Statistical Analysis Plan

Clinical Trial Protocol Identification No.	MS200588-0004
Title:	A multicenter, randomized, double-blind, phase III trial to evaluate the safety, immunogenicity, and efficacy of MSB11022 compared with Humira® in patients with moderately to severely active rheumatoid arthritis
Trial Phase	III
Investigational Medicinal Product(s)	MSB11022
Clinical Trial Protocol Version	11 July 2017/Version 2.0
Statistical Analysis Plan Author	PPD
Statistical Analysis Plan Date and Version	06 September 2018/Version 3.0 Final
Statistical Analysis Plan Reviewers	PPD.
	PPD, Medical Responsible, Fresenius Kabi SwissBioSim.
	PPD, PPD, Fresenius Kabi SwissBioSim.
	PPD, PPD, Fresenius Kabi PPD.
	PPD, PPD, Fresenius Kabi SwissBioSim.
	PPD, PPD, PPD.
	PPD, PPD, Fresenius Kabi SwissBioSim.
	PPD, PPD, EMD Serono.

PPD .

PPD , Medical Writer, PPD .

PP , PPD

, PPD .

1 Signature Page

Statistical Analysis Plan: MS200588-0004

A multicenter, randomized, double-blind, phase III trial to evaluate the safety, immunogenicity, and efficacy of MSB11022 compared with Humira® in patients with moderately to severely active rheumatoid arthritis

Approval of the SAP by Merck/EMD-Serono Responsible is documented within ELDORADO. Wet ink signature outside Eldorado – for PPD use only.

PPD

Trial Biostatistician   Signature

PPD   Date

PPD

_____

PPD

PPD   Signature

EMD Serono Research & Development Institute, Inc.   Date

45A Middlesex Turnpike, Billerica, United States of America

PPD

PPD   Signature

PPD   Date

PPD

PPD

PPD   Signature

PPD   Date

EMD Serono Research & Development Institute, Inc.

45A Middlesex Turnpike, Billerica, United States of America

2　　　　　　　Table of Contents

CCI

CCI

3   List of Abbreviations and Definition of Terms

ACR	American College of Rheumatology
ACR20	American College of Rheumatology 20% Response Criteria
ACR50	American College of Rheumatology 50% Response Criteria
ACR70	American College of Rheumatology 70% Response Criteria
ADA	Anti-drug Antibody
ADaM	Analysis Data Model
AE	Adverse Event
AESI	Adverse Event of Special Interest
ANA	Anti-Nuclear Antibody
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BOA	Biostatistical Output Assembly
CDAI	Clinical Disease Activity Index
CI	Confidence Interval
C_{max}	Maximum Serum Concentration
CRP	C-reactive protein
CTMS	Clinical Trial Management System
CTP	Clinical Trial Protocol
CSR	Clinical Study Report
C_{trough}	Trough Serum Concentration
CV%	Coefficient of Variation
DAS28-ESR	Disease Activity Score Based on a 28 Joint Count
DBP	Diastolic Blood Pressure
dsDNA	Double stranded deoxyribonucleic acid
eCRF	Electronic Case Report Form
EMA	European Medicines Agency
ECG	Electrocardiogram
EQ-5D-5L	European Quality of Life 5-dimensions and 5-levels questionnaire
ESR	Erythrocyte Sedimentation Rate
ETA	Greek letter η

EULAR	European League Against Rheumatism
FDI	Functional Disability Index
GBSOS	Global Biostatistics Output Standards
GCP	Good Clinical Practice
HAQ-DI	Health Assessment Questionnaire Disability Index
ICH	International Council on Harmonisation
IDDI	International Drug Development Institute
IDMC	Independent Data Monitoring Committee
IMP	Investigational Medicinal Product
ITT	Intent-to-Treat
IWRS	Interactive Web Response System
LLOQ	Lower Limit of Quantification
LTBI	Latent Tuberculosis Infection
MedDRA	Medical Dictionary for Regulatory Activities
MCS	Mental Component Score
NAb	Neutralizing Antibodies
NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse Events
PCS	Physical Component Score
PGA	Physician's Global Assessment
PK	Pharmacokinetic(s)
PP	Per-Protocol
PT	Preferred Term
QoL	Quality of Life
RA	Rheumatoid Arthritis
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
sc	Subcutaneous
SD	Standard Deviation
SDAI	Simplified Disease Activity Index
SDTM	Study Data Tabulation Model

SEM	Standard Error of the Mean
SF-36	36-item Short Form Health Survey
SJC	Swollen Joint Count
SMQ	Standardized MedDRA Query
SOC	System Organ Class
SBP	Systolic Blood Pressure
TB	Tuberculosis
TJC	Tender Joint Count
T_{max}	Time To Reach Maximum Serum Concentration
TNF	Tumor Necrosis Factor
TEAE	Treatment Emergent Adverse Event
ULOQ	Upper Limit of Quantification
VAS	Visual analogue score
WHODrug	World Health Organization Drug

4 Modification History

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
Draft 0.1	27 January 2017	PPD	N/A - original
Draft 0.2	24 February 2017	PPD	Sponsor comments incorporated
Draft 0.3	28 April 2017	PPD	Sponsor comments incorporated
Draft 0.4	11 August 2017	PPD	Sponsor comments incorporated. Updated as per Protocol amendment 2
Draft 0.5	06 October 2017	PPD	Sponsor comments incorporated
Final 1.0	08 December 2017	PPD	Sponsor comments incorporated
Final 2.0	29 June 2018	PPD	New cutoff rule for Final analysis added
Final 3.0	06 September 2018	PPD	Update to take into account one unique database lock after 4-month safety follow-up

5 Purpose of the Statistical Analysis Plan

The purpose of this statistical analysis plan (SAP) is to document technical and detailed specifications for the analysis of data collected for protocol MS200588-0004. Results of the analyses described in this SAP will be included in the Clinical Study Report (CSR). The CSR will include the 4-week safety follow-up visit (including Week 52 data analysis) and the 4-month safety follow-up visit. Any post-hoc or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective SAP will be clearly identified in the given CSR. Additionally, the planned analyses identified in this SAP will be included in future manuscripts.

The SAP is based upon Section 8 (Statistics) of the Clinical Trial Protocol (CTP) Version 2.0, dated 11 July 2017. It is prepared in compliance with the International Council on Harmonisation (ICH) E9.

The Independent Data Monitoring Committee (IDMC) SAP is also included in this document in Section 18.

6 Summary of Clinical Trial Features

Objectives:

Primary objective

The primary objective of this study is to evaluate the safety profile of MSB11022-CCI

compared to Humira in patients with moderately to severely active rheumatoid arthritis (RA) up to Week 52.

Secondary objectives

The key secondary objective is to compare the efficacy of MSB11022-CCI to Humira at Week 12 in patients with moderately to severely active RA.

Other secondary objectives include:

- To evaluate the immunogenicity profile of MSB11022-CCI compared to Humira in patients with moderately to severely active RA up to Week 52

- To further compare the efficacy and safety of MSB11022-CCI compared to Humira in patients with moderately to severely active RA up to Week 52

- To compare quality of life (QoL) and physical function on MSB11022-CCI with Humira in patients with moderately to severely active RA

- To compare injection site pain levels of MSB11022-CCI versus Humira

Exploratory objective

The exploratory objective is to evaluate population pharmacokinetics (PK) of MSB11022-CCI and Humira in patients with moderately to severely active RA

Methodology: This trial is a two-arm, randomized, multicenter, double-blind, parallel group trial designed to compare the safety, immunogenicity, and efficacy of MSB11022-CCI with Humira in approximately 260 randomized patients with moderately to severely active RA during a 52 week period.

Baseline is defined as the day on which the first dose of the investigational medicinal product (IMP) (blinded study drug) is administered (Week 0). The trial will include a pre-trial evaluation period (screening period, from 28 days to 3 days prior to drug administration), a double-blind 48-week treatment period, a 4-week safety follow-up visit, and a 4-month safety evaluation period. The primary safety endpoint will be assessed up to Week 52, key secondary efficacy endpoint will be assessed at Week 12.

Re-screening will be allowed once for those patients who do not meet the inclusion/exclusion criteria within the above specified time limits. The Medical Monitor must promptly inform the Sponsor about requests and reasons for re-screening. An IDMC will review the safety data from this trial on an ongoing basis.

Eligible patients will be randomized in permuted blocks in a 1:1 ratio by an interactive web response system (IWRS) to receive either MSB11022-CCI subcutaneous (sc) or Humira sc at a dose of 40 mg every other week starting at baseline up to and including Week 48. Randomization will be stratified by the type of systemic therapy previously received: non-biological (biologic naïve patients) versus biological (biologic experienced patients). Patients who previously received both (biological and non-biological systemic therapies) will be assigned to the "biological" group. The exposure to previous biological agents will be limited to one tumor necrosis factor (TNF) inhibitor other than adalimumab. The

participation of patients in the previous biological systemic therapy stratum will be capped at 20% of the total number of patients randomized. Capping will be enforced through application of an IWRS.

The first 2 doses of IMP will be administered on site by the Investigator or other qualified personnel and the patient will be educated on the correct process. The third dose of IMP will be administered on-site by the patient. Patients will be monitored for 1 hour following the first 3 IMP administrations. If the Investigator judges it appropriate after proper training, the patient (or a caregiver) may continue to self-inject/inject the treatment for the remaining doses. Injection site pain will be assessed after doses 3 to 5. If a patient self-administers, the injection should not be given in the arm. The remaining doses of IMP will be dispensed, and written instructions on proper dosage, administration, storage, and recording will be provided to the patient/care-giver. Administrations performed at home will be recorded in a diary with the accurate dosing information (dosing date and time).

The 52-week double-blind period will allow for the collection of long-term comparative safety, immunogenicity, and efficacy data for MSB11022-CCI versus Humira.

Subjects who achieve less than a 20% improvement in both swollen and tender joint counts at Week 24 (non-responders) will be discontinued from the IMP. These subjects will remain in the trial safety analysis set and participate in the safety and immunogenicity assessments. After Week 24, patients with less than 20% improvement in both swollen and tender joint counts at any scheduled visit up to Week 52 will also be discontinued. Subjects may continue at Investigator discretion from Week 24 to Week 52 as long as they maintain 20% or more improvement in both swollen and tender joint counts.

Patients who discontinue from study medication (but remain in the trial and continue for the safety/immunogenicity visits) before the Week 52 visit will immediately complete an Early Termination visit and return for a Safety/Follow-up visit four weeks after the last dose of IMP. During this 4-week Safety Follow-up period, no excluded treatment for RA (as defined in the protocol) should be administered. Patients will return for an additional safety evaluation four months after the last dose of IMP. Treatment for RA from the 4-week Safety Follow-up visit to the 4-month Safety Evaluation visit will be at the discretion of the Investigator following institutional standard of care. The 4-week and 4-month Safety Follow-up visits will not be duplicated if either of these visits falls within two weeks of an otherwise scheduled visit. If the 4-week and 4-month Safety Follow-up since last IMP are completed within the 52 week visit schedule, these will not be repeated thereafter.

All subjects will have PK samples taken pre-first dose, and then prior to dosing on Weeks 2, 4, 12, 24, 36 and 52. In addition, a subset of 60 subjects (30 per arm) will have additional samples to support the population PK analysis taken on Days 2 (24 h post first dose), 4, and 9, and Days 2 (24 h post Week 24 dose), 4, 9, and 14 (Week 26 pre dose) after the Week 24 dose.

The anticipated duration of the entire trial is approximately two years.

Visit schedules for safety, immunogenicity, and efficacy assessments are detailed in the

Schedule of Assessments in the protocol.
Planned number of patients: Approximately 260 randomized subjects
Primary endpoint: • Treatment-emergent adverse events of special interest (AESI) including and up to Week 52

Secondary endpoints:

Key secondary endpoint

- American College of Rheumatology 20% response criteria (ACR20) at Week 12. An ACR20 response is defined as an improvement of at least 20% in the number of tender joints and swollen joints, and at least 20% improvement in three out of the remaining five ACR core-set measures: patient assessment of arthritis pain, Patient's and Physician's Global Assessment of Disease Activity, physical function via the Health Assessment Questionnaire-Disability Index [HAQ-DI] and acute-phase reactant as measured by C-reactive protein (CRP).

Other secondary endpoints

Safety

- Occurrences of treatment-emergent adverse events (TEAEs) and serious adverse events (SAEs) including and up to Week 52

- Clinical laboratory values including hematology, chemistry, and urinalysis

- Vital signs

- Physical exam

- 12-lead electrocardiogram (ECG)

Immunogenicity

- Anti-drug Antibody (ADA) to adalimumab and ADA titer including and up to Week 52

- Neutralizing antibodies (NAbs) to adalimumab including and up to Week 52

Efficacy

- ACR20 at Weeks 2, 4, 8, 24, and 52

- American College of Rheumatology 50% Response Criteria (ACR50) and American College of Rheumatology 70% Response Criteria (ACR70) at Weeks 2, 4, 8, 12, 24, and 52

- Disease Activity Score based on a 28 joint count DAS28-ESR mean change from baseline at Weeks 2, 4, 8, 12, 24, and 52

- Proportion of patients with low disease activity as measured by DAS28-ESR, and remission at Weeks 2, 4, 8, 12, 24, and 52

- Clinical Disease Activity Index (CDAI) and Simplified Disease Activity Index (SDAI)

mean change from baseline at Weeks 2, 4, 8, 12, 24, and 52

- ACR/European League Against Rheumatism (EULAR) Boolean remission rates at Weeks 2, 4, 8, 12, 24, and 52

Quality of life

- HAQ-DI questionnaire at screening, baseline, and Weeks 2, 4, 8, 12, 24, and 52

- 36-item Short Form Health Survey (SF-36) questionnaire at baseline and Weeks 12, 24, and 52

- Euro-Quality of Life – 5 Dimensions and 5-levels (EQ-5D-5L) questionnaire at baseline and Weeks 12, 24, and 52

Injection-site pain

- Mean change in injection site pain on a visual analog scale (VAS), evaluated during three administrations of IMP (doses 3 to 5) (immediately post-injection, 15 minutes post-injection, and 1 hour post-injection)

Exploratory endpoints

Population pharmacokinetics

- Absorption profile characterization, if it is supported by the data

- Apparent clearance (CL/F)

- Volume of distribution (Vz/F)

- C_{trough} levels at Day 14 after first dose, at Week 4, 12 and 24 pre-dose, and at Day 14 after dose of Week 24

Planned trial and treatment duration per subject: The planned trial duration per subject is approximately 17 months: a screening period of 28 days to 3 days prior to IMP administration, a double-blind 48-week treatment period to evaluate long-term safety and immunogenicity, a safety follow-up visit four weeks after the last dose of IMP, and an additional safety evaluation at four months following the last dose of IMP. During the 4-week Safety Follow-up period, no excluded treatment for RA (as defined in the protocol) should be administered. Treatment for RA from the 4-week Safety Follow up visit to the 4-month Safety Evaluation visit will be at the discretion of the Investigator following institutional standard of care.

7 Sample Size/Randomization

7.1 Sample Size

According to the European Medicine Agency (EMA) recommendation (EMA/CHMP/SAWP/ 200743/2016), a total of 100 patients/arm should be adequate for a study primarily focusing on safety. To account for drop outs, this descriptive study will include 260 randomized patients in total (130/arm) in order to ensure 200 subjects are in the study at Week 52. In fact, based on averaged originator's dropout rates at Week 52, a 20% rate of drop-out overall is included in the sample size calculation (1, 2, 3). This sample size will permit an informative assessment of the

occurrence of most common AESIs (> 1/10) such as injection site reactions and rash with a precision of 12.0% and 9.0% for the AESI incidence of 26.0% and 11.0% respectively (0).

A subset of 60 patients (30 per treatment arm) will be randomly selected across all sites to participate in population PK analysis.

7.2 Randomization

Eligible subjects will be randomized in permuted blocks in a 1:1 ratio by an IWRS to receive either MSB11022-CCI sc or Humira sc at a dose of 40 mg every other week starting at baseline up to and including Week 48.

Randomization will be stratified by previous disease-modifying therapy use. Subjects will be stratified by the type of systemic therapy previously received: non-biological (biologic naïve subjects) versus biological (biologic experienced subjects). Subjects who previously received both (biological and non-biological systemic therapies) will be assigned to the "biological" group. The exposure to previous biological agents will be limited to one TNF inhibitor other than adalimumab. The participation of subjects in the previous biological systemic therapy stratum will be capped at 20% of the total number of subjects randomized.

8 Overview of Planned Analyses

Once the last subject has completed the 4-month safety follow-up visit, the database will be locked. The analysis will be performed using cleaned study data collected in electronic case report form (eCRF) or transferred from laboratories performing laboratory tests and biosample analysis until the date of the 4-month safety follow-up visit of the last subject. Once the data cleaning is complete, a data review meeting will be held prior to the database lock. In addition, no data can be frozen or locked and no randomization code should be unblinded until this Statistical Analysis Plan (SAP) has been approved.

The final analysis will include all data collected during the study up to week 52 date (or end of treatment date+28 days, whichever occurs later) for subjects who completed the study, and theoretical week 52 (ie 52 weeks after first intake) or 4 months safety follow-up visit, whichever occurs first, for subjects who early terminated the treatment.

All AE tables and listings will be re-produced to reflect any additional AEs and any updates to the reported AEs in final analysis. Additionally the concomitant medication and procedure tables and listings will be re-produced to reflect any updates.

In addition, an independent data monitoring committee (IDMC) will review safety and tolerability data of this study on an ongoing basis. The methodology and presentation of results for the IDMC are covered in the IDMC Charter and in Section 18.

9 Changes to the Planned Analyses in the Clinical Trial Protocol

Not applicable.

10 Protocol Deviations and Analysis Sets

10.1 Definition of Protocol Deviations

Protocol deviations occurring during the study will be determined for all enrolled subjects, mainly from the clinical database by either clinical and/or medical review processes, and entered into the Clinical Trial Management System (CTMS).

The mapping of the protocol deviations from the CTMS to analysis will be performed as per the table below:

PD Guidance	CTMS	SAP	SDTM	ADaM
Minor	Minor	Not required	Not included	Not included
Important	Major	Important	Important - flagged with PDEVXXX code	Important - flagged with PDEVXXX code
Priority (subset of Important)	Critical (subset of Major)	Important	Mapped to Important	Mapped to Important
NA	NA	Clinically Important (lead to exclusion from PP pop)	flagged with PDEVXXX code	additional flag = Y/N in ADDV (may also include conditions not considered to be PD's)

Important protocol deviations are protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being.

Important protocol deviations may also be recorded as 'major' protocol deviations in the CTMS but will presented only as important in the analysis output.

Important protocol deviations include:

- Subjects that are dosed on the study despite not satisfying the entry criteria;
- Subjects that develop withdrawal criteria whilst on the study but are not withdrawn;
- Subjects that receive the wrong treatment or an incorrect dose;
- Subjects that receive an excluded concomitant medication;
- Deviation from Good Clinical Practice (GCP).

Clinically Important Protocol Deviations are a subset of important protocol deviations which lead to the exclusion of a subject from the Per-Protocol (PP) analysis set.

The following deviations will be identified and confirmed prior to the partial database lock for the final analysis.

- Important protocol deviations including

  - Deviations from the inclusion and exclusion criteria

  - Deviations post inclusion

Protocol deviations may be identified by the data-managers, clinical and medical staff either by programmed validation checks or data listings/reports or manual verification of data sources.

Some important/major protocol deviation criteria may be identified in the clinical database via biostatistical programs. All important protocol deviations will be documented in SDTM datasets whether identified through sites monitoring, medical review or programming. Important protocol deviations to be identified by programming as well as all clinically important protocol deviations are listed and described in Appendix I. The number and percentage of subjects in the ITT Analysis Set with important protocol deviations leading to exclusion from the PP Analysis Set will be presented by treatment arm and reason for exclusion. All important protocol deviations will be presented in a data listing.

10.2   Definition of Analysis Sets and Subgroups

The following table provides an overview of the analysis sets to be used for the different analyses:

Table 1: Overview of Analysis Set Use

Analyses	Enrolled	ITT	PP	SAF	PK
Disposition • Before Randomization • After Randomization	✓	✓			
Analysis Sets	✓				
Reason for exclusion from PP set		✓			
Demographics and Other Baseline Characteristics		✓			
Previous and Concomitant Medications/ Procedures		✓	✓		
Compliance and Exposure			✓	✓	
Key Efficacy: ACR20		✓	✓		
Other efficacy endpoints		✓	✓		
QoL		✓	✓		
Injection site pain		✓		✓	
Immunogenicity				✓	
AESIs				✓	
Other safety data				✓	
Pharmacokinetics				✓	✓

ITT = Intent-to-treat, PP = Per-Protocol, SAF = Safety, PK = Pharmacokinetic

10.2.1   Enrolled Analysis Set

The Enrolled Analysis Set includes all subjects who signed the informed consent form.

10.2.2   Intent-to-Treat Analysis Set

The Intent-To-Treat (ITT) Analysis Set includes all subjects randomly allocated to a treatment, based on the intention to treat "as randomized" principle (ie, the planned treatment regimen rather than the actual treatment given in case of any difference). Subjects will be analyzed according to their randomized treatment.

In the event that subjects are incorrectly stratified, ie subjects are randomized according to the incorrect stratification, these subjects will be analyzed under the randomized treatment but with the correct stratification value.

10.2.3   Per-Protocol Analysis Set

The PP analysis set is defined as a subset of the ITT and will include subjects who are compliant with the Clinical Trial Protocol and do not meet criteria that could impact the key objective of the study.

Subjects will be analyzed according to their randomized and received treatment, as receipt of a different treatment from that assigned is an important protocol deviation.

Subjects who meet one of the following criteria will be excluded from the PP set:

- Incorrect treatment group allocation, different to assignment at randomization.
- Unavailability of measurements of the key secondary endpoint.
- Non-compliance with key entry criteria.
- Presence of clinically important protocol deviations.
- Not adequate compliance with and not sufficient exposure to study medication;
  - o Drop out before Week 12
  - o Not fully compliant to IP administration during the first 12 weeks of treatment(< 80% compliant).
- Incorrect stratification at randomization with respect to previous systemic therapy.

The criteria for exclusion from the PP analysis set are defined in detail in Appendix I.

10.2.4   Safety Analysis Set

The Safety Analysis Set (SAF) will include all randomized subjects who receive at least one dose of study treatment. Subjects will be analyzed according to the actual treatment they receive during the study. All safety and subject-level PK analyses will be performed using this analysis set.

10.2.5   PK Analysis Set

The Pharmacokinetic (PK) Analysis Set will include all subjects who:

- receive at least one dose of study treatment and

- have at least one valid post-dose pharmacokinetic assessment.

Important events which may impact the quality of the PK data include, but may not be limited to, sample processing errors that lead to inaccurate bioanalytical results, and/or inaccurate dosing on the day of PK sampling. In the case of such events, PK samples collected during the affected treatment will be excluded from descriptive statistics, but will still be included in the listings.

Analyses exploring associations between (observed) PK and immunogenicity, and between (observed) PK and efficacy, will also use the PK Analysis Set.

Subjects in the PK Analysis Set will be analyzed according to the actual treatment received during the study.

10.2.6 Subgroups

Subgroup analyses will be performed for the secondary endpoints ACR20 and ACR50 (insufficient data expected for ACR70) on subgroups as defined below, using the ITT analysis set.

The following subgroups will be defined:

- Previous Systemic Therapy Use

  - Biological

  - Non-biological only

- Age

  - < 65 years

  - ≥ 65 years

- Sex

  - Male

  - Female

- Weight

  - < 90kg

  - ≥ 90 kg

- BMI

  - < 25 kg/m^2

  - ≥ 25 kg/m^2 - < 30 kg/m^2

  - ≥ 30 kg/m^2

- Race

  - White

  - Other

- Ethnicity

  - Hispanic or Latino

  - Not Hispanic or Latino

- Time since first diagnosis of rheumatoid arthritis

  - <5 years

  - ≥ 5 years

11 General Specifications for Statistical Analyses

Descriptive statistics will be given with reference to the analysis set deemed appropriate for each particular endpoint. Unless otherwise indicated, all analyses/summaries will be presented by the two treatment arms, MSB11022 and Humira. Note that this document refers to treatment arms; however, the output shells present 'treatment group', as defined in the standard templates.

All statistical analyses will be performed using SAS® Version 9.2 or higher.

11.1 Descriptive Methods

Descriptive statistics for continuous data will include the following:

- Number of subjects with non-missing data (n)

- Number of subjects with missing data

- Mean, standard deviation (SD)

- Median, first and third quartiles (Q1 and Q3)

- Minimum and maximum

- 95% confidence interval (CI).

All statistics will be rounded according to Merck's Global Biostatistics Output Standards (GBSOS) Booklet (Version 4.0 or later), ie, percentages will be rounded to one decimal. The mean, median, Q1, Q3, and 95% CI will be presented with one additional decimal compared to the raw data; SD will be presented with two additional decimals compared to the raw data; and minimum and maximum will be presented with the same number of decimals as the raw data.

All data will be presented by nominal visit (including the safety follow-up visits – four weeks and four months after the last dose of IMP) and no visit windowing rules will be implemented. Data from unscheduled visits will be listed but will not be included in summaries presented by

visit. The exceptions will be in the calculation of the baseline values. Data from unscheduled visits will also be included in the calculation of the worst on-treatment values and grades.

Categorical data will be summarized by presenting the number of subjects with non-missing data (n), the number of subjects with missing data (missing), and the number and percentage of subjects in each category. Percentages will be based on the number of subjects with non-missing data in the analysis set of interest, unless otherwise specified.

Time to event variables will be summarized by presenting Kaplan Meier estimates and 95% CIs.

Change from baseline will be calculated as the post-baseline value minus the baseline value. Percent change from baseline is defined as (100 x (post-baseline – baseline) / baseline).

If all baseline values are missing for a particular variable, the change from baseline and percent change/percent improvement from baseline will not be calculated.

A 2-sided significance level of 5% will be used for all analyses. Two-sided 95% confidence intervals will be calculated.

11.2   Inferential Methods

There will be no adjustment for multiple testing because there are only two treatment arms and a single primary endpoint. All other analyses (key secondary and other secondary endpoints) are considered as supportive to the primary endpoint.

11.3   Listings

All data will be presented in subject listings. All listings produced for the analyses will be ordered by treatment arm (MSB11022, Humira), subject identifier, visit, and date, unless otherwise specified.

Planned treatment arms will be presented except for treatment exposure and safety listings where actual treatment will be shown.

11.4   Handling of Missing Data

For historical data collected at screening, partial dates will not be imputed. The exception is deriving time from diagnosis in months where the 1st of the month is imputed if day is missing, and the 1st January is imputed if both day and month are missing. These imputed dates will not be shown in the corresponding listing.

The handling of partial dates of birth is covered in the Demographics Section 12.2. Handling of partial or missing dates for concomitant medications and AEs is covered in Sections 13 and 16.1 of this SAP, respectively. Handling of missing date/time for study drug related injection site reaction analysis is covered in Section 16.1.1.

No imputation will be performed for PK or immunogenicity data.

11.4.1 Imputation Methods for Key Secondary Endpoint

The analysis of ACR20 at Week 12 on the ITT analysis set will be repeated using a non-responder method to impute any missing data. This method will classify a subject with missing ACR20 at Week 12 as a non-responder at that timepoint.

An imputation for withdrawal will also be performed, defined as follows: if a subject discontinues prematurely prior to Week 12, the ACR20 will be imputed as a non-responder, while subjects who complete the study but have a missing assessment for ACR20 at Week 12 will not be imputed.

11.5 Data handling after cut-off date

Details can be found in Section 8.

11.6 Trial Day

Trial Day 1 is defined as the day on which the first dose of study treatment is administered (Day 1/Week 0), or the day of randomization for ITT subjects who did not receive any study treatment. All other trial days will be calculated relative to that date. For example, the day before the start of study treatment is Day -1 (there is no Day 0).

11.7 Baseline Definitions

Baseline is defined as the last non-missing assessment prior to or on Trial Day 1. Exceptions are assessments on the date of first dose of trial treatment where assessment time is or should be reported. In these situations: if the time is before the treatment start time then the assessment is baseline, if the time is after the treatment start time then the assessment is post baseline. If the time is at the same time as starting treatment and is from a scheduled Screening or Day 1/Week 0 assessment then the assessment is baseline, otherwise it is post baseline. If the time is missing and there is no available baseline value meeting the prior criteria then the value will be used as baseline if it is from a scheduled Screening or Day 1/Week 0 assessment. These baseline data will be used in all demographic and baseline summaries.

For vital signs that have been measured twice at the same visit, the average value per visit will be calculated before determining the last non-missing assessments to use as the baseline.

11.8 Data Derivation

The following conversion factors will be used to convert days to months or years, where applicable:

- 1 month = 30.4375 days
- 1 year = 365.25 days.

See Section 12.2 for the calculation of subject age.

Duration will be calculated by the difference of start and stop date + 1 (eg AE duration (days) = stop date of AE – start date of AE + 1) (if not otherwise specified). Durations are calculated only when both dates are available (imputed dates cannot be used for the duration computation).

See Section 14 for the calculation of treatment duration.

The time since an event (eg time since first diagnosis) will be calculated as reference date minus date of event.

12 Trial Subjects

The subsections in this section include specifications for reporting subject disposition and treatment/study discontinuations. Additionally, procedures for reporting protocol deviations are provided.

12.1 Disposition of Subjects and Discontinuations

Summaries of the number and percent of subjects to enter, complete or discontinue study and treatment will be presented.

The number and percent of subjects for the following will be presented by original randomized treatment arm and overall:

- Number of subjects screened

- Number of subjects not randomized, and by reason

- Number of subjects randomized

- Number of subjects who have not received treatment after being randomized

- Number of subjects who completed or prematurely discontinued treatment, and by reason

- Number of subjects who prematurely discontinued the study after randomization, and by reason

- Number of subjects included in the PK sub-trial and the number who discontinued from the PK sub-trial.

Further summaries will also present the number of subjects in each of the analysis sets overall and by site. A summary and a listing of the reasons for exclusion from the PP analysis set will also be presented.

The results of the randomization algorithm according to IWRS will be summarized as follows:

- Number of randomized subjects overall, by country

- Number of randomized subjects by the previous systemic therapy use (non-biological or biologic naïve subjects versus biological or biologic experienced subjects) randomization strata (IWRS).

12.2　　　　　　Demographics

Age at informed consent, sex, race and ethnicity will be summarized by treatment arm and overall for the ITT and the PP analysis sets. No formal comparison will be made between the treatment arms.

Age (years) at informed consent will be calculated as:

(Date of informed consent - date of birth + 1)/ 365.25

Specifications for computation of partial dates:

- In case of missing day only: Age [years]: (year/month of given informed consent - year/month of birth)

- In case only year of birth is given: Age [years]: (year of given informed consent - year of birth).

12.3　　　　　　Medical History

Medical history data will be summarized by system organ class (SOC) and preferred term (PT) by treatment arm and overall for the ITT and PP analysis sets. Medical history terms will be coded using version 19.1 (or later) of the Medical Dictionary for Regulatory Activities (MedDRA). Listings of medical history will be sorted by treatment arm, subject identifier, start date of medical history event, SOC, PT and end date of medical history event.

12.4　　　　　　Other Baseline Characteristics

History of rheumatoid arthritis (time (in months) since diagnosis relative to informed consent) will be summarized by treatment arm and overall for the ITT and PP analysis sets. Baseline assessments for the ACR component variables (tender joint count (TJC), swollen joint count (SJC), Patient's and Physician's Global Assessment of Disease Activity, Patient's Assessment of Arthritis Pain, Patient's Assessment of Physical Function as measured by the HAQ-DI and CRP), DAS28-ESR, CDAI, SDAI and tuberculosis assessment at baseline will also be summarized by treatment arm and overall for the ITT and PP analysis sets.

Methotrexate dose at baseline will be summarized by treatment arm and overall for the ITT and PP analysis sets.

Height (cm), weight (kg), BMI (kg/m^2), vital signs (body temperature, heart rate, respiration rate and blood pressure), and ECG interpretation at baseline will be summarized by treatment arm for the ITT and PP analysis sets.

Height will be presented in centimeters (cm) and weight will be presented in kilograms (kg). The conversion factors used will be:

- Height in cm = Height in inches x 2.54

- Weight in kg = Weight in pounds / 2.2046

ADA status, ADA titer and NAbs status at baseline will be summarized by treatment arm for the ITT and the PP analysis sets.

13   Previous or Concomitant Medications/Procedures

Medications which are taken by subjects other than study medications (MSB11022 or Humira) will be defined as previous or concomitant medication.

Previous medications will be defined as those that started and stopped before the first dose of study medication.

Concomitant medications will be defined as those that started before or during the treatment period (and up to the 4 week Safety Follow-up visit or 28 days after the last dose of study drug, whichever comes later) and finished during the treatment period or are ongoing.

In case the date values will not allow to unequivocally allocate a medication as previous or concomitant, the medication will be considered as a concomitant medication.

Previous and concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) Level 2 and PT. Medications will be coded by the Enhanced World Health Organization Drug (WHODrug) dictionary, version September 2016 or later. In case multiple ATC's are assigned to a drug, all ATC-2nd level will be used for reporting.

Procedures will also be classed as concomitant using the same criteria defined above. Procedures will not be coded but the number of procedures overall and the name of procedure and reason for procedure will be summarized for each treatment arm.

14   Treatment Compliance and Exposure

Compliance and exposure data will be summarized by treatment arm. The following will be derived:

- Treatment duration (in weeks): calculated as the number of weeks from the first dose to the last dose in the study. Treatment duration (weeks) is defined as:

$$\text{duration} = \left(\frac{\text{date of last dose - date of first dose} + y}{7 \text{ days}} \right)$$

where

$y = 7$ for subjects whose last dose was the Week 0 dose

$y = 14$ otherwise

- Total number of injections administered: Any administration of study medication regardless of the actual dose (any dose > 0 mg) will be counted as one injection. These data will come from the eCRF.

- Compliance will be estimated as the proportion of the planned number of injections actually administered. Compliance will be summarized as a continuous variable, as well as the number and percent of subjects with $\leq 80\%$, $> 80 - \leq 100\%$, $> 100\% - \leq 120\%$ and $> 120\%$ compliance.

Compliance (%) = 100 * (total number of injections received/ number of planned injections). For subjects who withdraw from study treatment prematurely, the number of planned injections will be the number planned up to the point of withdrawal.

15   Endpoint Evaluation

The subsections in this section include specifications for analyzing clinical trial endpoints specified in the CTP to meet the trial objectives.

15.1   Primary Endpoint Analyses

The primary endpoint is the occurrence of treatment-emergent AESIs (hypersensitivity) up toWeek 52.

The primary analysis will be based on the safety analysis set.

The following is considered pre-defined AESI for this trial:
Hypersensitivity (defined by Standardized MedDRA Queries [SMQs] narrow as per the latest MedDRA version).

The proportions of subjects with AESIs (hypersensitivity) and accompanying 95% CIs (Exact 95% CIs, calculated with the Clopper-Pearson method) will be provided for each treatment arm. Half width of 95% CIs for the difference between treatment arms in occurrence for the most commonly observed clinical events of AESI (hypersensitivity) will also be presented. The cut-off point for the selection of the most commonly observed clinical events will be decided prior to the database lock and the treatment unblinding.

Listings of all AESIs (hypersensitivity) and listings of AESIs (hypersensitivity) with NCI-CTCAE Grade ≥ 3 will be provided.

15.2   Secondary Endpoint Analyses

15.2.1   Key Secondary Endpoint Analyses

The key secondary endpoint is ACR20 at Week 12.

The ACR20 is a key efficacy measure for which a subject must have at least 20% improvement from baseline in the following ACR Core Set values:

- TJC (68 joint count) and

- SJC (66 joint count) and

- An improvement of at least 20% from baseline in at least three of the following five assessments:

  - Patient's Global Assessment of Disease Activity on a visual analog scale (VAS)

  - Patient's Assessment of Arthritis Pain

  - Patient's Assessment of Physical Function as measured by the HAQ-DI

  - Physician's Global Assessment of Disease Activity

  - Acute phase reactant as measured by CRP.

The proportion of subjects with an ACR20 response at Week 12 in the two treatment arms will be reported along with Exact 95% CIs, calculated with the Clopper-Pearson method. The 95% CI for the difference between treatment arms in the proportion of subjects with an ACR20 response at Week 12, calculated with the Newcombe method, will also be presented.

The analyses will be carried out on both the ITT and the PP analysis sets.

Proportions for the two treatment arms will also be presented broken down by the subgroups for the ITT analysis set as described in Section 10.2.6.

Missing data will be handled as described in Section 11.4.1.

15.2.2   Other Secondary Efficacy Endpoint Analyses

Other secondary efficacy endpoints will be summarized descriptively according to the type of outcome. ACR50 at Week 12 will be presented by subgroups as defined in Section 10.2.6 for ITT analysis set, but there will be no breakdown by subgroup for the remaining efficacy endpoints. All analyses will be carried out on both the ITT and the PP analysis sets.

15.2.2.1   ACR20 at Weeks 2, 4, 8, 24, and 52

These endpoints will be analyzed as per the key secondary endpoint (see Section 15.2.1).

Additionally, time to achieve ACR20 response will be will be measured from Trial Day 1. The subjects who did not achieve ACR20 before or at Week 52 will be censored at the date of their last ACR assessment prior to or at Week 52. Subjects who did not have any post-baseline ACR assessment will be censored at time = 1 day.
Results will be summarized by means of Kaplan-Meier curves together with a summary of associated statistics (median, confidence interval) and the number of subjects under risk on the ITT analysis set.

15.2.2.2   ACR50 and ACR70

ACR50 and ACR70 are defined as an improvement of at least 50% and 70%, respectively, in the number of tender joints and swollen joints, and at least 50% and 70% improvement, respectively, in three out of the remaining five ACR core-set measures (patient pain, Patient's and Physician's Global Assessment of Disease Activity, HAQ-DI and CRP).

These endpoints will be analyzed at Weeks 2, 4, 8, 12, 24, and 52 as per the key secondary endpoint (see Section 15.2.1).

15.2.2.3   ACR Component Variables

Summary statistics will be presented for the actual values for the ACR component variables (tender joint count (TJC), swollen joint count (SJC), Patient's and Physician's Global Assessment of Disease Activity, Patient's Assessment of Arthritis Pain – divided by 10, Patient's Assessment of Physical Function as measured by the HAQ-DI and CRP) and change from baseline values, by visit and treatment arm.

Line graphs will also be provided to display the change from baseline over time for the ACR component variables, by treatment arm.

15.2.2.4   DAS28-ESR

The Disease Activity Score calculated on 28 joints (DAS28-ESR) is a composite score derived from four measures. Components of DAS28-ESR are:

- Number of tender joints (out of the 28)

- Number of swollen joints (out of the 28)

- Erythrocyte sedimentation rate (ESR)

- Patient's Global Assessment of Disease Activity.

For DAS28-ESR, the 28 joints to be examined and assessed as tender or not tender for TJC and to be examined and assessed as swollen or not swollen for SJC include 14 joints on each side of the subject's body: 2 shoulders, 2 elbows, 2 wrists, 10 metacarpophalangeal joints, 2 interphalangeal joints of the thumb, 8 proximal interphalangeal joints, and 2 knees.

DAS28 will be derived using the following formula:

$$DAS28 = 0.56*\sqrt{(TJC28)} + 0.28*\sqrt{(SJC28)} + 0.014*GH + 0.70*ln(ESR)$$

where:

- TJC28 = 28 joint count for tenderness

- SJC28 = 28 joint count for swelling

- ln(ESR) = natural logarithm of ESR

- GH = the general health component of the DAS (i.e., Patient's Global Assessment of Disease Activity, assessed using a scale of 1 to 100 where 100 is maximal activity; For analyses, GH will be divided by ten and converted to a 0.5 scale, i.e. 0, 0.5, 1, 1.5 etc. where [0-0.25] = 0, [0.25-0-75] = 0.5, [0.76-1.25] = 1, etc.).

A DAS28-ESR of > 5.1 implies active disease, < 3.2 low disease activity, and < 2.6 remission. A change of ≥1.2 (twice the measurement error) is defined as a significant change of the disease activity state.

The proportion of subjects with DAS28 low disease activity and remission at Weeks 2, 4, 8, 12, 24 and 52 will be reported along with Exact 95% CIs calculated with the Clopper-Pearson method.

Summary statistics will be presented for the actual DAS28-ESR values and change from baseline, by visit and treatment arm. Similarly for the DAS28-ESR component variables.

Line graphs will also be provided to display the change from baseline over time for DAS28-ESR and its component variables, by treatment arm.

15.2.2.5   CDAI

The CDAI is a composite index (without acute-phase reactant) for assessing disease activity. CDAI is based on the simple summation of the swollen and tender joint counts based on 28 joints along with the Patient's and Physician's Global Assessment of Disease Activity (VAS) for estimating disease activity. The CDAI ranges from 0 to 76.

The CDAI is calculated based on the following formula:

$$CDAI = SJC28 + TJC28 + GH + PGA$$

where:

- TJC28 = 28 joint count for tenderness

- SJC28 = 28 joint count for swelling

- GH = the general health component of the DAS (i.e., Patient's Global Assessment of Disease Activity assessed using a scale of 1 to 100 where 100 is maximal activity; For analyses, GH will be divided by ten and converted to a 0.5 scale, i.e. 0, 0.5, 1, 1.5 etc. where [0-0.25] = 0, [0.25-0-75] = 0.5, [0.76-1.25] = 1, etc.)

- PGA = Physician's Global Assessment of Disease Activity assessed using a scale of 1 to 100 where 100 is maximal activity. For analyses, PGA will be divided by ten and converted to a 0.5 scale, i.e. 0, 0.5, 1, 1.5 etc. where [0-0.25] = 0, [0.25-0-75] = 0.5, [0.76-1.25] = 1, etc.

Summary statistics will be presented for the actual CDAI values and change from baseline, by visit and treatment arm, at Weeks 2, 4, 8, 12, 24 and 52.

Line graphs will also be provided to display the change from baseline over time for CDAI, by treatment arm.

15.2.2.6   SDAI

The SDAI is the numerical sum of five outcome parameters: tender and swollen joint counts (based on a 28-joint assessment), Patient's and Physician's Global Assessment of Disease Activity (VAS) and level of CRP (mg/dL).

The SDAI is calculated based on the following formula:

$$SDAI = SJC28 + TJC28 + GH + PGA + CRP$$

where:

- TJC28 = 28 joint count for tenderness

- SJC28 = 28 joint count for swelling

- GH = the general health component of the DAS (i.e., Patient's Global Assessment of Disease Activity assessed using a scale of 1 to 100 where 100 is maximal activity; For analyses, GH will be divided by ten and converted to a 0.5 scale, i.e. 0, 0.5, 1, 1.5 etc. where [0-0.25] = 0, [0.25-0-75] = 0.5, [0.76-1.25] = 1, etc.)

- PGA = Physician's Global Assessment of Disease Activity assessed using a scale of 1 to 100 where 100 is maximal activity. For analyses, PGA will be divided by ten and converted to a 0.5 scale, i.e. 0, 0.5, 1, 1.5 etc. where [0-0.25] = 0, [0.25-0-75] = 0.5, [0.76-1.25] = 1, etc.

- CRP = C-reactive protein in mg/dL.

Summary statistics will be presented for the actual SDAI values and change from baseline, by visit and treatment arm, at Weeks 2, 4, 8, 12, 24, and 52.

Line graphs will also be provided to display the change from baseline over time for SDAI, by treatment arm.

15.2.2.7   ACR/EULAR Boolean Remission

Following the Boolean-based definition of remission of ACR/EULAR, at any timepoint, a subject must satisfy all of the following: tender joint count ≤ 1, swollen joint count ≤ 1, CRP ≤ 1 mg/dL, and Patient's Global Assessment Of Disease Activity ≤ 1 (0-10 VAS).

The proportion of subjects with ACR/EULAR remission at Weeks 2, 4, 8, 12, 24, and 52 will be reported along with 95% CIs.

15.3   Other Secondary Endpoint Analyses

15.3.1   Quality of life

The three QoL questionnaires completed are the Health Assessment Questionnaire-Disability Index (HAQ-DI), the European Quality of Life – 5 Dimensions and 5 Levels (EQ-5D-5L) and the 36-item Short-Form Health Survey (SF-36).

A summary of QoL questionnaire completion, including reason for non-completion will be presented by visit and treatment arm.

Summary statistics for absolute value and absolute change from baseline will be presented for domains and/or total scores at baseline and at Weeks 12, 24, and 52 by treatment arm.

The summaries will be presented using both the ITT and the PP analysis sets.

15.3.1.1   European Quality of Life - 5 Dimensions and 5 Levels (EQ-5D-5L)

The EQ-5D-5L consists of two components - the EQ-5D descriptive system and the EQ visual analogue scale (EQ VAS). The descriptive system comprises the following five dimensions:

- Mobility

- Self-care

- Uusual activities

- Pain/discomfort

- Anxiety/depression.

Each dimension has five levels coded from 1 to 5 as follows:

- No problems (1)

- Slight problems (2)

- Moderate problems (3)

- Severe problems (4)

- Unable to function (mobility, self-care, and usual activities) or extreme problems (pain/discomfort and anxiety/depression) (5).

Subject response to this component results in a 1-digit number expressing the level selected for each dimension. The digits for five dimensions can be combined into a 5-digit number.

The 5-digit health state represents the scores for mobility, self-care, usual activities, pain/discomfort and anxiety/depression, strictly in that order of the dimensions.

Given that there are five domains and five scores per domain, there are 3125 potential health states.

For example, if a subject had no problems for any dimension, their 5-digit health state would be 11111, while a subject with no problems for mobility, slight problems for self-care, severe

problems with usual activities, extreme pain and moderate anxiety/depression would have a 5-digit state of 12453.

Each of these 3125 potential health states will be converted into a single summary index value using the EQ-5D-5L Crosswalk Index Value Calculator, downloaded from the EuroQol website.

An index value of 1 represents perfect health, the 5-digit health state would be 11111; the lowest, and therefore worst, value is -0.109, where the 5-digit health state would be 55555.

Respondents indicate their self-rated health on the EQ VAS, a 20 cm vertical, visual analogue scale numbered from 0 to 100 where the endpoints are labelled 'Best health you can imagine' (100) and 'Worst health you can imagine' (0). This information can be used as a quantitative measure of health outcome as judged by the individual respondents.

Summaries of the absolute values and absolute change from baseline for the index value and VAS score will be presented as described in Section 15.3.1.

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

This questionnaire consists of 20 questions, referring to the following eight domains:

- Dressing/grooming
- Getting up
- Eating
- Walking
- Personal care
- Reach
- Grip
- Common daily activities.

Each domain has at least two component questions. There are four possible responses for each component.

- 0 = without any difficulty
- 1 = with some difficulty
- 2 = with much difficulty
- 3 = unable to do.

A domain score is determined from the highest score of the components in that domain (except when aids and devices are taken into account). For example, if there are three components in a domain and the responses were 1, 2 and 0 to the components in the domain, then the score for the domain would be 2.
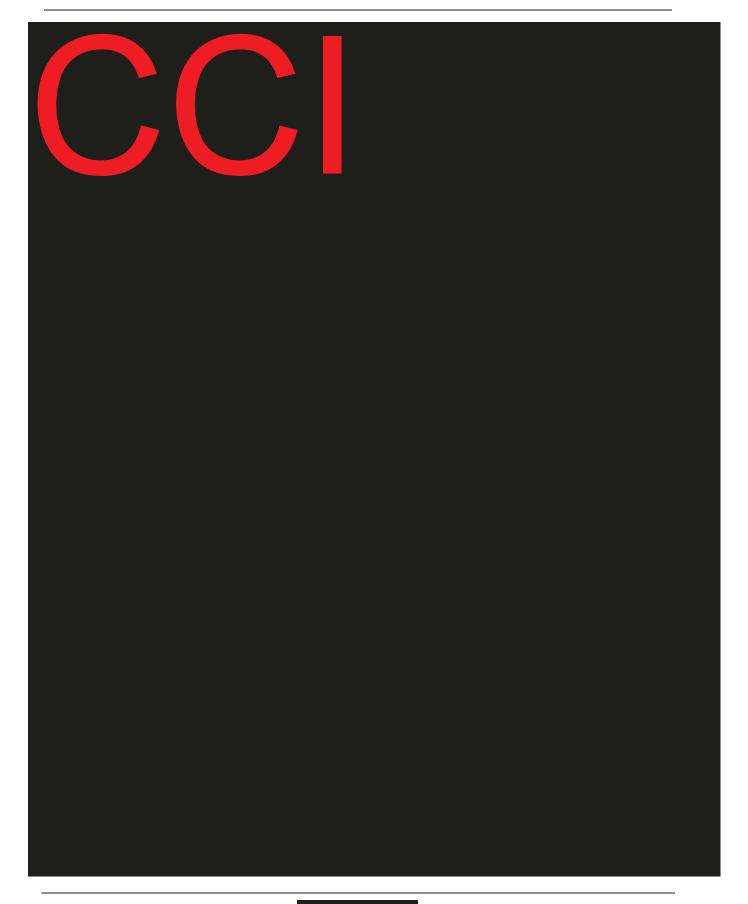
To calculate the HAQ-DI, the subject must have a domain score for at least six of the eight domains. The HAQ-DI score, or functional disability index (FDI), is the sum of the domain scores, divided by the number of domains that have a score (denominator is in the range 6 to 8).

The explanations below describe how the HAQ-DI score is calculated when we have missing answers to questions or when more than one answer is given for a question.
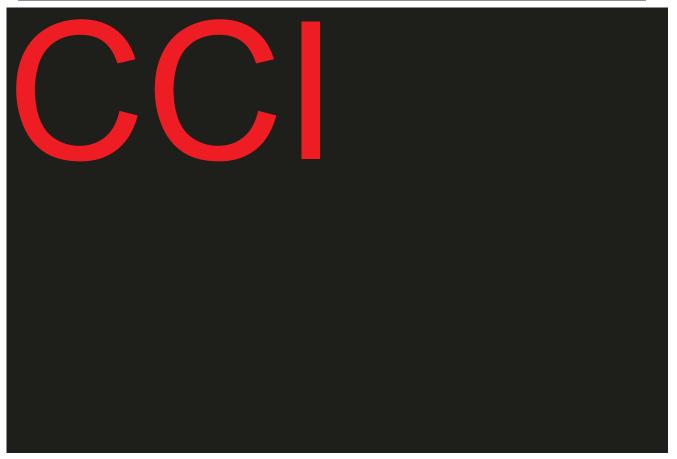
- If any domain consists of two questions and has one missing response, then the non-missing response will be used as the value for the domain. If a domain consists of three questions and has one missing response then the domain score will be the higher of the two responses. If a domain consists of three questions and has two missing responses then the domain will be considered missing.

- When there are no aids or devices used, the domain score is not modified. If the domain responses are missing and no aids or devices are listed then the domain will be considered missing.

- If aid or device and/or help for another person are checked for a domain:

  o If the domain score is missing, then the score for that domain will reflect the use of the aid or device, ie it will be scored as 2.

  o If the domain score is not missing, the score for that domain is raised from 0 or 1 to 2 and unchanged if already scored at 2 or 3

- The HAQ-DI will be set to missing if more than two of the domain scores are missing. In the event that one or two domain scores are missing, the final HAQ-DI score will be derived by dividing the total of the remaining domain scores by the number of non-missing domains.

- If a subject marks two responses adjacent to each other for an item on the HAQ-DI, the 'worst case' (ie response with the greater difficulty) will be used in the analysis. However, if a subject marks two responses which are not adjacent to each other, or marks more than two responses, the item will be coded as missing in the analysis.

Summaries of the absolute values and absolute change from baseline for HAQ-DI score (FDI) will be presented as described in Section 15.3.1.

CCI

CCI

CCI

CCI

CCI

15.3.1.4 Injection Site pain

The subject's reported perception of injection site pain will be measured on a VAS immediately after, 15 minutes after, and 1 hour after the 3rd, 4th and 5th injection administrations. The VAS ranges from 0 (no pain) to 100 (worst possible pain).

Summaries of the absolute values and absolute change from Week 4 (3rd dose) for the pain scores will be presented by timepoint (immediately after, 15 minutes after, and 1 hour after) and treatment arm for the SAF analysis set.

15.3.2 Immunogenicity Analysis

Immunogenicity data will be summarized by treatment arm for the SAF analysis set. Overall ADA status (negative or positive, where positive is defined as at least one positive result post-dose) will be summarized as well as ADA titers and presence or absence of neutralizing antibodies.

Time to positive ADA status, calculated from Trial Day 1, will also be analyzed. Subjects who discontinued the study or were lost to follow-up, or had not reached the endpoint of interest at the time of analysis, will be censored at the date of last sample collection. Results will be summarized by means of Kaplan-Meier curves together with a summary of associated statistics (median, 95% confidence interval) and the number of subjects under risk.

Summary statistics will be presented for the actual titer values, by treatment arm and visit up to Week 52. Boxplots will be produced to present the ADA titer results graphically. The boxplots will present Q1, Q3, median, minimum and maximum.

The number and percentage of ADA positive and NAb positive subjects will be presented for each visit up to Week 52.

The analysis exploring the correlation between ADA and efficacy, PK and safety will be described in Sections 15.3.2.1, 15.3.3.1, and 16.6 respectively.

15.3.2.1   Efficacy/ADA Correlation Analysis

Summary tables will present ACR20 and ACR50 by ADA status. A line graph will also be provided to display the ACR20 response rate over time by treatment arm and ADA status.

Boxplots displaying *DAS28-ESR* change from baseline at Week 12 by both ADA status and lower and upper quartiles of ADA titers will be produced. The change from baseline for *DAS28-ESR* will be the value for the specified visit, the ADA status or ADA titers will be the values for the treatment period.

The ADA titer quartiles will be defined using the highest ADA titer value from each subject for the period from baseline to Week 12 where subjects with ADA negative for the whole period will be assigned to a titer value of 0. ADA titers will be computed by splitting the subjects by treatment arm, taking their highest ADA titer values for the treatment period, arranging the values from smallest to largest and splitting into quartiles.

Line graphs will also be provided to display the change from baseline over time for continuous efficacy endpoints (TJC28, SJC28, ESR, DAS28-ESR, CDAI, SDAI), split by treatment arm and ADA status at a specified visit.

In addition, scatterplots of percentage change from baseline in continuous efficacy endpoints (DAS28-ESR, CDAI, SDAI) vs titer on log2 scale (if appropriate) will be generated for each treatment with separate symbols (if appropriate) for each study week.

15.3.3   Pharmacokinetic Analysis

15.3.3.1   Serum Concentrations of Adalimumab

All PK analysis will be performed using the PK analysis set (see Section 10.2).

All subjects will have PK samples taken pre-first dose, and then prior to dosing on Weeks 2, 4, 12, 24, 36, and 52. In addition, a subset of 60 subjects (30 per arm) will have additional samples to support the population PK analysis taken on Days 2 (24 h post first dose), 4, and 9, and Days 2 (24 h post Week 24 dose), 4, 9, and 14 (Week 26 pre-dose) after the Week 24 dose.

A listing of nominal and actual PK blood sample collection times as well as the derived sampling time deviations (as applicable) will be provided for all subjects.

Concentrations for adalimumab will be listed and summarized by treatment arm and visit using descriptive statistics for the subset of 60 subjects (30 per arm) with additional samples taken on the following timepoints: Week 0: Day 1 pre-dose, Days 2 (24 h post first dose), 4, and 9, and 14 (Week 2 pre-dose); Week 24: Day 1 pre-dose, Days 2 (24 h post Week 24 dose), 4, 9, and 14 (Week 26 pre-dose).

In order to facilitate direct evaluation of adalimumab trough serum concentrations across visits, an additional presentation of adalimumab data with summary statistics will be produced for trough samples at visits Week 0, 2, 4, 12, 24, 26, 36, 52.

Below the lower limit of quantitation (< LLOQ) concentrations will be treated as zero for the computation of descriptive statistics. Missing concentrations will be omitted from the calculation of descriptive statistics. Missing PK concentrations (eg no sample, insufficient sample volume for analysis, no result or result not valid) should be reported generally as "N.R.".

Pharmacokinetic concentrations will be summarized using descriptive statistics, including number of non-missing observations (n), mean, SD, coefficient of variation (CV%), minimum, median, and maximum. The geometric mean and geometric coefficient of variation will also be reported for visits where there are no concentrations below the limit of quantitation. The following conventions will be applied when reporting descriptive statistics of PK concentration data: mean, geometric mean, minimum, median, and maximum will be presented with 3 significant digits; SD 4 significant digits; CV% and geometric CV% 1 decimal place.

Figures of mean concentration-time data for trough adalimumab serum concentrations will be presented on linear (± SD) and semi logarithmic scales by treatment arm, and with treatments overlaid together as appropriate. Individual and mean trough adalimumab concentration-time data by treatment arm and visit week will be graphically presented on linear scale.

Concentrations for adalimumab will be listed and summarized by treatment arm and by ADA status (positive/negative) to assess the correlation between PK and ADA data. Also, a figure of mean (± SD) concentration-time data for adalimumab samples by ADA status (positive/negative) on linear scale will be presented by treatment arm, as appropriate.

15.4　　　　Exploratory Endpoint Analysis

15.4.1　　　　Population Pharmacokinetics

Full details of the Population PK Analysis planned to be performed on data of the current study will be described in a separate Data Analysis Plan. The results of the M&S work will be described in a dedicated report, separately from the study CSR.

It will be eventually up to the CRO that does the M&S and the supervisor (Merck or not) to decide how to structure all the details of the planned modeling, in agreement with the team.

16   Safety Evaluation

Safety and tolerability of study treatment will be assessed by the examination of AEs, laboratory data, ECGs, vital signs, and PK data.

All summaries of safety data will be based on the SAF analysis set.

16.1   Adverse Events

Adverse events will be collected throughout the study and coded using version 19.1 of MedDRA or later.

Treatment-emergent adverse events (TEAEs) are those events that start on or after the day of first administration of study treatment. If an AE started on the same date as the first administration of the first study treatment given, the timing related to study treatment as recorded on the adverse events page of the eCRF will be taken into account when determining treatment emergence. A worst case approach will be used, ie if time is missing or if the AE occurred at the same time as the first administration of the first study treatment given, then the AE will be classified as treatment-emergent.

Incomplete AE-related dates will be handled as follows:

- In case the onset date is missing completely or missing partially but the onset month and year, or the onset year are equal to the start of study treatment then the onset date will be replaced by the minimum of start of study treatment and AE resolution date.

- In all other cases the missing onset day or missing onset month will be replaced by 1.

- Incomplete stop dates will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of subject's death. In the latter case the date of death will be used to impute the incomplete stop date.

- In all other cases the incomplete stop date will not be imputed.

Only TEAEs will be included in tables and will be summarized by treatment arm. All AEs will be listed.

Unless otherwise stated, TEAEs will be displayed in terms of frequency tables: primary SOC in alphabetical order, then PT in order of decreasing frequency across both treatment arms. If a subject has more than one event for a particular SOC or PT, then the event will be counted once only for the subject. For example, if a subject has two different TEAEs but these are within the same SOC, then they will be counted once for each PT but only once for the SOC. If at any time a summary is being prepared on data that are not fully coded, uncoded TEAEs will be summarized with a SOC as 'Uncoded' and the PT term will be set to the TEAE verbatim text.

If an AE is reported for a given subject more than once during treatment, the worst relationship to study treatment will be tabulated.

Adverse events related to study treatment are those events with relationship missing, unknown or yes.

Treatment-emergent AEs will be summarized by primary SOC, PT and worst toxicity grade (according to National Cancer Institute-Common Terminology Criteria for Adverse Events [NCI-CTCAE], version 4.03) per subject. In case a subject had events with missing and non-missing toxicity grades, the maximum of the non-missing toxicity grades will be displayed.

16.1.1 All Adverse Events

An overall summary table will present the number of subjects with at least one of the following TEAEs, summarized by treatment arm:

- TEAE
- Study drug related TEAE
- Serious TEAE
- Study drug related serious TEAE
- TEAE of NCI-CTCAE toxicity Grade 1
- TEAE of NCI-CTCAE toxicity Grade 2
- TEAE of NCI-CTCAE toxicity Grade 3
- TEAE of NCI-CTCAE toxicity Grade 4
- TEAE of NCI-CTCAE toxicity Grade 5
- TEAE of NCI-CTCAE toxicity Grade ≥ 3
- TEAE of NCI-CTCAE toxicity Grade ≥ 4
- Study drug related TEAE of NCI-CTCAE toxicity Grade 1
- Study drug related TEAE of NCI-CTCAE toxicity Grade 2
- Study drug related TEAE of NCI-CTCAE toxicity Grade 3
- Study drug related TEAE of NCI-CTCAE toxicity Grade 4
- Study drug related TEAE of NCI-CTCAE toxicity Grade 5
- Study drug related TEAE of NCI-CTCAE toxicity Grade ≥ 3
- Study drug related TEAE of NCI-CTCAE toxicity grade ≥ 4
- TEAE of special interest
- TEAE of special interest Grade 1
- TEAE of special interest Grade 2
- TEAE of special interest Grade 3
- TEAE of special interest Grade 4

- TEAE of special interest Grade 5

- Study drug related AESI

- Study drug related AESI Grade 1

- Study drug related AESI Grade 2

- Study drug related AESI Grade 3

- Study drug related AESI Grade 4

- Study drug related AESI Grade 5

- TEAE leading to death

- Study drug related TEAE leading to death.

- Injection Site Reactions

Incidence rates will be included in these overall summaries as the number of events per 100 subject-years of exposure to IMP, calculated as the number of subjects with AEs divided by the total time at risk across all subjects in the given treatment arm, multiplied by 100. Time at risk for a subject is calculated as time from first dose to 28 days following the last dose of IMP. Two-sided 95% CIs for incidence rates will be provided using the exact Poisson method.

The following frequency tables will be prepared by treatment arm, primary SOC in alphabetical order, then PT in order of decreasing frequency across both treatment arms:

- TEAEs

- Study Drug related TEAEs

- Serious TEAEs

- Non-serious TEAEs with frequency of $\geq$ 5% in any treatment arm

- Non-serious TEAEs with frequency of $\geq$ 1% in any treatment arm

- TEAE leading to death

- Study drug related TEAE leading to death

- TEAEs by NCI-CTCAE toxicity grade

- Study Drug related TEAEs by NCI-CTCAE grade

- TEAEs by worst NCI-CTCAE toxicity grade.

- Injection Site Reactions- Local Tolerability

- Study Drug Related Injection Site Reactions by NCI-CTCAE grade and duration

- Study Drug Related Injection Site Reactions by NCI-CTCAE grade and time since last study administration

Summary statistics will be presented for the following:

- Duration of study drug related injection site reactions (hours) calculated as the difference of start date/time and stop date/time divided by 3600 [(stop date/time - start date/time)/3600]. If start time is missing then start time will be set to 00:00. If stop time is missing then stop time will be set to 23:59. Duration will not be calculated if start date or stop date are missing. If a subject has more than one study drug related injection site reactions then the event with the longest duration will be used.

- Number of subjects with duration of study drug related injection site reaction <24 hours, $\geq$ 24 - < 48 hours and $\geq$ 48 hours. If a subject has more than one study drug related injection site reactions then the event with the longest duration will be used.

- Time since the first study treatment administration to the study drug related injection site reaction start date/time calculated as the difference of first study drug administration date/time and the injection site reaction start date/time divided by 3600 [(Injection site reaction start date/time - First study drug administration date/time)/3600]. Time since the first study treatment administration will not be calculated if injection site reaction start date or first study drug administration date are missing. If first study treatment administration time is missing then it will be set to 23:59. If study drug related injection site reaction start time is missing then it will be set to 00:00. In the special cases that any time is missing and first study treatment administration date equals the study drug related injection site reaction start date then time since first study treatment administration will be set to 0. If a subject has more than one study drug related injection site reactions then the event with the shortest duration will be used.

- Number of subjects with time since the first study treatment administration to the study drug related injection site reaction <24 hours, $\geq$ 24 - < 48 hours and $\geq$ 48 hours. If a subject has more than one study drug related injection site reactions then the event with the shortest duration will be used.

- Time since last study treatment administration to the study drug related injection site reaction start date/time calculated as the difference of the last study drug administration date/time and the study drug related injection site reaction start date/time divided by 3600 [(Injection site reaction start date/time - Last study drug administration date/time)/3600]. Time since the last study treatment administration will not be calculated if study drug related injection site reaction start date or last study drug administration date are missing. If last study treatment administration time is missing then it will be set to 23:59. If injection site reaction start time is missing then it will be set to 00:00. In the special cases that any time is missing and last study treatment administration date equals injection site reaction start date then time since last study treatment administration will be set to 0. If a subject has more than one study drug related injection site reactions then the event with the shortest duration will be used.

- Number of subjects with time since the last study treatment administration to the injection site reaction <24 hours, $\geq$ 24 - < 48 hours and $\geq$ 48 hours. If a subject has more than one study drug related injection site reactions then the event with the shortest duration will be used.

- Medication or procedure taken for study drug related injection site reactions. Frequencies and percentages will be based on the number of subjects.

Additionally the study drug related injection site reactions rate based on the overall number of study treatment injections administered will be calculated by SOC and PT.

The following frequency tales will be prepared by PT in order of decreasing frequency across both treatment arms:

- TEAEs with frequency of ≥ 5% in any treatment arm

- TEAEs with frequency of ≥ 1% in any treatment arm

- Treatment emergent AESIs (as per section 15.1).

16.1.2 TEAEs leading to Discontinuation / Dose Reduction of Treatment

An overall summary table will present the number of subjects with at least one of the following, summarized by treatment arm:

- TEAE leading to temporary discontinuation of study drug

- Study drug related TEAE leading to temporary discontinuation of study drug

- TEAE leading to permanent discontinuation of study drug

- Study drug related TEAE leading to permanent discontinuation of study drug

- TEAE leading to study termination

The above listed summaries will also be presented for AESIs.

The following frequency tables will be prepared by treatment arm, primary SOC in alphabetical order, PT in order of decreasing frequency across both treatment arms and worst grade:

- TEAEs leading to temporary discontinuation of study drug

- TEAEs leading to permanent discontinuation of study drug

- Study drug related TEAEs leading to permanent discontinuation of study drug

- TEAEs leading to study termination

- TEAEs leading to death.

16.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

16.2.1 Deaths

Adverse events leading to death and treatment-related TEAEs leading to death will be summarized by primary SOC and PT, see Section 16.1.1.

The number of deaths and primary reason for death will be summarized based on the Death eCRF page by treatment arm. Also a listing of deaths for each treatment received will be provided including first and last dose date. In addition, a listing of TEAEs leading to death will be provided.

16.2.2   Serious Adverse Events

A separate listing will be provided of the subset of events indicated as serious. Serious TEAEs will also be summarized by primary SOC and PT, see Section 16.1.1.

16.2.3   Other Significant Adverse Event

16.2.3.1   Adverse Events of Special Interest (AESIs)

The occurrence of treatment-emergent adverse events of special interest (hypersensitivity) up to Week 52 is the primary endpoint (see Section 15.1 for the analysis of these AEs).

The following is considered pre-defined AESI for this study:

Hypersensitivity (defined by Standardized MedDRA Query [SMQ] narrow as per the latest MedDRA version).

16.3   Clinical Laboratory Evaluation

Laboratory results will be graded using the NCI-CTCAE, version 4.03, except where common terminology criteria (CTC) grades are not available. Laboratory tests that are not part of the NCI-CTCAE grading system will be summarized using the normal / above normal limit / below normal limit categories, unless otherwise specified.

In this study, there are three categories of laboratory evaluations for safety entered into the database as described below:

(1) Hematology and biochemistry parameters, assessed by the central laboratory:

- The hematology laboratory panel includes the following: White blood cell count, red blood cell count, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin concentration, erythrocyte sedimentation rate, platelet count, neutrophils (absolute and percentage), lymphocytes (absolute and percentage), monocytes (absolute and percentage), eosinophils (absolute and percentage), basophils (absolute and percentage).
- The biochemistry panel includes the following: Sodium, potassium, blood urea nitrogen, creatinine, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, gamma-glutamyl transferase, lactate dehydrogenase, serum albumin, calcium, phosphate, glucose, creatine kinase, uric acid, total bilirubin, total serum protein, cholesterol, triglycerides, C-reactive protein, coagulation (activated partial thromboplastin time and prothrombin time).

(2) Urinalysis will be assessed by the central laboratory and includes the following: Dipstick, including macroscopic appearance, bilirubin, blood, color, glucose, ketones, leukocyte esterase, nitrite, pH, protein, specific gravity, and urobilinogen. Full urinalysis (dipstick plus microscopic evaluation) at the Investigator's discretion only if warranted by an abnormal dipstick finding.

(3) Anti-nuclear antibody (ANA) and anti-double stranded deoxyribonucleic acid (dsDNA).

For category (1), absolute values and change from baseline will be summarized by visit and treatment arm. Absolute values will also be summarized overall, with percentage of subjects below lower limit / 10 and lower limit / 3, and above 10 x upper limit and 3 x upper limit. Box-plots will also be produced for absolute change from baseline, by visit and by treatment arm. Normal ranges will be provided by the central laboratory, and out of range flags will be calculated based on the normal ranges. Shift tables will be produced of the number of subjects with low, normal and high values based on normal ranges by visit and baseline for all parameters with normal ranges available. Laboratory values will also be summarized by worst on-treatment value and shift from baseline to worst on-treatment value, and by worst on-treatment NCI-CTCAE toxicity grade and shift from baseline to highest grade. For summaries of laboratory values, "on-treatment" will be defined as all available post-baseline values up to and including the Week 52/4-week safety follow-up safety visit or 28 days after the last dose of study drug, whichever occurs later. Direction of abnormality and NCI-CTCAE gradable parameters are presented in Appendix II: Direction of Abnormality for Worst On-treatment Values and NCI-CTCAE Gradable Parameters.

For category (2), absolute values and change from baseline will be summarized by visit and treatment arm. Shift tables by visit and baseline will be produced for parameters with categorical values.

For category (3), ANA and anti-dsDNA status will be listed and summarized.

16.4   Vital Signs

Absolute values and change from baseline in blood pressure, heart rate, weight, body temperature and respiratory rate will be summarized by visit and treatment arm. The maximum change from baseline will be listed for each vital sign parameter for each subject.

Temperature will be presented in degrees Celsius (°C). The conversion factor used will be:

- Temp (°C) = (5/9)*(Temp(°F) - 32)

See Section 12.4 for height and weight conversions.

For vital signs that have been measured twice at the same visit the average value at the visit will be calculated and used. (All individual values per visit will be listed.) Shift tables (including number of subjects and percentages) of vital signs (systolic blood pressure, diastolic blood pressure, pulse rate, body temperature, and respiratory rate), displaying baseline value versus worst on-treatment value will be provided for both increases and decreases. For summaries of vital signs, "on-treatment" will be defined as all available post-baseline values up to and including the Week 52/4-week safety follow-up safety visit or 28 days after the last dose of study drug, whichever occurs later.

The following parameter categories will be used for the vital signs shift tables:

Parameter	Baseline categories	Post-baseline categories
Systolic blood pressure (SBP)	< 140 / ≥ 140 mmHg	Absolute change of: ≤ 0 / > 0 - ≤ 20 / > 20 - ≤ 40 / > 40 mmHg
Diastolic blood pressure (DBP)	≤ 90 / > 90 mmHg	Absolute change of: ≤ 0 / > 0 - ≤ 20 / > 20 - ≤ 40 / > 40 mmHg
Heart rate	< 100 / ≥ 100 bpm	Absolute change of: ≤ 0 / > 0 - ≤ 20 / > 20 - ≤ 40 / > 40 bpm
Temperature	< 37°C / ≥ 37 - < 38°C / ≥ 38 - < 39°C / ≥ 39 - <40°C / ≥ 40°C	Absolute change of: ≤ 1 / > 1 - ≤ 2 / > 2 - ≤ 3 / > 3°C
Respiratory rate	< 20 / ≥ 20 bpm	≤ 5 / > 5 - ≤ 10 / > 10

16.5 Other Safety or Tolerability Evaluations

16.5.1 ECGs

Summary results will be presented for Week 12, 24 and 52. Shifts from baseline in ECG results will be presented by treatment arm. For summaries, "on-treatment" will be defined as all available post-baseline values up to and including the Week 52/4-week safety follow-up safety visit or 28 days after the last dose of study drug, whichever occurs later.

16.5.2 Tuberculosis Assessments

After initial TB assessments are carried out at screening, QuantiFERON-TB Gold tests are carried out at Weeks 24 and Week 52 during the study.

Completion status (done/not done) and result (negative/positive) will be summarized by visit and treatment arm.

16.6 Safety/ADA Correlation Analysis

The following frequency tables by treatment arm, status of ADA (negative or positive, where positive is defined as at least one positive result post-dose), primary SOC and PT will be provided for:
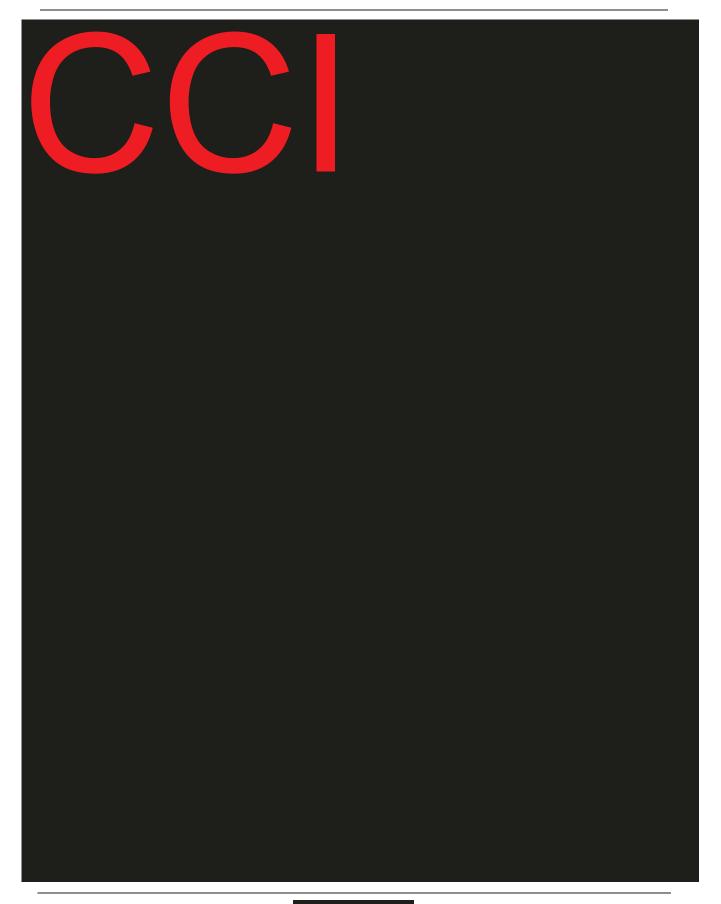
- TEAEs
- AESIs (Hypersensitivity)
- Injection Site Reactions

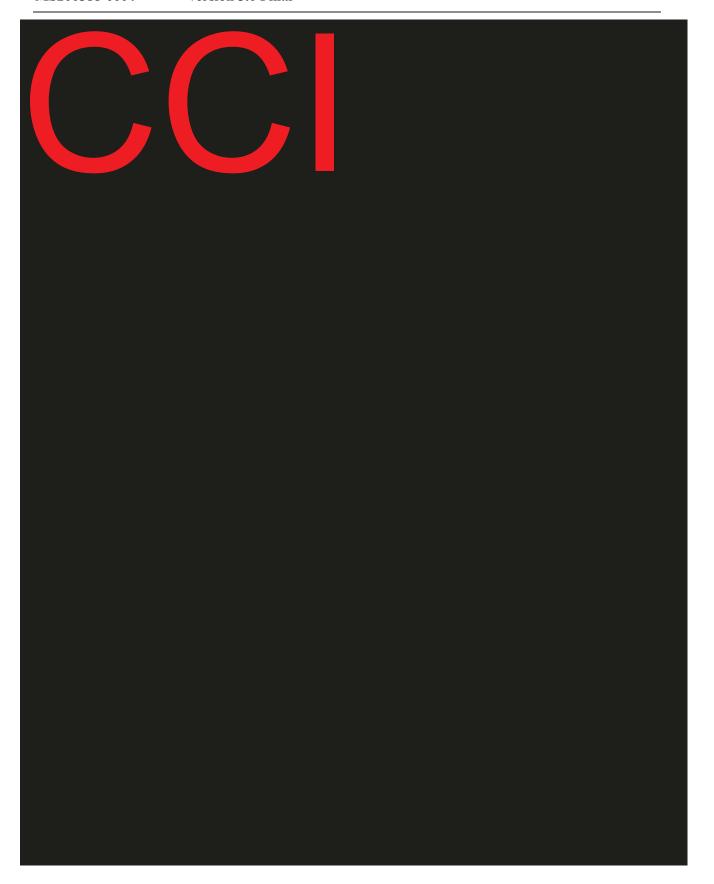The tables will be repeated for neutralizing antibody positive versus negative status.

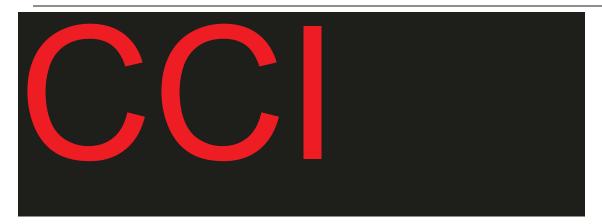## 17 Reporting Conventions and Data Handling

The reporting conventions will follow those detailed in the Biostatistical Output Assembly (BOA) v2.0 document.

CCI

CCI

CCI

CCI

CCI

19   References

1   Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. Arthritis Rheum. 2004; 50(5):1400-1411.

2   FDA, Guidance for Industry - RA: Developing Drug Products for Treatment, 2013.

3   Aletaha D, Landewe R, Karonitsch T, Bathon J, Boers M, Bombardier C, et al. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. Arthritis Rheum. 2008; 59(10):1371-1377.

4   Ware JE Jr and Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selections. Med Care. 1992 ; 30(6) :473-483.

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI