Protocol Number: FKS456-001

Official Title: A Randomized, Double-Blind, Multiple-Dose, Parallel-Group, Two-Arm Study to Evaluate the Efficacy, Safety and Immunogenicity of MSB11456 Compared to European Union-approved RoActemra® in Patients with Moderately to Severely Active Rheumatoid Arthritis (APTURA I study)

Document Date: 15 June 2022

NCT Number: NCT04512001

Statistical Analysis Plan (SAP)

Protocol Title:	A Randomized, Double-Blind, Multiple-Dose, Parallel-Group, Two-Arm Study to Evaluate the Efficacy, Safety and Immunogenicity of MSB11456 Compared to European Union-approved RoActemra® in Patients with Moderately to Severely Active Rheumatoid Arthritis (APTURA I study)
Protocol Version No. / Date:	Version 3.0 / 01 February 2021 and local Czech v3.1 / 08 February 2022 Version 2.0 (incorporating Global Amendment 1 - Substantial) 06 May 2020 Version 1.0/ 12 February 2020
CRF Version No. / Date:	8.0 / 28 March 2022
SAP Version No. / Date:	Final 2.0 / 15 June 2022

1.0 Approvals

Sponsor	
Sponsor Name:	Fresenius Kabi SwissBioSim GmbH
Representative / Title:	[redacted]
Signature / Date:	
Representative / Title:	[redacted]
Signature / Date:	
Representative / Title:	[redacted]
Signature / Date:	
Representative / Title:	[redacted]
Signature / Date:	
[redacted]	
Biostatistician / Title:	[redacted]
Signature / Date:	
Biostatistician / Title:	[redacted]
Signature / Date:	

1.0 Approvals

Sponsor

Sponsor Name: Fresenius Kabi SwissBioSim GmbH

Representative / Title:

Signature / Date:

Representative / Title:

Signature / Date:

Representative / Title:

Signature / Date:

Representative / Title:

Signature / Date:

[redacted]

Biostatistician / Title:

Signature / Date:

Biostatistician / Title:

Signature / Date:

1.0 Approvals

Sponsor

Sponsor Name: Fresenius Kabi SwissBioSim GmbH

Representative / Title: [redacted]

Signature / Date: [redacted]

Representative / Title: [redacted]

Signature / Date: [redacted]

Representative / Title: [redacted]

Signature / Date: [redacted]

Representative / Title: [redacted]

Signature / Date: [redacted]

[redacted]

Biostatistician / Title: [redacted]

Signature / Date: [redacted]

Biostatistician / Title: [redacted]

Signature / Date: [redacted]

2.0 Change History

Version / Date	Change Log
Final Draft 1.0 / 30Dec2020	Draft submitted for review by Health Authorities
Draft 1.2/ 15Sep2021	Draft
Draft 1.3/ 12Nov2021	Draft
Final 1.1/ 13DEC2021	Final SAP prior to Week 30 Database Lock
Final 2.0/ 15Jun2022	Final SAP prior to Week 55 Database Lock

3.0 Table of Contents

4.0 Purpose

The Statistical Analysis Plan (SAP) describes the statistical methods to be used during the reporting and analyses of data collected under Fresenius Kabi SwissBioSim GmbH Protocol FKS456-001.This SAP describes all analyses planned to be included in the Week 30 clinical study report (CSR), Week 55 CSR, and Week 63 CSR Addendum.

For details of the data included in the Week 30 CSR please see the Week 30 Data Cut Plan which is a separate document.

5.0 Scope

The SAP outlines the following:

- Study Objectives
- Study Design
- Study Estimands
- Applicable Study Definitions
- Statistical Methods

6.0 Introduction

This SAP should be read in conjunction with the study protocol (current version: Version 3.0 / 01 February 2021 and local Czech v3.1 / 08 February 2022), case report form (CRF) (current version: 8.0 dated 28-Mar-2022) and the Data Access Plan (current version: Version 2.0 dated 08 December 2021). Any changes to the protocol or CRF may necessitate updates to the SAP.

The SAP will be approved (by the Sponsor and [redacted]) at two timepoints. The first prior to W30 database lock and the second timepoint prior to W55 database lock. For W63 (safety follow-up) database lock, if further updates are needed, it will again be approved. Unblinding of study treatment codes to the unblinded team are defined in the Data Access Plan.

6.1 Changes from Protocol

Changes in study setup from protocol:

- Active/latent tuberculosis will only be listed if at least one case is reported.

Sub study post-dose PK concentrations analysis will not be conducted due to low recruitment during Covid-19 pandemic. Only listings will be provided.

- Correction has been made to the null hypothesis of the primary analysis stated in the protocol as H_0: $(\mu_T - \mu_R) \leq -0.6$ or $(\mu_T - \mu_R) \geq 0.5$. This has been updated to H_0: $(\mu_T - \mu_R) < -0.6$ or $(\mu_T - \mu_R) > 0.5$.

7.0 Study Objectives

7.1 Primary

The primary objective of the study is to demonstrate equivalent efficacy of proposed biosimilar tocilizumab MSB11456 and EU--approved RoActemra (both administered subcutaneously) in patients with moderately to severely active rheumatoid arthritis (RA).

7.2 Secondary

The secondary objective of the study is to compare the safety, immunogenicity and long-term efficacy of MSB11456 to EU--approved RoActemra.

7.3 Exploratory

The exploratory objectives of the study are:

- To explore the effects of a single treatment transition (i.e., in patients who transitioned from EU--approved RoActemra to MSB11456) on efficacy, safety and immunogenicity and
- To describe pharmacokinetic (PK) parameters of MSB11456 and EU-approved RoActemra

8.0 Study Design

This is a multicenter, randomized (1:1), active-controlled, double-blind, multiple fixed-dose, multinational, two-arm, parallel-group study to compare the efficacy, safety and immunogenicity of the proposed biosimilar candidate MSB11456 versus EU-approved RoActemra (hereinafter referred to as "EU-RoActemra" only) in patients with moderately to severely active RA. In a separate sub-study, the population pharmacokinetics of MSB11456 and EU-RoActemra will be described.

This study will enroll patients with moderately to severely active RA who have an inadequate response to ≥ 1 disease-modifying anti-rheumatic drug(s) which may include biologic disease--modifying anti-rheumatic drugs and who are currently receiving a stable dose of methotrexate. Patients who have previously received 1 or 2 biologic treatments in total for RA may be allowed to participate in the study. However, the population of patients with previous exposure to any biologic treatment will be capped at 10% of the total study population. Patients who have previously received more than 2 biologic treatments for RA will not be permitted to participate in the study.

Patients must have received methotrexate for at least 12 consecutive weeks immediately prior to randomization and are on a stable dose between 10 and 25 mg / week methotrexate for the last 8 weeks prior to screening. Patients will continue to take the same stable dose of methotrexate during the 52-week treatment period of the study. All other disease-modifying- anti-rheumatic drugs must be withdrawn prior to randomization.

The study will have a duration of up to 67 weeks and will include a Screening Period of a maximum of 28 days prior to first study drug administration, a double-blind 24-week Core Treatment Period (Day 1 to Day 169 / W24), an additional 28-week double-blind Extended Treatment Period (Day 169 / W24 to Day 365 / W52) and a 12-week Safety Evaluation Period. A Safety Follow-Up Visit will be performed at W55, 4 weeks after the last dose of study drug at W51 and a final End of Study Visit will be conducted at W63, 12 weeks after the last dose of study drug at W51. (Note: The 12-week Safety Evaluation starts at W51, after the last study drug injection and the Extended Treatment Period ends at W52.) The end of the study for each patient is defined as the patient's last visit.

Patients whose eligibility is confirmed at baseline will be randomized in a 1:1 ratio by an Interactive Response Technology (IRT) system to receive either MSB11456 or EU-RoActemra at a dose of 162 mg delivered by subcutaneous injection starting at Day 1, then weekly up to W51 inclusive. Randomization will be stratified by previous exposure to biologic treatment for RA [yes/no].

The first 3 doses of study drug (162 mg at Day 1, Day 8 and Day 15) will be administered on-site. The subsequent 2 doses (on Day 22 and Day 29) will also be administered on-site in the Czech Republic only. Patients will be monitored for 2 hours following the on-site administrations of study drug. The following weekly doses of study drug can be self--administered if the patient is able to self-administer him / herself at the discretion of the Investigator. Patients will check their axillary temperature prior to the self-administration of study drug. In the case of an elevated temperature, it is the Investigator's decision to determine any further actions according to local practice and their medical judgment.

The primary efficacy endpoint will be evaluated at W24 (7 days after the 24th study drug administration), along with other endpoints. At W24, after all efficacy and safety assessments have been performed, patients will enter the Extended Treatment Period. Patients who were originally randomized to receive EU-RoActemra will be re-randomized in a 1:1 ratio to continue their weekly treatment with EU-RoActemra or to switch to MSB11456 starting at W24. Patients who were originally randomized to MSB11456 will continue this treatment for the complete 52 weeks of study treatment. During the Extended Treatment Period, efficacy, safety and immunogenicity data will be analyzed up to W52 with the last assessment performed 7 days after the last study drug administration (i.e., on W51).

Safety data will be collected and analyzed at W55 and W63 (i.e., 4 and 12 weeks after the last study drug administration, respectively). Samples for immunogenicity analysis will be collected at W55.

In order to minimize missing data in the evaluation of the treatment effect under the treatment policy strategy, patients who discontinue treatment early or violate the protocol should continue to be followed for all regularly scheduled visits for safety and efficacy assessments up to the end of the corresponding treatment period.

Pharmacokinetic trough concentration samples will be collected from all study patients at scheduled visits.

Patients may be asked to participate in a population PK sub-study performed during the 24-week Core Treatment Period (Day 1 to Day 169 / W24) and contribute additional PK samples. Participation in the population PK sub-study is optional.

A Data Monitoring Committee or Data Safety Monitoring Board will not be involved in this study.

8.1 Sample Size Considerations

8.1.1 Background Information

In the absence of innovator studies comparing the weekly subcutaneous dosing regimen versus placebo, the dataset used to build the statistical assumptions has been extrapolated from the pivotal studies with the intravenous presentation. The subcutaneous 162 mg weekly regimen was shown to have comparable efficacy and safety to the intravenous presentation at a dose of 8 mg / kg every 4 weeks (Burmester, 2014).

In order to be in line with both the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) requirements, different acceptance margins are predefined for the primary endpoint mean absolute change from baseline in Disease Activity Score 28-Erythrocyte Sedimentation Rate (DAS28-ESR) at W24.

Equivalence margins of ± 0.6, correspond to the retention of approximately 65% of conservative estimates of treatment effect sizes relative to placebo for EU-RoActemra, based on the upper bound of 95% confidence interval from a meta-analysis performed by the Sponsor. The meta-analysis was performed using the following publications Genovese, 2008; Smolen, 2008; Kremer, 2011; Emery, 2008. The equivalence margins defined respectively for FDA and EMA are specified below:

- For the FDA, the equivalence margins are set to [- 0.6, 0.5]. The FDA has recommended a stricter margin using 0.5 as the upper bound (FDA electronic correspondence 19 October 2016). To help with the study feasibility the lower bound is set to - 0.6. A change of 0.6 in favor of MSB11456 is seen as a non-clinically significant difference as long as the safety and immunogenicity profile of MSB11456 demonstrates no more risk than EU-RoActemra
- For the EMA, the equivalence margins are set to ± 0.6

The primary efficacy analysis is the change from baseline at W24 in DAS28-ESR which will be analyzed using an analysis of covariance with study drug and previous exposure to biologic treatment for RA [yes/no] stratum as fixed effects and baseline DAS28-ESR as a covariate. The difference between treatments is estimated by the least squares mean difference between MSB11456 and EU-RoActemra, with its 95% confidence interval for the EMA and its 90% confidence interval for the FDA.

- For the FDA: MSB11456 will be considered equivalent to EU-RoActemra if the 90% confidence interval for the difference in mean change from baseline to W24 in DAS28-ESR between MSB11456 and EU-RoActemra lies entirely within the equivalence interval of [- 0.6, 0.5] in the Intent-To-Treat (ITT) Analysis Set

- For the EMA: MSB11456 will be considered equivalent to EU-RoActemra if the 95 % confidence interval for the difference in mean change from baseline to W24 in DAS28-ESR between MSB11456 and EU-RoActemra lies entirely within the equivalence interval of [-0.6, 0.6] in the ITT Analysis Set

Specificities of Secondary Efficacy Endpoint "ACR20 at W24": The treatment difference in American College of Rheumatology (ACR)20 at W24 will be assessed versus predefined margins. The equivalence margins are set to ± 15%. These proposed equivalence margins correspond to 50% retention of conservative estimates of the effect size of EU-RoActemra relative to placebo (derived from Genovese, 2008; Smolen, 2008; Kremer, 2011; Emery, 2008 using meta-analysis).

The primary and key secondary analyses will be performed on the ITT Analysis Set and repeated on the Per Protocol (PP) Analysis Set.

8.1.2 Sample Size Information

A sample size of 542 randomized patients (271 patients per arm) is chosen to provide approximately 460 patients (230 per arm) in the PP Analysis Set at W24, assuming a 15% drop-out rate (including major protocol deviations).

For the FDA:

- A total of 460 evaluable patients (230 per arm) will provide 90% power to demonstrate equivalence between treatments for the primary endpoint, with equivalence margins of [- 0.6, 0.5] and a type I error of 5%, assuming no difference between the 2 treatment groups and a common standard deviation of 1.76 (derived from Genovese, 2008; Smolen, 2008; Kremer, 2011).

For the EMA:

- A total of 460 evaluable patients (230 per arm) will provide 90% power to demonstrate equivalence between treatments for the primary endpoint, with equivalence margins of ± 0.6 and a type I error of 2.5%, assuming no difference between the 2 treatment groups and a common standard deviation of 1.76.
- In addition, this sample size provides more than 80% power to demonstrate that the 95% confidence interval for the difference between treatments in the secondary efficacy endpoint ACR20 response rate at W24 will be included in the equivalence interval [- 15 %, + 15 %], assuming no difference between the 2 treatment groups and that both MSB11456 and EU-RoActemra have an ACR20 response rate of 60% at W24 (Genovese, 2008; Smolen, 2008; Kremer, 2011).

The drop-out/protocol deviation rate will be monitored on blinded data throughout the first 24 weeks. If larger than anticipated (> 15 %), an investigation on the reasons for dropping out will be conducted. The number of randomized patients may be adjusted as a consequence. With the exception of a potential increase in subjects in safety and ITT analysis sets, this blinded monitoring does not impact ITT or PP allocation.

The sample size for the population PK sub-study will be a subset of approximately 30 patients per group, for a total of 60 patients (to ensure that approximately 50 patients complete the PK sub-study).

8.2 Randomization

Randomization will be performed via a centralized IRT system. Eligible patients will be randomly assigned to either MSB11456 or EU-RoActemra in a 1:1 ratio, stratified by previous exposure to biological treatment for RA [yes/no], which is a prognostic factor (Strata for geographical region were planned in the IRT tool as well according to the study protocol, but all subjects will be randomized in the same European stratum.)

Patients with previous exposure to any biologic treatment will be capped at 10% of the total study population.

Randomization will be conducted in permuted blocks.

Although the randomization is stratified, [redacted] created only 1 random list that incorporates all strata. To ensure a balanced randomization the system will only assign full blocks to each stratum combination.

At the W24 Visit, patients who were initially randomized to the EU-RoActemra group will be re-randomized in a 1:1 ratio to receive either MSB11456 or EU-RoActemra at the W24 Visit and weekly thereafter until W51. Re-randomization will use the same stratification factors as the original randomization. Patients who were initially randomized to the MSB11456 group will remain on weekly treatment with MSB11456 throughout the entire study. Patients who discontinue study drug before or at W24 will not be re-randomized at W24 and will discontinue the study.

As the existence of a re-randomization number will unblind the reader to the patient's treatment in the Core Treatment Period the re-randomization number will remain invisible to IRT users. As the nature of the random number might reveal details on the block size this is also not visible to IRT users. Patients will be identified by their screening numbers.

Although only patients randomized to EU-RoActemra during the Core Period will be re-randomized, in order to preserve the blind, it will appear that all patients have been re-randomized. Re-randomization will be silent, i.e., the re-randomization step will be hidden on the end user system. Hence, re-randomization will not impact the double-blind nature of the study as blinding will be kept.

9.0 Study Estimands

This section describes the primary and secondary estimands, their attributes, and the proposed strategy to address the specified intercurrent events. The estimands described in this SAP, although not defined in the protocol, are based on the planned objectives, endpoints, and analysis sets from the protocol. For data not collected (missing) or excluded from the assessment of a given estimand (e.g. hypothetical strategy after the onset of an intercurrent event), specific data imputation rules are defined in Section 12.5.

9.1 Efficacy Estimands and their Attributes

The following intercurrent events will be considered for the definition of the estimands in the core period;

1. Treatment discontinuation or interruption due to Adverse Events with a compliance of < 80%: This includes subjects with either of the following two scenarios.

Scenario 1	AE record of: dose interruption or dose withdrawal.
Scenario 2	Disposition page record of discontinuation due to AE.

2. Treatment discontinuation due to Lack of Efficacy: This includes subjects with a record of discontinuations due to lack of efficacy in the disposition page.
3. Treatment discontinuation due to any other reason: This includes all discontinuations due to reasons other than those in 1) and 2) and also includes treatment interruptions or withdrawal due to adverse events as reported in AE page and treatment discontinuation as reported in disposition page but with a compliance of ≥ 80%.
4. Prohibited medications: Prohibited medications (as defined in protocol) will be determined using the protocol deviation log by the medical team. In the determination of prohibited medications, COVID-19 related vaccinations will not be considered as a prohibited medication.
5. Dose modification of Methotrexate: Subjects with a dose modification of Methotrexate will be determined by the medical team with the use of the Concomitant Medication dataset. Only Methotrexate used post baseline will be considered.
6. COVID-19 vaccination: COVID-19 vaccination will be considered as a separate potential intercurrent event. Subjects with COVID-19 vaccinations will be determined programmatically with the use of the concomitant medication dataset. Only COVID-19 vaccinations on or after first treatment date will be considered.

9.1.1 Primary and Associated Supportive Estimands

Primary Estimand 1.0 (DAS28-ESR ITT)

- Treatment of interest: Subcutaneous injection of MSB11456 compared to EU-RoActemra
- Population of interest: Patients with moderately to severely active RA who were randomized: ITT Analysis Set
- Variable / endpoint of interest: Absolute change from baseline at W24 for Disease Activity Score 28-Erythrocyte Sedimentation Rate (DAS28-ESR)
- Potential intercurrent events / strategy to address:
  - Treatment discontinuation or interruptions prior to W24, regardless of reason / treatment policy strategy
  - Use of prohibited medication or change in permitted medications (other than Methotrexate), with the potential to impact W24 efficacy assessments / treatment policy strategy
  - Modification in Methotrexate, with the potential to impact W24 efficacy assessments/ treatment policy strategy
  - Covid-19 vaccination prior to W24 / treatment policy strategy
- Population level summary: Difference between treatment groups on mean absolute change from baseline at W24

This estimand targets the effect of treatment on the variable measurement regardless of adherence to the treatment or to the protocol, including use of prohibited medication prior to W24 and follows a 'treatment policy' strategy for all intercurrent events.

Supportive Estimand 1.1 (DAS28-ESR Per Protocol)

- Treatment of interest: Subcutaneous injection of MSB11456 compared to EU-RoActemra
- Population of interest: Patients with moderately to severely active RA who were randomized, completed the Core Period and fulfilled all criteria to be included in the PP Analysis Set (see section 9.2.3)
- Variable / endpoint of interest: Absolute change from baseline at W24 for Disease Activity Score 28-Erythrocyte Sedimentation Rate (DAS28-ESR)
- Potential intercurrent events / strategy to address:
  - Treatment discontinuation or interruptions prior to W24, regardless of reason / treatment policy strategy
  - Covid-19 vaccination prior to W24 / treatment policy strategy
- Population level summary: Difference between treatment groups on mean absolute change from baseline at W24

This estimand targets the effect of treatment on the variable measurement in the target population of those subjects who adhered to the protocol and did not experience any clinically important protocol deviations impacting primary endpoint.

Supportive Estimand 1.2 (DAS28-ESR Hypothetical return-to-Baseline ITT)

- Treatment of interest: Subcutaneous injection of MSB11456 compared to EU-RoActemra

- Population of interest: Patients with moderately to severely active RA who were randomized: ITT Analysis Set
- Variable / endpoint of interest: Absolute change from baseline at W24 for Disease Activity Score 28-Erythrocyte Sedimentation Rate (DAS28-ESR)
- Potential intercurrent events / strategy to address:
    - Treatment discontinuation or interruption prior to W24
        - Due to lack of efficacy or due to Adverse Event (AE) / hypothetical strategy
        - Due to any other reason / treatment policy strategy
    - Use of prohibited medication or change in permitted medications (other than Methotrexate) with the potential to impact W24 efficacy assessments / hypothetical strategy
    - Modification in Methotrexate, with the potential to impact W24 efficacy assessments/ hypothetical strategy
    - Covid-19 vaccination prior to W24 / treatment policy strategy
- Population level summary: Difference between treatment groups on mean absolute change from baseline at W24

This estimand targets the effect of treatment on the variable measurement in the target population should the patient who discontinued treatment due to study related events (lack of efficacy or AE) or took prohibited medication or had any dose modification would no longer benefit from treatment and measurements are projected as a worst case scenario as if the patient DAS28-ESR value returns to baseline levels immediately after the time of the intercurrent event ('hypothetical' strategy). For other intercurrent events a treatment policy strategy is followed.

Supportive Estimand 1.3 (DAS28-ESR Hypothetical continuing per protocol ITT)

- **Treatment of interest:** Subcutaneous injection of MSB11456 compared to EU-RoActemra
- Population of interest: Patients with moderately to severely active RA who were randomized: ITT Analysis set.
- Variable / endpoint of interest: Absolute change from baseline at W24 for Disease Activity Score 28-Erythrocyte Sedimentation Rate (DAS28-ESR)
- Potential intercurrent events / strategy to address:
    - Treatment discontinuation or interruption prior to W24
        - Due to lack of efficacy or due to Adverse Event (AE) / hypothetical strategy
        - Due to any other reason / hypothetical strategy
    - Use of prohibited medication with the potential to impact W24 efficacy assessments / hypothetical strategy
    - New or modification in permitted medications for management of the underlying disease, with the potential to impact W24 efficacy assessments/ hypothetical strategy
    - Covid-19 vaccination prior to W24 / treatment policy strategy
- Population level summary: Difference between treatment groups on mean absolute change from baseline at W24.

This estimand targets the effect of treatment on the variable measurement in the target population should the patient did not discontinued treatment for any reason and should the patient did not have taken prohibited

medication nor any dose modification and follows a 'hypothetical' strategy for the relevant intercurrent events. For these intercurrent events a 'hypothetical' strategy is followed where measurements are projected as a per protocol scenario i.e as if the patient had continued to follow the protocol after the time of the intercurrent event. For other intercurrent events a treatment policy strategy is followed

9.1.2 Key Secondary and Associated Supportive Estimands

Key Secondary Estimand 2.0 (ACR20 ITT)

- Treatment of interest: Subcutaneous injection of MSB11456 compared to EU-RoActemra
- Population of interest: Patients with moderately to severely active RA who were randomized: ITT Analysis Set
- Variable / endpoint of interest: ACR20 response at W24
- Potential intercurrent events / strategy to address:
  - Treatment discontinuation prior to W24, regardless of reason/ treatment policy strategy
  - Use of prohibited medication or change in permitted medications (other than Methotrexate) with the potential to impact W24 efficacy assessments / treatment policy strategy
  - Modification in Methotrexate with the potential to impact W24 efficacy assessments/ treatment policy strategy
  - Covid-19 vaccination prior to W24 / treatment policy strategy
- Population level summary: Difference between treatment groups on ACR20 response rate at W24

This estimand targets the effect of treatment on the variable measurement regardless of adherence to the treatment or to the protocol, including use of prohibited medication prior to W24 and follows a 'treatment policy' strategy for all intercurrent events.

Supportive Estimand 2.1 (ACR20 Per Protocol)

- Treatment of interest: Subcutaneous injection of MSB11456 compared to EU-RoActemra
- Population of interest: Patients with moderately to severely active RA who were randomized, completed the Core Period, and included in the PP Analysis Set
- Variable / endpoint of interest: ACR20 response at W24
- Potential intercurrent events / strategy to address:
  - Treatment discontinuation or interruptions prior to W24, regardless of reason / treatment policy strategy
  - Covid-19 vaccination prior to W24 / treatment policy strategy
- Population level summary: Difference between treatment groups on ACR20 response rate at W24

This estimand targets the effect of treatment on the variable measurement in the target population of those subjects who adhered to the protocol and did not experience any clinically important protocol deviations.

Supportive Estimand 2.2 (ACR20 Hypothetical Non-Responder ITT)

- Treatment of interest: Subcutaneous injection of MSB11456 compared to EU-RoActemra

- Population of interest: Patients with moderately to severely active RA who were randomized: ITT Analysis Set
- Variable / endpoint of interest: ACR20 response at W24
- Potential intercurrent events / strategy to address:
    - Treatment discontinuation prior to W24
        - Due to lack of efficacy or due to AE / hypothetical strategy
        - Due to any other reason / treatment policy strategy
    - Use of prohibited medication or change in permitted medications (other than Methotrexate) with the potential to impact W24 efficacy assessments / hypothetical strategy
    - Modification in Methotrexate, with the potential to impact W24 efficacy assessments/ hypothetical strategy
    - Covid-19 vaccination prior to W24 / treatment policy strategy

- Population level summary: Difference between treatment groups on ACR20 response rate at W24

This estimand targets the effect of treatment on the variable measurement in the target population should the patient who discontinued treatment due to study related events (lack of efficacy or AE) or took prohibited medication or had any dose modification would no longer benefit from treatment and measurements are projected as a worst case scenario as if the patient ACR20 is considered as non-responder immediately after the time of the intercurrent event ('hypothetical' strategy). For other intercurrent events a treatment policy strategy is followed.

9.1.3 Early efficacy (Week 12) Estimands

The following estimands will be constructed based on DAS28-ESR at week 12;

- Supportive Estimand 3.0 (Early DAS28-ESR ITT) will have the same attributes as Primary Estimand 1.0, except that week 24 (W24) will be replaced by Week 12 (W12) in all relevant attributes' description.
- Supportive Estimand 3.1 (Early DAS28-ESR PP) will have the same attributes as Supportive Estimand 1.1, except that W24 will be replaced by W12 in all relevant attributes' description.
- Supportive Estimand 3.2 (Early DAS28-ESR Hypothetical Continuing Per Protocol ITT) will have the same attributes as Supportive Estimand 1.3, except that W24 will be replaced by W12 in all relevant attributes' description.

9.1.4 Summary of Intercurrent Events

Analyses	Intercurrent Events Treatment Policy	Intercurrent Events Hypothetical Policy
Primary Estimand 1.0 (DAS28-ESR ITT)	- Treatment discontinuation or interruptions prior to W24, regardless of reason - Use of prohibited medication or change in permitted medications (other than Methotrexate), with the potential to impact W24 efficacy assessments - Modification in Methotrexate, with the potential to impact W24 efficacy assessments - Covid-19 vaccination prior to W24	
Supportive Estimand 1.1 (DAS28-ESR PP)	- Treatment discontinuation or interruptions prior to W24, regardless of reason - Covid-19 vaccination prior to W24	
Supportive Estimand 1.2 (DAS28-ESR Hypothetical return-to-Baseline ITT)	- Treatment discontinuation or interruption prior to W24, due to any other reason - Covid-19 vaccination prior to W24	- Treatment discontinuation or interruption prior to W24, due to lack of efficacy or due to Adverse Event (AE - Use of prohibited medication or change in permitted medications (other than Methotrexate) with the potential to impact W24 efficacy assessments - Modification in Methotrexate, with the potential to impact W24 efficacy assessments
Supportive Estimand 1.3 (DAS28-ESR Hypothetical continuing per protocol ITT)	- Covid-19 vaccination prior to W24	- Treatment discontinuation or interruptions prior to W24, regardless of reason - Use of prohibited medication or change in permitted medications (other than Methotrexate), with the potential to impact W24 efficacy assessments - Modification in Methotrexate, with the potential to impact W24 efficacy assessments

Analyses	Intercurrent Events Treatment Policy	Intercurrent Events Hypothetical Policy
Key Secondary Estimand 2.0 (ACR20 ITT)	Same as in Estimand 1.0	Same as in Estimand 1.0
Supportive Estimand 2.1 (ACR20 PP)	Same as in Estimand 1.1	Same as in Estimand 1.1
Supportive Estimand 2.2 (ACR20 Hypothetical Non-Responder ITT)	Same as in Estimand 1.2	Same as in Estimand 1.2
Supportive Estimand 3.0 (Early DAS28-ESR ITT)	Same as in Estimand 1.0 (*)	Same as in Estimand 1.0 (*)
Supportive Estimand 3.1 (Early DAS28-ESR PP)	Same as in Estimand 1.2 (*)	Same as in Estimand 1.2 (*)
Supportive Estimand 3.2 (Early DAS28-ESR Hypothetical Continuing Per Protocol ITT)	Same as in Estimand 1.3 (*)	Same as in Estimand 1.3 (*)
Note: *) will have the same attributes as Supportive Estimand 1.3, except that W24 will be replaced by W12 in all relevant attributes' description		

9.2 Analysis Sets

9.2.1 Study Periods

- The Core Period begins on the day of initial randomization (Day 1) and includes assessments performed prior to the W24 dosing time. (see Appendix 15.1 for details and special cases). Data summaries for this period will be presented by initial treatment group – MSB11456 and EU-RoActemra – and overall, when applicable.
- The Extended Period for subjects that are re-randomized and re-dosed begins at the W24 dosing time and includes the ensuing 28 weeks of treatment and the 12 weeks safety evaluation period (up to Week 63 visit). Subjects that are not re-randomized do not have an extended period. Data summaries for this period will be presented by treatment group after re-randomization - MSB11456, EU-RoActemra/MSB11456, and EU-RoActemra – and overall, when applicable.
- The Overall Period comprises all study weeks from Day 1 to W63. Data summaries for this period will be presented by treatment group after re-randomization – MSB11456, EU-RoActemra/MSB11456, and EU-RoActemra – and overall, when applicable. Patients who stopped treatment before re-randomization or were never dosed post-re-randomization will be allocated to the MSB11456 or EU-RoActemra treatment group according to their initial randomization.

- The Week 12 Period is a subset of the Core Period. It begins on the day of initial randomization (Day 1) and includes assessments performed not later than W12 analysis visit (see Section 10.4). Data summaries for this period will be presented by Week 12 treatment group – MSB11456 and EU-RoActemra – and overall, when applicable.

9.2.2 Intent-to-Treat Analysis Set

The ITT Analysis Set includes all randomized patients. Patients will be analyzed according to their randomized treatment.

In the Extended Period (EP), the EP-ITT Analysis Set includes all patients who are re-randomized into the Extended Period. Patients will be analyzed according to their randomized treatment group; for patients initially randomized to EU-RoActemra, this will include both their initial and re-randomization assignment. Subjects who are re-randomized but not dosed afterwards will have no data contributing to the Extended Period apart from the extended baseline data.

9.2.3 Per Protocol Analysis Set

In the Core Period, the PP Analysis Set includes all randomized and treated patients (hence a subgroup of the ITT Analysis Set) who completed the Core Period: attended the Week 24 visit with no clinically important protocol deviations before the primary efficacy endpoint analysis time point (W24) and who have a treatment compliance of ≥ 80% (including Methotrexate compliance) in the Core Treatment Period (refer to Section 12.3.1 for a definition of treatment compliance). A clinically important protocol deviation is a major protocol deviation likely to affect the efficacy of treatment. Patients will be analyzed according to their randomized and received treatment, as receipt of a different treatment from that assigned is a clinically important protocol deviation. Methotrexate compliance will be determined by the clinical team prior to W30 database lock and W55 database lock. Since Methotrexate detailed dosing data is not captured in the case report forms, the medical team will review concomitant medication data and flag subjects with Methotrexate compliance of less 80% prior to Week 12 and Week 24 for the W30 and W55 CSR, before unblinding.

In the Week 12 Period, the W12-PP (Week 12 Per Protocol) Analysis Set includes all randomized and treated patients who attended the W12 visit, with no clinically important protocol deviations before Week 12 visit and who have a treatment compliance of ≥ 80% (including Methotrexate compliance) up to week 12 visit.

In the Extended Period, the EP-PP Analysis Set includes all re-randomized and treated patients (hence a subgroup of the EP-ITT Analysis Set) who do not have any clinically important protocol deviations before W52 and who have a treatment compliance of ≥ 80% (including Methotrexate compliance) in both the Core Treatment Period and in the Extended Treatment Period. This is not applicable to the W30 CSR. Patients will be analyzed according to their randomized and received treatment after re-randomization, as receipt of a different treatment from that assigned is a clinically important protocol deviation.

For the overall Period, an Overall PP Analysis set includes all randomized and treated patients who do not have any clinically important protocol deviations before W52 and who have a treatment compliance of ≥ 80% (including Methotrexate compliance) in both the Core Treatment Period and

in the Extended Treatment Period. The EP-PP Analysis Set and the Overall PP Analysis Set are the same. Patients will be presented by treatment group after re-randomization. For patients who were not re-randomized, they will be presented by treatment group according to initial randomization.

Any subject that receives an incorrect treatment during the study will be excluded from the PP analysis set corresponding to the period in which the error took place.

9.2.4 Safety Analysis Set

The Safety Analysis (SAF) Set includes all patients who receive at least 1 dose of study drug (MBS11456 or EU-RoActemra) during the Core Treatment Period. Patients will be analyzed according to the actual treatment they receive.

In the Extended Period, the EP-SAF Set includes all patients who receive at least 1 dose of study drug (MBS11456 or EU-RoActemra) during the course of the Extended Treatment Period. Patients will be analyzed according to the actual treatment they receive.

Should a subject be assigned multiple treatments by error (both MSB11456 and EU-RoEctemra) after randomization or re-randomization, the treatment assigned the majority of times will be used as the actual treatment.

9.2.5 Pharmacokinetic Analysis Set

The PK Analysis Set includes all treated patients who have at least 1 measurement of trough concentration in the Core Period and without major protocol deviations impacting pharmacokinetics during the Core Treatment Period. The protocol deviations that impact pharmacokinetics data will be flagged by the clinical team in the protocol deviation log.

In the Extended Period, the EP-PK Analysis Set includes all patients who have at least 1 measurement of trough concentration during the Extended Period and without major protocol deviations impacting pharmacokinetics until (and including) W52.

For the overall period, an Overall PK Analysis Set includes all subjects having at least one measurement of trough concentration during the study without major protocol deviations impacting pharmacokinetics during the treatment period. Patients will be presented by treatment group after re-randomization. For patients who were not re-randomized, they will be presented by treatment group according to initial randomization.

Patients will be analyzed according to the actual treatment they receive.

9.2.6 Summary

The table below presents an overview of the parameters that will be analyzed for each Analysis Set.

Analyses	All Screened	ITT	EP-ITT	PP	W12-PP	EP-PP	SAF	EP-SAF	PK	EP-PK
Patient disposition	✓	✓	✓							
Demographic and baseline characteristics		✓		✓						
Prior and concomitant medications		✓	✓							
Study drug exposure		✓	✓	✓		✓	✓	✓		
Major protocol deviations		✓	✓							
Primary and secondary estimands		✓*		✓*	✓*					
Secondary efficacy		✓	✓	✓		✓				
Other efficacy		✓	✓	✓		✓				
Immunogenicity							✓	✓		
Pharmacokinetics									✓	✓
Safety							✓	✓		

* corresponding estimand
Abbreviations:
ITT = Intent-to-Treat Analysis Set,
EP-ITT = Extended Period Intent-to-Treat Analysis Set,
PP = Per Protocol Analysis Set,
W12-PP = W12 Per Protocol Analysis Set,
EP-PP = Extended Period Per Protocol Analysis Set,
SAF = Safety Analysis Set
EP_SAF = Extended Period Safety Analysis Set
PK = Pharmacokinetic Analysis Set
EP-PK = Extended Period Pharmacokinetic Analysis Set

10.0 Conventions and Standards

Irrespective of being explicitly mentioned, all data collected on the eCRF will be presented in data listings. In those listings that are displayed by treatment group all patients who have started the Core Period on MSB11456 will be found in the MSB11456 section. Patients who have been on EU-RoActemra in the Core Period and have been re-randomized to MSB11456 will be found in the EU-RoActemra/MSB11456 section. Patients who have started the Core Period on EU-RoActemra and have either discontinued before re-randomization or have been re-randomized to EU-RoActemra will be found in the EU-RoActemra section.

An in-depth compendium on further conventions, standards, derivations and definitions can be found in the appendix.

10.1 Baseline and Change from Baseline

Unless stated otherwise, for each endpoint of interest, the baseline is defined as the last non-missing assessment taken prior to randomization and the first dose of study drug. In cases where baseline assessments are taken on the same day as the first dose of study drug and no times are reported, it will be assumed that these assessments are taken prior to study drug being administered. For patients who are randomized but not dosed after randomization, the baseline is defined as the last non-missing assessment prior to or on the date of randomization.

Change from baseline is defined as (value at post-baseline visit – value at baseline).

For analyses performed on some safety data in the Extended Period, baseline will be reset as the last non-missing assessment prior to or on the W24 treatment. This will be referred to as “extended baseline” to distinguish from the Core Period baseline described above.

10.2 Study Day

The first day of study drug administration is defined as Day 1. In this study, the first study drug administration occurs on the day of randomization. For patients who are randomized but not dosed after randomization, Day 1 is defined only as the randomization date.

Study day is defined as the number of days from Day 1, with the day prior to Day 1 defined as Day -1.

- Before Day 1: Study Day = (Date of interest – Date of Day 1)
- On or after Day 1: Study Day = (Date of interest – Date of Day 1) + 1

10.3 Reporting Time Points

There will be 3 analysis and reporting time points in the study:

- W30 CSR: The primary efficacy analysis at W24 will be conducted after all patients have completed the W30 assessments or have terminated the study before W30. The report will include all data up to and including data from the W30 Visit. Extended period will not be completed for all subjects at this time, and hence the Extended Period PP cannot be defined for the W30 CSR. Refer to the Data Access Plan for further details regarding unblinding in the context of the W30 CSR.

- W55 CSR: The full clinical study analysis will be conducted at W55 after all patients have completed both the Extended Treatment Period (W24 to W52) and the Safety Follow-up Visit (W55) or have terminated the study before W55. This CSR will include all results, including the outputs already produced for the W30 CSR (rerun). Refer to the Data Access Plan for further details regarding unblinding in the context of the W55 CSR
- W63 Safety Follow-up CSR Addendum: This CSR Addendum will include the safety data collected up to W63 (Physical examination, Vital signs, electrocardiogram (ECG), Chest X-ray, Concomitant medications and procedures, AEs and Laboratory data). Outputs will be provided on the overall treatment period.

10.4 Windows

For the purpose of analysis, the actual visit date will be mapped as follows:

Visit	Target Day	Study Day	Interval (Days)
2 (Baseline)	1	≤ 1	N/A
3 (W1)	8	2-11	10
4 (W2)	15	12-22	11
5 (W4)	29	23-43	21
6 (W8)	57	44-71	28
7 (W12)	85	72-99	28
8 (W16)	113	100-141	42
9 (W24)	169	142-190	49
10 (W30)	211	191-253	63
11 (W42)	295	254-330	77
12 (W52)	365	331-375	45
13 (W55)	386	376 -414	39
14 (W63)	442	≥ 415	N/A

Abbreviations: W=week; N/A = not applicable

If more than 1 visit within the same period (including unscheduled visits) falls within the same defined window, the following rules will apply:

- If 2 actual visit dates with non-missing data have the same time interval from the target day, the latest visit with non-missing data will be considered for analysis
- The worst results within each window will be considered for cases when the analysis of a worst result is applicable (e.g. for shift tables or NCI-CTCAE analyses of laboratory results)
- In all other cases the visit closest to the target day with non-missing data will be considered for analysis

10.5 Efficacy Measurements

10.5.1 DAS28-ESR

The DAS28-ESR is a measure of disease activity in 28 joints that consists of a composite numerical score of the following variables: Tender Joint Count, Swollen Joint Count, erythrocyte sedimentation rate and Patient's Global Assessment of Disease Activity. DAS28-ESR will be derived using the following formula:

$$DAS28\_ESR = 0.56*\sqrt{TJC28}+0.28*\sqrt{SJC28}+0.014*GH+0.7*\ln(ESR)$$

Note: If ESR=0 then ESR=1 will be used for DAS28ESR derivation. It is clinically justified to carry out this imputation because an ESR value of 0 or 1 represents negligible inflammation.

Where:

- TJC28 = 28 joint count for tenderness
- SJC28 = 28 joint count for swelling
- ln(ESR) = natural logarithm of erythrocyte sedimentation rate [mm/hr]
- GH = the general health component of the DAS (i.e., Patient's Global Assessment of Disease Activity on a scale of 1 to 100 where 100 is maximal activity)

The level of disease activity will be categorized as follows:

- Remission: DAS28-ESR < 2.6
- Low: 2.6 ≤ DAS28-ESR < 3.2
- Moderate: 3.2 ≤ DAS28-ESR ≤ 5.1
- High: DAS28-ESR > 5.1

The imputation method for missing individual joint assessments is explained in Section 15.9.1 in the appendix.

10.5.2 DAS28-CRP

The Disease Activity Score 28-C-reactive protein (DAS28-CRP) is a measure of disease activity in 28 joints that consist of a composite numerical score of the following variables: Tender Joint Count, Swollen Joint Count, C-reactive protein and Patient's Global Health Assessment of Disease Activity.

$$DAS28\_CRP = 0.56*\sqrt{TJC28}+0.28*\sqrt{SJC28}+0.014*GH+0.36*\ln(CRP+1)+0.96$$

Note: CRP values below the LoQ (<0.3) will be imputed using the LLoQ.

Where:

- TJC28 = 28 joint count for tenderness
- SJC28 = 28 joint count for swelling
- CRP = C-reactive protein [mg/L]
- GH = General Health of the DAS (i.e., Patient's Global Assessment of Disease Activity on a scale of 1 to 100 where 100 is maximal activity)

A patient will be further classified as a responder using The European League Against Rheumatism (EULAR) Boolean-based response criteria definition:

At any timepoint, a patient must satisfy all of the following:

- Tender joint count 28 ≤1
- Swollen joint count 28≤1
- C-reactive protein ≤1 mg/dL
- Patient Global Assessment ≤1 (on a 0-10 scale)

The imputation method for missing individual joint assessments is explained in Section 15.9.1 in the appendix.

10.5.3 ACR20/50/70

The ACR20 is an efficacy measure for which a patient must have at least 20 % improvement in the following ACR Core Set values:

- Tender Joint Count (68 joint count)
- Swollen Joint Count (66 joint count)
- An improvement of at least 20 % in at least 3 of the following 5 assessments:
    - Patient's Global Assessment of Disease Activity
    - Patient's Assessment of Arthritis Pain
    - Patient's Assessment of Physical Function as measured by the Health Assessment Questionnaire-Disability Index
    - Physician's Global Assessment of Disease Activity
    - Acute phase reactant as measured by erythrocyte sedimentation rate or C-reactive protein

ACR50 = 50 % improvement in at least 3 of the 5 measures and 50 % improvement in the Swollen and Tender Joint Count.

ACR70 = 70 % improvement in at least 3 of the 5 measures and 70 % improvement in the Swollen and Tender Joint Count.

The imputation method for missing individual joint assessments is explained in Section 15.9.1 in the appendix.

10.5.4 Simplified Disease Activity Index

The Simple Disease Activity Index (SDAI) includes 5 items:

- 28–Swollen Joint Count
- 28–Tender Joint Count
- Patient's Global Assessment of Disease Activity on a Visual Analog Scale (0 to 10.0 scale)
- Physician's Global Assessment of Disease Activity on a Visual Analog Scale (0 to 10.0 scale)
- C-reactive protein level in mg / dL

$$SDAI = SJC + TJC + PGA+ PhGA + CRP$$

Since the visual analog scale used in the study is a 0 to 100 mm scale, the corresponding items will be divided by 10 before summing.

The imputation method for missing individual joint assessments is explained in Section 15.9.1 in the appendix.

The level of disease activity can be interpreted as:

- Remission: $SDAI \leq 3.3$
- Low: $3.3 < SDAI \leq 11$
- Moderate: $11 < SDAI \leq 26$
- High: $26 < SDAI$

10.5.5 Clinical Disease Activity Index

The Clinical Disease Activity Index (CDAI) includes 4 items:

- 28–Swollen Joint Count
- 28–Tender Joint Count
- Patient's Global Assessment of Disease Activity on a Visual Analog Scale (0 to 10.0 scale)
- Physician's Global Assessment of Disease Activity on a Visual Analog Scale (0 to 10.0 scale)

$$CDAI = SJC + TJC + PGA + PhGA$$

Since the visual analog scale used in the study is a 0 to 100 mm scale, the items will be divided by 10 before summing.

The imputation method for missing individual joint assessments is explained in Section 15.9.1 in the appendix. The level of disease activity can be interpreted as:

- Remission: $CDAI \leq 2.8$
- Low: $2.8 < CDAI \leq 10$
- Moderate: $10 < CDAI \leq 22$
- High: $22 < CDAI$

10.6 Handling of Missing Dates

In general, incomplete dates will not be imputed. If an imputation is required for analysis:

- For dates where day only is missing, the first day of the month will be imputed for start dates and the last day of the month will be imputed for stop dates
- For dates where day and month are missing, the first day of the year will be imputed for start dates and the last day of the year will be imputed for stop dates

If start dates are missing or incomplete for an AE (including deaths) or medications, the following algorithm will be used for imputation:

Start Date		Stop Date						
		Complete: yyyymmdd		Partial: yyyymm		Partial: yyyy		Missing
		<1st dose	≥1st dose	<1st dose yyyymm	≥1st dose yyyymm	<1st dose yyyy	≥1st dose yyyy	
Partial: yyyymm	=1st dose yyyymm	2	1	N/A	1	N/A	1	1
	≠1st dose yyyymm		2	2	2	2	2	2
Partial: yyyy	=1st dose yyyy	3	1	3	1	N/A	1	1
	≠1st dose yyyy		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1=impute as the date of first dose; 2=impute as the first of the month; 3=impute as January 1 of the year; 4=impute as January of the stop year

Imputation rules for missing or incomplete stop dates for AEs and medications:

Initial imputation

- For partial stop date “mmyyyy”, impute the last day of the month
- For partial stop date “yyyy”, impute December 31 of the year
- For completely missing stop date, do not impute
- If the stop date imputation leads to a stop date that is after the death date, then impute the stop date as the death date
- If the stop date imputation leads to a stop date that is before the start date, then there is a data error and do not impute the stop date

Imputation rules for partial or missing death dates:

- If death year and month are available but day is missing:
  - If “mmyyyy” for last contact date = “mmyyyy” for death date, set death date to the day after the last contact date

  - If "mmyyyy" for last contact date < "mmyyyy" for death date, set death date to the first day of the death month
  - If "mmyyyy" for last contact date > "mmyyyy" for death date, data error and do not impute
- If both month and day are missing for death date or a death date is totally missing, set death date to the day after the last contact date

The imputed dates will be used to assess whether AEs should be considered as treatment-emergent and if medications should be included in the safety summaries as prior or concomitant, however the original partial dates will be included in data listings.

11.0 Interim Analyses

An interim analysis is not planned for this study.

As stated in Section 10.3, the primary analysis (including the primary efficacy analysis at W24) will be conducted after all patients have completed the W30 assessments or have terminated the study before W30. This analysis is considered a final analysis from a statistical perspective even though the study continues until W63.

12.0 Statistical Methods

All analyses will use SAS version 9.4 or higher.

The coding dictionaries available at date of first patient in are WHODRUG Dictionary Version WHODRUG2020MAR01_GLOBAL_B3 and MedDRA Version 23.0. For further details on coding dictionaries, please see the Data Management Plan and the Coding Conventions Specifications document.

Unless otherwise noted, categorical variables will be summarized using counts and percentages and include a missing category. Percentages will be rounded to 1 decimal place, except 100 % will be displayed without any decimal places and percentages will not be displayed for zero counts. Percentages will be based on the number of patients in the analysis set of interest (N). If the analysis is done by visits, percentages will be based on the number of patients still present in the study at the visit.

Continuous variables will be summarized using the number of patients in the analysis set (N), number of observations (n), number of missing observations (N-n), mean, standard deviation (SD), median, minimum and maximum. The minimum and maximum values will be displayed to the same level of precision as the raw data; the mean and median to a further decimal place; and the SD to 2 additional decimal places.

Unless otherwise described, all tables will either summarize

- the Core Period (displayed by the 2 originally randomized treatment groups)
- the Extended Period (displayed by the 3 treatment groups after re-randomization), or
- the Overall Period (displayed by the 3 treatment groups after re-randomization but including data from the Core Period and the Extended Period)

Core and Overall Period outputs that are referring to a baseline will always use the Core Period baseline. Unless otherwise specified, extended period tables referring to a baseline will use the extended baseline.

12.1 Patient Disposition

The following information will be summarized for patient disposition by treatment group and in total:

- Number of patients randomized at the initial randomization (2 treatment groups) tabulated by country, and center using the ITT analysis set
- Patient disposition for the Core Period and the Extended Period (including, number screened, number of screen failures, number of patients not randomized and reasons, the number of patients who were randomized, treated with study drug, completed treatment, discontinued treatment with reason for discontinuation, completed period/study, and discontinued period/study with reason for discontinuation). The denominator of these percentages is based on the ITT Analysis Set for study discontinuation and the SAF Analysis Set for treatment discontinuation
- The number of subjects receiving treatment at Week 12, Week 24 and Week 52 will be summarized
- The reason for exclusion from the PP set at Week 12 and Week 24 will be summarized
- Summaries of analysis sets with reason for exclusion using all screened subjects.
- Note that EP-PP and EP-PK analysis sets will not be determined at the Week 30 lock and will not appear in the Week 30 CSR TFLs
- Major protocol deviations prior to Week 12 and prior to Week 24 leading to exclusion from the PP and/or PK Analysis Sets
- Randomization list of patients and their actual versus randomized treatment group and strata (previous exposure to biological treatment for RA) for all randomized and re-randomized patients

12.2 Demographic and Baseline Characteristics

Where permitted by local regulation the following demographic and baseline data will be summarized for the ITT and PP Analysis Set for the Core Period by treatment group and in total.

- Age [in years, at day of patient signature on informed consent]
- Race
- Sex
- Ethnicity
- Height
- Weight and
- Body mass index

General medical history will be coded according to MedDRA and will be summarized for the ITT Analysis Set and PP Analysis Set by system organ class (SOC) and preferred term (PT) for the Core Period by treatment group and in total.

RA history will be summarized using the ITT and PP Analysis Set. RA history variables include:

- RA functional class
- Previous exposure to biologic treatment for RA [yes/no]
- Time since symptom onset
- Time since first RA diagnosis
- Time from first symptoms to diagnosis and
- Presence and type of extra-articular manifestations

In addition, baseline efficacy endpoint data as well as Antidrug Antibody (ADA) status will be summarized using the ITT and PP Analysis Set.

General prior medications and prior medications for disease under study will be coded using WHODRUG and will be summarized for the ITT Analysis Set by medication preferred name and medication category (Anatomic Therapeutic Classification (ATC) levels 2 and 4) for the Core Period by treatment group and in total.

Prior medications are defined as any medication discontinued prior to the first dose of study treatment.

Tables will be sorted by decreasing order of frequency in the ATC groups and PTs within ATC groups in the total across all treatment groups.

Prior procedures (i.e. surgeries etc.), defined as any procedure that occurred prior to randomization or the first dose of study treatment will be displayed in a data listing.

Smoking status will be listed.

12.3 Treatments

12.3.1 Extent of Study Drug Exposure

Summary statistics will be provided on the total number of doses administered and the total duration of exposure (date of last dose – date of first dose +7 day) for the Core Period (ITT, PP, and SAF Analysis Sets), Extended Period (EP-ITT, EP-PP, EP-SAF Analysis Sets), and Overall Period (ITT, EP-PP, SAF Analysis Sets) by treatment group and in total. It should be noted that for the extended period related outputs in the Week 30 CSR, only the following exposure summaries will be presented, exposure from Week 24 to Week 30 (EP-ITT and EP-SAF) and exposure from Day 1 to Week 30 (EP-ITT and EP-SAF).

In addition, treatment compliance will be estimated per patient by the proportion of the planned number of injections actually administered and be summarized as both a continuous and categorical variable for the Core Period (Week 12 and 24), Extended Period, and Overall Period by treatment group. Treatment compliance will be categorized as follows: $< 80\%$, $\geq 80\%$ to $\leq 100\%$ and $> 100\%$. If a patient takes the treatment that they were not randomized to, for the purposes of compliance, that injection will be considered as missing for calculations.

For the Core Period, 24 injections are planned for all patients. In the Extended Period 28 injections are planned. These planned numbers apply to all patients, irrespective of discontinuation of study or study drug. Therefore, the denominator for Core Period treatment compliance is 24, for

Extended Period treatment compliance is 28 (7 for the Week 30 CSR) and for Overall Period treatment compliance is 52 (31 for the Week 30 CSR).

Additionally, cumulative exposure, calculated in patient-years, will be summarized. The duration of 1 year is defined as 365.25 days for that purpose. Cumulative exposure is defined as follows:

$$Years\ on\ Study = \left(\sum_{\substack{applicable \\ analysis\ set}} (last\ day - day\ of\ first\ injection + 7) \right) / 365.25$$

12.3.2 Concomitant Medications and Procedures

Concomitant medications, defined as those medications taken at any time on or after the first dose of study treatment, will be coded using WHODRUG and will be summarized for the ITT Analysis Set by medication group and medication subgroup (ATC Level 2 and 4) for the Core Period, Extended Period, and Overall Period by treatment group and overall. Core and Overall Period tables will utilize the ITT Analysis Set. Extended Period tables will utilize the EP-ITT Analysis Set.

The number and percentage of patients using each medication will be displayed together with the number and percentage of patients using at least 1 medication within each medication group and subgroup.

This display of medications will be provided for general concomitant medications and for when the intake reason is the disease under study.

If a medication or procedure date is missing or partially missing and it cannot be determined whether the medication/procedure was prior or concomitant, it will be considered as concomitant.

Tables will be sorted by decreasing order of frequency in the ATC groups and PTs within ATC groups in the total across all treatment groups.

Concomitant procedures, defined as those procedures happening concurrently after the first dose of study treatment, will be displayed in a data listing.

12.4 Major Protocol Deviations

All protocol deviations data will be entered into the Clinical Trials Management System (CTMS). The study team and the Sponsor will conduct on-going reviews of the deviation data from CTMS. Details concerning handling and rating of protocol deviations can be found in the "Protocol Deviation Guidance" document.

The list of major and clinically important protocol deviations will be finalized before the Week 30 partial database lock for any deviations occurring before/at the Week 30 visit and again before the Week 55 database lock for deviations occurring during the remainder of the Extended Period.

The PP (W12 and W24) and PK Analysis Sets must be finalized at the post-freeze data review meeting, prior to unblinding for the W30 CSR. The EP-PP and EP-PK Analysis Set must be

finalized at the post-freeze data review meeting and follow-up, prior to unblinding for the W55 CSR, with the understanding that post lock discovery of any patient taking the incorrect treatment will exclude the patient from the EP-PP and EP-PK populations.

Major protocol deviations will be summarized for the Core Period and Extended Period using the ITT and the EP-ITT Analysis Sets, respectively, by treatment group and in total.

A listing of all protocol deviations (minor and major) will be provided.

A table and listing detailing all COVID-19 related protocol deviations, summarized by the Core and Extended Period, will be provided. Categories will be based on missed visits or endpoints not collected in the manner described in the protocol. The COVID-19 situation continues to be monitored closely and additional COVID-19 related analyses may be warranted, in which case a SAP amendment would be prepared. The number of subjects with COVID-19 related vaccinations prior to Week 24 will be identified using the concomitant medication data and summarized.

12.5 Efficacy Analyses

All questionnaires and assessments that are conducted to support the below described analyses will be listed. These include:

- Swollen and tender joint counts 28 and 66/68
- Patient's and Physician's Global Assessment of Disease Activity
- Health Assessment Questionnaire-Disability Index
- Patient's assessment of arthritis pain

12.5.1 Study Hypotheses

The hypothesis of the primary analysis in this study is as follows:

Primary hypothesis for the FDA:

H_0: $(\mu_T - \mu_R) < -0.6$ or $(\mu_T - \mu_R) > 0.5$

H_1: $-0.6 \leq (\mu_T - \mu_R) \leq 0.5$

where

- μ_T = mean absolute change from baseline in DAS28-ESR at W24 for MSB11456
- μ_R = mean absolute change from baseline in DAS28-ESR at W24 for EU-RoActemra
- hypothesis test is based on 90% confidence intervals

Primary hypothesis for the EMA:

H_0: $(\mu_T - \mu_R) < -0.6$ or $(\mu_T - \mu_R) > 0.6$

H_1: $-0.6 \leq (\mu_T - \mu_R) \leq 0.6$

where

- μ_T = mean absolute change from baseline in DAS28-ESR at W24 for MSB11456

- μ_R = mean absolute change from baseline in DAS28-ESR at W24 for EU-RoActemra
- hypothesis test is based on 95% confidence intervals

12.5.2 Hypothesis Testing Strategy and Multiplicity

There is no adjustment for multiplicity because this study has only one primary endpoint, and study success will be evaluated according to the success criterion defined for each agency.

12.5.3 Time to Intercurrent Events

A Kaplan-Meier (K-M) plot will be provided to display the time of occurrence of intercurrent event(s). The following will be provided:

- A plot, for each intercurrent event, displaying K-M curve for time to occurrence of the intercurrent event
- A plot displaying the time to occurrence of first intercurrent event for patients with at least one intercurrent event.

12.5.4 Primary Estimand 1.0 (DAS28-ESR ITT)

The trial will compare subcutaneous injections of MSB11456 to EU-RoActemra. The primary estimand to demonstrate equivalent efficacy of MSB11456 and EU-RoActemra is based on the DAS28-ESR at W24 (see Section 9.1.1). The primary comparison of interest is the change from baseline at Week 24. The primary comparison will be made according to treatment policy regardless of whether patients discontinue treatment prior to W24, used any prohibited medications, had any Methotrexate dose modification prior to W24, or received a COVID-19 related vaccine prior to Week 24, i.e., all available information at W24 will be included in the analysis.

12.5.4.1 Imputation Methods

According to the protocol patients who discontinue study treatment are asked to attend the W24 visit, with the intention of reducing the amount of missing data for patients who discontinue treatment. Missing DAS28-ESR scores at W24 will be imputed by a multiple imputation (MI) procedure as described below and further detailed in Section 15.9.2 of the Appendix.

For a subject with multiple IEs that are addressed with the hypothetical strategy the date of the first IE occurrence for the imputation of that estimand will be used for censoring. IEs that are addressed with treatment policy will not be considered in determining the date for censoring.

Non-monotone missing data (i.e. interim missing DAS28-ESR scores for patients who have missed visits/endpoints but have returned for next visit) is assumed to be missing at random (MAR) and will be imputed separately for each treatment group using the Markov chain Monte Carlo (MCMC) option of SAS PROC MI.

Then the monotone missing values for each treatment group will be imputed via the chained equation method, using SAS PROC MI option MONOTONE REG. First, all missing data for the first post-baseline visit are imputed; then missing data for the next visit are imputed using observed data plus the just imputed missing data; and so on to the W24 visit. The number of imputations

will be set to 500 and the PROC MI steps will always use the seed of 135984. Missing values for DAS28-ESR score will be imputed at each post-baseline visit. Baseline DAS28-ESR score and randomization stratification variable will be used to model the distribution of trajectory values.

Imputed DAS28-ESR scores will be restricted such that the values are greater than zero. Sample SAS code can be found in Section 15.9.2 of the appendix.

12.5.4.2 Primary Analysis

Change from baseline at W24 in DAS28-ESR will be analyzed using an analysis of covariance with treatment group and previous exposure to biologic treatment for RA [yes/no] as fixed effects and baseline DAS28-ESR as a covariate. Stratification variables will be used as entered in IRT. The difference between treatments will be estimated by the least squares mean difference between MSB11456 and EU-RoActemra, with its 95% confidence interval for the EMA and its 90% confidence interval for the FDA.

- For the FDA: MSB11456 will be considered equivalent to EU-RoActemra if the 90% confidence interval for the difference in mean change from baseline to W24 in DAS28-ESR between MSB11456 and EU-RoActemra lies entirely within the equivalence interval of [-0.6, 0.5]
- For the EMA: MSB11456 will be considered equivalent to EU-RoActemra if the 95% confidence interval for the difference in mean change from baseline to W24 in DAS28-ESR between MSB11456 and EU-RoActemra lies entirely within the equivalence interval of [-0.6, 0.6]

The primary analysis is based on the Core Period using the ITT Analysis Set.

A forest plot showing the 90% and 95% confidence intervals around the mean treatment difference of change from baseline to W24 in DAS28-ESR will be created for overall and the subgroup populations.

12.5.4.3 Supportive Estimand 1.1 (DAS28-ESR Per-Protocol)

The same ANCOVA analysis used for the primary estimand will be performed on the PP Analysis Set. All data available will be included in the analysis. No or little missing data is foreseen and hence it is not planned to impute missing values. However, if more than 5% of subjects included in the PP analysis set have missing DAS28-ESR assessments at Week 24 then the same MI applied to Estimand 1.0 will be performed.

12.5.4.4 Supportive Estimand 1.2 (DAS28-ESR Hypothetical return-to-Baseline ITT)

A similar analysis as for the primary estimand will be performed with the exception that measurements that occur after treatment discontinuation due to lack of efficacy or due to AE as well as after use of prohibited medication (or modification of permitted medications) that could impact W24 efficacy assessments will not be considered in the analysis. Imputation of missing DAS28-ESR at W24 and imputation of discounted values due to the occurrence of relevant intercurrent events as described above will be based on a pattern-mixture model combining a baseline observation carried forward (BOCF) MI approach with MAR sequential imputation.

The BOCF MI approach will impute multiple values drawn from the baseline distribution of the endpoint values and will be used for DAS28-ESR assessments after patients' discontinuation from study treatment prior to W24 due to lack of efficacy or due to an AE and/or for patients who used prohibited medication or modified their permitted medications prior to W24. For those patients, the assumption will be made that W24 DAS28-ESR would regress to the baseline value. The distribution of baseline values will be estimated using a regression model with baseline DAS28-ESR score and randomization stratification variable as covariates while predictions from this model will be used to impute missing and discounted (observations affected by IE) DAS28-ESR at W24. For other missing data, MAR sequential imputation will be used (as for the primary estimand).

12.5.4.5 Supportive Estimand 1.3 (DAS28-ESR Hypothetical continuing per protocol ITT)

A similar analysis as for the primary estimand 1.0 will be performed with the exception that the following DAS28-ESR assessments will be discounted and those now missing assessments will be imputed using MI as in the primary analysis:

- DAS28-ESR values that occurred after treatment discontinuation due any reason
- DAS28-ESR values that occurred after use of prohibited medication (or modification of permitted medications) that could impact W24 efficacy.

For these intercurrent events a 'hypothetical' strategy is followed where measurements are projected as a per protocol scenario as if the patient had continued to follow the protocol after the time of the intercurrent event

12.5.4.6 Sensitivity Analyses for the primary estimand 1.0

The following sensitivity analyses to be performed on the ITT Analysis Set will be confirmed at the blinded data review meeting prior to the W30 partial database lock:

- As little missing data is expected, the primary analysis of the primary estimand (DAS28-ESR) will be performed with no imputation of missing values.
- If data is not normally distributed as expected, a sensitivity analysis will be performed using appropriate non-parametric methods.
- A listing will be produced comparing the previous exposure to biologic treatment for RA strata used in IRT versus the patient information recorded on the eCRF at the randomization visit. If randomized stratum versus actual has substantial differences, a sensitivity analysis for the primary estimand will be performed that uses the actual strata information instead of the those used in IRT.
- To assess the robustness of the MAR assumption for the primary estimand, a tipping point analysis will be performed, based on the same analysis of covariance model as for the primary analysis. Missing changes from baseline at W24 in DAS28-ESR will be imputed by *proc mi* in SAS, accounting for treatment arm, randomization stratum and baseline DAS28-ESR as a continuous covariate (if baseline is also missing then the subject will be excluded from the analysis).

For the complete datasets obtained by *proc mi* the so-called δ-method will be applied (Cro, 2020). For pairs of the shift parameters, denoted by ($\delta 1$, $\delta 2$), the shift parameter $\delta 1$ will be added to each imputed value in the MSB11456 treatment arm, and the shift parameter $\delta 2$ will be added to each imputed value in the RoActemra treatment arm.
500 series of imputations will be performed for each pair of shift parameters using the same seed as with the primary estimand 1.0, each complete dataset will then be analyzed by the original ANCOVA model, and the results of the analysis will be combined by *proc mianalyze* in SAS. 90% and 95% CIs will be calculated for the combined results.
$\delta 1$ and $\delta 2$ values will be incremented from -3 to 3 by steps of 1 (imputed value should not result in negative value for DAS28-ESR at week 24). Results for each pair (49 pairs in total) will be tabulated. The random seed initializing *proc mi* will be 135984.

12.5.5 Key Secondary Estimand

12.5.5.1 Key Secondary Estimand 2.0 (ACR20 ITT)

12.5.5.1.1 Imputation Methods

The comparison will be made according to treatment policy and all available information at W24 will be included in the analysis (even if collected after occurrence of an intercurrent event).

Missing ACR20 response data will be imputed using the LOCF method. All missing ACR20 assessments will be imputed using the last non-missing assessment. Subjects who have just a baseline assessment will have their post baseline assessments imputed as non-responders.

LOCF is proposed because few missing ACR assessments are expected at Week 24. However, a sensitivity analysis using MI is proposed to address any situation of a large number of missing values at week 24.

12.5.5.1.2 Analysis Method

The difference in ACR20 response rate at W24 will be compared using a 95 % stratified Newcombe confidence interval to adjust for the stratification factor previous exposure to biologic treatment for RA [yes/no] and assessed against an equivalence margin of [-15 %, 15 %]. Mantel-Haenszel weights will be used to combine the stratum components

This key secondary analysis will be performed on the ITT Analysis Set.

12.5.5.2 Supportive Estimand 2.1 (ACR20 Per-Protocol Analysis)

The same analysis method as the key secondary estimand **(ACR20 ITT)** will be performed on the PP Analysis Set. Very little missing data is expected and it is therefore not planned to impute missing values. However, if at least 5% of subjects in the per protocol analysis set have missing ACR20 assessments at Week 24 then the same LOCF method applied to Estimand 2.0 will be performed.

12.5.5.3 Supportive Estimand 2.2 (ACR20 Hypothetical Return to Baseline ITT)

A similar analysis as that for the primary estimand will be carried out with the exception that measurements that occurred after treatment discontinuation due to lack of efficacy or due to AE as well as after use of prohibited medication (or modification of permitted medications) that could impact W24 efficacy assessments will not be considered in the analysis. Imputation of missing ACR response data at W24 and imputation of discounted values due to the occurrence of relevant intercurrent events as described above will be based on a combination of a non-responder approach with LOCF. The non-responder approach will be used for patients who discontinued from study treatment prior to W24 due to lack of efficacy or due to an AE and/or for patients who used prohibited medication or modified their permitted medications prior to W24. For other missing data, LOCF will be used (as for the key secondary estimand).

12.5.5.4 Sensitivity Analyses for the key secondary estimand 2.0

The following sensitivity analyses to be performed on the ITT Analysis Set will be confirmed at the blinded data review meeting prior to the W30 partial database lock:

- The ACR20 response rate at week 24 analysis will be performed on the ITT Analysis Set with no imputation.
- Missing ACR20 values will be imputed using the MI approach.

  This key secondary analysis will be performed on the ITT Analysis Set. Multiple imputation (MI) will be used to impute missing ACR20 response data at Week 24. Multiple imputation of missing data ACR20 response will be performed by fully conditional specification (FCS) using PROC MI in SAS with the FCS option. The FCS MI specifies the multivariate imputation model on a variable-by-variable basis by a set of conditional densities. For each missing value at Week *X*, the FCS MI will generate values from a conditional distribution for the missing data given the other data prior to Week *X*. The FCS MI model will include treatment group and stratification factors for randomization. A total of 500 imputed (complete) datasets will be generated for the MI analysis.

  The difference in ACR20 response rate at W24 will be compared using a 95 % stratified Newcombe confidence interval to adjust for the stratification factor previous exposure to biologic treatment for RA [yes/no] and assessed against an equivalence margin of [-15 %, 15 %] for each imputed dataset. SAS PROC MIANALYZE will be used to pool the results of the 95 % stratified Newcombe confidence intervals and generate an overall result by Rubin's rules (1987). See code in Section 15.9.3.2
- Tipping point analysis will further be used to assess the robustness of the primary study results. Different number of events between treatment groups will be assessed until the study conclusion is changed. All combinations of the number of responders and non-responders among the missing values will be evaluated. The stratification factor will not be taken into account in the generation of the combinations of responders and non-responders among missing data. Consequently, the selection of responders and non-responders will be only performed once for each combination. Each imputed value is initially imputed as a responder. The imputed values in the EU-RoActemra group will remain as responders while the MSB11456 imputed values are replaced as a nonresponders one at a time, therefore changing

the number of events between groups. Once all imputed values in the MSB11456 group have been replaced with nonresponders values, the data will reset to where the MSB11456 group imputed values are all responders and one by one the imputed values in the EU-RoActemra group are replaced with a nonresponder until all EU-RoActemra are nonresponders and all MSB11456 are responders.

Furthermore, every pair of imputations between the EU-RoActemra and MSB11456 groups will be assessed similarly, creating a matrix of possible patterns. At each iteration, a similar analysis to that of the key secondary analysis is assessed, and the direction of the analysis is recorded.

A graphical method will be used to present the tipping point analysis for ACR20 at week 24 (Liublinska, 2014). On the two axes of the graph all the possible numbers of successes in between two treatment arms for subjects with missing data will be presented. For each pair of supposed outcome values the common risk difference across strata and its 95% CI will be estimated by the method specified for the LOCF key secondary analysis.

12.5.6 Secondary Estimand 3.0

12.5.6.1 Main Analysis 3.0

The main analysis is similar to the primary analysis of Primary Estimand (DAS28-ESR ITT) but uses Change from baseline at W12 in DAS28-ESR. This will be analyzed using an analysis of covariance with study drug, and previous exposure to biologic treatment for RA [yes/no] as fixed effects and baseline DAS28-ESR as a covariate. Stratification variables will be used as entered in IRT. The difference between treatments will be estimated by the least squares mean difference between MSB11456 and EU-RoActemra, with its 95% confidence interval for the EMA and its 90% confidence interval for the FDA.

If the proportion of subjects with missing W12 change from baseline DAS28-ESR scores is above 5%, then missing change from baseline in DAS28-ESR scores at W12 will be imputed by a multiple imputation (MI) procedure similar to the one described in 12.5.3.1 for the primary analysis. However, the proportion of missing data at W12 is expected to be low.

12.5.6.2 Supportive Analysis 3.1 (Early DAS28-ESR Per Protocol)

The same analysis as for the main analysis will be performed on the Week 12 PP Analysis Set. All data available will be included in the analysis. No, or very little, missing data is foreseen, and no imputation of missing values is planned.

12.5.6.3 Supportive Analysis 3.2 (Early DAS28-ESR Hypothetical Continuing Per Protocol ITT)

The same analysis described in Section 12.5.4.5 will be performed on the ITT Analysis Set but for Week 12.

12.5.6.4 Sensitivity Analysis – for main Estimand at Week 12 DAS28-ESR

The same analysis for estimand 3.0 will be performed on DAS28-ESR at Week 12 with no imputation.

12.5.7 Subgroup Analyses

Analyses on the primary and key secondary estimands (Estimand 1.0, and Estimand 2.0 as described in Section 9.1) will be performed for the following subgroups using the same methods as the main analyses, removing the strata from the model, when applicable:

- Previous exposure to biologic treatment for RA [yes/no] as captured in IVRS
- ADA positive/ADA negative
  - Neutralizing antibody (NAb) positive/NAb negative
- Non COVID-19 vaccinated/COVID-19 vaccinated prior to Week 24
- Non COVID-19 vaccinated/COVID-19 vaccinated prior to Week 12 for Estimand 3.0 (DAS28 ESR at Week 12).

Inferential statistics will be provided for each subgroup provided there are sufficient subjects to complete the analysis.

ADA or NAb status will be defined as positive for a complete study period if the patient has at least 1 confirmatory positive result post-dose any time during this period. Otherwise the status is defined as negative for this complete study period.

The estimate of the difference with the standard error, 90% and 95% confidence intervals will be presented.

A forest plot will be displayed with the analysis by each of the subgroups as well as overall.

12.5.8 Secondary Analyses

Unless stated otherwise, secondary analyses will be performed on both the ITT and PP Analysis Sets.

12.5.8.1 DAS28-ESR Change from Baseline

The DAS28-ESR mean absolute change from baseline will be summarized descriptively over time and analyzed using a mixed-effect repeated measures model assuming an unstructured covariance matrix with treatment, visit, treatment-by-visit interaction, previous exposure to biologic treatment for RA (yes/no) included as factors and baseline DAS28-ESR from the Core Period as a covariate. For further details, see Section 15.5.1 of the Appendix.

The mixed models will be performed on the Core Period, Extended Period (using the extended baseline data) and overall (using the Core Period baseline data). The Core Period tables will utilize the ITT and PP Analysis Sets. The Extended Period tables will utilize the EP-ITT and EP-PP Analysis Sets. The overall period tables will utilize the ITT and overall PP (combination of PP and EP-PP) analysis sets. The 95% confidence interval of the least-squared mean will be provided for each time point for the Core Period.

A plot of the mean (SD) DAS28-ESR over time will be provided on the Overall Period.

To provide an exploratory overview over the distribution of DAS28-ESR values over time, aligned histograms will be prepared for the Overall Period.

12.5.9 Other Efficacy Analyses

No imputation of missing data will be performed on these other efficacy analyses.

12.5.9.1 DAS28-CRP

The DAS28-CRP mean absolute change from baseline will be summarized descriptively over time and analyzed using a mixed-effect repeated measures model with treatment, visit, treatment-by-visit interaction, and previous exposure to biologic treatment for RA (yes/no) included as factors and baseline DAS28-CRP from the Core Period as a covariate.

The mixed effects models on the Core Period and on the Extended Period, will be carried out as has been previously described for the DAS28-ESR change from baseline endpoint. The 95% confidence interval of the least-squared mean will be provided for each time point for the Core Period.

The Core Period tables will utilize the ITT and PP Analysis Sets. Extended Period tables will utilize the EP-ITT and EP-PP Analysis Sets.

12.5.9.2 ACR20 / ACR50 / ACR70

The ACR20/ACR50/ACR70 response rate over time will be summarized descriptively using the LOCF imputed dataset and compared during the Core Period using 95% Newcombe confidence intervals for the difference stratified on previous exposure to biologic treatment for RA (yes/no).

These summaries will be performed on the Core Period, and the Extended Period comparing to baseline from the Core Period. The Core Period tables will utilize the ITT and PP Analysis Sets. The Extended Period tables will utilize the EP-ITT and EP-PP Analysis Sets.

To provide an exploratory overview over the distribution of ACR values over time, aligned bar charts will be prepared for the Overall Period.

12.5.9.3 DAS28-ESR and DAS28-CRP Categorical Response

The number and percentage of patients within each DAS28-ESR and DAS28-CRP disease activity/categorical response, as described in Section 10.5, will be provided for the Core Period, and Extended Period, by treatment group and visit. Categorical disease activity for DAS28-ESR is categorized as [remission/low/moderate/high], categorical response for DAS28-CRP is categorized as [yes/no] by the Boolean response criterion. The Core Period tables will utilize the ITT and PP Analysis Sets. The Extended Period tables will utilize the EP-ITT and EP-PP Analysis Sets.

To provide an exploratory overview over the distribution of the DAS categorical responses over time, aligned bar charts will be prepared for the Overall Period.

12.5.9.4 CDAI and SDAI

The number and percentage of patients within each CDAI and SDAI level of disease category (remission/low/moderate/high), as described in Section 10.5, will be provided for the Core Period,

and Extended Period, by treatment group and visit. The Core Period tables will utilize the ITT and PP Analysis Sets. The Extended Period tables will utilize the EP-ITT and EP-PP Analysis Sets.

To provide an exploratory overview over the distribution of the CDAI and SDAI over time, aligned bar charts will be prepared for the categorical responses and aligned histograms for the CDAI and SDAI values for the Core and Extended Periods on same graph.

12.6 Immunogenicity and Pharmacokinetic Analyses

Immunogenicity analyses will be performed on the SAF and EP-SAF Analysis Sets; PK analyses (i.e. trough concentration analyses) will be performed on the PK and EP-PK Analysis Sets.

Participation in the population PK sub-study is optional.

12.6.1 Antidrug Antibody Incidence Rate, Antidrug Antibody Titer, NAb Incidence Rate

ADA or NAb status will be defined as positive for a given study period if the patient has at least 1 positive result post-dose in the ADA confirmatory assay any time during this period. Otherwise the status is defined as negative for this study period.

Treatment-induced ADA positive status is defined as:

- In patients with an ADA negative pre dose sample, a treatment-induced ADA response is defined as any post dose sample being positive in the ADA confirmatory assay
- In patients with an ADA positive pre dose sample, a treatment-induced ADA response is defined as a 1.808-fold increase in titers from the pre dose assessment to a post dose assessment.

The ADA and NAb incidence rates over time as well as ADA titer over time will be summarized descriptively by visit (and overall for incidence) and treatment arm. The denominator for each visit will be the number of subjects with a valid immunogenicity assessment at that visit. This descriptive summary will be provided for the Core Period, the Extended and the Overall Period. The geometric mean, median minimum, and maximum values of ADA titer will be presented for overall and by visit.

An additional table “treatment-induced” ADA positive status (without the pre dose time point) will be presented for the Core Period and the Extended Period.

To provide a visual overview of overall ADA and NAb positive rates, aligned bar charts will be prepared on the Overall Period using the Safety Analysis Set. For the purpose of the bar chart a patient is defined as positive when he has at least 1 confirmatory positive result post-dose. ADA incidence over time and NAb incidence over time will also be presented in separate bar charts on the Overall Period.

A descriptive summary (including the lower (T1) and upper tertile (T2)) of the maximum post baseline ADA titer of each subject will presented. Subjects with at least one post baseline confirmed ADA positive results and subjects with all negative ADA titer results will also be

summarized. The maximum post baseline ADA titer for all subjects will be categorized into three groups: Low titer (Maximum ADA Titer < T1), Mid titer (T1≤Maximum ADA Titer ≤T2), High Titer (Maximum ADA Titer>T2)

Confirmed ADA titer for subjects with treatment-related hypersensitivity reactions will be summarized using the safety analysis set. In addition, a boxplot of ADA titer over time on the log 2 scale will be provided for the Core and Overall Periods.

12.6.2 Trough Concentration

Trough concentration will be summarized descriptively on the PK Analysis Set for the Core Period, Extended Period, and Overall Period for each treatment group and timepoint. Core Period table will utilize the PK Analysis Set. Extended Period tables will utilize the EP-PK Analysis Set. Overall Period tables will utilize the Overall PK Analysis Set.

For trough concentration the descriptive summary statistics additionally comprise:

- Coefficient of variation
- Geometric mean
- Geometric coefficient of variation

For the purpose of these analyses, values below the lower limit of quantitation (LLOQ) will be imputed as the ½LLOQ of the assay. Missing PK concentrations (e.g., no sample, insufficient sample volume for analysis, no result or result not valid) will be reported in a listing and classed as "N.R." They will be excluded from summary statistics.

Descriptive statistics for C_{trough} will also be created by ADA status, ADA tertile subgroups, and NAb status (if sufficient data is available for the latter, to be confirmed at the first blinded data review meeting).

To provide a visual overview over the distribution of trough concentration over time, the following figures will be displayed:

- Mean concentration-over time by Overall Period on linear (+/- standard deviation) and semi logarithmic scales
- Mean concentration-over time by Core Period on linear (+/- standard deviation) and semi logarithmic scales
- Individual patient concentration-time data on the Overall Period by treatment group and visit week on linear and semi logarithmic scales
- Mean concentration-over time on linear (+/- standard deviation) by ADA status and ADA tertile subgroups
- Box plot of serum drug concentration at Week 24 and at Week 52 for the ADA negative vs. ADA positive in each treatment group

12.7 Safety Analyses

All Safety Analyses will be provided for the Core Period, the Extended Period and Overall Period on the SAF Set. Core and Overall Period tables will utilize the SAF Analysis Set. Extended Period tables will utilize the EP-SAF Analysis Set.

12.7.1 Adverse Events

All AEs reported in the CRF will be coded using MedDRA.

A summary of Treatment emergent adverse events (TEAEs) will be presented, including:

- The number of events reported
- The number and percentage of
  - Patients reporting at least 1 TEAE
  - Patients reporting at least 1 study drug related TEAE
  - Patients reporting at least 1 serious TEAE
  - Patients reporting at least 1 TEAE ≥ grade 3
  - Patients reporting at least 1 study drug related TEAE ≥ grade 3
  - Patients reporting at least 1 TEAE ≥ grade 4
  - Patients reporting at least 1 study drug related TEAE ≥ grade 4
  - Patients reporting any TEAE of special interest
  - Patients reporting any study drug related TEAE of special interest
  - Patients with withdrawal of study drug due to a TEAE
  - Patients with withdrawal of study drug due to a study drug-related TEAE
  - Patients with interruption of study drug due to a TEAE
  - Patients with interruption of study drug due to a study drug-related TEAE
  - Patients discontinuing the study due to a TEAE
  - Patients discontinuing the study due to a study drug-related TEAE
  - Patients with any TEAE leading to death
  - Patients with any study drug-related TEAE leading to death
  - Patients with at least 1 serious injection site reaction (ISR) or ISR leading to treatment interruption or discontinuation.
  - Non-Serious TEAEs Experienced by ≥ 5% of Patients by Preferred Term
  - TEAEs Experienced by ≥ 2% of Patients by Preferred Term

- TEAE incidence rate with corresponding 95% confidence intervals (calculated as LnRate= log(Rate), and SE of LNRate=1/sqrt(number of events); 95%CI Rate = exp(LnRate +/- 1.96*SE)) per patient year for all TEAE categories above

Incidence rate per patient year is calculated as follows:

$$Days\ on\ Study = \sum_{all\ patients\ in\ SA\ Set} (last\ day - day\ of\ first\ injection + 7)$$

$$Incidence\ rate\ per\ patient\ year = \frac{Incidence * 365{,}25}{Days\ on\ Study}$$

where last day is the 28th day after the day of treatment discontinuation if applicable and the date of data extraction for those patients that have not yet ended the study.

TEAEs are defined as occurring or worsening (in severity or relationship to study drug) with an onset at the time of or following dosing time at Visit 2 (initial injection at Day 1). In reality, all AEs which change in severity or relationship to study drug are assigned a new start date and captured as a new record.

The start of TEAEs (or worsening of an existing AE) that start on a patient's re-randomization day and for which dosing and/or start time of the AE are not available will be assigned to the Extended Period. In case dosing and TEAE start time are both available they should be considered for allocating the TEAE start to a study period.
In case the end day of an AE falls on the re-randomization day and dosing and AE stop time do not clearly prove that the AE has ended before the W24 dosing the AE stop falls in the Extended Period. All AE summaries are presented by SOC and PT unless otherwise specified.

The following displays of TEAE related information will be provided:

- A breakdown of the number and percentage of patients reporting each TEAE, categorized by body system and PT. Note: Counting will be by patient, not by event and patients are only counted once within each body system or PT.
- A further tabulation of these data categorized by relationship to study drug. Patients with multiple events within a particular body system or PT will be counted under the category of their most drug-related event within that body system or PT. Relationship to study drug is categorized as related and not related
- A summary of TEAEs reported, categorized by NCI-CTCAE Toxicity Grades 1 to 5. Patients with multiple events within a particular body system or PT will be counted under the category of their most severe event within that body system or PT
- A summary of study drug-related TEAEs reported, categorized by NCI-CTCAE Toxicity Grades 1 to 5.
- A summary of TEAEs leading to withdrawal of study drug
- A summary of TEAEs leading to discontinuation of the study
- A summary of TEAEs of special interest (TEAEs of special interest will be defined according to the protocol and will be identified via the tick box on the CRF)
- A summary of non-serious TEAEs that exceed a frequency threshold of 5 % within any treatment group. (The frequency threshold is to be calculated as number of patients in a treatment group who had this TEAE at least once divided by number of patients in this treatment group.) In case the 5 % threshold is only exceeded for less than 5 TEAE terms the threshold will be reduced to 2 %
- A further tabulation presenting the PTs for the events in descending order of frequency for the MSB11456 patients (all data from MSB11456 only patients and Extended Period data for the switchers) will also be presented
- TEAEs by SOC and PT and by ADA Status
- TEAEs by SOC and PT and by NAb Status (if sufficient data available)
- MedDRA (version 23.0) SMQ defined Hypersensitivity and Anaphylaxis reactions tabulated by ADA status

- TEAEs of special interest and SAEs presented by ADA and NAb status (if sufficient data available)

All AEs (including non-treatment-emergent events) recorded on the CRF will be listed.

12.7.2 Deaths and Serious Adverse Events

The following displays of SAE and death related information will be provided:

- A breakdown of the number and percentage of patients reporting each SAE, categorized by body system and PT coded according to the MedDRA dictionary. Note: Counting will be by patient, not by event and patients are only counted once within each body system or PT.
- A further tabulation of these data categorized by relationship to study drug. Patients with multiple events within a particular body system or PT will be counted under the category of their most drug-related event within that body system or PT. Relationship to study drug is categorized as related and not related
- Any TEAE leading to death
- Any study drug related TEAE leading to death

Tables will be sorted by decreasing order of frequency in the SOCs and PTs within SOCs in the total across all treatment groups.

A listing of SAEs and a listing of deaths will be provided in the Safety Tables section.

12.7.3 Laboratory Data

Laboratory test results are reported in International System of Units (SI) units. However, due to the Covid-19 situation usage of the central laboratory may be limited. If this is the case, certain safety assessments may be performed at a local laboratory. Local lab results for all parameters except ESR and urinalysis (dipstick) will be included in data listings and shift tables but not in summaries of continuous values.

Summary statistics on observed values and changes from baseline will be provided for continuous laboratory variables for the Core Period, Extended Period, and Overall Period by treatment group and visit. For the Extended Period summaries of change from baseline tables will be provided for change from Core Period baseline. Shift tables for the Core, Extended and Overall Period will be used to present changes in categorical laboratory variables by visit. Core and Overall Period shift tables will use the Core Period baseline, Extended Period shift tables will be provided for Core Period and Extended Period baseline. Categories should be low, normal, high and missing.

Further a summary and shift tables of worst post-baseline value by categories (low, normal, high, missing) will be provided for Core, Extended and Overall Period. For the shift tables, Core Period and Overall Period will use Core Period baseline, Extended Period will use Core Period and Extended Period baseline.

Where applicable, hematology and clinical chemistry laboratory results will be graded programmatically using version 5.0 of the NCI-CTCAE criteria (see Section 15.11 of the appendix). These laboratory NCI-CTCAE toxicity grades will be summarized by worst grade and by shift tables from baseline to worst NCI-CTCAE toxicity grade. These summaries will all use the Core Period baseline for the core, extended and overall period and the worst value. An additional shift table for the extended period which utilizes the extended period baseline line will be presented.

Abnormal values will be displayed in a listing. This listing will additionally present the rest of the respective patient's values for the laboratory parameter in question in order to display the conspicuous values in a chronological context.

Laboratory data will be displayed in a pyramid table for the Core Period, Extended Period and for the Overall Treatment Period (see Section 15.4.4 of the Appendix).

The laboratory data will be displayed in a box and whisker plot for the Core Period and Extended Periods together in one graph.

Completion status and results of QuantiFERON-TB Gold tests will be summarized by visit and treatment group for the Overall Period.
Anti-double-stranded DNA data will be listed only.

12.7.4 Vital Signs

The following vital sign data (observed values and changes from baseline) will be summarized using descriptive statistics over time for the Core Period, Extended Period, and Overall Period:

- Respiratory rate
- Heart rate after a 5-minute rest, and
- Arterial blood pressure after a 5-minute rest

- Temperature (^{0}C)

For vital signs that have been measured twice at the same date the average value at the visit will be calculated and used (All individual values per visit will be listed.). Shift tables of vital signs (systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse rate, temperature, and respiratory rate) will be provided for Core, Extended and Overall Period. Core Period and Overall Period will use Core Period baseline, Extended Period will use both Core Period and Extended Period baseline.

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: ≤0 / >0 - ≤1 / >1 - ≤2 / >2 - ≤3 / >3°C
Respiratory rate	<20 / ≥20 bpm	≤0 / >0 - ≤5 / >5 - ≤10 / >10

The maximum change from baseline will be listed for each vital sign parameter for each patient.

12.7.5 Physical Examinations, Electrocardiograms, and Other Observations Related to Safety

12.7.5.1 Physical Examination

Physical examination results will be listed containing the investigators' assessments of normal/abnormal and clinical significance.

An additional listing will be provided containing only the abnormal results in physical examination.

12.7.5.2 Electrocardiograms

Shift tables of clinically significant 12-lead electrocardiogram abnormalities will be provided for the Core Period, Extended Period, and Overall Period. Categories will be normal, abnormal non clinically significant, abnormal clinically significant and missing.

All ECG data will be listed. An additional listing will be provided containing the ECG details (all assessments of these patients in chronological order) from those patients who had at least 1 clinically significant abnormality.

12.7.5.3 Local Tolerability

Injection site reactions (ISRs) will be summarized for the Core Period and Extended Period by treatment group and in total. Frequency will be provided by term provided on eCRF (including "Other"), then by severity and by relationship to treatment.

The event rate will be analyzed by number of injections. The total number of ISR for a patient will be divided by the number of injections the patient has received during the period. These normalized event rates will be summarized as described above. ISRs will be presented by ADA and NAb status (if enough patients with Nab positive status) to assess the impact of ADA and NAb results on safety.

Action taken with study treatment and outcome will be presented in a listing only.

An additional listing will be provided that only contains those cases of ISRs of CTCAE grade 3 and above.

13.0 References

Burmester, GR et al. A randomized, double-blind, parallel-group study of the safety and efficacy of subcutaneous tocilizumab versus intravenous tocilizumab in combination with traditional disease-modifying anti-rheumatic drugs in patients with moderate to severe RA (SUMMACTA) study. Ann Rheum Dis 2014;73:69-74.

Genovese, MC et al. Interleukin-6 receptor inhibition with tocilizumab reduces disease activity in RA with inadequate response to disease-modifying antirheumatic drugs. Arthritis Rheum 2008;58:2968-80.

Smolen, JS et al. Effect of interleukin -6 receptor inhibition with tocilizumab in patients with RA (OPTION study): a double-blind, placebo controlled, randomized trial. Lancet 2008;371:987-97.

Kremer, JM et al. Tocilizumab inhibits structural joint damage in RA patients with inadequate responses to methotrexate. Arthritis Rheum 2011;63:609-21.

Emery, P et al. IL-6 receptor inhibition with tocilizumab improves treatment outcomes in patients with RA refractory to anti-tumor necrosis factor biologicals: results from a 24-week multicenter randomized placebo controlled trial. Ann Rheum Dis 2008;67:1516-23.

Cro S et al. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Statistics in Medicine. 2020;39: 2815–2842.

Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Statistics in Medicine. 2014 October 30; 33(24): 4170–4185.

Siddiqui O. MMRM versus MI in dealing with missing data--a comparison based on 25 NDA data sets. J Biopharm Stat. 2011 May;21(3):423-36. doi: 10.1080/10543401003777995. PMID: 21442517.

14.0 Glossary of Abbreviations

Glossary of Abbreviations:	
ACR	American College of Rheumatology
ADA	Antidrug Antibody
AE	Adverse Event
ATC	Anatomic Therapeutic Classification
BOCF	Baseline Observation Carried Forward
CDAI	Clinical Disease Activity Index
CRF	Case Report Form
CSR	Clinical Study Report
CTMS	Clinical Trials Management System
DBP	Diastolic Blood Pressure
DAS28-CRP	Disease Activity Score 28-C-reactive Protein
DAS28-ESR	Disease Activity Score 28-Erythrocyte Sedimentation Rate
ECG	Electrocardiogram
EMA	European Medicines Agency
EP	Extended Period
EULAR	European League Against Rheumatism
FDA	Food and Drug Administration
ISR	Injection Site Reaction
ITT	Intent-to-Treat
K-M	Kaplan-Meier
LLOQ	Lower Limit of Quantitation
LOCF	Last Observation Carried Forward
MCMC	Markov Chain Monte Carlo
MI	Multiple Imputation
NAb	Neutralizing Antibody
PK	Pharmacokinetic
PP	Per Protocol
PT	Preferred Term
RA	Rheumatoid Arthritis
SAF	Safety
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan

SBP	Systolic Blood Pressure
SDAI	Simple Disease Activity Index
SI	International System of Units
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event