variables, the model based mean and standard error will be presented. The Baseline, visit means, change from baseline at each visit will also be presented by treatment group for subjects who have both Baseline and post Baseline visit values. For continuous variables, the mixed model repeated measurement (MMRM) model with treatment regimen, clinical visit and stratification factors (prior TNF and concurrent MTX use) as fixed factors, and baseline score as a fixed continuous covariate, will be used. In addition, participants will be included in the model as a random effect. Furthermore, the

treatment by clinical visit interaction will be included in the model. An unstructured covariance structure will be used to model the within-patient measurements. Parameter estimation of the MMRM will be based on the residual maximum likelihood method (REML).

At Week 16 analysis, MMRM models will only include visits through Week 16. At final analysis, MMRM models will include all visits through Week 24. Any observations beyond Week 24 will be summarized using descriptive statistics and p-value will be based on ANOVA.

Some endpoints may require additional analyses other than the general ones listed in Table 10 – these additional analyses will be described in detail in later sections.

10.2 Primary Efficacy Analysis

The primary endpoint to assess the efficacy of risankizumab is ACR20 response at Week 16.

The primary null hypothesis is the proportion of subjects who achieve ACR 20 at Week 16 in the combined Arm 1 and Arm 2 of risankizumab is less than or equal to that of the placebo.

Alternative hypothesis is that the proportion of participants who achieve ACR 20 at Week 16 in the combined arms of Arm 1 and Arm 2 of risankizumab is superior to that of the placebo.

The difference in proportion of participants that achieve ACR 20 between the combined groups of risankizumab (Arm 1 and Arm 2) and the placebo arm (Arm 5) will be estimated and tested using the stratified Cochran-Mantel-Haenszel risk difference estimate, stratified based on prior TNFi use and concurrent MTX use. Pairwise comparisons of the risankizumab dose groups versus the placebo, as well as comparison of the combined groups of risankizumab (Arm 2 and Arm 3) versus the placebo, will be

conducted using the same stratified Cochran-Mantel-Haenszel methods. There will be no adjustments for multiplicity in these analyses.

The primary analysis will be performed when the last subject completes the Week 16 visit, when the primary endpoint (ACR 20 at Week 16) is assessed. The trial will be unblinded to the sponsor project and trial team to perform this analysis. Blinded treatment assignments will not be disseminated to the sites, investigators, and subjects until the end of study database lock.

10.3 Sensitivity Analysis of Primary Efficacy Variable

The following sensitivity analyses for the primary endpoint ACR20 at Week 16 will be conducted:

- The primary analysis for point estimate and treatment comparisons will be repeated using as observed cases without any imputation.

10.4 Dose Response Modelling

The dose-response relationship among the 4 Risankizumab dose groups (weighted cumulative doses between Week 0 and Week 12 are 600 mg (Arm 1), 300 mg (Arm 2), 200 mg (Arm 3), and 75 mg (Arm 4) and the placebo group will be characterized for the primary endpoint ACR20 at Week 16 using the Multiple Comparison Procedure – Modeling (MCP-Mod)[1,2] method. The response based on the primary analysis approach NRI will be used, and ADDPLAN DF software will be used to perform the MCP-Mod analyses.

A set of 5 pre-specified standardized candidate dose-response models, as described in Table 13, will be utilized to examine the dose-response relationship. A statistically significant dose response relationship will be declared if at least one model is identified by the MCP-Mod method to be statistically significant at $\alpha = 0.05$ one-sided. The fitted dose response curves will be presented graphically for all statistically significant models along with confidence bands. The minimum effective dose (MED) will be identified for

each statistically significant model based on the pre-specified clinical meaningful target 20% target difference from placebo. The weighted MED across all significant models will be calculated, with weight being inverse of model AIC.

Table 13. Candidate Models

Model	$f(d, \theta)$ d = dose, θ = Model Parameters	$f^0(d, \theta)$ Standardized Model	Initial Value(s) for Parameter(s)
Linear	$E_0 + \delta d$	d	NA
Logistic	$E_0 + \frac{E_{max}}{1 + exp\left(\frac{ED_{50} - d}{\delta}\right)}$	$\frac{1}{1 + exp\left(\frac{ED_{50} - d}{\delta}\right)}$	$ED_{50} = 150, \delta = 95$
E_{max}	$E_0 + \frac{E_{max} d}{ED_{50} + d}$	$\frac{d}{ED_{50} + d}$	$ED_{50} = 150$
sigE_{max}	$E_0 + \frac{E_{max} d}{ED_{50} + d}$	$\frac{d^h}{{ED_{50}}^h + d^h}$	$ED_{50} = 150, h = 0.8$
Quadratic	$E_0 + \beta_1 d + \beta_2 d^2$	$d + \frac{\beta_2}{\lvert\beta_1\rvert} d^2$	$\delta = -0.001$

Note: The candidate model and model parameters provided in the above table are just examples and should be adjusted case by case based on input from CPPM.

10.5 Secondary Efficacy Analyses

The secondary endpoints are as follows:

- ACR 50 response at Week 16
- ACR 70 response at Week 16
- Change in Tender Joint Count at Week 16 as compared to baseline
- Change in Swollen Joint Count at Week 16 as compared to baseline
- Change in HAQ-DI at Week 16 as compared to baseline
- Change in SF-36 at Week 16 as compared to baseline
- Change in Dactylitis count at Week 16 as compared to baseline (in subjects with dactylitis at baseline)

- Change in SPARCC Enthesitis Index at Week 16 as compared to baseline (in subjects with enthesitis at baseline)
- Change in mNAPSI at Week 16 as compared to baseline (in subjects with nail psoriasis)
- $PASI_{90}$ response at Week 16 assessed in subjects with a $\geq 3\%$ baseline PsO BSA

Binary endpoints will be analyzed in a similar fashion as the primary endpoint.

For continuous endpoints, the baseline, visit mean, change from baseline within treatment group will be presented by treatment group. The comparison between combined Arm 1 and Arm 2 vs placebo will be analyzed using a mixed model repeated measures model (MMRM) which is valid under the missing at random (MAR) assumption. Under the MMRM model, there's no explicit imputation of missing data, rather, the future statistical behavior of those participants who drop out given their past measurements is assumed to be the same as those who remain with the same history.

Specifically, for continuous endpoints, such as HAQ-DI at Week 16, a MMRM which is valid under the missing at random (MAR) assumption will be implemented. In particular, between-treatment differences in the change in HAQ-DI at Week 16 (from baseline) will be evaluated using a MMRM with treatment regimen, clinical visit and stratification factors such as prior TNF and MTX use as fixed factors, and baseline HAQ-DI score as a fixed continuous covariate. In addition, participants will be included in the model as a random effect. Furthermore, the treatment by clinical visit interaction will be included in the model. An unstructured covariance structure will be used to model the within-patient measurements. Parameter estimation of the MMRM will be based on the residual maximum likelihood method (REML). If all post-baseline values are missing then the missing value will not be imputed and the data for the respective participant will be removed from the analysis, thus, the total number of participants providing data for analysis could be smaller than the total number of participants in the FAS.

10.6 Further Efficacy Analysis

Further endpoints are as follows:

- ACR 20/50/70 at all other measured time points
- Change in Physician's Global (VAS) at all measured time points as compared to baseline
- Change in Patient's Pain (VAS) at all measured time points as compared to baseline
- Change in Patient's Global Activity (VAS) assessments at all measured time points as compared to baseline
- Change in C-Reactive Protein (CRP) at all measured time points as compared to baseline
- Change in minimal disease activity (MDA) at at all measured time points compared to baseline
- Change in DAS28(CRP) at all measured time points as compared to baseline
- PsO endpoints assessed at all measured time points in subjects with a $\geq 3\%$ baseline PsO BSA:
  - $PASI_{75}$ response
  - $PASI_{100}$ response
- Change in sPGA clear and almost clear
- EULAR (European League Against Rheumatism) response at all measured time points
- Change in PsARC (Psoriatic Arthritis Response Criteria) at all measured time points as compared to baseline
- Presence of dactylitis (yes/no) at all measured time points
- Change in LDI at all measured time points as compared to baseline (in subjects with dactylitis at baseline)
- Change in LEI at all measured time points as compared to baseline (in subjects with enthesitis at baseline)
- Change in FACIT-F at all measured time points as compared to baseline

- Change in BASDAI at Weeks 16 and 24 (in subject with baseline inflammatory spondylitis) as compared to baseline
- Change in mTSS at Week 24 as compared to baseline
- Change in PsAMRIS parameters at Week 16 as compared to baseline

Analyses at all measured time points will be conducted in a similar fashion as for Week 16 endpoints except for EULAR response, mTSS, and PSAMRIS. EULAR response will be analysed using ordinal logistic regression with treatment group, prior TNFi use, and concurrent MTX use as main factors and baseline DAS value as a covariate. mTSS and PSAMRIS will be analysed using ANCOVA with treatment group, prior TNFi use, and concurrent MTX use as main factors and baseline value as a covariate. 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.

10.7 Handling of Multiplicity

There will be no adjustments for multiplicity in these analyses.

10.8 Efficacy Subgroup Analysis

The following subgroup analyses will be performed to evaluate the consistency of efficacy in the primary endpoint across demographic data, baseline disease characteristics, and drug history for PsA.

Subgroup Factor	Categories
Age	< 65, ≥ 65
Sex	Male or Female
BMI	< 30 or ≥ 30
Race	White or Others
Prior TNFi use	Naïve or experienced
Concurrent MTX use	Yes or no

.

10.9 Efficacy Variables Definition and Conventions

This section introduces the detailed definition of the efficacy variables. For detailed example and information, please refer to Section 10.0 Appendix of the protocol.

10.9.1 ACR Criteria

ACR criteria is a commonly used standard criteria mentioned in the guidance of American College of Rheumatology to evaluate the effectiveness of investigation drug in PsA clinical trials. It is a composite measurement calculated based on the improvement over aset 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)
   - Patient'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 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 be scientifically as robust as possible, the principle to calculate ACR response is to minimize imputation whenever possible. "As Observed" ACR response will be calculated first based on a derived visit window instead of the nominal visit identifier (e.g., Week 6 visit) 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.

10.9.2 Joint Evaluation

Anatomical joints are evaluated for swelling and tenderness at every study visit. The 34 anatomical joints in Table 14 are assessed in this study for both the left and right side of the body.

Table 14. 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.

10.9.3 Patient's Global Assessment of Disease Activity for Arthritis (PtGA)

The patient global assessment VAS will be self-administered by the patient at Baseline, Week 2,4,8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS).

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."

10.9.4 Physician's Global Assessment of Disease Activity for Arthritis (PhGA)

The physician will assess Patient's disease activity at the time of visit using a Physician's Global Assessment of Disease VAS at Baseline, Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS). The range is 0 to 100 mm with no activity being indicated by 0 and severe activity by 100.

10.9.5 Patient's Assessment of Pain (VAS)

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."

The pain-VAS will be self-administered by the patient at baseline, Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS).

10.9.6 Disease Activity Score Based on DAS28

The Disease Activity Score (DAS) is a combined index used to measure the joint disease activity in patients with PsA. The DAS provides a score between 0 and 10, indicating how active the joint disease is in psoriatic arthritis is at the time of measurement.

DAS28(CRP) score will be determined based on a continuous scale of combined measures of TJC28, SJC28, PtGA (in mm), and CRP (in mg/L).

DAS28(CRP) = 0.56 √("TJC28") + 0.28 √("SJC28") + 0.36 ln (CRP + 1) + 0.014 PtGA + 0.96, where √ is square root and ln is natural log.

Low Disease Activity (LDA) by DAS28 (CRP) is defined as 2.6 ≤ DAS28 (CRP) < 3.2. Clinical Remission (CR) by DAS28 (CRP) is defined as DAS28 (CRP) < 2.6.

Table 15. Anatomical Joints for DAS28(CRP) Calculation

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

Proportional extrapolation will be used to compute joint counts for the joints that are replaced or not assessed.

As-Observed DAS28 Scores

To calculate "as observed" DAS28 scores, the "as observed" component value will be calculated first. Then the components will be included in the calculation per the DAS formula selected. If any "as observed" component is missing in a window, then the "as observed" DAS28 score will be missing.

The DAS28 will be evaluated at baseline, Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS).

10.9.7 European League Against Rheumatism (EULAR) Response Criteria

EULAR response will be evaluated according to the following table at Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS).

<table>
<tr><th rowspan="2">DAS28 at endpoint</th><th colspan="3">Improvement in DAS28 from baseline:</th></tr>
<tr><th>> 1.2</th><th>> 0.6 and ≤ 1.2</th><th>≤ 0.6</th></tr>
<tr><td>≤ 3.2</td><td>Good</td><td rowspan="2">Moderate</td><td rowspan="3">None</td></tr>
<tr><td>> 3.2 and ≤ 5.1</td><td rowspan="2"></td></tr>
<tr><td>> 5.1</td><td colspan="2"></td></tr>
</table>

For missing data use Non-Responder Imputation using the score of "None" as non-response.

10.9.8 Minimal Disease Activity (MDA) in PsA

Table 16. Minimal Disease Activity (MDA) Criteria in Psoriatic Arthritis

A subject 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-DI ≤ 0.5
- Tender entheseal points ≤ 1 (Presence of Enthesitis in the 18 sites examined for the LEI and SPARCC evaluation)

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. See Section 6.0 for details on imputation. The MDA will be evaluated at baseline, Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS).

10.9.9 Psoriatic Arthritis Response Criteria (PsARC) Response

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. The PsARC will be evaluated at Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS).

10.9.10 Psoriasis Area and Severity Index (PASI) and Body Surface Area (BSA)

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 PASI75/90/100 (defined as at least 75% and 90% reduction in PASI score compared to baseline) will be summarized by treatment group in subjects with ≥ 3% BSA (Body Surface Area) psoriasis involvement at baseline.

The PASI will be evaluated at baseline, Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS) only for subjects with ≥ 3% BSA of psoriatic plaques at baseline and/or at current visit.

If any "as observed" component is missing in a window, then the "as observed" PASI score will be missing.

10.9.11 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 17).

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 17. 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

If any "as observed" component is missing in a window, then the "as observed" sPGA score will be missing.

sPGA Responder

The sPGA will be dichotomized into Responders, 0 (clear) and 1 (almost clear) in one category, and Non-Responders, ≥ 2 in the other. The sPGA will be evaluated at baseline, Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS) only for subjects with ≥ 3% BSA of psoriatic plaques at baseline and/or at current visit.

10.9.12 Leeds Dactylitis Index (LDI)

The LDI basic 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). 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 the 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 both affected and tender, 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. The imputation rule for presence of dactylitis is that if all sites are missing then presence of dactylitis will be imputed as "present."

The dactylitis will be evaluated at baseline, Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS) for subjects with dactylitis at baseline.

10.9.13 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.

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. The imputation rule for presence of enthesitis is that if all sites are missing then presence of enthesitis will be imputed as "present."

10.9.14 Research Consortium of Canada (SPARCC) Enthesitis Index

The SPARCC Enthesitis index is an outcome measure for enthesitis in SpA created by the Spondyloarthritis Research Consortium of Canada.[3] 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 18. Enthesial Sites Examined for SPARCC Calculation

Medial epicondyle	Lateral epicondyle	Supraspinatus insertion into greater tuberosity of humerus	Greater trochanter
Quadriceps insertion into superior border of patella	Atellar 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 LEI and SPARCC:

If enthesitis is present with any of the 18 sites for LEI or SPARCC, the subject is counted as a subject with enthesitis based on LEI and SPARCC. Accordingly, "subjects with ehthesitis at baseline" will also be defined using presense at any of the 18 LEI and

SPARCC sites. The imputation rule for presence of enthesitis is that if all sites are missing then presence of enthesitis will be imputed as "present."

10.9.15 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 Section 10.13. 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.

The mNAPSI will be evaluated at Baseline, Week 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS) for patients with nail psoriasis at baseline.

10.9.16 Health Assessment Questionnaire-Disability Index (HAQ-DI)

HAQ-DI 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.

The HAQ-DI will be self-administered by the subject at Baseline, Week 2, 4, 8, 12, 16 (EOT), 20, 24, 28 and 32 (EOS).

10.9.17 SF-36v2

The 36-Item Short Form, Version 2 (SF-36v2) Questionnaire with 1 week recall will be completed by the subject at Baseline, Weeks 4, 16 and 24 or at PD. The SF-36v2 health survey 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.[4]

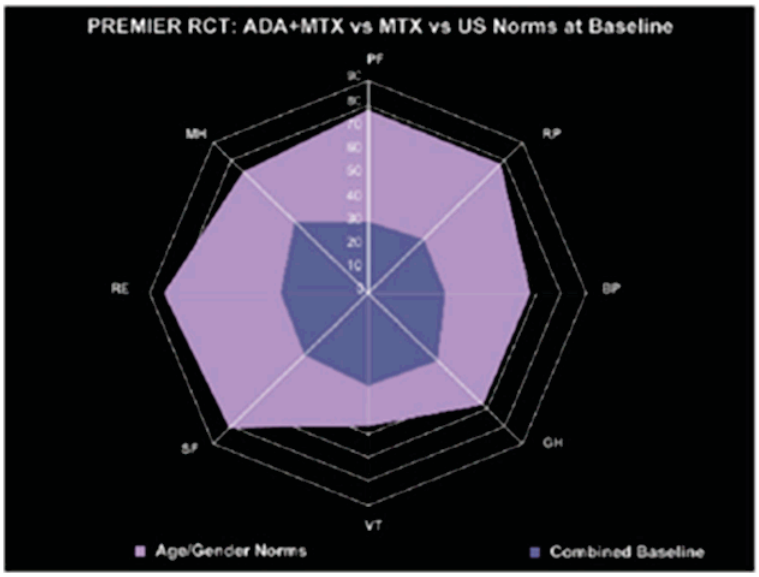
The coding and scoring for the SF-36 will use the software provided by Optimum.

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 2A, B). 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.

For a given time point, the values from the five treatment groups will be plotted in one spydergram. Each time point (baseline, Week 4, 16 and 24) will have its own spydergram showing difference between treatment groups.

Figure 2. An Example of Spydergram of SF-36



Note: Strand 2009[6]

10.9.18 Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)

The FACIT-Fatigue is a 13-item questionnaire[7] 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. FACIT-F will be evaluated at Weeks 0, 4, 16, and 24.

10.9.19 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 10 cm Visual Analogue 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 are more than one mark for an item, take the average.

Below are the 6 questions.

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

The BASDAI will be evaluated at Baseline, 16 and 24.

10.9.20 Modified Total Sharp Score (mTSS)

To obtain the total mTSS score, scores for erosions and joint space narrowing (JSN) in both the hands and feet will be added together. If a joint or bone is scored as not visible (e.g., poor film quality, missing imaging, severe misalignment, flexion deformity, dislocation) or if radiographs show a joint or bone with surgical fusion, replacement (prosthesis), or amputation, the joint will be counted as missing in the calculation of the total mTSS score.

Erosions

There are in total 52 erosion scoring locations. The range of the total erosion score for both hands and feet for 1 patient per time point is from 0 to 320. Maximum total erosion score of hands is 200 and the maximum total erosion score of the feet is 120. An erosion total score is computed by summing the erosion values at each of the scored locations.

Gross osteolysis and "pencil in cup" is scored separately. In the final summary score, joints with one of these abnormalities get the maximum score assigned for erosions.

When a joint location has not been scored due to inadequate coverage or poor image quality, that location will be designated as NA and the total erosion score will be calculated using the following rules:

- 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 will be calculated as Total Erosion Score = average of all available scores*52
- 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

Erosion will be scored at the following 20 hand/wrist locations in each hand/wrist: 1 IP joint of the thumb; 4 DIP joints; 4 PIP joints; 5 MCP joints; the first metacarpal (MC1) bone; the radius, ulna, and trapezium/trapezoid (scored as 1 unit; multangular [Mul]); the navicular (Nav); and the lunate.

The following 6 locations will be scored in each foot: 5 MTP joints and 1 IP of the great toe. Each side of each foot joint (i.e., proximal and distal) is graded and added together for a joint score; therefore, a range of 0 to 10 is the range per foot joint.

Since two independent readers evaluate each film, the mean score will be calculated for the two readers from the individual erosion scores.

Joint Space Narrowing (JSN)

There are 52 total JSN scoring locations. The range of the total JSN Score for both hands and feet per patient per time point is from 0 to 208. The maximum possible score for joint space narrowing is 160 for the hands and 48 for the feet. A total JSN score is computed by summing the joint narrowing values at each of the scored locations.

Gross osteolysis and "pencil in cup" are scored separately. In the final summary score, joints with one of these abnormalities get the maximum score assigned for joint space narrowing.

When a bone location has not been scored because of inadequate coverage or image quality, that location will be scored as NA and the total JSN score will be calculated using the following rules:

- 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

The JSN Score will be determined for the following 20 hand/wrist locations in each hand/wrist: 1 IP joint; 4 DIP joints; 4 PIP joints; 5 MCP joints; carpometacarpal (CMC) joints 3, 4, and 5; the radiocarpal (RC) joint, the multangular navicular (MAN) joint; and the capitonaviculolunate (CNL) joint.

The following 6 locations will be scored in each foot: 5 MTP joints and the IP joint of the great toe.

Since two independent readers evaluate each film, the mean score will be calculated for the two readers from the individual JSN scores.

Modified Total Sharp Score

The modified Total Sharp Score (mTSS) is the sum of the Erosion Score (i.e., Total Erosion Score) combined with the sum of the JSN Score (i.e., Total JSN Score). The range of the mTSS for both hands and feet, per patient per timepoint is 0 to 528. Each of the Total Change Score for Erosion, JSN and mTSS will be calculated by substracting respective post-treatment from baseline scores. Total Change Scores will not be calculated for patients with baseline only visit.

The average modified TSS from the two readers will be used in statistical analysis.

Adjudication

Cases for adjudication will be determined based on the difference in Change Score for mTSS between Reviewer 1 and 2.

Queries to the assessment database will be performed on an ongoing basis to designate cases for adjudication. For each patient dataset, the query will compare primary Reviewers' results to determine if differences necessitate adjudication. The queries will be defined by the thresholds described below:

1. The difference in mTSS change from Week 0 to Week 24 is greater than ± 6 mTSS units between Reviewer 1 and Reviewer 2

2. Change in mTSS that represents a difference in direction (one positive, one negative) by 3 or more mTSS units between Reviewer 1 and Reviewer 2.
3. Reviewer 1 scores adequate number of scoring locations resulting in calculation of mTSS and the Reviewer 2 scores less than adequate scoring locations resulting in no calculation of mTSS in either baseline of post-treatment assessments.

Once the system has determined a case for adjudication the case will be presented to a third reader for adjudication. The average modified TSS from the adjudicator and the reader whose result is closest to the adjudicator will be used in the analysis.

The total mTSS will be evaluated at Baseline and Week 24. Change from baseline in total mTSS score at Week 24 will be summarized.

10.9.21 Psoriatic Arthritis Magnetic Resonance Image Scoring System (PsAMRIS)

The international "MRI in inflammatory arthritis" group of OMERACT (Outcome Measures in Rheumatology) has developed the OMERACT PsAMRIS for evaluation of inflammatory and destructive changes in PsA hand[6] (or foot, if no swollen joint is available in hands at baseline) MRI will be performed as presented in the first approximately 90 subjects from the sites selected for MRI sub-study. Images will be read centrally for each subject at two time points (Weeks 0 and 16) in a chronological order and without knowledge of subject identity or treatment assignment. If a subject has to be rescued by a biologic treatment after Week 4 and prior to Week 16, MRI images need to be obtained prior to the start of rescue biologic treatment. If the rescue treatment is initiated prior to Week 4, post-treatment MRIs are not required.

An MRI Reviewer will assess images from baseline and Week 16 to produce MRI data on inflammation and damage for each given patient. The Reviewer will be presented with 2 time points from a given patient (as well as unscheduled if acquired and submitted). The MRI efficacy review assessments will be conducted as a side by side paired review. The Reviewer will not know the chronological order of visits belonging to the same

subject. The MRIs will be read in a random manner with the Reviewer being blinded to the site-subject number, visit and MRI scan date.

Each feature will be scored by following the criteria listed in Table 19.

The scored regions of the fingers are: D, the distal interphalangeal (DIP) joint region; P, the proximal interphalangeal joint (PIP) region; and M, the metacarpophalangeal (MCP) joint region. Each joint region is further delineated into distal (di), proximal (pr), dorsal (do) and volar (vo) joint regions. See the appendix for a diagram of the joints of the hands.

The regions of the toes are: D, the distal interphalangeal (DIP) joint region (DIP joints are not scored in the foot); P, the proximal interphalangeal joint (PIP) region; and M, the metacarpophalangeal (MCP) joint region. Each joint region is further delineated into distal (di), proximal (pr), dorsal (do) and volar (vo) joint regions. See the appendix for a diagram of the joints of the foot.

Table 19. PsAMRIS Scoring Guidance

Feature	Location	Scoring Criteria
Synovitis	M, P, and D	0: Normal 1: Mild 2: Moderate 3: Severe NA
Tenosynovitis	M, P, and D	0: None 1: < ½ tendon thickness 2: > ½ and < 1 tendon thickness 3: > 1 tendon thickness NA
Periarticular Inflammation	M, P, and D (dorsal and volar)	0: Present 1: Absent NA
Osteitis	M1, M2, P1, P2, D1, and D2 (distal and proximal)	0: Normal 1: Mild 2: Moderate 3: Severe NA
Bone Erosion	M1, M2, P1, P2, D1, and D2 (distal and proximal)	0 – 10 or NA, based on proportion of eroded bone compared to assessed bone volume from the articular surface
Bone Proliferation	M, P, and D	0: Present 1: Absent NA

The PsAMRIS will be evaluated at Baseline and Week 16. Change from baseline in PsAMRIS score at Week 16 will be summarized. PSAMRIS will have six separate scores for Synovitis (0 – 72), Tenosynovitis (0 – 72), Periaticular Inflammation (0 – 48), Osteitis (0 – 144), Bone Erosion (0 – 480), and Bone Proliferation (0 – 24).

11.0 Safety Analysis

11.1 General Considerations

Safety analyses will include reporting of adverse events, laboratory, and vital signs. Safety summaries will be provided using the Safety Analysis Set as defined in Section 5.1.

Mean changes from baseline of continuous laboratory and vital sign variables will be summarized for the Safety Analysis Set. Categorical data will be summarized using frequencies and percentages. Continuous variables will be analyzed using one-way ANOVA and categorical variables will be analyzed using Fisher's exact test. Number of non-missing values will be given. Missing safety data will not be imputed.

11.2 Analysis of Adverse Events

Treatment-Emergent Adverse Events (TEAE)

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 15 weeks (105 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).

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 Adverse Events**

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

   - Any TEAEs
   - Any severe TEAEs
   - Any serious AEs (SAE)
   - Any related TEAEs
   - Any TEAEs leading to discontinuation of study drug
   - Any TEAEs leading to death
   - Any deaths
   - Any TEAEs in areas of safety interest (Refer to Table 20)

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 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

Adverse events 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 present. The only exception is if the subject has another occurrence of the same adverse event with the most extreme severity (Life threatening). In this case, the subject will be counted under the "Life-threatening" category.

4. Adverse Events by "Reasonably Possibly Related" Relationship

TEAEs will also be summarized by relationship defined by "reasonably possibly related" to drug, as assessed by the investigator. If a subject has an AE with an unknown relationship, then the subject will be counted in as 'related.'

5. TEAEs by Preferred Term in Descending Frequency

TEAEs, SAEs, and TEAEs leading to discontinuation of study drug will be summarized by treatment group in decreasing order of frequency of MedDRA PT in the total risankizumab doses as well as the highest dose arm. 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 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

Areas of safety interest 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 20.

Table 20. 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 as described in Table 6 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 as described in Table 6 (for MACE +) using CECAT and CETERM from the CE SDTM dataset.	Display underlined terms from MACE and underlined terms below: • Hospitalization for Unstable Angina • PCI • CBG	N

Table 20. 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 as described in Table 6 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

Table 20. Areas of Safety Interest Listing (Continued)

ASI Grouping	Categories (ASI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Serious infections, TB, fungal and opportunistic infections (including herpes zoster)	Serious infections	Serious PTs of the CMQ (company MedDRA query) Infections (CMQ 80000018)	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
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

Table 20. Areas of Safety Interest Listing (Continued)

ASI Grouping	Categories (ASI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
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
Drug induced liver injury (DILI)	Drug induced liver injury (DILI)	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 A5 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

The treatment-emergent adverse events occurring during the entire study will be presented by event rate per 100 patient years. 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 as date of last dose – date of first dose + 105 days (5 half-lives)) of all subjects normalized by 365.25, and rounded to one decimal place.

10. 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 Events
- 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 continuous laboratory parameters will be summarized by n, mean, standard deviation, minimum value, median, and maximum value for each treatment group.

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 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 21 and Table 22 will be summarized.

Table 21. 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 22. Criteria for Potentially Clinically Important Hematology Values

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

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, 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 23.

Table 23. Laboratory Tests

Category	Test Name
Hematology	Activated Partial Thromboplastin Time
	Basophils/Leukocytes
	Eosinophils/Leukocytes
	Neutrophils/Leukocytes
	Fibrinogen
	Hematocrit
	Hemoglobin
	Hemoglobin A1C
	Prothrombin Intl. Normalized Ratio
	Lymphocytes/Leukocytes
	Monocytes/Leukocytes
	Platelets
	Erythrocytes
	Reticulocytes/Erythrocytes
	Leukocytes
Chemistry	Alkaline Phosphatase
	Alanine Aminotransferase
	Amylase
	Aspartate Aminotransferase
	Bicarbonate
	Direct Bilirubin
	Bilirubin
	Indirect Bilirubin
	Calcium
	Cholesterol
	Creatine Kinase
	Creatine Kinase MB
	Chloride
	Creatinine
	C Reactive Protein
	Gamma Glutamyl Transferase
	Glucose
	Hepatitis B DNA
	Hepatitis B Virus Surface Antigen
	Hepatitis C Virus Antibody

Table 23. Laboratory Tests (Continued)

Category	Test Name
Chemistry (continued)	Choriogonadotropin Beta
	HDL Cholesterol
	HIV-1/2 Antibody
	Potassium
	Lactate Dehydrogenase
	Lipase
	M. tuberculosis IFN Gamma Response
	Protein
	Rheumatoid Factor
	Sodium
	Triiodothyronine, Free
	Thyroxine, Free
	Triglycerides
	Troponin I
	Thyrotropin
	Urate
	Albumin
	Urea Nitrogen
Urinalysis (dipstick)	Albumin/Creatinine
	Bacteria
	Casts
	Creatinine Clearance
	Crystals
	Squamous Epithelial Cells
	Ketones
	Neutrophils/Leukocytes
	Nitrite
	pH
	Urobilinogen

11.3.2 Statistical Methods

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

Changes from Baseline 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:

- $< 1.5 \times$ ULN
- $\geq 1.5 \times$ ULN – $< 3.0 \times$ ULN
- $\geq 3.0 \times$ ULN – $< 5.0 \times$ ULN
- $\geq 5.0 \times$ ULN – $< 10.0 \times$ ULN
- $\geq 10.0 \times$ ULN – $< 20.0 \times$ ULN
- $\geq 20.0 \times$ ULN

where ULN is the upper normal limit.

Shift tables of Baseline 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 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 $\geq 3 \times$ ULN, or
- AST $\geq 3 \times$ ULN, or
- Alkaline phosphatase $\geq 1.5 \times$ ULN, or
- Total bilirubin $\geq 2 \times$ ULN.

A listing of possible Hy's Law cases will be provided which includes subjects with any elevated ALT of > 3 × ULN or AST of > 3 × ULN, Alkaline phosphatase < 2 × ULN, and associated with an increase in bilirubin ≥ 2 × ULN).

In addition, the Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot[9] will be generated to evaluate the liver safety profile of risankizumab. eDISH plot is a log/log display of correlation between peak total bilirubin (TBIL) vs. ALT, both in multiples of ULN, with horizontal and vertical lines indicating Hy's law thresholds, i.e., ALT = 3 × ULN and total bilirubin = 2 × ULN. The eDISH plot makes immediately evident subjects potentially matching Hy's law laboratory criteria, all located in the upper right quadrant of the graph. Data points in the lower right quadrant, i.e., exceeding 3 × ULN for ALT, but being below 2 × ULN for total bilirubin, suggest an increased risk for liver injury as well, if incidence is differing between active treatment and control groups, however, not to the same extent and with less specificity as compared to Hy's law.

Figure 3. An Example of eDISH Plot

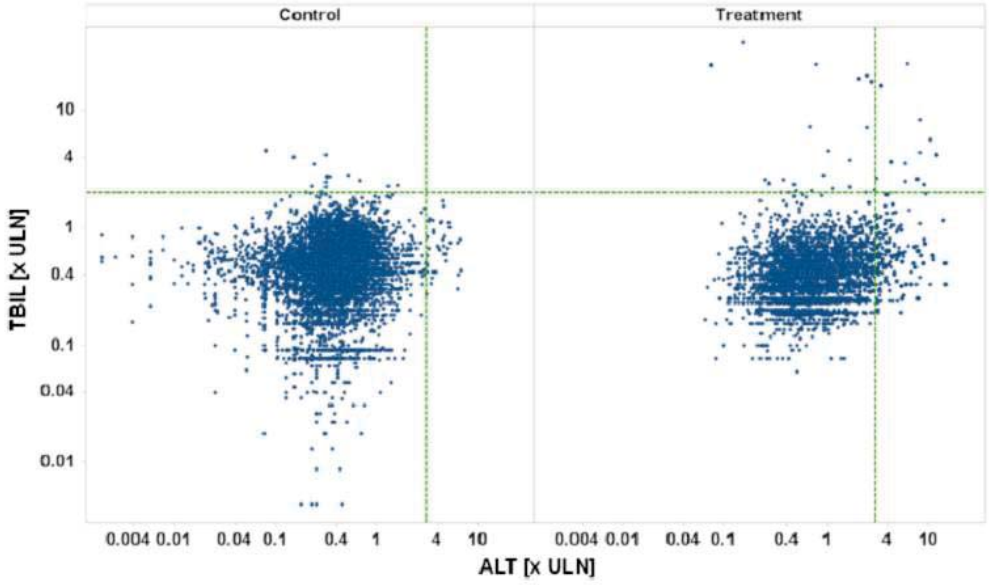


Fig. 1 eDISH plot, TBIL [× ULN] vs. ALT [× ULN] on a log/log scale, treatment by panel, pooled active versus control. *ULN* upper limit of normal, *ALT* alanine aminotransferase, *TBIL* total bilirubin

Note: Merz 2014[9]

11.4 Analysis of Vital Signs and Weight

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 24 presents the Criteria for Potentially Clinically Significant Vital Sign Findings.

Table 24. Criteria for Potentially Clinically Significant Vital Sign Findings

Vital Signs	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg or decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 180 mmHg or increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg or decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 105 mmHg or increase ≥ 15 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm or decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm or increase ≥ 15 bpm from Baseline

11.4.2 Statistical Methods

Changes from Baseline 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).

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

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.

11.4.3 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).

11.5 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

Soluble, cellular blood biomarkers and Ribonucleic Acid (RNA) will be collected in this study. Biomarker analysis will be performed by Exploratory Statistics group in a separate statistical analysis plan and results will be summarized in a separate report.

13.0 Pharmacokinetic Analysis

Descriptive statistics of risankizumab trough concentration measurements and risankizumab anti-drug antibody titer by treatment group and visit will be provided.

14.0 Summary of Changes

14.1 Summary of Changes Between the Latest Version of Protocol and the Current SAP

- The SAP includes Arm 2 + 3 vs Placebo in the evaluation and the protocol's Section 7.3.1 does not.
- The protocol Section 7.3 plans to evaluate dose response by comparing Arm 1 vs 2 and Arm 2 vs 3 at various time points. The SAP proposes to evaluate dose response by MCP-Mod approach which is more comprehensive and takes various dose response curves into consideration.

- Protocol Section 7.3.2 states an unstructured covariance structure will be used in MMRMs in secondary endpoint analysis. SAP goes on to say that other structures will be tested if it fails to converge.
- In the safety section the SAP mentions MACE, adverse events by decreasing frequency, and the use of an eDISH plot for laboratory parameters, whereas the protocol does not.
- Protocol Section 5.3.3 states that abnormal laboratory values will be also graded for intensity by using RCTC Version 2.0 criteria. In the SAP, the more comprehensive CTCAE 4.0 criteria is used instead to evaluate abnormal laboratory values.
- The SAP rephrases the missing-data handling paragraph of Protocol Section 7.5 for clarity.

15.0 References

1. Pinheiro J, Bornkamp B, Bretz F. Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. J Biopharm Stat. 2006;16(5):639-56.
2. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics. 2005;61(3):738-48.
3. 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.
4. 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.
5. 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(6890):1429-30.

6. 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):1800-4.
7. Yellen SB, Cella DF, Webster K, et al. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage. 1997;13(2):63-74.
8. Active Psoriatic Arthritis Review Charter version Final_v1.0_08-SEP-2016 for 1311.5.
9. Merz M, Lee K, Kullak-Ublick G, et al. Methodology to assess clinical liver safety data. Drug Saf. 2014;37 (Suppl 1):S33-45.
10. Greenland S, Robins M. Estimation of a common effect parameter from sparse follow-up data. Biometrics. 1985;41(1):55-68.

16.0 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.

1. 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.

2. Splinter hemorrhages: Small, longitudinal, linear, dark brown hemorrhage under the fingernail.
3. Nail bed hyperkeratosis: Thickened keratin in the nail bed.
4. Red spots in the lunula: Small pink or red macules in the lunula.

A3. PSAMRIS Joints

Figure A3.1 PsAMRIS Scoring Regions – Hand

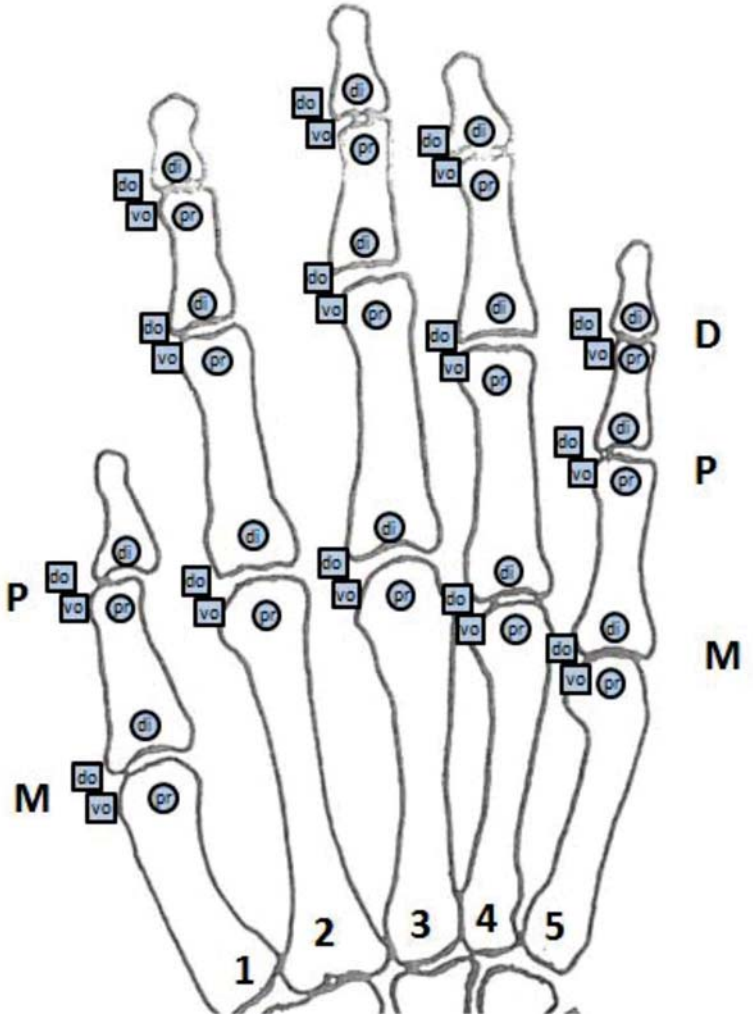
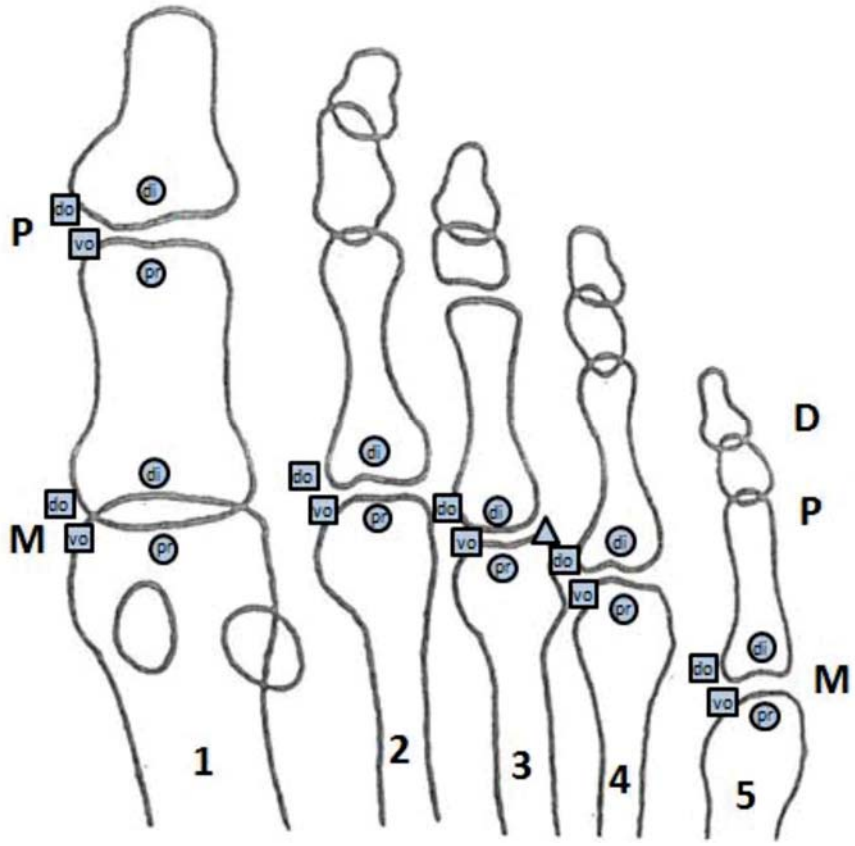


Figure A3.2 PsAMRIS Scoring Regions – Foot



A4. Important Protocol Violations

Category/ BI Code		Description	Automatic/Manual	AbbVie Criterion
A		**Entrance Criteria Not Met**		
	A1	Inclusion criterion not met	Automatic	1
	A2	Exclusion criterion met	Automatic	1
B		**Informed Consent**		
	B1	Informed consent not available/not done	Automatic and manual	1
	B2	Informed consent too late	Automatic and manual	1
C		**Trial Medication and Randomization**		
	C1	Incorrect trial medication taken	Manual	2
	C3	Non-compliance with study medication	Manual	2
D		**Concomitant Medication**		
	D1	Prohibited medication use before and during the treatment period of the trial	Manual	3
	D2	Violation of concomitant medication	Manual	3
E	E	Incorrect study procedure(s) performed.	Manual	4

Table A5. 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 A5. 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 A5. 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

Document Approval

Study 1311-5 - Statistical Analysis Plan Version 1 - 12May2017 (E3 16.1.9)

Version: 1.0 **Date:** 12-May-2017 07:47:17 PM **Company ID:** 05122017-00F9F681558114-00001-en

Signed by:	Date:	Meaning Of Signature:
[redacted]	12-May-2017 04:31:31 PM	Approver
	12-May-2017 07:47:17 PM	Approver