1.0 Title Page

Statistical Analysis Plan

Study M14-500

A Phase 4 Trial Assessing the ImPact of Residual Inflammation Detected via Imaging TEchniques, Drug Levels and Patient Characteristics on the Outcome of Dose TaperIng of Adalimumab in Clinical Remission Rheumatoid ArThritis (RA) Subjects (PREDICTRA)

Date: 19 Jul 2018

Version 1.0

2.0 Table of Contents

List of Tables

List of Figures

List of Appendices

3.0 Introduction

This statistical analysis plan (SAP) describes the statistical analyses to be completed by the Abbvie Global Statistics Department for Adalimumab (D2E7) study Protocol M14-500 dated 25 February 2016, which incorporates Amendment 1 (original Protocol: 15 May 2014).

This SAP provides further details on the statistical methods outlined in the protocol and describes analysis conventions to guide the statistical programming work. The SAP will be signed off before the study database is locked.

This is the first version of the statistical analysis plan based on Protocol M14-500. Derived Data Sets will be programmed using SAS® Version 9.2 (SAS Institute, Inc., Cary, NC 27513) or later under the UNIX operating system. Analyses will be performed using SAS® Version 9.3 (SAS Institute, Inc., Cary, NC 27513) or later under the UNIX operating system.

4.0 Study Objectives, Design and Procedures

4.1 Objectives

The primary objective of this study is to investigate the association between residual disease activity at Baseline as detected by magnetic resonance imaging (MRI) and the occurrence of flares in RA subjects randomized to an adalimumab dose tapering regimen. In addition, outcomes from the adalimumab withdrawal arm will be presented descriptively.

The secondary objectives are:

- To assess the occurrence and severity of flares and the time to flares in both tapering and withdrawal arms.
- To investigate the association between double-blind (db) Baseline subject demographic and disease characteristics and the occurrence of flares.

- To investigate the association between dbBaseline adalimumab trough concentrations and the occurrence of flares.
- To evaluate the effectiveness of rescue therapy with open-label adalimumab 40 mg every other week (eow) over 16 weeks in subjects experiencing a flare.
- To assess the change in rheumatoid arthritis MRI scoring system (RAMRIS) scores from Baseline to Final visit in the taper, withdrawal and Open-Label Rescue Period.
- To investigate the MRI-flare associations in sub-groups of subjects who meet additional clinical remission criteria at dbBaseline including simplified disease activity index (SDAI) ≤ 3.3, clinical diseases activity index (CDAI) ≤ 2.8 and ACR/EULAR 2011 boolean-based remission, as well as to describe the course of disease and patient reported outcome (PRO) measures in the taper, withdrawal and Open-Label Rescue Period overall and per dbBaseline subgroup.
- To assess the rate of anti-adalimumab antibodies (AAA) positive subjects in the tapering and withdrawal arms.

The study has also the following exploratory objectives:

- In the subgroup of subjects with a dbBaseline Ultrasound (US) assessment:
  - To investigate the association between dbBaseline ultrasound scores and the occurrence of RA flares.
  - To investigate the association between the dbBaseline ultrasound scores and Baseline MRI RAMRIS scores.
  - To describe the change in the ultrasound scores from dbBaseline to the time of RA flare in the tapering arm.
- To investigate the association between biomarker values at dbBaseline (or their change over time) and the occurrence of RA flares.

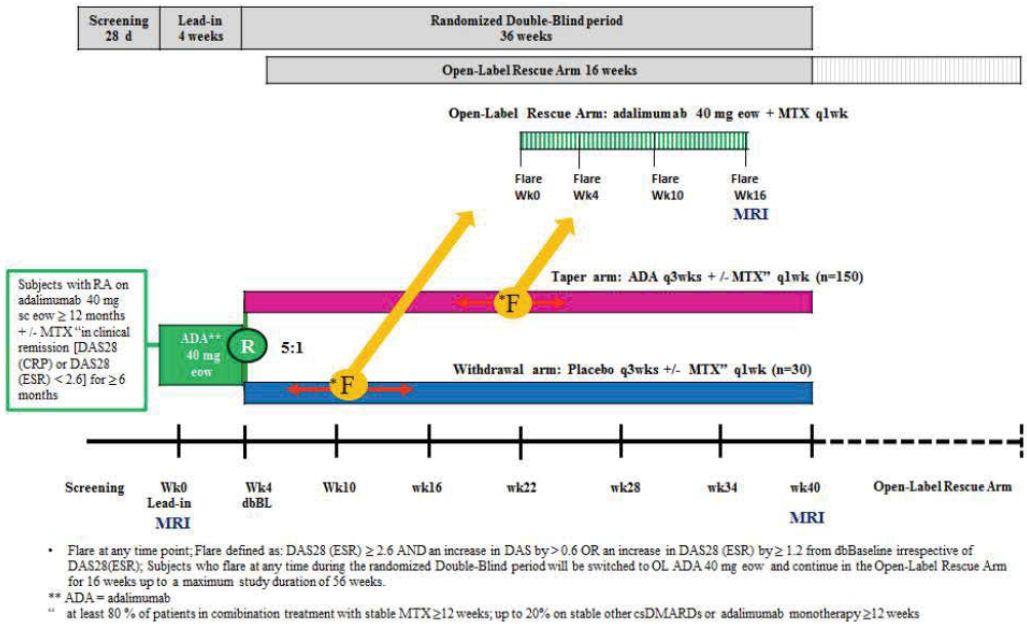
4.2 Study Design

This is a Phase 4, multicenter, randomized, double-blind, parallel-group study in subjects with RA who are in stable clinical remission defined as DAS28(ESR) or

DAS28(CRP) < 2.6 for at least 6 months prior to the Screening Visit. The study duration will include a Screening period of up to 28 days, a 4-week Lead-In Period (can be extended up to 2 more weeks if needed for procedural compliance) with open-label 40 mg adalimumab eow and a 36-week Double-Blind Period with 40 mg adalimumab/placebo q3wks; subjects who experience a flare (defined as an increase from dbBaseline in DAS28 [ESR] of > 0.6 AND a DAS28 [ESR] > 2.6, OR an increase in DAS28(ESR) of ≥ 1.2 irrespective of the resulting DAS28 [ESR]) at any time will enter a Rescue Arm and will be followed for 16 weeks.

The study schematic is presented in Figure 1.

Figure 1. Study Design Schematic



The study activities will start with a Screening Period of up to 28 days to confirm inclusion/exclusion criteria including a DAS28(ESR) assessment of < 2.6.

Subjects who fulfill all Screening criteria will enter the study with a 4-week Lead-In OL Period: subjects will receive adalimumab 40 mg sc eow starting at Week 0 of the Lead-In Period. Week 0 will occur 2 weeks after the subject's last dose of commercial Humira® (± 2 days) to keep in line with eow dosing; the second injection will be given 14 days thereafter (± 2 days).

A confirmed adequate MRI from the Lead-In Period is required before the dbBaseline visit. At the end of the Lead-In Period, subjects will be assessed again for DAS28(ESR) at the dbBaseline Week 4 visit. Subjects with a confirmed DAS28(ESR) of < 2.6 at the dbBaseline visit, and as such fulfilling confirmed DAS28(ESR) remission at 2 time points (Day 1 of Lead-In and 4 weeks later at the dbBaseline visit) will be randomized in a 5:1 ratio to one of two double-blind arms and followed for an additional 36 weeks in the Double-Blind Period:

1. A reduced frequency of adalimumab 40 mg sc every 3 weeks (q3wks): tapering arm, or
2. Placebo sc q3wks: withdrawal arm.

The randomized Double-Blind period will begin at the dbBaseline Visit (Week 4) and will continue with visits every 6 weeks until the Week 40 Visit. The Week 40 visit will be considered the Final Visit. During the Double-Blind Period, subjects will be evaluated for efficacy, including detection of flares, PROs, safety and laboratory assessments at scheduled visits on: Weeks 4, 10, 16, 22, 28, 34 and 40.

Subjects with a confirmed flare at any time point will undergo Flare Week 0 visit procedures and will be immediately switched to an Open-Label Rescue Period starting with the first OL injection of adalimumab 40 mg sc at the visit. In the Open-Label Rescue Period, subjects will be further evaluated at Flare Weeks 4, 10 and 16 for efficacy, PROs, safety and laboratory assessments.

4.3 Sample Size

The planned total number of subjects enrolled into the Lead-In Period is approximately 150.

A recently published study (DOSERA) showed that 56% (15/27 patients) of patients who were previously in low disease activity had treatment failure within 48 weeks after reducing the dose of etanercept; the median time to failure was 36 weeks (95% CI:15.6-NE).[1] The STRASS study, also published in 2015, showed in a similar RA patient population, that progressive tapering of TNFi therapy (adalimumab or etanercept) was associated with flare in 76.6% (49/64) of patients throughout the study duration; importantly, 28.8% of patients flared after the first step of tapering (adalimumab 40 mg q3 weeks or etanercept 50 mg q10 days) and the median time to flare was 9 months.[2] In another tapering study (DRESS) a cumulative incidence of short-lived flares in a progressive tapering of TNFi of 55% at 9 months (and 73% at 18 months) was observed. Among these studies, the DRESS study had the higher sample size with 121 patients in the tapering arm but only 43% successfully tapered the TNFi. None of these studies was designed or powered to evaluate predictors of flare neither found definite predictors upon further statistical analysis. A proof of concept study conducted in 44 RA patients in clinical remission receiving treatment with a bDMARD has shown that residual synovial inflammation determined by comprehensive ultrasound assessment predicted relapse within a short term after discontinuation of the treatment.[3]

These data provided the rationale for assuming a conservative flare rate of 30% and for our sample size calculation.

To estimate the odds ratio for the occurrence of flare with baseline MRI score based on historical MRI data, an effective sample size of 100 subjects in the dose tapering group will ensure a precision for the estimation with the width of 95% CI no more than 0.19 for an OR of 1.1, and no more than 0.26 for an OR of 1.2. Such sample size also ensures a precision for the estimation of a point-biserial correlation coefficient[4] ρ with the width of 95% CI of ρ no more than 0.39 for a mild correlation coefficient of 0.26, no more than

0.33 for a moderate correlation coefficient of 0.43. The confidence intervals are determined by using bootstrap methods (percentile bootstrap, N = 10,000 bootstrap samples).

Based on a higher than assumed preliminary flare rate (40%) in Study M14-500, a re-evaluation of the sample size calculation was conducted. Assuming 40% flare rate, a sample size of 100 subjects in the tapering arm ensures a precision for the estimation with the width of 95% CI no more than 0.19 for an OR of 1.1, and no more than 0.26 for an OR of 1.2. Such sample size also ensures a precision for the estimation of a correlation coefficient ρ with the width of 95% CI of ρ no more than 0.36 for a correlation coefficient of 0.25, no more than 0.31 for a correlation coefficient of 0.41.

Precision levels under various assumptions (different flare rates and True ORs) are presented in Table 1 below.

Under a 5:1 randomization ratio (dose tapering versus withdrawal), a total of 120 subjects will be randomized. Accounting for a 10 - 20% discontinuation rate during the Lead-In Period, approximately 150 subjects will need to be enrolled into the Lead-In Period.

Table 1. Sample Size Assumptions

True Odds Ratio	Flare Rate	Sample Size (Dose Tapering Arm)	OR from Logistics Regression (95% CI): CI Width	Correlation (95% CI): CI Width
1.1	30%	100	1.1 (1.02, 1.21): 0.19	0.26 (0.06, 0.45): 0.39
1.2	30%	100	1.2 (1.11, 1.36): 0.26	0.43 (0.25, 0.58): 0.33
1.1	40%	100	1.1 (1.03, 1.21): 0.19	0.25 (0.07, 0.43): 0.36
1.2	40%	100	1.2 (1.10, 1.36): 0.26	0.41 (0.25, 0.55): 0.31

4.4 Interim Analysis for Double-Blind Baseline Characteristics

There are few data describing patient disease characteristics including markers of residual inflammation such as sensitive imaging assessments and potential biomarkers as well as drug levels in RA patients who are in sustained clinical remission on a stable treatment with a TNFi and MTX. An interim analysis was conducted to describe the dbBaseline characteristics of the RA population providing overall observations only. It was performed after the dbBaseline assessments of the entire study population have been completed. The study team remained blinded to individual data. The details of the interim analyses/evaluation were described separately in the interim analysis SAP.

5.0 Analysis Populations, Stratification Variables, and Subgroup Definition

5.1 Definition for Analysis Populations

All Lead-In-Treated Subjects

The All Lead-In-Treated Subject population will comprise all subjects who are enrolled into the study and received at least one dose of study medication in the Lead-In Period.

All Double-Blind-Treated Subjects

The All Double-Blind-Treated Subject population will comprise all subjects who have received at least one dose of double-blind study medication.

Unless otherwise specified, efficacy and other assessment will be primarily analyzed for subjects who are randomized to the adalimumab 40 mg q3 weeks arm (tapering arm) and have received at least one dose of double-blind study medication. Descriptive statistics will be provided for the subjects who are randomized to the placebo arm (withdrawal arm) and have received at least one dose of study medication.

All Open-Label-Rescue-Treated Subjects

The All Open-Label-Rescue-Treated Subject population will comprise all subjects who have received at least one dose of study medication in the Open-Label Rescue Period.

5.2 Variables Used for Stratification

No stratification was performed for this study.

5.3 Subgroup Definition

The study will perform exploratory analysis based on the subgroup of subjects with a dbBaseline Ultrasound (US) assessment.

The MRI-flare associations will be also investigated in subgroups of subjects who meet additional clinical remission criteria at dbBaseline including simplified disease activity index (SDAI) ≤ 3.3, clinical diseases activity index (CDAI) ≤ 2.8 and ACR/EULAR 2011 Boolean-based remission.

6.0 Analysis Conventions

6.1 Definition of Baseline

Baseline

The Baseline value will be determined by the last non-missing measurement recorded on or before the date of the first dose of study drug. If dosing times and observation times are collected, the observation time also needs to be before the dosing time for it to be Baseline. When dosing times or observation times are not collected, it is assumed that observations obtained on the day of the first dose of study drug (Day 1) are prior to the first dose of study drug and therefore Baseline values. If multiple measurements are recorded on the same day and measurement times are collected, the measurement recorded at the latest time prior to dosing will be used as Baseline (assuming measurement times are prior to doing time). If multiple measurements are recorded on the

same day and measurement times are not collected, the average of values will be used as Baseline.

dbBaseline

The dbBaseline value will be determined by the last non-missing measurement recorded on or before the first dose of study drug is received during the Double-Blind (DB) period.

If dosing times and observation times are collected, the observation time also needs to be before the dosing time for it to be dbBaseline. When dosing times or observation times are not collected, it is assumed that observations obtained on the day of the first dose of study drug in DB period are prior to the first dose of study drug in DB period and therefore dbBaseline values. If multiple measurements are recorded on the same day and measurement times are collected, the measurement recorded at the latest time prior to dosing will be used as dbBaseline (assuming measurement times are prior to doing time). If multiple measurements are recorded on the same day and measurement times are not collected, the average of values will be used as dbBaseline.

6.2 Definition of Final Observation

The final observation for each subject will be defined as the last non-missing value excluding information collected after 70 days of the last dose of study drug. Any data collected beyond 70 days after the last dose of study drug will not be used for analysis but will be displayed in the data listings.

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

Rx Days relative to the first dose of study drug are calculated for each time point related to study visits. They are defined as the number of days between the day of the first dose of study drug and the specific time point. 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 (i.e., there is no Rx Day 0). Specifically,

Rx Day = Visit Date – (first dose of study drug date) + 1, if Visit Date ≥ first dose of study drug date

Rx Day = Visit Date – (first dose of study drug date), if Visit Date < first dose of study drug date

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. Rx Days are used to map actual study visits to the protocol specified study visits as described in Section 6.4.

Rx Days in a particular period (Lead-In, Double-Blind and Open-Label Rescue Period) are defined as number of days between the day of the first dose of study drug in the respective period and the specific time point.

6.4 Definition of Analysis Windows

Since subjects do not always adhere to the study visit schedule, the following rules will be applied to assign actual visits to protocol-specified visits including early termination visits. For each study visit mentioned in the protocol, a nominal or target day will be selected to represent the corresponding visit along with a window around the target day. Windows will be selected in a non-overlapping fashion so that a date collected on the case report form (CRF) does not correspond to multiple visit windows. Moreover, windows will not discard any post-baseline measurement recorded on the CRF. If a subject had 2 or more actual visits in one visit window, the visit closest to the target will be used as the study visit for that window. If 2 visits are equidistant from the target day, then the later visit will be used for reporting.

The visit windows and the target/nominal day for each study visit are shown in Table 2 through Table 4. For some parameters, data may not be collected at all study visits; separate visit windows will be used for such parameters as listed underneath (Table 5, Table 6, Table 7, Table 8, Table 9, Table 10, Table 11, Table 12, Table 13, Table 14, Table 15, Table 16).

Table 2. Visit windows for Analysis of all endpoints during Lead-In Period

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 0 (Baseline)	–999	1	1

a. Visit windows calculated from first dose of study drug during Lead-In Period.

Table 3. Visit windows for Analysis of all endpoints during Double-Blind Period excluding FACIT-fatigue, SF-36, TSQM, WPAI, chemistry/hematology, RF, ACPA, Urinanalysis, MRI, Ultrasound, ANA, Anti-dsDNA

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 4 (dbBaseline)	–999	1	1
Week 10	2	43	64
Week 16	65	85	106
Week 22	107	127	148
Week 28	149	169	190
Week 34	191	211	232
Week 40	233	253	273

a. Visit windows calculated from first dose of study drug during Double-Blind Period.

Table 4. Visit windows for Analysis of all endpoints during Open-Label Rescue period[b] excluding WPAI, chemistry/hematology, RF, ACPA, Urinanalysis, MRI, Ultrasound, Flare Severity, ANA, Anti-dsDNA, RAPID-3 at home assessment

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Flare Week 0	–999[c]	1	1
Flare Week 4	2	29	50
Flare Week 10	51	71	92
Flare Week 16	93	113	133

a. Visit windows calculated from first dose of study drug during Open-Label Rescue Period.

b. If there are no multiple measurements, the observation of the last visit in Double-Blind Period should be the same as that of Flare Week 0 in Open-Label Rescue Period. If there are multiple measurements, the observation closest to the nominal day of the last visit in Double-Blind Period and the observation closest to the nominal day of Flare Week 0 in Open-Label Rescue Period should be used.

c. –999 or last dose in Double Blind Period.

Table 5. Visit windows for Analysis of FACIT-fatigue, TSQM, Urinanalysis during Double-Blind Period

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 4 (dbBaseline)	–999	1	1
Week 16	2	85	127
Week 28	128	169	211
Week 40	212	253	294

a. Visit windows calculated from first dose of study drug during Double-Blind Period.

Table 6. Visit windows for Analysis of WPAI, SF-36 during Double-Blind Period

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 4 (dbBaseline)	–999	1	1
Week 28	2	169	211
Week 40	212	253	294

a. Visit windows calculated from first dose of study drug during Double-Blind Period.

Table 7. Visit windows for Analysis of chemistry/hematology during Double-Blind Period

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 10	1	43	64
Week 16	65	85	127
Week 28	128	169	211
Week 40	212	253	294

a. Visit windows calculated from first dose of study drug during Double-Blind Period.

Table 8. Visit windows for Analysis of MRI, ANA, Anti-dsDNA during Double-Blind Period

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 4 (dbBaseline)	–999	1	1
Week 40	2	253	504

a. Visit windows calculated from first dose of study drug during Double-Blind Period.

Table 9. Visit windows for Analysis of ANA, Anti-dsDNA during Double-Blind Period

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 40	1	253	506

a. Visit windows calculated from first dose of study drug during Double-Blind Period.

Table 10. Visit windows for Analysis of Ultrasound during Double-Blind Period

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 4 (dbBaseline)	–999	1	1

a. Visit windows calculated from first dose of study drug during Double-Blind Period.

Table 11. Visit windows for Analysis of RF, ACPA during Double-Blind Period

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Week 4 (dbBaseline)	–999	1	1
Week 40	2	253	504

a. Visit windows calculated from first dose of study drug during Double-Blind Period.

Table 12. Visit windows for Analysis of WPAI, chemistry/hematology during Open-Label Rescue Period[b]

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Flare Week 0	–999[c]	1	1
Flare Week 10	2	71	92
Flare Week 16	93	113	133

a. Visit windows calculated from first dose of study drug during Open-Label Rescue Period.

b. If there are no multiple measurements, the observation of the last visit in Double-Blind Period should be the same as that of Flare Week 0 in Open-Label Rescue Period. If there are multiple measurements, the observation closest to the nominal day of the last visit in Double-Blind Period and the observation closest to the nominal day of Flare Week 0 in Open-Label Rescue Period should be used.

c. –999 or last dose in Double Blind Period.

Table 13. Visit windows for Analysis of RF, ACPA, MRI, ANA, Anti-dsDNA during Open-Label Rescue Period[b]

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Flare Week 16	1	113	226

a. Visit windows calculated from first dose of study drug during Open-Label Rescue Period.

b. If there are no multiple measurements, the observation of the last visit in Double-Blind Period should be the same as that of Flare Week 0 in Open-Label Rescue Period. If there are multiple measurements, the observation closest to the nominal day of the last visit in Double-Blind Period and the observation closest to the nominal day of Flare Week 0 in Open-Label Rescue Period should be used.

Table 14. Visit windows for Analysis of Urinanalysis during Open-Label Rescue Period[b]

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Flare Week 0	–999[c]	1	1
Flare Week 16	2	113	224

a. Visit windows calculated from first dose of study drug during Open-Label Rescue Period.

b. If there are no multiple measurements, the observation of the last visit in Double-Blind Period should be the same as that of Flare Week 0 in Open-Label Rescue Period. If there are multiple measurements, the observation closest to the nominal day of the last visit in Double-Blind Period and the observation closest to the nominal day of Flare Week 0 in Open-Label Rescue Period should be used.

c. –999 or last dose in Double Blind Period.

Table 15. Visit windows for Analysis of Ultrasound, Flare Severity during Open-Label Rescue Period[b]

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Flare Week 0	–999[c]	1	1

a. Visit windows calculated from first dose of study drug during Open-Label Rescue Period.

b. If there are no multiple measurements, the observation of the last visit in Double-Blind Period should be the same as that of Flare Week 0 in Open-Label Rescue Period. If there are multiple measurements, the observation closest to the nominal day of the last visit in Double-Blind Period and the observation closest to the nominal day of Flare Week 0 in Open-Label Rescue Period should be used.

c. –999 or last dose in Double Blind Period.

Since RAPID-3 at home assessment will be completed between the flare week visits for this questionnaire there are additional visit windows required.

Table 16. Visit windows for Analysis of RAPID-3 at home assessment during Open-Label Rescue Period[b]

Protocol Specified Visit Week	Lower Bound[a]	Nominal Day[a]	Upper Bound[a]
Flare Week 1	2	8	11
Flare Week 2	12	15	18
Flare Week 3	19	22	29
Flare Week 5	30	36	39
Flare Week 6	40	43	46
Flare Week 7	47	50	53
Flare Week 8	54	57	60
Flare Week 9	61	64	71
Flare Week 11	72	78	81
Flare Week 12	82	85	88
Flare Week 13	89	92	95
Flare Week 14	96	99	101
Flare Week 15	102	106	110[c]

a. Visit windows calculated from first dose of study drug during Open-Label Rescue Period.

b. If there are no multiple measurements, the observation of the last visit in Double-Blind Period should be the same as that of Flare Week 0 in Open-Label Rescue Period. If there are multiple measurements, the observation closest to the nominal day of the last visit in Double-Blind Period and the observation closest to the nominal day of Flare Week 0 in Open-Label Rescue Period should be used.

c. Maximum (110, Day prior to Flare Week 16 site visit).

6.5 Missing Data Imputation

Missing data will be imputed for efficacy variables only. All data will be summarized as "observed" or "imputed".

Efficacy data will be imputed using the Last Observation Carried Forward (LOCF) approach where only post-baseline values will be carried forward up to each time point of evaluation for subjects who had missing clinical assessments. For Double-Blind Period, missing values are only imputed up to the time when the subject enters Open-Label Rescue Period.

Categorical outcome variables: For proportions (flare, clinical remission, regain disease control, low disease activity index, subjects with HAQ-DI normal) the continuous variable will be imputed before categorical outcome variable is calculated, e.g., the DAS28(ESR) score will be imputed as defined in Section 10.2.7 before calculation of the flare outcome variable.

For subjects who falsely entered the Open-Label Rescue Period: These subjects are treated as if they had discontinued from the study at the time when they falsely entered the Open-Label Rescue Period before reaching a flare, i.e., have missing data afterwards. Their data collected during the Open-Label Rescue Period will thus not be included in analyses.

For subjects who flared according to DAS28(ESR) criteria but falsely remained in the Double-Blind Period: These subjects are treated as if they flared, i.e., all Double-Blind information collected after the flare will be discarded and they would have no data in the Open-Label Rescue Period.

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

7.1 Demographic and Baseline Characteristics

The analysis populations for demographics and dbBaseline characteristics are the Lead-In-Treated Subject population and the Double-Blind-Treated Subject population.

The following variables will be summarized by treatment arm and overall at Week 0 (Baseline) and Week 4 (dbBaseline), whenever there is an available observation.

Demographics

- Gender [male, female]
- Age [years]
- Race [White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander]

- Ethnicity [Hispanic or Latino, non-Hispanic/Latino]
- Body weight [kg]
- Height [cm]

Clinical Patient Characteristics

- Rheumatoid Factor (RF) level or serum concentration
- Rheumatoid Factor (RF) status: Positive or Negative level or serum concentration
- Anti-cyclic citrullinated peptide antibody (Anti-CCP, also called ACPA) level or serum concentration
- Anti-cyclic citrullinated peptide antibody (Anti-CCP, also called ACPA) status: Positive or Negative level or serum concentration
- Disease duration in years
- Duration of adalimumab therapy in years defined as the duration between date of first adalimumab therapy as collected in the Medical History Case Report Form and date of first dose of study drug
- Duration of remission in years after initiating adalimumab defined as the duration between date of first remission as collected in the Medical History Case Report Form and date of first dose of study drug
- Previous treatment with conventional synthetic Disease Modifying Anti-rheumatic Drugs (csDMARDs) or biologic Disease Modifying Anti-rheumatic Drugs (bDMARDs, excluding adalimumab) or both: yes or no

Efficacy Measurements at dbBaseline

- Tender joint count (TJC68) defined as the number of tender joints out of 68 assessed joints
- Swollen joint count (SJC66) defined as the number of swollen joints out of 66 assessed joints
- Tender joint count (TJC28) defined as the number of tender joints out of 28 assessed joints used for DAS28(ESR) calculation

- Swollen joint count (SJC28) defined as the number of swollen joints out of 28 assessed joints used for DAS28(ESR) calculation
- Patient's global assessment of disease activity (mm on a 100-mm visual analogue scale [VAS])
- Physician's global assessment of disease activity (mm on a 100-mm VAS)
- Patient's global assessment of RA pain (mm on a 100-mm VAS)
- ESR (mm/hr)
- C-reactive protein (CRP) (mg/L)
- DAS28(ESR)
- DAS28 remission: DAS28(ESR) < 2.6
- Clinical Disease Activity Index (CDAI)
- CDAI remission: defined as CDAI ≤ 2.8
- Simplified Disease Activity Index (SDAI)
- SDAI remission: defined as SDAI ≤ 3.3
- Hand and wrist rheumatoid arthritis MRI scoring system (RAMRIS) scores
  - Synovitis
  - Bone marrow edema (BME)/osteitis
  - Composite of synovitis and BME
  - Tenosynovitis
  - Overall inflammation score as composite of synovitis, BME and tenosynovitis score
  - Erosions
- Ultrasound assessment using Gray Scale Ultrasonography (GSUS) and Power Doppler Ultrasonography (PDUS) (described in Appendix A)

Patient Reported Outcomes at dbBaseline

- Health Assessment Questionnaire – Disability Index (HAQ-DI)
- Normal HAQ-DI: HAQ-DI ≤ 0.5
- Functional Assessment of Chronic Illness Therapy-fatigue (FACIT-fatigue)
- Routine Assessment of Patient Index Data (RAPID-3)

- Morning stiffness assessment (duration [min], severity [0 - 10])
- Sleep disturbance assessment
- Treatment Satisfaction Questionnaire for Medication (TSQM)
- Work Productivity and Activity Impairment (WPAI)
- Short Form-36 Health Survey Questionnaire (SF-36)

Vital Signs at dbBaseline

- Systolic and diastolic blood pressure [mmHg]
- Pulse [bpm]
- Respiratory rate [rpm]
- Temperature [°C]

12-Lead Electrocardiogram (ECG)

- Overall ECG assessment [Normal, Abnormal – not clinically significant, Abnormal – Clinically significant, Unable to evaluate]
- Abnormal ECG details [Rhythm, Conduction, Axis, Ventricular hypertrophy, Q waves, ST segment, T waves, Abnormal U waves, Corrected QT interval using Bazett Correction (QTcB), and Other]

Chest X-ray, Purified Protein Derivative (PPD), Tuberculosis (TB) Prophylaxis

- Chest x-ray
  - Normal, Abnormal
  - Calcified granulomas [Absent, Present]
  - Pleural scarring [Absent, Present]
  - Pleural thickening [Absent, Present]
- PPD Skin Test [Positive, Negative, Missing] (Note that a positive or negative PPD skin test will be determined by the PPD mm induration result on the CRF. An induration of less than 5 mm is considered negative.)

- QuantiFERON-TB Gold or equivalent test (positive, negative, indeterminate): If QuantiFERON-TB Gold or equivalent test is determined repeatedly the result closest and before the date of first dose of study drug should be used.
- TB test (positive, negative, indeterminate) – in the event that both a PPD test and QuantiFERON-TB gold or equivalent test are performed, the result of the QuantiFERON-TB gold or equivalent test will supersede the result of the PPD test
- Enrolled under TB prophylaxis [Yes, No]
  - If enrolled [INH, Other]

Tobacco/Nicotine and Alcohol Use

- Tobacco/Nicotine use [user, ex-user, non-user, unknown]
- Alcohol use [drinker, ex-drinker, non-drinker, unknown]

7.1.1 Analysis of Demographic Data and Baseline Characteristics

Demographic and baseline characteristics will be summarized for Lead-In-Treated Subjects and All Double-Blind-Treated Subjects by treatment arm and overall if appropriate. The number of observations, mean, standard deviation, minimum, first quartile, median, third quartile and maximum will be summarized for continuous variables. Categorical or discrete variables will be summarized via counts and percentages. For categorical data, the number of subjects with missing information will also be presented. No statistical comparison will be performed on demographic data and baseline characteristics.

7.2 Medical and Surgical History

Medical history data will be summarized and presented by count and percentage for All Double-Blind-Treated Subjects broken down by body systems and diagnoses as captured on the CRF.

The body systems will be presented in alphabetical order and the diagnoses will be presented in alphabetical order within each body system. The number and percentage of subjects with a particular diagnosis will be summarized for each treatment group and overall. Subjects reporting more than 1 condition/diagnosis within a body system will be counted only once for that body system. No statistical comparison will be performed on medical history.

7.3 Vaccination History

The number and percentage of subjects that had taken Bacille Calmette-Guerin (BCG) vaccination will be summarized for each treatment group and overall for the All Double-Blind-Treated Subjects analysis population. No statistical comparison will be performed on vaccination history.

7.4 Prior Treatment and Concomitant Medications

Prior and concomitant medications will be summarized by each treatment group as well as overall for All Double-Blind-Treated Subjects. 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, but not after the last dose of study drug + 14 days.

The number and percentage of subjects who received a prior or concomitant medication will be tabulated separately by the generic name assigned by the most current version of the World Health Organization (WHO) Drug Dictionary.

7.4.1 Reporting Special Medications

The following classes of medications will be summarized for both prior (including ongoing at baseline) and concomitant treatment by Generic and/or brand name:

- Disease-modifying antirheumatic drugs (DMARDs) - conventional synthetic and biologic (excluding adalimumab for prior medication)

- Nonsteroidal anti-inflammatory drugs (NSAIDs)
- Systemic corticosteroids
- Other medications (excluding DMARDs, NSAIDs, and systemic corticosteroids)

Number of different DMARDs (excluding adalimumab)/NSAIDs/Steroids taken per subject prior to or at dbBaseline will be reported.

Average dose of Concomitant medication for MTX and other allowed csDMARDs, and oral Corticosteroids at the dbBaseline will be also summarized by treatment and overall.

7.5 Protocol Deviation

Protocol deviations are categorized as follows:

- Those who entered the study even though they did not satisfy the entry criteria
- Those who developed withdrawal criteria during the study and were not withdrawn
- Those who received an excluded or prohibited concomitant medication.
- Those who received wrong treatment or incorrect dose

7.6 Screening failures

Number of screening failures and reasons for screening failure will be summarized among all screened subjects in a subject screening status table.

8.0 Patient Disposition

The number of subjects will be tabulated by investigator site and overall for the following sets: all subjects in the Lead-In Period, all subject treated in the Lead-In Period, all randomized subjects, all subjects treated in the Double-Blind Period, all subjects in the Open-Label Rescue Period (excluding subjects who falsely entered the Open-Label Rescue Period), all subjects who falsely entered the Open-Label Rescue Period, all

subjects treated in the Open-Label Rescue Period (excluding subjects who falsely entered the Open-Label Rescue Period), all in Open-Label Rescue Period treated subjects who falsely entered the Open-Label Rescue Period, subjects who completed study, and subjects who discontinued, subjects who discontinued during the Double-Blind Period, subjects who discontinued during the Open-Label Rescue Period for each treatment group and overall, as appropriate.

Reasons and primary reasons for discontinuation of study drug will be summarized as recorded on the CRF by the following categories: adverse event (AE), withdrew consent, lost to follow-up, and other. Subjects may have more than one reason for discontinuing study drug, but they will be counted only once for the total number of discontinuations.

A listing of subjects who falsely entered the Open-Label Rescue Period even if they had no flare, as well as of subjects who flared according to DAS28(ESR) criteria but falsely remained in the Double-Blind Period will be provided. This listing will contain subject numbers and the individual DAS28(ESR) values, DAS28(ESR) change to dbBaseline and dichotomous variable indicating a flare over time.

9.0 Study Drug Exposure and Compliance

Study drug exposure will be summarized per treatment group and overall as follows:

- Exposure in the Lead-In Period
  - Duration of treatment [days]: for subjects who did not enter the Double-Blind Period calculated as time between first injection in the Lead-In-Period and last injection in the Lead-In Period + 14 days; for subjects who did enter Double-Blind Period, calculated as time between first injection in the Lead-In Period and first injection in the Double-Bind Period.
  - Number of study drug injections received in Lead-In Period [categorical and quantitative].
- Exposure in the Double-Blind Period
  - Duration of treatment [days]: for subjects who did not enter the Open-Label Rescue Period calculated as time between first injection in the

Double-Blind Period and last injection in the Double-Blind Period + 14 days; for subjects who did enter the Open-Label Rescue Period, calculated as time between first injection in the Double-Blind Period and first injection in the Open-Label Rescue Period.

- Number of study drug injections received in the Double-Blind Period [categorical and quantitative].

- Exposure in the Open-Label Rescue Period
  - Duration of treatment [days] calculated as time between first injection in the Open-Label Rescue Period Rescue and last injection in the Open-Label Rescue Period + 14 days.
  - Number of study drug injections received in the Open-Label Rescue Period [categorical and quantitative].
- Total Exposure to any adalimumab
  - Duration of treatment [days] calculated as time between first adalimumab injection and last adalimumab injection + 14 days for subjects received adalimumab during double-blind period. If subject received placebo during double-blind period, deduct the double-blind placebo exposure
  - Number of study drug injections [categorical and quantitative].

These will be presented by the number of subjects, mean, standard deviation, minimum, first quartile, median, third quartile and maximum. In addition, the total number of patient years per period and total will be presented.

In addition, the number and percentage of subjects exposed to study drug will be summarized for the following exclusive duration intervals:

<u>Exposure in the Lead-In Period</u>

- ≥ 1 day
- ≥ 15 days
- ≥ 29 day
- ≥ 43 days
- ≥ 57 days

Exposure in the Double-Blind Period

- ≥ 1 day
- ≥ 15 days
- ≥ 29 day
- ≥ 43 days
- ≥ 57 days
- ≥ 71 days
- ≥ 85 days
- ≥ 99 days
- ≥ 113 days
- ≥ 127 days
- ≥ 141 days
- ≥ 155 days
- ≥ 169 days
- ≥ 183 days
- ≥ 197 days
- ≥ 211 days
- ≥ 225 days
- ≥ 239 days
- ≥ 253 days
- ≥ 267 days

Exposure in the Open-Label Rescue Period

- ≥ 1 day
- ≥ 15 days
- ≥ 29 day
- ≥ 43 days
- ≥ 57 days
- ≥ 71 days

- ≥ 85 days
- ≥ 99 days
- ≥ 113 days
- ≥ 127 days

Exposure to any Adalimumab

- ≥ 1 day
- ≥ 15 days
- ≥ 29 day
- ≥ 43 days
- ≥ 57 days
- ≥ 71 days
- ≥ 85 days
- ≥ 99 days
- ≥ 113 days
- ≥ 127 days
- ≥ 141 days
- ≥ 155 days
- ≥ 169 days
- ≥ 183 days
- ≥ 197 days
- ≥ 211 days
- ≥ 225 days
- ≥ 239 days
- ≥ 253 days
- ≥ 267 days
- ≥ 281 days
- ≥ 309 days
- ≥ 337 days

- ≥ 365 days
- ≥ 393 days
- ≥ 421 days

Compliance

Compliance in the Lead-In Period

- Compliance with injections [%] calculated as (number of injections received in the Lead-In Period/number of injections planned in the Lead-In Period, i.e., from first injection in the Lead-In Period until last injection in the Lead-In Period)*100.

Compliance in the Double-Blind Period

- Compliance with injections [%] calculated as (number of injections received in the Double-Blind Period/number of injections planned in the Double-Blind Period, i.e., from the first injection in the Double-Blind Period until last injection in the Double-Blind Period)*100.

Compliance in the Open-Label Rescue Period

- Compliance with injections [%] calculated as 100*(number of injections received in the Open-Label Rescue Period/number of injections planned in the Open-Label Rescue Period i.e., from first injection in the Open-Label Rescue Period until last injection).

Total Compliance

- Compliance with injections [%] calculated as 100*(number of injections received/number of injections planned).

10.0 Efficacy Analysis

10.1 General Considerations

Unless otherwise specified, efficacy and other assessment will be primarily analyzed for subjects who are randomized to the adalimumab 40 mg q3 weeks arm (tapering arm) and have received at least one dose of study medication. Descriptive statistics will be provided for the subjects who are randomized to the placebo arm (withdrawal arm) and have received at least one dose of study medication.

All statistical inference (see Section 10.3 and Section 10.4) will be based on a 2-sided alpha-level of 0.05, unless otherwise stated. No multiplicity adjustment will be conducted. Descriptive statistics will be provided. These include the number of observations, mean, standard deviation, minimum, first quartile, median, third quartile and maximum for continuous variables; and counts and percentages for discrete variables.

For time to event analyses Kaplan-Meier curve and descriptive information including number of subjects with events, number of subjects censored, quartiles, median will be presented.

The last available pre-treatment value recorded on or before first dose of study drug received in the Double-Blind Period will be considered as baseline. This includes all efficacy and safety assessments at dbBaseline (Week 4). All changes and percent changes will be calculated based on the dbBaseline unless otherwise stated.

For details on handling missing data imputation refer to Section 6.5.

For subjects who falsely entered the Open-Label Rescue Period: These subjects are treated as if they had discontinued from the study at the time when they falsely entered the Open-Label Rescue Period before reaching a flare, i.e., have missing data afterwards. Their data collected during the Open-Label Rescue Period will thus not be included in analyses.

For subjects who flared according to DAS28(ESR) criteria but falsely remained in the Double-Blind Period: These subjects are treated as if they flared, i.e., all Double-Blind information collected after the flare will be discarded and they would have no data in the Open-Label Rescue Period.

10.2 Efficacy Variables

The following sections describe the efficacy variables that are collected and derived for this study's reporting purpose. The efficacy variables will be analyzed for the All Double-Blind-Treated Subject population. For endpoints in the Open-Label Rescue Period only subjects who have received at least one dose of the open-label study drug will be included (Open-Label-Rescue-Treated Subject population).

Primary Efficacy Variables

- The primary endpoint is the occurrence of flare up to Week 40 in the tapering arm (and its association to the wrist synovitis and bone marrow edema [BME] RAMRIS scores as well as a composite of both).

Secondary Variables

- Time to flare
- Flare severity
- Proportion of subjects with a flare
- Proportion of subjects in clinical remission (defined as DAS28°[ESR] < 2.6) in the Open-Label Rescue Period over time
- Proportion of subjects who regain disease control (defined as DAS28[ESR]°<°2.6 if DAS28[ESR] ≥ 2.6 at flare and defined as DAS28[ESR] decrease > 1.2 if DAS28[ESR] < 2.6 at flare) in the Open-Label Rescue Period over time
- Time to clinical remission from the occurrence of flare
- Time to regain disease control from the occurrence of flare
- Proportion of subjects with low disease activity (defined as DAS28[ESR]°< 3.2) in the Open-Label Rescue Period over time

- Change from dbBaseline in DAS28(ESR), Clinical Disease Activity Index (CDAI), Simplified Disease Activity Index (SDAI) over time
- Proportion of subjects maintaining clinical remission (defined by DAS, SDAI and CDAI [DAS28(ESR) < 2.6; SDAI ≤ 3.3; CDAI ≤ 2.8]) over time
- Change from dbBaseline to Week 40 or final Visit in MRI synovitis, BME(Osteitis), tenosynovitis, composite of synovitis and BME, overall inflammation score and erosions RAMRIS scores
- Change from dbBaseline in Health Assessment Questionnaire – Disability Index (HAQ-DI) over time
- Proportion of subjects with HAQ-DI normal (HAQ-DI ≤ 0.5) over time
- Change from dbBaseline in RAPID-3 scores assessed over time
- Change from Flare Week 0 in RAPID-3 at home assessments over time
- Change from dbBaseline in Swollen Joint Count (both 28 and 66 joints) over time
- Change from dbBaseline in Tender Joint Count (both 28 and 68 joints) over time
- Change from dbBaseline in Patient's Global Assessment of Disease activity over time
- Change from dbBaseline in Patient's Global Assessment of RA pain over time
- Change from dbBaseline in Physician's Global Assessment of Disease activity over time
- Change from dbBaseline in morning stiffness assessment over time (duration and severity)
- Change from dbBaseline in Sleep disturbance assessment over time
- Change from dbBaseline in Treatment Satisfaction Questionnaire for Medication (TSQM) over time
- Change from dbBaseline in Work Productivity and Activity Impairment (WPAI) over time
- Change from dbBaseline in Short Form-36 over time
- Change from dbBaseline in Functional Assessment of Chronic Illness Therapy – fatigue (FACIT-fatigue) over time

- Change from dbBaseline in CRP over time
- Change from dbBaseline in ESR over time

Exploratory Variables

- dbBaseline and change from dbBaseline to the time of flare in PDUS and GSUS individual and composite scores of synovitis, synovial hypertrophy and tenosynovitis

10.2.1 MRI of the hand and wrist (RAMRIS scores)

If MRI score is assessed by two Readers, the total MRI score is calculated as average between total MRI score of Reader 1 and 2. If in addition to those an adjudicator was used for the evaluation, then no averaging is needed. Results of the adjudicator are used to calculate total MRI score. MRI data of the hand and wrist were collected and recorded using the Outcomes Measures in Rheumatology Clinical Trials' Rheumatology Arthritis MRI Scoring System (OMERACT-RAMRIS). This scoring instrument measures Bone Erosions, Osteitis (BME), Synovitis, and Tenosynovitis in the most affected or dominant hand (2 – 5 metacarpophalangeal joints [MCP]) and wrist (distal radioulnar, radiocarpal, and intercarpal-carpometacarpal) joints.

The hand deemed to be in the worst condition was scanned, and images were sent to the central imaging laboratory designated by the Sponsor for reading and scoring. Bone erosions, osteitis, synovitis scores were assigned based on the OMERACT-RAMRIS for the MCP and wrist joints listed in Table 17 and Table 18, respectively.

Table 17. MCP Joints Assessed for MRI Endpoints

Endpoint	Scoring Range	Joints Assessed
Erosions	0 - 10	Distal metacarpal I, II, III, IV, V, Proximal ends of proximal phalanges I, II, III, IV, V
Osteitis	0 - 3	Distal metacarpal I, II, III, IV, V, Proximal ends of proximal phalanges I, II, III, IV, V
Synovitis	0 - 3	Metacarpophalangeal I, II, III, IV, V

Table 18. Wrist Joints Assessed for MRI Endpoints

Endpoint	Scoring Range	Joints Assessed
Erosions	0 - 10	Proximal metacarpal I, II, III, IV, V, Trapezium, Trapezoid, Capitate, Hamate, Scaphoid, Lunate, Triquetrum, Pisiform, Distal radius, Distal ulna
Osteitis	0 - 3	Proximal metacarpal I, II, III, IV, V, Trapezium, Trapezoid, Capitate, Hamate, Scaphoid, Lunate, Triquetrum, Pisiform, Distal radius, Distal ulna
Synovitis	0 - 3	Distal radioulnar, Radiocarpal, Intercarpal and Carpometacarpal [CMC]

Tenosynovitis of the flexor and extensor tendons is evaluated on the wrist. Ten groups of extensor and flexor tendons sheaths are scored for tenosynovitis, with range of 0 - 3 score for each group.

Total wrist and MCP scores will be calculated for each endpoint based on the sum of scores from all corresponding joint assessments. The total erosion score can range from 0 to 250. The total osteitis score can range from 0 to 75. The total synovitis score can range from 0 to 24. The total tenosynovitis score can range from 0 to 30. If all joints assessed for each endpoint have missing values, then the corresponding endpoint will be missing. In addition, a composite of synovitis and BME/osteitis and an overall total inflammation score by summing the osteitis, synovitis, and tenosynovitis scores will be obtained. The composite score will be missing when two endpoints are missing. The overall total inflammation score will be missing when all three endpoints are missing.

The MRI RAMRIS scores will be assessed at dbBaseline, Week 40, Flare Week 16, and at early termination.

Procedures for performing MRIs and scoring are provided in the MRI charter Version 1.0.

10.2.2 Joint Evaluation

Anatomical joints are evaluated for swelling and tenderness at every study visit. The 34 anatomical joints in Table 19 are assessed in this study for both the left and right side of the body. TJC68 and SJC66 represent total tender joint count and total swollen joint count, respectively, based on the joints (including the left and right side of the body) listed in the table below.

Table 19. Anatomical Joints Assessed for TJC68 and SJC66

Temporomandibular	Sternoclavicular	Acromio-clavicular	Shoulder
Elbow	Wrist	Metacarpophalangeal I	Metacarpophalangeal II
Metacarpophalangeal III	Metacarpophalangeal IV	Metacarpophalangeal V	Thumb Interphalangeal
Proximal Interphalangeal II	Proximal Interphalangeal III	Proximal Interphalangeal IV	Proximal Interphalangeal V
Distal Interphalangeal II	Distal Interphalangeal III	Distal Interphalangeal IV	Distal Interphalangeal V
Hip[a]	Knee	Ankle	Tarsus
Metatarsophalangeal I	Metatarsophalangeal II	Metatarsophalangeal III	Metatarsophalangeal IV
Metatarsophalangeal V	Great Toe/Hallux	Interphalangeal II	Interphalangeal III
Interphalangeal IV	Interphalangeal V		

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," provided the joint was not replaced ("9") or could not be assessed ("NA") due to other reasons (e.g., post-corticosteroid joint injection). The total tender joint count (TJC68) will be derived as the sum of all "1s" thus collected and no penalty will be considered for the joints not assessed or those which have been replaced. A similar method will be followed for the derivation of total swollen joint count (SJC66). Thus, the range for TJC68 will be 0 to 68 and 0 to 66 for SJC66.

TJC28 and SJC28 represent total tender joint count and total swollen joint count, respectively, based on the 28 joints (including the left and right side of the body) listed in the table below.

Table 20. Anatomical Joints Assessed for TJC28 and SJC28

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

No imputation will be done for missing joint assessments.

10.2.3 Visual Analogue Scale Measurements

Four visual analogue scale (VAS) measurements are collected at every visit.

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

This VAS measurement is scored by the subject on a 100 mm horizontal scale to assess the subject's overall functionality assessment considering the disease activity within the past 24 hours. The score ranges from 0 (no activity) to 100 (severe activity).

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

This VAS measurement is scored by the investigator on a 100 mm horizontal scale to assess the subject's overall disease activity irrespective of the subject's own assessment. The score ranges from 0 (no activity) to 100 (severe activity).

10.2.3.3 Patient's Global Assessment of Pain

This VAS measurement is scored by the subject on a 100 mm horizontal scale to assess the subject's overall assessment of pain due to the disease condition during the past week. The score ranges from 0 (no pain) to 100 (severe pain).

10.2.3.4 Subject's Assessment of Sleep Disturbance

This VAS measurement is scored by the subject on a 100 mm horizontal scale to assess the subject's overall assessment of severity of sleep disturbance due to the disease condition during the past week. The score ranges from 0 (sleep is no problem) to 100 (sleep is a major problem).

10.2.4 Morning Stiffness Assessment

This measurement is scored by the subject on a numeric rating-scale with 0 to 10 scores range on a self-reported questionnaire to assess the severity of subject's morning stiffness due to disease condition during the past week in minutes (from time of awaking to time of maximal improvement). The score ranges from 0 (not severe) to 10 (very severe).

The morning stiffness assessment is collected at every visit.

10.2.5 Erythrocyte Sedimentation Rate (ESR)

ESR is a laboratory parameter and considered as an efficacy variable for the RA indication. ESR is a general marker of inflammation that is sensitive to changes in inflammatory response.

The ESR is measured at all visits.

10.2.6 C-Reactive Protein (CRP)

C-reactive protein (CRP) is a laboratory parameter and a general marker of inflammation that is sensitive to acute changes in inflammatory response. The abnormal and normal

values will be determined according to the normal ranges provided by the laboratory for the CRP.

The CRP is measured at all visits.

10.2.7 Disease Activity Score (DAS28(ESR))

Disease Activity Score (DAS28) is a combined index used to measure the disease activity in patients with Rheumatoid Arthritis (RA). DAS28(ESR) provides a score between 0 and 10, indicating how active the rheumatoid arthritis is at the time of measurement.

The calculation of the DAS28(ESR) score is based on four variables (including ESR). DAS28(ESR) is calculated using the following formula:

$$\text{DAS28(ESR)} = 0.56 \times \sqrt{(\text{TJC28*})} + 0.28 \times \sqrt{(\text{SJC28**})} + 0.70 \times \text{natural logarithm (ln)(ESR}^{\#}\text{) [mm/hg]} + 0.014 \times \text{Global Health (GH}^{\text{»}}\text{) [mm]}$$

* TJC28 refers to the Subject's total Tender Joint Count out of the provided 28 evaluated joints.

** SJC28 refers to the Subject's total Swollen Joint Count out of the provided 28 evaluated joints.

\# ESR refers to Erythrocyte Sedimentation Rate evaluated at the site at current visit and expressed in mm/hg (1^{st} hour).

» GH refers to the Subject's Global Assessment of Disease Activity (PGA).

Note for programming: Please calculate the DAS28(ESR) based on the different components from the original eCRF pages because of loading and calculation issues in EDC.

DAS28(ESR) score would be considered missing for a given visit window if any of the components is missing in that window. For LOCF imputation, LOCF would be implemented for each component first, and then LOCF DAS28(ESR) scores will be calculated based on LOCF component values.

Additional variables will be derived based on the thresholds of DAS28(ESR):

- Clinical remission defined as DAS28(ESR) < 2.6
- Low disease activity defined as DAS28(ESR) < 3.2

- Time to clinical remission from the occurrence of flare, where clinical remission defined as DAS28(ESR) < 2.6
- Time to regain disease control from the occurrence of flare, where regain disease control defined as DAS28(ESR) < 2.6 if DAS28(ESR) ≥ 2.6 at flare, or DAS28(ESR) decrease > 1.2 if DAS28(ESR) < 2.6 at flare

The DAS28(ESR) score, clinical remission and low disease activity using DAS28(ESR) will be measured at all visits.

10.2.8 Flare and flare severity assessment

Flare was defined as DAS28(ESR) > 2.6 AND an increase from dbBaseline in DAS28(ESR) > 0.6, or increase of DAS28(ESR) ≥ 1.2 from dbBaseline DAS28(ESR) score irrespective of DAS28(ESR).

If the calculated DAS28(ESR) score meets flare criteria described above, the visit then becomes a Flare Week 0 visit.

At the Flare Week 0 Visit, severity of flare as assessed by a numeric rating-scale (NRS) with a 0 to 10 score range is recorded:

- *Physician's assessment of Flare severity*
  Physician will rate the severity of flare from 0 to 10 as follows:

0	1	2	3	4	5	6	7	8	9	10
Not Severe										Very Severe

- Subject's assessment of Flare severity
  Subjects will rate the severity of their flare from 0 to 10 in a numeric rating-scale where 0 is "not severe" and 10 is "very severe" (see example below):

Please make an "X" on the box below to indicate how severe your flare is:

0	1	2	3	4	5	6	7	8	9	10
Not Severe										Very Severe

Time to the occurrence of flare will be analyzed.

10.2.9 Simplified Disease Activity Index (SDAI)

SDAI is a measure of disease activity derived as follows:

$$SDAI = TJC28^{*} + SJC28^{**} + GH^{\#}\ [cm] + PhGA^{\gg}\ [cm] + CRP\ [mg/dL];$$

* TJC28 refers to the Subject's total Tender Joint Count out of the provided 28 evaluated joints.

** SJC28 refers to the Subject's total Swollen Joint Count out of the provided 28 evaluated joints.

\# GH refers to the Subject's Global Assessment of Disease Activity (PGA).

» PhGA refers to the Physician's Global Assessment of Disease Activity (PhGA).

An additional variable will be derived based on the threshold of SDAI:

- Remission (SDAI $\leq$ 3.3)

The SDAI and remission using SDAI will be measured at all visits.

10.2.10 Clinical Disease Activity Index (CDAI)

CDAI is a measure of disease activity derived as follows:

$$CDAI = TJC28^{*} + SJC28^{**} + GH^{\#}\ [cm] + PhGA^{\gg}\ [cm];$$

* TJC28 refers to the Subject's total Tender Joint Count out of the provided 28 evaluated joints.

** SJC28 refers to the Subject's total Swollen Joint Count out of the provided 28 evaluated joints.

\# GH refers to the Subject's Global Assessment of Disease Activity (PGA).

» PhGA refers to the Physician's Global Assessment of Disease Activity (PhGA).

An additional variable will be derived based on the threshold of CDAI:

- Remission (CDAI ≤ 2.8)

The CDAI and remission using CDAI will be measured at all visits.

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

The HAQ-DI is completed by the subject at each visit. HAQ is a self-reported subject-oriented outcome measure. The Disability Index of HAQ for a subject is calculated as the mean of the following eight category scores (range: 0 to 3): Dressing and Grooming, Rising, Eating, Walking, Hygiene, Reach, Grip, and Activities. A lower HAQ-DI score is better.

The score of each category is calculated as the maximum of the scores for the questions within that category. A lower category score will be adjusted to 2 if aids and devices or help from another person were used in that category. If a component question is left blank or the response is too ambiguous to assign a score, then the score for that category is determined by the remaining completed questions. The total HAQ-DI score is the mean of the eight category scores. A subject must have scores for at least six categories in order to compute the HAQ-DI score.

Additional variable will be derived based on the threshold of HAQ-DI, which designates normal physical function:

- HAQ-DI normal (HAQ-DI ≤ 0.5)

10.2.12 36-Item Short Form (SF-36) Questionnaire

The 36-Item Short Form, Version 2 (SF-36v2) Questionnaire with 4 week recall is completed by the subject at dbBaseline, Weeks 28, 40, Flare Weeks 0, 4, 10, 16, and at early termination. 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 sub-domains: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health.

The coding and scoring for the SF-36 is done as described in Appendix B.

10.2.13 Treatment Satisfaction Questionnaire for Medication (TSQM)

TSMQ is completed by the subject at dbBaseline, Weeks 16, 28, 40, Flare Weeks 0, 4, 10, 16, and at the early termination. TSQM (Version 1.4) is a general measure of satisfaction with pharmaceutical treatment that is suitable for use across a wide variety of medication types and illness conditions. The instrument includes 14 items. The scoring method for this questionnaire is presented in Appendix C.

TSQM responses are used to derive scores for scales measuring Effectiveness, Side-Effects, Convenience, and Global Satisfaction. Scores for each of the four scales range from 0 to 100 with higher scores indicating a better state or outcome (e.g., greater perceived effectiveness or satisfaction).

10.2.14 Work Productivity and Activity Impairment Questionnaire: Specific Health Problem for RA Version 2.0 (WPAI)

The Work Productivity and Activity Impairment Questionnaire: Specific Health Problem for RA Version 2.0 (WPAI) was developed to measure the effect of overall health and specific symptom on work productivity and non-work activity limitation. WPAI is completed by the subject at dbBaseline, Weeks 28, 40, Flare Weeks 0, 10, 16, and at early termination.

The coding and scoring for the WPAI is done as described in Appendix D.

10.2.15 Routine Assessment of Patient Index Data (RAPID-3)

Routine Assessment of Patient Index Data (RAPID-3) a validated disease activity score assessing physical function, patient global assessment for pain and patient global health is completed at each scheduled visit.

The scoring method for RAPID-3 is presented in Appendix E.

The RAPID-3 is completed at every visit.

Additionally, at the Flare Week 0 visit, subjects are given a RAPID-3 questionnaire assessment and asked to complete the answers weekly at home without calculating the scores (and return at each visit thereafter) until Flare Week 16.

10.2.16 FACIT-Fatigue Scale (version 4)

The FACIT-Fatique Scale score is completed by the subject at dbBaseline, Weeks 16, 28, 40, Flare Weeks 0, 4, 10, 16, and at early termination. The FACIT-Fatigue scale is a symptom-specific subscale of the Functional Assessment of Chronic Illness Therapy scales (FACIT), which are self-reported measures of quality of life. It is composed of following 13 individual questions measuring the fatigue of patients with responses 0 = Not at all, 1 = A little bit, 2 = Somewhat, 3 = Quite a bit, 4 = Very much.

FACIT-Fatigue questions

1. I feel fatigued
2. I feel weak all over
3. I feel listless ("washed out")
4. I feel tired
5. I have trouble starting things because I am tired
6. I have trouble finishing things because I am tired
7. I have energy

8. I am able to do my usual activities
9. I need to sleep during the day
10. I am too tired to eat
11. I need help doing my usual activities
12. I am frustrated by being too tired to do the things I want to do
13. I have to limit my social activity because I am tired

All of the items (except item 7 "I have energy" and item 8 "I am able to do my usual activities") will be reversed (reversed score = 4 – raw score) in order to have higher values representing lower level of fatigue. The FACIT-fatigue score will be calculated as follows:

FACIT-Fatigue score = 13 × [Sum of answered item scores/Number of items answered]

The score ranges from 0 to 52, with 52 indicating the lowest level of fatigue.

Computational Note

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.

Usually, the following conventions are already implemented in OC by data handling conventions therefore they should not show up as issues in SAS datasets. If they should, however, the following rules will be applied:

- If two responses that are not adjacent (i.e., 0-"Not at all" and 3-"Somewhat") are circled on the CRF, the question response will be left blank in the database.
- For questions "I have energy" and "I am able to do my usual activities," if two adjacent responses are provided on the CRF, the response with the lowest

number will be entered into the database. For instance, if both 0-"Not at all" and 1-"A little bit" are circled, 0 will be entered into the database.

- For all other questions, if two adjacent responses are provided on the CRF, the response with the highest number will be entered into the database. For instance, if both 0-"Not at all" and 1-"A little bit" are circled, 1 will be entered into the database.

10.2.17 Ultrasound Assessment Using Gray Scale Ultrasonography (GSUS) and Power Doppler Ultrasonography (PDUS)

In the subgroup of subjects with a dbBaseline Ultrasound (US) assessment:

- To investigate the association between dbBaseline ultrasound scores and the occurrence of RA flares.
- To investigate the association between the dbBaseline ultrasound scores and Baseline MRI RAMRIS scores.
- To describe the change in the ultrasound scores from dbBaseline to the time of RA flare in the tapering arm.

10.2.18 Biomarker

Biomarkers Analysis is not covered in this SAP.

10.3 Analysis for Primary Objective

The primary objective is to investigate the association between dbBaseline hand and wrist

- Synovitis,
- BME RAMRIS scores as well as
- a composite of synovitis and BME RAMRIS scores

and the occurrence of flares up to Week 40 in the tapering arm only. This association will be examined using logistic regression, which is deemed as the main analysis for the primary assessment. The 95% confidence interval of the odds ratio will be calculated.

Additional analyses will be performed to assess this association using various statistical methods. Specifically, descriptive statistics of dbBaseline MRI scores (synovitis, BME, composite of synovitis and BME score) will be provided for the two groups of subjects who flare and who do not flare, and the between-group difference in the mean scores will be computed together with a 95% confidence interval. Linear regression will be used to model the relationship between the DAS28(ESR) at flaring (or at the end of the study for subjects who do not flare) and the dbBaseline MRI RAMRIS scores (synovitis, BME, composite of synovitis and BME score). Receiver operating characteristic (ROC) curve approach will also be utilized to investigate the potential flare prediction criteria based on MRI RAMRIS scores.

For subjects who prematurely discontinue in the Double-Blind Period (i.e., prematurely discontinue prior to the detection of flaring), their last DAS28(ESR) score will be used to impute their flare status.

10.4 Efficacy Analyses for Secondary Objectives

To investigate the association between the primary endpoint occurrence of flare and the specified dbBaseline parameters, as subject demographics and clinical disease characteristics at dbBaseline including

- Smoking status, co-morbidities, anti-citrullinated peptide antibody (ACPA) status, Rheumatoid Factor (RF) status, disease duration, previous treatment with csDMARDs or bDMARDs (excluding adalimumab) or both, duration of adalimumab therapy, remission duration, DAS28(ESR), CRP and HAQ-DI

similar analyses (one factor per analysis) will be conducted as for the primary assessment.

To characterize the development of flares, the proportion of subjects who flare in the tapering arm and withdrawal arm and overall, and the 95% confidence interval of this proportion will be calculated. Time to flare (from first dose in the Double-Blind Period) in these two arms will be summarized using Kaplan-Meier survival techniques, which includes Kaplan-Meier estimator of time to flare curve. The proportion of subjects within

each level of flare severity will also be computed for the tapering arm and withdrawal arm and overall, accompanied by a 95% confidence interval.

To evaluate the effectiveness of re-treatment with adalimumab 40 mg eow as a rescue therapy after flaring, the proportion of flared subjects in clinical remission and subjects who regain disease control will be calculated over time with a 95% confidence interval, separately for subjects from the tapering arm and from the withdrawal arms. Additionally, the time to clinical remission and time to regain disease control from the occurrence of flare will be summarized using Kaplan-Meier methodology, separated for subjects from the tapering arm and from the withdrawal arm. To consider disease activity after flare, proportion of subjects with low disease activity (defined as DAS28[ESR] < 3.2) in the Open-Label Rescue Period will be summarized over time, separately for subjects from the tapering and withdrawal arm.

The proportion of subjects who maintain clinical remission (DAS28[ESR] < 2.6; SDAI ≤ 3.3, CDAI ≤ 2.8) will be summarized over time for each variable in the tapering and withdrawal arm.

To describe the MRI change for subjects during the taper, withdrawal and Open-Label Rescue Period, the change from dbBaseline to end of study (DB Week 40, Flare Week 16 or early termination visit) in MRI RAMRIS scores and its individual component scores (MRI synovitis, BME, tenosynovitis, composite of synovitis and BME, overall inflammation score and erosions RAMRIS scores) will be summarized separately for flared and non-flared subjects originated in the tapering and withdrawal arm respectively.

To describe the disease course of subjects in the tapering, withdrawal and Open-Label Rescue Period, different clinical and health reported outcome measures described in Section 10.2 under secondary endpoints, will be summarized over time separately for flared and non-flared subjects originated in the tapering and withdrawal arm respectively.

In addition, in order to describe the course of disease and patient reported outcome (PRO) measures in subgroups of subjects who meet additional clinical remission criteria at

dbBaseline including simplified disease activity index (SDAI) ≤ 3.3, clinical diseases activity index (CDAI) ≤ 2.8 and ACR/EULAR 2011 boolean-based remission (TJC28 ≤ 1, SJC28 ≤ 1, CRP ≤ 1 mg/dL and patient global assessment (PGA) ≤ 1 cm),[5] different clinical and health reported outcome measures, will be summarized over time separately for flared and non-flared subjects originated in the tapering and withdrawal arm per each subgroup respectively. See below the list of clinical and health reported outcome measures, which will be summarized at all time points or as specified.

- Change from dbBaseline in DAS28(ESR), Clinical Disease Activity Index (CDAI) and Simplified Disease Activity Index (SDAI) over time
- Change from dbBaseline to final Visit (DB Week 40, Flare Week 16 or early termination visit) in synovitis, BME, tenosynovitis, composite of synovitis and BME, overall inflammation score and erosions RAMRIS scores
- Change from dbBaseline in Health Assessment Questionnaire – Disability Index (HAQ-DI) over time
- Proportion of subjects with HAQ-DI normal (HAQ-DI ≤ 0.5) over time
- Change from dbBaseline in RAPID-3 scores over time
- Change from Flare Week 0 in RAPID-3 at home assessments over time
- Change from dbBaseline in Swollen Joint Count (both 28 and 66 joints) over time
- Change from dbBaseline in Tender joint Count (both 28 and 68 joints) over time
- Change from dbBaseline in Patient's Global Assessment of Disease activity over time
- Change from dbBaseline in Patient's Global Assessment of RA pain over time
- Change from dbBaseline in Physician's Global Assessment of Disease activity over time
- Change from dbBaseline in morning stiffness assessment over time (duration and severity)
- Change from dbBaseline in Sleep disturbance assessment over time

- Change from dbBaseline in Treatment Satisfaction Questionnaire for Medication (TSQM) over time
- Change from dbBaseline in Work Productivity and Activity Impairment (WPAI) over time
- Change from dbBaseline in Short Form-36 (SF-36) over time
- Change from dbBaseline in Functional Assessment of Chronic Illness Therapy – fatigue (FACIT fatigue) over time
- Change from dbBaseline in CRP over time
- Change from dbBaseline in ESR over time

To investigate the MRI-flare associations in sub-groups of subjects who meet additional clinical remission criteria at dbBaseline including

- simplified disease activity index (SDAI) ≤ 3.3,
- clinical diseases activity index (CDAI) ≤ 2.8 and
- ACR/EULAR 2011 boolean-based remission,

similar statistical approaches as described in Section 10.3 for assessing the association between dbBaseline MRI and occurrence of flares will be applied for each subgroup.

10.5 Efficacy Analysis for Exploratory Objectives

All exploratory analyses will only be conducted in the tapering arm.

The ultrasonographic assessments are performed at dbBaseline and at flare. Descriptive statistics of ultrasonography (US) scores (Grey Scale [GS] synovial hypertrophy [SH], synovial inflammation PD, combined synovial PD and SH GS, GS tenosynovitis, PD tenosynovitis scores) at dbBaseline will be summarized for the flared and non-flared subjects to assess their association with the occurrence of RA flares. In addition, the scores at flare as well as change from dbBaseline will be summarized.

Additional analyses to assess the association between dbBaseline ultrasonographic scores and the occurrence of a flare will be performed if deemed necessary.

To assess the association between dbBaseline US synovial scores (GS SH, synovial inflammation PD and for combined GS SH and synovial PD scores) and dbBaseline MRI synovitis RAMRIS score, linear regression approach will be used to model the relationship between each dbBaseline US synovial score and dbBaseline MRI synovitis RAMRIS score.

To assess the association between dbBaseline PDUS tenosynovitis and GSUS tenosynovitis scores and dbBaseline MRI tenosynovitis RAMRIS score, linear regression approach will be used to model the relationship between the dbBaseline PDUS tenosynovitis and GSUS tenosynovitis scores and dbBaseline MRI tenosynovitis RAMRIS score.

The occurrence and time to flare from dbBaseline in the concomitant MTX, other csDMARDs and no csDMARDs patients will be analyzed for potential differential effect by descriptive analyses.

10.6 Handling of Multiplicity

Not applicable.

10.7 Pharmacokinetic Analyses

Pharmacokinetic analysis is not covered in this SAP.

11.0 Safety Analysis

11.1 General Considerations

Safety analyses will include reporting of adverse events, laboratory, and vital signs measurements. Safety analyses will be carried out using the All Lead-In-Treated Subject population and the All Double-Blind-Treated Subject population, respectively. For safety analyses in the Open-Label Rescue Period only subjects who have received at least one dose of the open-label study drug will be included (Open-Label-Rescue-Treated

Subject population). Each subject's treatment group will be the actual treatment received by the subject even if this differs from his/her randomized treatment assignment.

Mean change in all continuous laboratory variables and vital signs variables at each visit will be summarized for all treated subjects. The last evaluation prior to the first dose of study drug will be used as dbBaseline for all analyses. The dbBaseline, minimum, maximum, and final value means will be presented for subjects who have both dbBaseline and post-dbBaseline values. Categorical data will be summarized using frequencies and percentages. Additionally, the number of non-missing values will be given.

All safety measurements including TEAE, Lab measurements and vital signs will be reported separately by different periods (Lead-In, Double-Blind and Open-Label Rescue Period).

For selected parameters, a listing of all subjects with any laboratory determination meeting Common Toxicity Criteria (CTC) of Grade 3 or higher will be provided. Shift tables for changes from dbBaseline according to the normal range will also be provided for lab variables.

No statistical comparison will be performed on safety endpoints.

11.2 Analysis of Adverse Events

Treatment-emergent adverse events (AEs) will be summarized and reported. In addition, serious adverse events (SAEs) collected between the signing of the informed consent and the first dose will be reported separately.

11.2.1 Treatment-Emergent Adverse Events

A Treatment-emergent Adverse Event (TEAE) is defined as an adverse event with an onset date that is on or after the first dose of study medication and no more than 70 days after the last dose of study drug or an adverse event with onset date before the first dose of study drug, but increased in severity on or after the first dose of study drug, and no more than 70 days after the last dose of study drug.

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

TEAE will be allocated to different study period (Lead-In, Double-Blind and Open-Label Rescue Period) based on the start date from the first dose in the respective period onwards.

Adverse event data will be presented by system organ classes (SOCs) and preferred terms (PTs) using the Medical Dictionary for Drug Regulatory Activities (MedDRA) Version 20.1 or most up to date version. All adverse event tables will be sorted in alphabetical order by SOC and PT.

11.2.1.1 Adverse Event Overview

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

- Any AEs
- Any AE that was assessed as related to study drug by the investigator
- Any severe AEs
- Any serious AEs (SAEs)
- Any AE of Special Interest
- Any AE leading to discontinuation of study drug.
- Any AE leading to death

Additional AEs may be considered for tabulation/summary based on recommendations from Clinical and Safety as deemed appropriate. Any event with an unknown severity will be considered as severe and any AE with an unknown relationship will be considered as drug-related.

11.2.1.2 Adverse Events by System Organ Class and Preferred Term

TEAEs will be summarized and presented using primary Medical Dictionary for Regulatory Activities (MedDRA) system organ classes (SOCs) and preferred terms (PTs) according to Version 20.1 or higher of the MedDRA coding dictionary. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

A subject who reports more than 1 AE in different SOCs will be counted only once in the overall total. A subject who reports 2 or more different PTs, which are in the same SOC, will be counted only once in the SOC total. A subject who reports more than 1 AE with the same PT will be counted only once for that PT using the most extreme incident (i.e., most severe for the severity tables and most related for the relationship tables).

11.2.1.3 Adverse Events by Maximum Severity

TEAEs will also be summarized by maximum severity. If a subject has an AE with an 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 that if the subject has another occurrence of the same AE with the most extreme severity – severe. In this case, the subject will be counted under the severe category.

11.2.1.4 Adverse Events by Maximum Relationship

TEAEs will also be summarized by maximum relationship to study drug, as assessed by the investigator. If a subject has an AE with an unknown relationship, then 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 present. The only exception is if the subject has another occurrence of the same AE with a relationship assessment of probably related or possibly related. In this case, the subject will be counted under the probably related or the possibly related category, respectively.

11.2.1.5 Adverse Events by "At Least Possibly Related" Relationship

TEAEs and TESAEs will also be summarized by relationship defined by "at least possibly related" to study 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.'

11.2.1.6 Serious Adverse Events (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

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

11.2.1.7 Frequent (≥ 5%) Adverse Events by System Organ Class and Preferred Term

TEAEs occurring for more than 5% of the subjects in any of the treatment arms will be summarized by MedDRA PT in decreasing frequency.

11.2.1.8 Adverse Events of Special Interests

The Adverse Events of Special Interests (AESI) categories will be summarized and presented for each treatment group using LMQ and PT. The following table includes all AESI for this study with the LMQ searches.

Table 21. AESI

AESI	LMQ Name	LMQ Search Term Definition
Infection		
All Infection	Infection LMQ	Terms clearly identifying an infection.
Serious Infection	Infection LMQ	All Infections – Subset for SAEs.
Legionella Infection	Legionella Infection LMQ	Terms Legionella infections. Culture and isolation of Legionella from respiratory secretions and/or pleural fluid. Direct fluorescent antibody (DFA) staining that identifies Legionella in respiratory secretions and/or pleural fluid. Four-fold increase in Legionella antibody titer or a single titer of 1:128.
Diverticulitis	Diverticulitis LMQ	Leukocytosis, CT Findings (Phlegmon, free air bubbles, abscess, colonic wall thickening, free fluid, obstruction), Fistulas (colovesica, rectovaginal).
Opportunistic Infection Excluding Oral Candidiasis and TB	Opportunistic Infection LMQ	Terms describing unusually severe forms of infections by a microbe typically seen only in immunocompromised individuals e.g., the term 'Herpes sepsis' is included. Investigations for infections that generally occur exclusively in immunocompromised patients e.g., JC virus test positive. Broader terms for infections typically only seen in immunocompromised persons when a more specific term for that infection type is not available e.g., the term 'stomatoccal infection' is included. Terms describing specific forms of fungal infections or tuberculosis infections which typically occur only in immunocompromised patients e.g., the term 'disseminated TB' is included.
Oral Candidiasis	Oral Candidiasis LMQ	Terms describing Oral Candidiasis only.
Tuberculosis		Active tuberculosis and latent tuberculosis combined.
Active Tuberculosis	Active Tuberculosis LMQ	Terms that describe active disease of tuberculosis.
Latent Tuberculosis	Latent Tuberculosis LMQ	Terms that describe positive results for TB screening tests (either skin or blood tests). Terms that describe TB infection or latent infection.
Parasitic Infection	Parasitic Infection LMQ	Protozoal infectious disorders, Helminthic Disorders, and Ectoparasitic infestations. Skin and subcutaneous arthropod and parasitic infestations, Parasitic lower respiratory tract infections.

Table 21. AESI (Continued)

AESI	LMQ Name	LMQ Search Term Definition
Parasitic Infection (continued)		Investigations terms for parasite identification and serology tests.
Reactivation of Hepatitis B	Reactivation of Hepatitis B Viral Infection LMQ	HBVsAg positive, HBeAb positive, PCR test for viral titer, HBVsAb positive, HBVc Ab positive.
Progressive Multifocal Leukoencephalopathy (PML)	Progressive Multifocal Leukoencephalopathy LMQ	PML definitive diagnosis.
Malignancies		
Malignancies	Malignnacies LMQ	All MedDRA terms for malignant or unspecified tumors: • Malignancy related conditions. • Terms for malignancy related conditions. • Terms for therapeutic and diagnostic procedures used to treat malignancies; treated malignancy to be specified. • Some of these procedures are also used for the treatment of non-malignant conditions. • Laboratory tests (incl. the results) that are specific for the malignancies. • To make agreement on "in situ" stage whether they fall under malignancies or not; currently assessed as malignancies. • To make agreement on general terms of "neoplasm;" whether considered as malignancies when no further details provided. • Tumour markers. • Terms related to tumor markers including the result. For European Group on Tumour Markers, please reference the web site at http://ar.iiarjournals.org/content/27/4A/1901.abstract: • Terms related to benign tumors. • Terms related to premalignant conditions. • Treatment and diagnostic procedures that do not specify a malignancy indicated. • Malignancy related therapeutic and diagnostic procedures includes a variety of terms which describe chemotherapy or radiotherapy treatment. Please be aware, however, that there are some terms which relate to chemotherapeutic. Such terms are not linked to chemotherapy or radiotherapy terms.

Table 21. AESI (Continued)

AESI	LMQ Name	LMQ Search Term Definition
Malignancies (continued)		
Lymphoma	Lymphoma LMQ	All terms for malignant or unspecified tumors. Terms for malignancy related conditions. Terms for therapeutic and diagnostic procedures used to treat malignancies. Some of these procedures are also used for the treatment of non-malignant conditions terms related to tumor markers.
Hepatosplenic T-Cell Lymphoma (HSTCL)	Hepatosplenic T-Cell Lymphoma (HSTLC) LMQ	The most specific inclusion criteria is the diagnosis based on T-cell histology for the lymphoma of T-cell receptor type.
Non-Melanoma Skin Cancer (NMSC)	Non-Melanoma Skin Cancer LMQ	"Skin Neoplasms, Malignant and Unspecified" And Add the Preferred Terms. "Squamous Cell Carcinoma" and "Bowen's Disease" to this Search.
Melanoma	Melanoma LMQ	The most common skin cancers are basal cell carcinoma and squamous cell carcinoma, but this category includes malignancies arising from other types of skin cells (e.g., Merkel cell carcinoma) and appendegeal cells (e.g., eccrine carcinoma).
Leukaemia	Leukemia LMQ	Terms for Lymphoid leukemia (e.g., Acute lymphoblastic leukemia, ALL, Chronic lymphocytic leukemia, CLL). Terms for Myeloid leukemia (e.g., Acute myeloblastic leukemia, AML, Chronic myeloid leukemia, CML, Myeloid sarcoma). Terms for Monocytic leukemia (e.g., Acute monoblastic leukemia, Acute monocytic leukemia, AML, Chronic myelomonocytic leukemia). Terms for Other leukemia s of specified cell type (e.g., Acute erythroid leukemia). Terms for Leukemia of unspecified cell type. Terms for Leukemia in remission.
Other Malignancies Except Lymphoma, HSTCL, Leukemia, NMSC, and Melanoma		If an AE is identified as a "malignancy" in the "Malignancies LMQ" search above but not a "Lymphoma," "HSTCL," "Leukemia," "NMSC," or a "Melanoma," then it belongs to "Other Malignancies" category.

Table 21. AESI (Continued)

AESI	LMQ Name	LMQ Search Term Definition
Immune Reactions		
Allergic Reactions Including Angioedema/ Anaphylaxis	Allergic Reactions LMQ	Any terms representing events which include anaphylaxis, allergic events, hypersensitivity, angioedema, angioneurotic edema, urticarial (hives).
Lupus-Like Reactions and Systemic Lupus Erythematosus (SLE)	Lupus-Like Reactions and Systemic Lupus Erythematosus (SLE) LMQ	Lupus erythematosus and associated conditions in both primary and secondary locations in this are included except Lupoid hepatic cirrhosis, which is excluded. Additional terms representing the diagnostic criteria of the American College of Rheumatology (ACR), e.g., Malar rash, Arthritis, Renal disorder, etc.
Vasculitis	Vasculitis LMQ	Terms for primary vasculitides (e.g., Henoch-Schonlein purpura, Behçet's syndrome, or Granulomatosis with polyangiitis (which includes Wegener's granulomatosis). Terms containing vasculitis (e.g., Cutaneous vasculitis, Lupus vasculitis, or Rheumatoid vasculitis). Terms containing arteritis (e.g., Arteritis coronary or Polyarteritis nodosa). Terms containing angiitis (e.g., Microscopic polyangiitis or Thromboangiitis obliterans). Terms for forms of purpura indicative of a vascular condition such as Henoch-Schonlein purpura and Chronic pigmented purpura (narrow, which includes Majocchi's purpura) and Palpable purpura (broad). Terms for laboratory test results that may indicate vasculitis (e.g., Antineutrophil cytoplasmic antibody increased).
Cutaneous Vasculitis	Cutaneous Vasculitis LMQ	Terms for vasculitides clearly identifiable as cutaneous (e.g., Cutaneous vasculitis, Skin vasculitis, Majocchi's purpura, Leukocytoclasitc, vasculitis, vasculitis legs, Necrotizing vasculitis, polyangiitis or Vasculitic rash).
Non-Cutaneous Vasculitis		Terms for vasculitides that are NOT in Cutaneous Vasculitis LMQ.
Sarcoidosis	Sarcoidosis LMQ	Sarcoidosis of any organ.
Autoimmune Hepatitis	Autoimmune Hepatitis LMQ	
Cardiovascular/Vascular		
Myocardial Infarction	Myocardial Infarction LMQ	Myocardial Infarction. Terms which included the words 'myocardial infarction, heart attack, cardiac enzymes increased, cardiac troponin increased, CK-MB increased, ECG ST elevation, Q wave abnormal.'

Table 21. AESI (Continued)

AESI	LMQ Name	LMQ Search Term Definition
Cardiovascular/Vascular (continued)		
Cerebrovascular Accident	Cerebrovascular Accident LMQ	Terms for conditions specific for cerebrovascular accident disorders irrespective of cause and irrespective of acuteness or chronicity.
Congestive Heart Failure	Congestive Heart Failure LMQ	Terms describing existing cardiac failure in its various forms with or without right/left ventricular specified. A small number of terms for symptoms, signs and investigational findings that are pathognomonic of the condition.
Pulmonary Embolism	Pulmonary Embolism LMQ	Terms that are specifically assigned to describe the condition where an embolic event occurred in the lung of a patient, chronic or acute, arterial or venous.
Respiratory		
Interstitial Lung Disease	Interstitial Lung Disease LMQ	ILDs include inflammatory and fibrotic diseases that ultimately disrupt the alveolar-capillary interface, leading to hypoxemia. ILDs usually manifest with a restrictive physiology, but airway involvement can cause obstruction or mixed physiology. ILDs are associated with many clinical settings, including connective tissue disease; occupational, environmental and drug exposures; and primary pulmonary disorders.
Gastrointestinal Events		
Intestinal Perforation	Intestinal Perforation LMQ	Terms for/related to any part of GI Tract, GI ulcers, obstruction, hemorrhage.
Pancreatitis	Pancreatitis LMQ	Terms with the word "pancreatitis" (other than those indicative of chronic conditions). Terms indicative of pancreatic dysfunction (such as pancreatorenal syndrome). Terms for typical complications, e.g., Pancreatic pseudocyst.
Skin and Subcutaneous Tissue Disorders		
Stevens-Johnson Syndrome	Severe Skin Reaction Stevens-Johnson Syndrome LMQ	This falls within the erythema multiforme-toxic epidermal necrolysis spectrum, representing an intermediate severity. All terms and their synonyms specifically linked to erythema multiforme, Stevens-Johnson syndrome, and toxic epidermal necrolysis are included. No procedure terms or appropriate investigation terms are specific.

Table 21. AESI (Continued)

AESI	LMQ Name	LMQ Search Term Definition
Skin and Subcutaneous Tissue Disorders (continued)		
Erythema Multiforme	Erythema Multiforme LMQ	Diagnosis terms that represent the condition Erythema Multiforme are included. Terms that represent the signs/symptoms that are specific for Erythema Multiforme and included in the criteria for the diagnosis of Erythema Multiforme are included.
Worsening and New Onset of Psoriasis	Worsening New Onset of Psoriasis LMQ	Psoriasis can present in guttate form, where the lesions are smaller, or involve flexural parts of the body (e.g., axilla, inframammary areas), or involve fingernails or toenails (i.e., nail psoriasis), or involve palms or soles. In pustular psoriasis, the presentation is characterized by sterile pustules on a brightly erythematous base. When pustular psoriasis involves the palms and soles, it is termed palmoplantar pustulosis. Other variants of psoriasis include psoriasiform dermatitis and acrodermatitis continua.
Nervous System Disorder		
Demyelinating Disorders Including Multiple Sclerosis, Guillain-Barré Syndrome, and Optic Neuritis and Others	Demyelinating Disorders LMQ	Terms for encephalomyelitis and leukoencephalopathies related to demyelination (e.g., Acute haemorrhagic leukoencephalitis and Progressive multifocal leukoencephalopathy). Trigeminal neuralgia is included in broad terms due to possible association with multiple sclerosis or other demyelinating conditions. Terms representing a disability scale which is highly specific for MS (e.g., Expanded disability status scale score increased).

Table 21. AESI (Continued)

AESI	LMQ Name	LMQ Search Term Definition
Nervous System Disorder (continued)		
Amyotrophic Lateral Sclerosis	Amyotrophic Lateral Sclerosis LMQ	There are no specific tests to provide a definite diagnosis of ALS although the presence of upper and lower motor neuron signs in a single limb is strongly suggestive. The diagnosis is based on symptoms and signs as well as a series of tests to assess muscle weakness and atrophy, hyperreflexia, and spasticity which gets progressively worse. For these reasons the search terms recommended for adjudication purposes would need to be consistent with the diagnosis of ALS and not with non-specific findings describing all the different types of motor neuron diseases. Patients may exhibit symptoms and signs of the disease such as hyperreflexia, and spasticity. Tests such as muscle biopsy, nerve conduction velocity, electromyography as well as MRI scans of the brain or spinal cord may also point to a possible diagnosis of ALS although none of these tests are definitive and will therefore not be included in the search.
Reversible Posterior Leukoencephalopathy Syndrome (RPLS)	Reversible Posterior Leukoencephalopathy Sydrome (RPLS) LMQ	Reversible Posterior Leukoencephalopathy Syndrome definitive diagnosis. Posterior Reversible Encephalopathy Syndrome definitive diagnosis.
Hematologic Events		
Hematologic Disorders Including Pancytopenia	Hematologic Disorders Including Pancytopenia LMQ	Terms referring to direct alterations of the hematopoiesis. Hematological signs and diagnoses of bone marrow depression. Hematological investigation results of bone marrow depression. Neonatal terms are included in the broad search (the term neonatal does not allow a conclusion whether the condition is of acquired or inherited origin). Certain abnormal terms are included in the broad searches.
Hepatic Events		
Liver Failure and Other Liver Events (Except Gall Bladder Related Events)	Liver Events LMQ	Terms representing liver related diagnoses or diagnoses that can be related histological descriptions. Also included are conditions associated with jaundice or cholestasis of possible hepatic origin and therefore excludes terms indicating jaundice caused by haemolytic conditions, such as Jaundice extrahepatic obstructive. Diagnoses associated with pregnancy are included.

Table 21. AESI (Continued)

AESI	LMQ Name	LMQ Search Term Definition
Other		
Humira Administration Related Medication Error	Medication Errors LMQ	Medication errors. Product Quality Issues. Terms which included the words Medication errors that relates to 'adalimumab use.'
Injection Site Reactions	Injection Site Reaction LMQ	Terms in the Injection Site Reactions, Infusion Site Reactions, Vaccination Site Reactions. Terms which included the words injection site, infusion site, application site, administration site, and 'vaccination site.' The search is focused on local reactions to injections.

Additional AEs may be considered for tabulation/summary based on recommendations from Clinical and Safety as deemed appropriate.

11.2.1.9 Adverse Events by 100 Patient Years

TEAEs will also be presented by event rate per 100 patient-years. These will be presented for any TEAEs.

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 of all subjects, as defined in Section 9.0, normalized by 365.25, and rounded to 1 decimal place.

11.3 Analysis of Laboratory Data

For the assessment of laboratory data, values observed more than 70 days after the last dose of study drug will be excluded.

11.3.1 Variables and their Units

All laboratory parameters to be collected in this study are listed in the Table 22. Laboratory parameters will be reported using the standard international (SI) units.

The laboratory test names and standard international (SI) units (as applicable) of all laboratory variables for the study are shown in Table 22.

Table 22. List of Lab Variables and their units

Lab Variables	SI Units
Hematology	
White Blood Cell (WBC) Count	$\times 10^9$/L
Red Blood Cell (RBC) Count	$\times 10^{12}$/L
Bands	
Hemoglobin	g/L
Hematocrit	Proportion of 1
Platelets count	$\times 10^9$/L
Neutrophils	$\times 10^9$/L
Basophils	$\times 10^9$/L
Eosinophils	$\times 10^9$/L
Lymphocytes	$\times 10^9$/L
Monocytes	$\times 10^9$/L
Chemistry	
Total Bilirubin	µmol/L
Alkaline Phosphatase (ALP)	U/L
Serum glutamic oxaloacetic transaminase (SGOT/AST)	U/L
Serum glutamic pyruvic transaminase (SGPT/ALT)	U/L
Total Protein	g/L
Albumin	g/L
Glucose	mmol/L
Cholesterol	mmol/L
Triglycerides	mmol/L
Blood Urea Nitrogen (BUN)	mmol/L
Creatinine	µmol/L
Uric acid	µmol/L
Sodium	mmol/L
Potassium	mmol/L
Calcium	mmol/L
Inorganic Phosphorus	mmol/L

Table 22. List of Lab Variables and their units (Continued)

Lab Variables	SI Units
Urinalysis	
Specific Gravity	-
pH	-
Protein	-
Glucose	-
Ketones	-
Blood	-
Nitrite	-
Leukocyte esterase	-
Pregnancy	
Pregnancy - serum	-
Pregnancy - urine	-
Serology	
CRP	mg/L
Human Chorionic Gonadatropin	-
Anti-Nuclear Antibody (ANA)	-
Anti-CCP (also called ACPA)	EU
Rheumatoid Factor (RF)	KU/L
Hepatitis B - surface Antigen (HBsAg) - surface Antibody (HBsAb) - core Antibody (HBcAb)	-
Hepatitis C Antibody	-
Anti-dsDNA	-
Erythrocyte Sedimentation Rate (ESR) 1	mm/hr
Additional Blood Samples collected	
Quanti-FERON TB Gold Test 2	-

1. ESR is optional and should only be performed at Screening and Baseline, if used to meet inclusion criterion.
2. QuantiFERON-TB Gold test (whole blood – required only if PPD skin test not performed).

11.3.2 Statistical Methods

11.3.2.1 Analysis for Continuous Laboratory Data

Analyses of laboratory data will presented based on the National Cancer Institute Common Toxicity Criteria for Adverse Event (NCI CTCAE) scale.

Summary statistics such as mean at dbBaseline, mean value at each visit, mean change from dbBaseline, standard deviation and median will be presented for each laboratory value by treatment group for the Double-Blind Period. An ANOVA model with only treatment as a factor, not controlling for dbBaseline, will be used to test statistical significance for the change from dbBaseline mean.

For the Open-Label Rescue Period similar summary statistics will be presented for each laboratory value for overall values.

11.3.2.2 Shift Table Analyses

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 dbBaseline values below/within/above the normal range versus minimum/maximum/final values below/within/above the normal range.

11.3.2.3 Analysis for Liver Elevations

Drug Induced Liver Injury (DILI) occurs infrequently and has been the leading cause of safety-related drug withdrawal for the past 50 years. It is desirable to have data early on to detect the signal or the potential of the investigation drug to cause severe liver injury.

A listing of potentially clinically significant liver function laboratory values is provided.

The listing includes all subjects who met any of the following four criteria:

- ALT $\geq 2.5 \times$ ULN, or
- AST $\geq 2.5 \times$ ULN, or

- Alkaline Phosphatase ≥ 2.5 × ULN, or
- Total Bilirubin ≥ 1.5 × ULN.

For ALT, AST, alkaline phosphatatse and Total Bilirubin, shift table categories will be < 1.5 × ULN, ≥ 1.5 × ULN – < 3.0 × ULN, ≥ 3.0 × ULN – < 5.0 × ULN, ≥ 5 × ULN – < 8 × ULN, ≥ 8 × ULN and missing. Shift table from dbBaseline to post-dbBaseline maximum value will be summarized for each treatment group.

11.3.2.4 Potentially Clinically Significant Laboratory Values

Frequencies and percentages of subjects with post dbBaseline lab values meeting the following criteria will be summarized.

Table 23. Criteria for Potentially Clinically Important Chemistry Values

		Definition of Potentially Clinically Important (Version 3.0) NCI CTCAE Grade 3 or Greater	
Chemistry Variables	**Units**	**Very Low**	**Very High**
Total Bilirubin	mcmol/L		> 3.0 × 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		> 5.0
Creatinine	mcmol/L		> 3.0 × ULN
Sodium	mmol/L	< 130	> 155
Potassium	mmol/L	< 3.0	> 6.0
Calcium	mmol/L	< 1.75	> 3.1
Total Cholesterol	mmol/L		> 10.34
ALP	U/L		> 5.0 × ULN
Inorganic Phosphorus	mmol/L	< 0.6	

Table 24. Criteria for Potentially Clinically Important Hematology Values

		Definition of Potentially Clinically Important (Version 3.0) Grade 3 or Greater
Hematology Variables	**Units**	**Very Low**
Hemoglobin	g/L	< 80.0
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

A separate listing will be provided that presents all of the subjects and values that are NCI CTCAE toxicity grade 3 or above. For each of these subjects, the whole course of the respective parameter will be listed. The NCI CTCAE grading is shown in Table 25 below:

Table 25. NCI CTCAE Grading

Laboratory Parameter	CTC Grade 1	CTC Grade 2	CTC Grade 3	CTC Grade 4
Total Bilirubin	> ULN – 1.5 × ULN	> 1.5 – 3.0 × ULN	> 3.0 – 10.0 × ULN	> 10.0 × ULN
SGOT/AST	> ULN – 2.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 20.0 × ULN	> 20.0 × ULN
SGPT/ALT	> ULN – 2.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 20.0 × ULN	> 20.0 × ULN
Albumin	< LLN – 30 g/L	< 30 g/L – 20 g/L	< 20 g/L	—
Glucose	> ULN – 8.9 mmol/L < LLN – 3.0 mmol/L	> 8.9 – 13.9 mmol/L < 3.0 – 2.2 mmol/L	> 13.9 – 27.8 mmol/L < 2.2 – 1.7 mmol/L	> 27.8 mmol/L < 1.7 mmol/L
Triglycerides	> ULN – 2.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 10 × ULN	> 10 × ULN
Creatinine	> ULN – 1.5 × ULN	> 1.5 – 3.0 × ULN	> 3.0 – 6.0 × ULN	> 6.0 × ULN
Sodium	> ULN – 150mmol/L < LLN – 130mmol/L	> 150 – 155mmol/L —	> 155 – 160mmol/L < 130 – 120 mmol/L	> 160 mmol/L < 120 mmol/L
Potassium	> ULN – 5.5 mmol/L < LLN – 3.0 mmol/L	> 5.5 – 6.0 mmol/L —	> 6.0 – 7.0 mmol/L < 3.0 – 2.5 mmol/L	> 7.0 mmol/L < 2.5 mmol/L
Calcium	> ULN – 2.9 mmol/L < LLN – 2.0 mmol/L	> 2.9 – 3.1 mmol/L < 2.0 – 1.75 mmol/L	> 3.1 – 3.4 mmol/L < 1.75 – 1.5 mmol/L	> 3.4 mmol/L < 1.5 mmol/L
Cholesterol	> ULN – 7.75 mmol/L	> 7.75 – 10.34 mmol/L	> 10.34 – 12.92 mmol/L	> 12.92 mmol/L
Hemoglobin	< LLN – 100 g/L	< 100 – 80 g/L	< 80 – 65 g/L	< 65 g/L
Platelets count	< LLN – 75.0 × 10^9/L	< 75.0 – 50.0 × 10^9/L	< 50.0 – 25.0 × 10^9/L	< 25.0 × 10^9/L
WBC count	< LLN – 3.0 × 10^9/L	< 3.0 – 2.0 × 10^9/L	< 2.0 – 1.0 × 10^9/L	< 1.0 × 10^9/L
Neutrophils	< LLN – 1.5 × 10^9/L	< 1.5 – 1.0 × 10^9/L	< 1.0 – 0.5 × 10^9/L	< 0.5 × 10^9/L
Lymphocytes	< LLN – 0.8 × 10^9/L	< 0.8 – 0.5 × 10^9/L	< 0.5 – 0.2 × 10^9/L	< 0.2 × 10^9/L
ALP	> ULN – 2.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 20.0 × ULN	> 20.0 × ULN
Inorganic Phosphorus	< LLN – 0.8 mmol/L	< 0.8 – 0.6 mmol/L	< 0.6 – 0.3 mmol/L	< 0.3 mmol/L

LLN = lower limit of normal range; ULN = upper limit of normal range

If there are multiple post-dbBaseline measurements on the same day, the worst value will be used.

11.4 Analysis of Vital Signs and Weight

11.4.1 Variables and Criteria Defining Abnormality

Vital sign variables include sitting systolic blood pressure, sitting diastolic blood pressure, pulse, body temperature, and weight. The criteria for potentially clinically significant vital sign findings are presented in Table 26.

Table 26. Criteria for Potentially Clinically Significant Vital Sign Findings

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and/or decrease ≥ 20 mmHg from dbBaseline
	High	Value ≥ 180 mmHg and/or increase ≥ 20 mmHg from dbBaseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and/or decrease ≥ 15 mmHg from dbBaseline
	High	Value ≥ 105 mmHg and/or increase ≥ 15 mmHg from dbBaseline
Pulse	Low	Value ≤ 50 bpm and/or decrease ≥ 15 bpm from dbBaseline
	High	Value ≥ 120 bpm and/or increase ≥ 15 bpm from dbBaseline
Respiratory Rate	Low	< 10 rpm
	High	> 24 rpm
Weight	Low	> 7% decrease from dbBaseline

11.4.2 Statistical Methods

The following summary statistics will be presented: dbBaseline mean, the mean at each visit, mean change from dbBaseline, standard deviation, and median. For the Double-Blind period values are presented by treatment group. An ANOVA model with only treatment as a factor, not controlling for dbBaseline, will be used to test statistical significance for the change from dbBaseline mean.

For the Open-Label Rescue Period similar summary statistics will be presented for overall values. For the Lead-In Period the dbBaseline mean, standard deviation and median will be presented (no additional vital signs data is collected in the Lead-In period).

The number and percentage of subjects meeting the criteria for potentially clinically significant vital sign values will be summarized for the Double-Blind (by treatment group) and Open-Label Rescue Period.

Vital sign results satisfying the criteria for potentially clinically significant vital sign findings will be identified in a listing.

11.5 Analysis of ECG Parameters

No post-baseline ECG obtained for this study.

11.6 Analysis for Other Safety Variables

Not applicable for the current planned analysis.

12.0 Summary of Changes

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

In the description of the sample size calculation in Section 4.3, the 95% confidence intervals were used instead of 90% confidence intervals. In addition to the conservative flare rate in the protocol (30%) a 40% flare rate is also considered.

All statistical inference will be based on a 2-sided alpha-level of 0.05 (instead of 0.1). So 95% CI are calculated in Section 10.3 and Section 10.4.

In Section 10.2 of the SAP, in addition to the considered RAMRIS scores in the protocol (synovitis, BME and erosions RAMRIS scores), the tenosynovitis, composite of synovitis and BME and the overall inflammation score (sum of synovitis, BME and tenosynovitis scores) are included for the analysis of change in MRI score.

In Section 10.2 and Section 10.4 the proportion of subjects with HAQ-DI normal (HAQ-DI $\leq$ 0.5) will be considered for all visit and not only for dbBaseline and Week 40 as defined in the protocol.

In Section 10.2 the variable regain disease control is introduced in the Open-Label Period defined as DAS28[ESR] < 2.6 if DAS28[ESR] ≥ 2.6 at flare and defined as DAS28[ESR] decrease > 1.2 if DAS28[ESR] < 2.6 at flare.

In Section of 10.5 of the SAP beyond the synovial hypertrophy [SH] Grey Scale [GS], synovial inflammation PD the combined synovial PD and SH GS, GS tenosynovitis, PD tenosynovitis are considered as well and in addition to the MRI synovitis RAMRIS score the tenosynovitis RAMRIS score is used to assess the association between PDUS and GSUS and the MRI RAMRIS scores.

Since biomarker analyses are not covered in this SAP this part is not included in the exploratory variables paragraph in Section 10.2.

It was specified that the exploratory analyses (Section 10.5) are only conducted for the tapering arm. This includes analyses of ultrasound data even though in one of the exploratory objectives Section 4.0 in the study protocol the withdrawal arm was mentioned.

13.0 Appendices

Appendix A. Ultrasound Assessment Using Gray Scale Ultrasonography (GSUS) and Power Doppler Ultrasonography (PDUS)

Grey-Scale (GS) Synovial hypertrophy (SH), synovial power Doppler (PD), GS and Doppler tenosynovitis is defined and scored according to the Outcome Measures in Rheumatology Clinical Trails (OMERACT).

For synovial US scores the following 46 joints are evaluated:

1. Glenohumeral (posterior and axillary recesses, and biceps sheath)
2. Elbow (anterior, lateral, medial, and posterior recesses)
3. Wrist (radiocarpal, midcarpal, and distal radioulnar joints; dorsal recesses)
4. 1st - 5th metacarpophalangeal (MCP) (dorsal recesses)
5. 1st - 5th proximal interphalangeal (PIP) of the hands (dorsal and palmar/volar recesses)
6. Hip (anterior recess)
7. Knee (suprapatellar and parapatellar recesses)
8. Ankle joints (dorsal, lateral, and medial recesses of tibiotalar joint; and medial and lateral recesses of subtalar joint)
9. 2nd - 5th metatarsopharlangeal (MTP) joints (dorsal recess)

At each intra-articular synovial recess, grey-scale (GS) synovial hypertrophy (SH) will be scored semiquantitatively on a scale of 0 – 3 as follows: 0, absent; 1, mild; 2, moderate; 3, marked.

At each intra-articular synovial recess, intra-synovial power Doppler (PD) signal will be also scored on a semiquantitative scale of 0 – 3 as follows: 0, no synovial PD signal; 1, mild [≤ 3 PD signals within the SH]; 2, moderate [≤ 3 PD signals in less than half of the SH area]; 3, marked [PD signals in more than half of the SH area]).

Each joint will be scored for GS SH and intra-synovial PD signal on a scale from 0 to 3. These global scores will correspond to the maximum score for GS SH and PD signal, respectively, obtained from any one of the synovial recess evaluated at each joint. Wrist GS SH and wrist synovial PD signal score will correspond to the maximum scores for GS SH and PD signal, respectively, obtained from either the radiocarpal, the midcarpal, or the ulnarcarpal joints. Ankle GS SH and ankle synovial PD signal score will correspond to the maximum scores for GS SH and PD signal, respectively, obtained from either the tibiotalar or the subtalar joints.

Afterwards, a combined GS SH and synovial PD signal global score for synovitis will be calculated for each joint as follows;

- 0, no GS SH, no PD signal
- 1, minimal GS SH (i.e., score = 1) with synovial PD score ≤ 1
- 2, moderate GS SH (i.e., score = 2) with synovial PD score ≤ 2 or minimal GS SH (i.e., score 1) with synovial PD score = 2
- 3, severe GS SH (i.e., grade = 3) with synovial PD score ≤ 3 or minimal GS SH (i.e., score = 1) or moderate GS SH (i.e., score = 2) with synovial PD score = 3.

For US tenosynovitis scores the following 18 tendon/tendon compartments are evaluated:

1. 2nd, 4th and 6th extensor compartments at the wrist
2. 2nd - 5th finger flexor tendons
3. Tibialis posterior
4. Peroneal tendons

Each tendon/tendon compartment will be scored semiquantitatively on a scale of 0 - 3 for PD tenosynovitis as follows;

- 0, no peri-tendinous signal

- 1, focal abnormal signal within the widened synovial sheath (i.e., signals in only one area of the widening sheath) seen in two perpendicular planes
- 2, multifocal abnormal signal within the widened synovial sheath (i.e., signals in more than one area of the widening sheath), seen in two perpendicular planes
- 3, diffuse abnormal signal within the widened synovial sheath (i.e., signals filling most of the widened sheath), seen in two perpendicular planes

The presence of abnormal intratendinous PD signal in addition to peri-tendinous PD signal will upgrade the scores 1 and 2 to 2 and 3, respectively.

GS SH and synovial PD signal will be scored longitudinally to the synovial recesses. GS and PD tenosynovitis will be scored in both transverse and longitudinal plane; the assigned scores should be confirmed in both plane.

Scoring system at the patient level

Total indices for GS SH, synovial PD signal, combined GS SH and synovial PD signal global score, GS tenosynovitis and PD tenosynovitis will be calculated as follows;

- Total index for GS SH: the sum of the SH scores obtained for each evaluated joint or joint region, 0 - 120
- Total index for synovial PD signal: the sum of synovial PD signal scores obtained for each evaluated joint or joint region, 0 - 120
- Total index for combined GS SH and synovial PD signal global score: the sum of combined synovitis score obtained for each evaluated joint or joint region, 0 - 120
- Total index for GS tenosynovitis: the sum of the grey-scale tenosynovitis scores
- Total index for PD tenosynovitis: the sum of the PD tenosynovitis scores obtained for each evaluated tendon/tendon compartment, 0 - 54

Appendix B. Scoring the SF-36v2

The SF-36v2™ consists of 36 questions covering 8 health domains: physical functioning, bodily pain, role limitations due to physical problems, role limitations due to emotional problems, general health perceptions, mental health, social function, and vitality. The SF-36v2™ will be scored for the 8 sub-domains according to the standard SF-36v2™ scoring algorithms (0 to 100 scale) explained in the SF-36v2™ Manual and Interpretation Guide.[6] The physical component summary (PCS) and mental component summary (MCS) will be scored according to the standard SF-36v2™ scoring algorithm (0 to 100 scale) explained in the SF-36v2™ Physical and Mental Health Summary Scales Manual.[6] A higher score indicates a better health state.

The coding and scoring for each component is done according to the scheme in Table 27.

Table 27. Coding and Scoring for the 8 SF-36 (Version 2) Scales

Scale	Item	Coding	Range of Raw Score	Normalization*
Physical functioning	*Items 3a to 3j*	Yes, limited a lot = 1 Yes, limited a little = 2 No, not limited = 3	10 – 30	(S-10)/20 × 100
Role-physical	*Items 4a to 4d*	All of the time = 1 Most of the time = 2 Some of the time = 3 A little of the time = 4 None of the time = 5	4 – 20	(S-4)/16 × 100
Bodily pain	*Item 7*	None = 6 Very mild = 5.4 Mild = 4.2 Moderate = 3.1 Severe = 2.2 Very severe = 1	2 – 12	(S-2)/10 × 100
	Item 8 if both items 7 and 8 are answered	Not at all and item 7 equals 'None' = 6 Not at all and item 7 not equal 'None' = 5 A little bit = 4 Moderately = 3 Quite a bit = 2 Extremely = 1		
	Item 8 if item 7 is not answered	Not at all = 6 A little bit = 4.75 Moderately = 3.5 Quite a bit = 2.25 Extremely = 1		

Table 27. Coding and Scoring for the 8 SF-36 (Version 2) Scales (Continued)

Scale	Item	Coding	Range of Raw Score	Normalization*
General health	*Item 1*	Excellent = 5 Very good = 4.4 Good = 3.4 Fair = 2 Poor = 1	5 – 25	(S-5)/20 × 100
	Items 11a and 11c	Definitely true = 1 Mostly true = 2 Don't know = 3 Mostly false = 4 Definitely false = 5		
	Items 11b and 11d	Definitely true = 5 Mostly true = 4 Don't know = 3 Mostly false = 2 Definitely false = 1		
Vitality	*Items 9a and 9e*	All of the time = 5 Most of the time = 4 Some of the time = 3 A little of the time = 2 None of the time = 1	4 – 20	(S-4)/16 × 100
	Items 9g and 9i	All of the time = 1 Most of the time = 2 Some of the time = 3 A little of the time = 4 None of the time = 5		

Table 27. Coding and Scoring for the 8 SF-36 (Version 2) Scales (Continued)

Scale	Item	Coding	Range of Raw Score	Normalization*
Social functioning	*Item 6*	Not at all = 5 Slightly = 4 Moderately = 3 Quite a bit = 2 Extremely = 1	2 – 10	(S-2)/8 × 100
	Item 10	All of the time = 1 Most of the time = 2 Some of the time = 3 A little of the time = 4 None of the time = 5		
Role-emotional	*Items 5a to 5c*	All of the time = 1 Most of the time = 2 Some of the time = 3 A little of the time = 4 None of the time = 5	3 – 15	(S-3)/12 × 100

Table 27. Coding and Scoring for the 8 SF-36 (Version 2) Scales (Continued)

Scale	Item	Coding	Range of Raw Score	Normalization*
Mental health	*Items 9b, 9c and 9f*	All of the time = 1 Most of the time = 2 Some of the time = 3 A little of the time = 4 None of the time = 5	5 – 25	(S-5)/20 × 100
	Items 9d and 9h	All of the time = 5 Most of the time = 4 Some of the time = 3 A little of the time = 2 None of the time = 1		

* S = raw score = sum of item scores after coding

The standardization of each scale score is shown in Table 28.

Table 28. Standardization of the SF-36 (Version 2) Scale Scores

Item	Standardization (Z-score)
Physical functioning	PF_Z = (PF-83.29094)/23.75883
Role-physical	RP_Z = (RP-82.50964)/25.52028
Bodily pain	BP_Z = (BP-71.32527)/23.66224
General health	GH_Z = (GH-70.84570/20.97821
Vitality	VT_Z = (VT-58.31411)/20.01923
Social functioning	SF_Z = (SF-84.30250)/22.91921
Role-emotional	RE_Z = (RE-87.39733)/21.43778
Mental heath	MH_Z = (MH-74.98685)/17.75604

Note: Based on the 1998 general United States population.

SF-36 physical and mental component transformed scores are calculated using the conventions in Table 29.

Table 29. SF-36 Aggregate Component and Transformed Scores

Score	Calculation
Aggregate Standardized Score	
Physical component	AGG_PHYS = (PF_Z * 0.42402) + (RP_Z * 0.35119) + (BP_Z * 0.31754) + (GH_Z * 0.24954) + (VT_Z * 0.02877) + (SF_Z * –0.00753) + (RE_Z * –0.19206) + (MH_Z * –0.22069)
Mental component	AGG_MENT = (PF_Z * –0.22999) + (RP_Z * –0.12329) + (BP_Z * –0.09731) + (GH_Z * –0.01571) + (VT_Z * 0.23534) + (SF_Z * 0.26876) + (RE_Z * 0.43407) + (MH_Z * 0.48581)
Transformed Summary	
Transformed physical	PCS = 50 + (AGG_PHYS * 10)
Transformed mental	MCS = 50 + (AGG_MENT * 10)

Missing Data in an Individual Questionnaire

In the event that data is missing for an individual item from the sub-domains of the SF-36v2™ Health Status Survey, the average value of the completed items in the

corresponding sub-domain will be used as an estimate of the missing item. If more than 50 percent of the items from a sub-domain are missing for an individual questionnaire, the corresponding deficient sub-domain(s) will be excluded from analyses. For example, if only 3 of the 5 general health questions are answered, the mean of those 3 answers will be used to fill in the responses of the remaining 2 general health questions. However, if only 1 or 2 general health questions are answered, the general health score would be set to missing. In addition, if the respondent is missing any 1 of the 8 scale scores, then the physical and the mental component scores will be set to missing.

Each of the 8 scales and the 2 components (physical and mental) will be reported using norm-based scoring as per the standard guidelines adopted by Abbvie.

Appendix C. Scoring the Treatment Satisfaction Questionnaire for Medication (TSQM)

Table 30. Responses and raw scores for TSQM items

Scales	Item #	Response	Raw Score	Scale Scores
Effectiveness	1, 2 and 3	Extremely Dissatisfied Very Dissatisfied Dissatisfied Somewhat Satisfied Satisfied Very Satisfied Extremely Satisfied	1 2 3 4 5 6 7	(Sum – 3)/0.18 where Sum = Sum of Raw Scores from items 1, 2 and 3.
Side Effects	4	Yes No	1 0	
	*5	Extremely Bothersome Very Bothersome Somewhat Bothersome A Little Bothersome Not at All Bothersome	1 2 3 4 5	If answer to Question 4 is "Yes" then (Sum – 4)/0.16 where Sum = Sum of Raw Scores of Items 5, 6, 7 and 8. OR if answer to Question 4 is "No," then 100
	*6, 7 and 8	A Great Deal Quite a Bit Somewhat Minimally Not at All	1 2 3 4 5	
Convenience	9 and 10	Extremely Difficult Very Difficult Difficult Somewhat Easy Easy Very Easy Extremely Easy	1 2 3 4 5 6 7	Sum = Sum of Raw Scores from items 9, 10 and 11. (Sum – 3)/0.18
	11	Extremely Inconvenient Very Inconvenient Inconvenient Somewhat Convenient Convenient Very Convenient Extremely Convenient	1 2 3 4 5 6 7	

Table 30. Responses and raw scores for TSQM items (Continued)

Scales	Item #	Response	Raw Score	Scale Scores
Global Satisfaction	12	Not at All Confident A Little Confident Somewhat Confident Very Confident Extremely Confident	1 2 3 4 5	Sum = Sum of Raw Scores from items 12, 13 and 14. (Sum – 3)/0.14
	13	Not at All Certain A Little Certain Somewhat Certain Very Certain Extremely Certain	1 2 3 4 5	
	14	Extremely Dissatisfied Very Dissatisfied Dissatisfied Somewhat Satisfied Satisfied Very Satisfied Extremely Satisfied	1 2 3 4 5 6 7	

* Questions 5 - 8 are answered only when the answer to question 4 is "Yes."

Appendix D. Scoring the WPAI:SHP

The Work Productivity and Activity Impairment Questionnaire: Specific Health Problem for AS Version 2.0 (WPAI) was developed to measure the effect of overall health and specific symptoms on productivity at work and outside of it. Six questions asked on this questionnaire can be described briefly as below:

Questions:

Q1. Currently employed? (Yes/No).

Q2. Hours missed from work due to Axial Spondyloarthritis.

Q3. Hours missed due to other reasons.

Q4. Hours actually worked.

Q5. Degree Axial Spondyloarthritis affected productivity while working.

Q6. Degree Axial Spondyloarthritis affected regular activities other than job.

The following 6 measures will be derived based on the responses from the 6 questions. The 4 main impairment scores (S1 to S4) are expressed as *percent impairment* based on the above questions.

Scores:

S0. Employment: defined below

S1. *Absenteeism*: Percent work time missed due to SpA:

$$100\times\left[\frac{Q2}{Q2+Q4}\right]$$

S2. *Presenteeism*: Percent impairment while working due to SpA:

$$100 \times \left[\frac{Q5}{10} \right]$$

S3. Percent overall work impairment due to AS:

$$100 \times \left[\frac{Q2}{Q2+Q4} + \left\{ 1 - \frac{Q2}{Q2+Q4} \right\} \times \frac{Q5}{10} \right]$$

S4. Percent activity impairment due to SpA:

$$100 \times \left[\frac{Q6}{10} \right]$$

S5. Did subject miss work (defined below). This is needed to derive the proportion of subjects who missed work.

Missing Data Handling Conventions

When calculating the WPAI:SHP scores, the following computational notes should be followed.

- Define Employment as a binary YES or NO variable where YES corresponds to "Employed" and NO corresponds to "Not Employed."
  - A subject will be considered "employed" at a given visit if Q1 = YES or Q2 > 0 or Q4 > 0.
  - A subject will be considered "unemployed" at a given visit if Q1 = NO and no positive hours recorded under Q2 and Q4 (i.e., if Q1 = NO AND Q2 ≤ 0 AND Q4 ≤ 0, then UNEMPLOYED).
  - Employment status for a subject will be considered "missing" at a given visit if Q1 = missing and no positive hours recorded under Q2 and Q4.
- If a subject is "unemployed" or employment status is "missing," then S1, S2, and S3 will be set to "missing."

- If Q2 = 0 and Q4 = 0 or missing then Q2/(Q2 + Q4) = missing (i.e., S1 = missing).
- If Q2 = 0 and Q4 = 0, then set S3 to missing.
- If Q2 is missing or Q4 is missing, then set S1 and S3 to missing.
- If Q4 = missing, then DO NOT set Q5 = missing.
- If Q5 is missing, then apply the following rules:
  - If Q2 > 0, Q4 = 0, and Q5 = missing, then S3 = 100%.
  - If Q2 = 0, Q4 > 0, and Q5 = missing, then S3 is missing.
  - If Q2 > 0, Q4 > 0, and Q5 = missing, then S3 is missing.
- Determine if a subject missed work (based on Q2) in order to analyze the proportion of subjects who missed work:
  - Create a binary (yes or no) "missed work" variable.
  - A subject will be considered as yes to missed work if Q2 is greater than 0.
  - If Q2 = missing, then MISSED WORK = missing.
  - If Q2 > 0, then MISSED WORK = "yes."
  - If Q2 = 0, then MISSED WORK = "no."
  - Therefore, the proportion of subjects who missed work will be counted based on the number of subjects with MISSED WORK = YES.

Appendix E. Scoring the RAPID-3

1. For patient's functional status (FN) sum up the scores in questions A - J only and use following formula to calculate formal score:

1 = 0.3	7 = 2.3	13 = 4.3	19 = 6.3	25 = 8.3
2 = 0.7	8 = 2.7	14 = 4.7	20 = 6.7	26 = 8.7
3 = 1.0	9 = 3.0	15 = 5.0	21 = 7.0	27 = 9.0
4 = 1.3	10 = 3.3	16 = 5.3	22 = 7.3	28 = 9.3
5 = 1.7	11 = 3.7	17 = 5.7	23 = 7.7	29 = 9.7
6 = 2.0	12 = 4.0	18 = 6.0	24 = 8.0	30 = 10

2. For patient's pain tolerance (PN) enter raw score of question 2.
3. For patient's global estimate (PTGE) enter raw score of question 3.
4. For Total score sum up FN, PN, PTGE.

Appendix F. ROC Analysis for the evaluation of RAMRIS scores

F.1. Index Score Variables

The index scores to predict flare status (with flare/without flare) are the Baseline hand and wrist synovitis ($P_{synovitis}$) and BME (P_{BME}) RAMRIS scores as well as a composite of both ($P_{composite\ of\ both}$).

F.2. Receiver Operating Characteristics (ROC) Analysis

The minimally acceptable prediction accuracy for the index scores will be assessed utilizing the Area Under the Curve (AUC) measure.

The value of AUC > 0.75 indicates a good predictive power of an index score to correctly predict absence and presence of flare.

ROC Analysis of Index Scores

For each RAMRIS index score (defined in Section 10.2.1) the AUC, its standard error, and the 95% confidence interval will be calculated. The AUC of an index score and the AUC of chance (AUC = 0.5) will be compared. The difference between AUCs will be calculated and its 95% confidence interval along with the p-value will be provided.

The following SAS code should be used to produce ROC curve, AUC estimates and to perform comparisons with the AUC of chance (AUC = 0.5), assuming DAS28(ESR) *flare_status* is coded as '0 = no flare vs. 1 = flare':

F.3. Optimal Cut-off and Index

Optimal Cut-off of Index Score

Optimal cutoff value providing the highest correct prediction rate will be determined for each index score $P_{synovitis}$, P_{BME}, $P_{composite}$ of both using logistic regression with a dichotomized DAS28(ESR) flare status variable as a dependent variable and the respective index score as an independent variable. From all index score values, the optimal cutoff value will be the one, which will provide the highest correct prediction rate according to the respective dichotomized DAS28(ESR) flare status variable. If more than one cutoff value is available with highest correct prediction rate, then for each of these candidate optimal cutoff values a frequency table with the dichotomized DAS28(ESR) flare status variable and the respective index score dichotomized at candidate optimal cutoff will be run and p-values of chi-square test will be displayed. The final optimal cutoff among the candidates will be the one with the smallest p-value.

The three optimal cut-offs $P^*_{synovitis}$, P^*_{BME}, $P^*_{composite}$ of both will be defined from this procedure.

Index of flare Status

The cut-offs $P^*_{synovitis}$, P^*_{BME}, $P^*_{composite\ of\ both}$ will dichotomize index scores and will define 3 indices of predicted flare status:

1. ($P_{synovitis} \leq P^*_{synovitis}$): Without flare versus ($P_{synovitis} > P^*_{synovitis}$): With flare
2. ($P_{BME} \leq P^*_{BME}$): Without flare versus ($P_{BME} > P^*_{BME}$): With flare
3. ($P_{composite\ of\ both} \leq P^*_{composite\ of\ both}$): Without flare versus ($P_{composite\ of\ both} > P^*_{composite\ of\ both}$): With flare

F.4. Statistical Measures of flare status Index resulting from RAMRIS scores performance

The following table contains the terms necessary to calculate index performance characteristics.

Table 31. Terms to define sensitivity, specificity and accuracy

Outcome predicted by index of RAMRIS index	Condition as determined by dichotomized DAS28(ESR) flare status variable		
	Positive (With flare)	Negative (Without flare)	Row Total
Positive (With flare)	True Positive (TP)	False Positive (FP)	TP + FP (Total number of subjects With flare predicted)
Negative (Without flare)	False Negative (FN)	True Negative (TN)	FN + TN (Total number of subjects Without flare predicted)
Column Total	TP + FN (Total number of subjects With flare condition)	FP + TN (Total number of subjects Without flare condition)	N = TP + TN + FP + FN (Total number of subjects)

Sensitivity = TP/(TP + FN) = (Number of true positive)/(Number of all positive conditions)

Specificity = TN/(TN + FP) = (Number of true negative)/(Number of all negative conditions)

Accuracy = (TN + TP)/(TN + TP + FN + FP) = (Number of correct predictions)/Number of all conditions)

Positive Predictive Value = TP/(TP + FP) = (Number of true positive)/(Number of all predictions of positive outcome)

Negative Predictive Value = TN/(TN + FN) = (Number of true negative)/(Number of all predictions of negative outcome)

Likelihood Ratio = Sensitivity/(1-Specificity)

Performance of flare status index resulting from RAMRIS scores

For each of the three dichotomized index scores of flare status defined in the Section 10.2.1, sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV) will be displayed along with 95% CIs, as well as likelihood ratio will be calculated.

14.0 References

1. van Herwaarden N, van der Maas A, Minten MJM, et al. Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis: open label, randomised controlled, non-inferiority trial. BMJ. 2015;350:h1389. doi:10.1136/bmj.h1389.
2. Fautrel B, Pham T, Alfaiate T, et al. Step-down strategy of spacing TNF-blocker injections for established rheumatoid arthritis in remission: results of the multicentre non-inferiority randomised open-label controlled trial (STRASS: Spacing of TNF-blocker injections in Rheumatoid ArthritiS Study). Ann Rheum Dis. 2016;75(1):59-67.
3. Iwamoto T, Ikeda K, Hosokawa J, et al. Prediction of relapse after discontinuation of biologic agents by ultrasonographic assessment in patients with rheumatoid arthritis in clinical remission: high predictive values of total gray-scale and power Doppler scores that represent residual synovial inflammation before discontinuation. Arthritis Care Res (Hoboken). 2014;66(10):1576-81.
4. Linacre, J. The expected value of a point-biserial (or similar) correlation. Rasch Measurement Transactions. 2008;22(1):1154.
5. Felson DT, Smolen JS, Wells G, et al. American College of Rheumatology/European League against Rheumatism preliminary definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum. 2001;63(3):573-86.
6. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A Users' Manual. Boston: The Health Institute. 1994.

Document Approval

Study M14500 - Statistical Analysis Plan Version 1 - 19Jul2018 (E3 16.1.9)

Version: 1.0 **Date:** 20-Jul-2018 02:12:57 PM **Company ID:** 07202018-00F9F683DF67D6-00001-en

Signed by:	Date:	Meaning Of Signature:
[redacted]	19-Jul-2018 02:02:40 PM	Author
	19-Jul-2018 11:19:48 PM	Approver
	20-Jul-2018 02:12:57 PM	Approver