Janssen Research & Development

Statistical Analysis Plan

A Phase 2a Multicenter, Randomized, Double Blind, Parallel, Proof of concept Study Evaluating the Efficacy and Safety of Nipocalimab and Certolizumab Combination Therapy in Participants with Active Rheumatoid Arthritis despite Prior Treatment with Advanced Therapies (bDMARD or tsDMARD)

Protocol 80202135ARA2002; Phase 2a

AMENDMENT 1

JNJ-80202135 (Nipocalimab)

Status:	Approved
Date:	15 July 2024
Prepared by:	Janssen Research & Development, LLC; Janssen Research & Development, a division of Janssen Pharmaceutica NV
Document No.:	EDMS-RIM-1197500,2.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

VERSION HISTORY

This statistical analysis plan for Study **80202135ARA2002** is based on the protocol amendment-3 dated 05 October 2023.

SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	06-December 2023	Not Applicable	Initial release
2.0	15 July 2024	Multiple imputation model was simplified to impute each individual component separately	MI was imputing all the 7 ACR components simultaneously
		To add an additional pre-planned subgroup analysis (baseline MTX use (Yes/No) to efficacy analysis, Section 5.6.3.3	Clinical request
		For secondary and tertiary endpoints, remaining missing data after ICE rules are applied will be imputed as non-responder imputation.	To be consistent with the previous IRIS study
		To include data handling rules for Pharmacodynamic analysis “A concentration not quantifiable (below LLOQ) will be replaced by half of the value of the lower limit of quantification (LLOQ) for numerical calculations in PD analysis, i.e. if Anti-CPP LLOQ <25, then value will be replaced by 12.5 U/ml (25*0.5)”	To plan any possible values that might be less than LLOQ
		To remove MMRM and CMH model descriptions.	Multiple imputation will be used to impute missing continuous endpoints. MMRM is not required. CMH analysis will not be performed, instead logistic regression will be used for all the binary endpoints.
		Throughout the SAP, a non-content related wording or formatting changes were made	To enhance clarity
		Safety Section 5.7.2 added the word “Potential” to Hypersensitivity reactions	To align with PSAP

LIST OF ABBREVIATIONS

ADA	anti-drug antibody
AE	adverse event
ALT/SGPT	alanine aminotransferase
ANCOVA	analysis of covariance
AST/SGOT	aspartate aminotransferase
ATC	anatomic and therapeutic class
AUC	area under the curve
BMI	body mass index
BSA	body surface area
CI	confidence interval
CL	total systemic clearance
Cmax	maximum concentration
CRF	case report form
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DMC	Data Monitoring Committee
DPS	data presentation specifications
ECG	electrocardiogram
eCRF	electronic case report form
F (%)	absolute SC bioavailability
FAS	full analysis set
FDA	Food and Drug Administration
ICH	International Conference on Harmonisation
IQ	interquartile
IVRS	interactive voice response system
IWRS	interactive web response system
LLOQ	lower limit of quantification
MedDRA	Medical Dictionary for Regulatory Activities
MRD	minimum required dilution
NAb	neutralizing antibodies
PD	pharmacodynamic(s)
PI	principal investigator
PK	pharmacokinetic(s)
PP	per protocol
QTc	QT corrected
QTL	quality tolerance limit
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SMQs	standardised MedDRA queries
TEAE	treatment-emergent adverse event
Tmax	time to maximum concentration
US NCI	United States National Cancer Institute
V	volume distribution
Vz	volume of distribution based on terminal phase
Vz/F	apparent volume of distribution based on terminal phase after extravascular administration
WHO	World Health Organization
WHO-DD	World Health Organization Drug Dictionary

1. INTRODUCTION

The Statistical Analysis Plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for all planned analyses of efficacy, safety, tolerability, pharmacokinetics (PK) and immunogenicity of combination therapy (Nipocalimab and Certolizumab) in the 80202135ARA2002 study.

Efficacy analysis for primary and secondary endpoints will be performed after all participants reach Week 12 (or have terminated study participation before the Week 12 visit). The database will be locked for primary and secondary efficacy endpoints at Week 12, and the final database lock for all other planned analyses will be thorough Week 30.

1.1. Objectives, Endpoints and Estimands

1.1.1. Objectives and Endpoints

Objectives	Endpoints
Primary	
• To evaluate the efficacy of combination therapy with nipocalimab and certolizumab compared to certolizumab monotherapy in participants with moderately to severely active RA despite ≥1 advanced therapy (bDMARDs or tsDMARDs).	• Change from baseline in DAS28-CRP at Week 12
Secondary	
• To evaluate the efficacy of combination therapy with nipocalimab and certolizumab compared to certolizumab monotherapy in participants with moderately to severely active RA	• ACR20, ACR50, ACR70, and ACR90 responses at Week 12 • DAS28-CRP remission at Week 12 • DAS28-CRP LDA at Week 12 • Change from baseline in HAQ-DI score at Week 12 • Change from baseline in CDAI at Week 12
Tertiary OR Exploratory	
• To evaluate the safety and tolerability of combination therapy with nipocalimab and certolizumab	• Treatment-emergent AE through Week 30 • Treatment-emergent SAEs through Week 30 • Treatment-emergent AEs leading to discontinuation of study intervention through Week 30 • Treatment-emergent AESIs through Week 30 • Laboratory parameters and change from baseline in laboratory parameters through Week 30 • Vital sign parameters and change from baseline in vital sign parameters through Week 30
• To evaluate the PK and immunogenicity of combination therapy with nipocalimab and certolizumab	• Serum nipocalimab and certolizumab concentrations through Week 30 in participants receiving active study intervention • Incidence and titers of antibodies to nipocalimab (ADA and NAbs) through Week 30

Objectives	Endpoints
• To evaluate the efficacy of combination therapy with nipocalimab and certolizumab in participants with baseline ACPA high (#1), baseline csDMARD use (#2) and baseline MTX use #3	• Change from baseline in DAS28-CRP through Week 24 • ACR20, ACR50, ACR70, and ACR90 responses through Week 24 • DAS 28-CRP remission through Week 24 • DAS28-CRP LDA through Week 24 • Change from baseline in HAQ-DI score through Week 24 • Change from baseline in CDAI through Week 24 • CDAI LDA through Week 24 • CDAI remission through Week 24
• To evaluate the efficacy of combination therapy with nipocalimab and certolizumab	• ACR20, ACR50, ACR70, and ACR90 responses through Week 24 • Percent improvement from baseline in ACR components through Week 24 • Change from baseline in ACR components through Week 24 • DAS28-CRP (and DAS28-ESR) LDA through Week 24 • DAS28-CRP (and DAS28-ESR) remission through Week 24 • Change from baseline in DAS28-CRP (and DAS28-ESR) through Week 24 • CDAI LDA through Week 24 • Change from baseline in CDAI through Week 24 • CDAI remission through Week 24 • Change from baseline in SDAI through Week 24 • SDAI LDA through Week 24 • SDAI-based ACR/EULAR remission through Week 24 • Boolean-based ACR/EULAR remission through Week 24
• To evaluate the efficacy on PROs of combination therapy with nipocalimab and certolizumab	• Decrease of ≥0.22 points from baseline in HAQ-DI score through Week 24 • Change from baseline in Joint Pain severity score through Week 24
• To evaluate the impact of combination therapy with nipocalimab and certolizumab on PD and disease biomarkers	• Change from baseline in serum Immunoglobulin profile (IgG/M/E/A) and IgG subtype levels through Week 30 • Change from baseline in biomarker levels through Week 30

Efficacy endpoint definitions are described in Section 5.4.1, Section 5.5.1 and Section 5.6.1.

A description of the primary estimand as a precise description of the intervention effect reflecting the clinical question posed by the trial objective is presented in Section 5.4.2. The secondary estimands are described in Section 5.5.3 and the exploratory estimands are described in Section 5.6.3

1.2. Study Design

1.2.1. Overall Design

This is a Phase 2a clinical development study for the combination therapy (nipocalimab and certolizumab) in Rheumatoid Arthritis will evaluate the safety and efficacy of the combination therapy compared to certolizumab monotherapy.

This study is a randomized, double blind, parallel, multicentre, PoC interventional study to evaluate the efficacy and safety of a combination therapy with nipocalimab and certolizumab in participants between 18 and 75 years old, with moderately to severely active RA despite ≥1 advanced therapy (bDMARDs or tsDMARDs) but are certolizumab naïve.

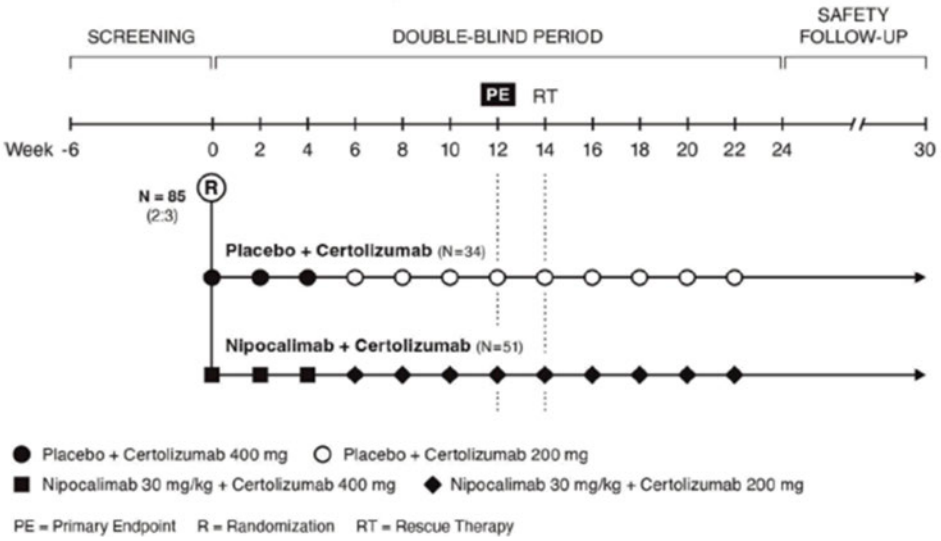
The total duration of the study is up to 36 weeks and consisting of 3 study periods: a 6-week screening period (rescreening is permitted once per participant), a 24-week double-blind period (22 weeks of treatment), and a 6-week safety follow-up period (8 weeks after the last study intervention administration). A diagram of study design is provided in Figure 1.

Participants will be screened for study eligibility within 6 weeks after informed consent form signing and interactive web response system (IWRS) registration. An IV administered placebo will be used in the certolizumab monotherapy group to allow for double blinded evaluation of efficacy and safety.

This study will randomize approximately 85 participants with 34 participants in the certolizumab monotherapy and 51 participants in the combination therapy nipocalimab and certolizumab. Eligible participants will then be randomized in a 2:3 ratio to one of the following:

- Combination therapy: Nipocalimab 30 mg/kg IV q2w and Certolizumab 400 mg at Week 0, 2, 4; then 200 mg Q2W
- Certolizumab Monotherapy: Certolizumab 400 mg at Week 0, 2, 4; then 200mg Q2W and Placebo IV q2w

Figure 1: Schematic Overview of the Study Through End of Study



Throughout the study, stable doses of concomitant NSAIDs, oral corticosteroids, and selected conventional synthetic DMARDs are allowed but limited to MTX, SSZ, HCQ and LEF, see Table 1. Participants should not initiate any new treatment for RA through Week 30.

Table 1: Permitted Concomitant Medications for 80202135ARA2002

Permitted Concomitant Medications for RA	**Maximum Allowable Dosage**
csDMARDs	MTX: ≤25 mg/week SSZ: 1000-2000 mg/day HCQ or CQ: 200-400 mg/day Leflunomide: 10-20 mg/day
Oral corticosteroids	Equivalent to average of ≤10 mg/day of prednisone
NSAIDs and other analgesics	No more than the usual marketed dosages approved in the country where the study is being conducted
CQ=chloroquine; csDMARD=conventional synthetic disease-modifying anti-rheumatic drugs; HCQ=hydroxychloroquine; MTX=methotrexate; NSAIDs=nonsteroidal anti-inflammatory drugs; RA=rheumatoid arthritis; SSZ=sulfasalazine	

Efficacy data will be collected through Week 24 and safety data will be collected through Week 30. All assessments will be performed according to the schedule of activities (SoA) detailed in Section 1.3 in the protocol. Participants who discontinue study agent administration prior to Week 22 will continue to be evaluated for efficacy and safety through Week 30 as per protocol.

Every reasonable effort should be made to keep concomitant medications stable as defined in the protocol. Beginning at the screening visit, all concomitant therapies and all changes in concomitant therapies should be recorded throughout the study.

Key safety assessments include adverse events (AEs), serious adverse events (SAEs), AEs of special interest (AESIs), clinical laboratory tests (including blood chemistry, hematology, urinalysis, and lipid panel), physical examinations and vital signs.

An independent, external Data Monitoring Committee (DMC) will be commissioned for this study. The DMC will monitor the safety of participants during the study. Refer to Section 9.5 of the study protocol for details. An early safety review will be conducted by the independent DMC for the first 15 participants randomized to any treatment arm and treated, see Section 5.10 for details.

Database locks (DBLs) are scheduled at Weeks 12, and end of study (Week 30). The first DBL will occur when all randomized participants have either completed the Week 12 assessments or terminated study participation prior to the Week 12 visit (referred to as Week-12 DBL). The final DBL will occur when all randomized participants have either completed their final safety visit or have terminated study participation (referred to as Week-30 DBL).

Additional DBLs may be added as necessary and will be specified in the SAP prior to the additional DBLs.

The primary endpoint is the change from baseline in DAS28-CRP at Week 12, primary analysis will be performed at Week 12 DBL.

1.2.2. Randomization and Blinding

Randomization at Week 0 will be used to minimize bias in the assignment of participants to intervention groups and to increase the likelihood that known and unknown participant attributes (eg, demographic and baseline characteristics) are evenly balanced across intervention groups. The maintenance of blinded treatment will be used to reduce potential bias during data collection and evaluation of clinical endpoints.

At Week 12 DBL, the data will be unblinded for analysis to some Sponsor personnel while participants are still participating in the study. Identification of Sponsor personnel who will have access to the unblinded participant-level data will be documented prior to unblinding. Investigative study sites and participants will remain blinded to initial treatment assignment until after the final database is locked.

1.2.2.1. Randomization

A dynamic central randomization based on biased-coin algorithm will be implemented in this study using an interactive web response system (IWRS). When a participant is eligible for randomization at a study site, the randomization requestor at that study site will contact the IWRS using the requester's own user identification and personal identification number and provide the relevant participant details to uniquely identify that participant. The site personnel will enter participant's baseline covariates listed below in IWRS, then the covariate adaptive randomization algorithm will run in the background to assign a unique treatment code, which will dictate the treatment assignment and matching study agent kit for that participant. Full details of the randomization algorithm are specified in Covariate Adaptive Randomization User Requirements Specification (CARUSR) in vTMF.

Dynamic central randomization targets to balance the distribution of participant to achieve the randomization ratio (2:3) at the study level and within the levels of each individual stratification factors:

1. csDMARDs usage at baseline (yes or no)
2. Screening ACPA level (high ≥400 U/mL; low <400 U/mL)
3. Baseline DAS28-CRP (high ≥5.8; low <5.8) using CRP at screening
4. Study country
5. Investigator site

1.2.2.2. Blinding

The investigator will not be provided with randomization codes. The codes will be maintained within the IWRS, which has the functionality to allow the investigator to break the blind for an individual participant.

Data that may potentially unblind the intervention assignment (ie, study intervention serum concentrations, anti-study intervention antibodies, study intervention preparation/accountability data, intervention allocation, biomarker or other specific laboratory data) will be handled with special care to ensure that the integrity of the blind is maintained and the potential for bias is minimized. This can include making special provisions, such as segregating the data in question from view by the investigators, clinical team, or others as appropriate until the time of database lock and unblinding.

Under normal circumstances, the blind should not be broken until after all participants have completed the study and the database is finalized (Week 30). Otherwise, the blind should be broken only if specific emergency intervention/course of action would be dictated by knowing the intervention status of the participant. In such cases, the investigator may in an emergency determine the identity of the intervention by contacting the IVRS/IWRS. It is recommended that the investigator contact the sponsor or its designee, if possible, to discuss the particular situation, before breaking the blind.

Telephone contact with the sponsor or its designee will be available 24 hours per day, 7 days per week. In the event the blind is broken, the sponsor must be informed as soon as possible. The date, time, and reason for the unblinding must be documented by the IVRS/IWRS, in the appropriate section of the eCRF, and in the source document. The documentation received from the IVRS/IWRS indicating the code break must be retained with the participant's source documents in a secure manner.

Participants who have had their intervention assignment unblinded should continue to return for scheduled evaluations.

In rare circumstances when a potential safety issue that may impact the overall benefit-risk assessment of the investigational product has been identified in this study, selected Sponsor personnel may be unblinded to safety-related data in order to investigate the safety issue and determine if additional actions are required. The safety data should be kept blinded to any personnel not essential to the safety review. If other rare, unforeseen circumstances arise that may necessitate unblinding of selected sponsor personnel, these will be assessed and documented on a case-by-case basis. The data should be kept blinded to any personnel not essential to the review or investigation. In general, randomization codes will be disclosed fully only if the study is completed, and the clinical database is closed.

2. STATISTICAL HYPOTHESIS

The primary hypothesis is that treatment with nipocalimab in combination with certolizumab is superior to certolizumab monotherapy in participants with moderately to severely active RA despite treatment with ≥1 advanced therapy (bDMARDs or tsDMARDs) as assessed by the mean change from baseline in DAS28-CRP at Week 12.

3. SAMPLE SIZE DETERMINATION

The sample size calculation is based upon the primary endpoint, change from baseline in DAS28-CRP at Week 12 to provide sufficient power to detect treatment differences in the primary endpoint, between the certolizumab monotherapy and combination therapy.

For the sample size determination, the following assumptions were made:

- For Certolizumab monotherapy: a mean change from baseline in DAS28-CRP at Week 12 of -1.75 was assumed, based on mean change from baseline in DAS28-ESR through Week 24 as reported in the REALISTIC study (Weinblatt 2012).
- For Nipocalimab and certolizumab combination therapy: a mean change from baseline in DAS28-CRP of -2.45 was assumed for the ACPA positive population. It is expected that participants in the combination therapy will achieve a high remission rate. A delta of -2.45 in change from baseline in DAS28-CRP could translate to approximately 45% remission rate based on the results of the golimumab combined treatment group (100 mg golimumab + MTX) at Week 52 in the GO-FORWARD study (Keystone 2010). From the above, the mean difference in change from baseline in DAS28-CRP between the combination therapy and certolizumab monotherapy for the ACPA positive population is assumed to be -0.70 (-2.45 minus -1.75).
- In study 80202135ARA2001, the ACPA negative patients achieved low efficacy (-0.4), but the number of ACPA negative patients was small (10%). With the expectation that a similar 10% of ACPA negative patients will be included in study 80202135ARA2002, the assumed overall treatment effect in change from baseline in DAS28-CRP is set to be -0.67 [(-0.7*90%) + (-0.4*10%)], to ensure sufficient power for the overall population.

A sample size of 85 participants (34 participants in certolizumab monotherapy group; and 51 participants in combination therapy group) will provide a power of approximately 80% to detect a difference of 0.67 in the change from baseline in DAS28-CRP at Week 12 with a pooled standard deviation of 1.2 between the groups, and alpha of 0.05 (1-sided).

Table 2 provides the power evaluation using various assumptions for a total sample size of 85 participants. The bolded assumptions are considered the base case.

Table 2: Statistical Power for Treatment Difference in Change from Baseline in DAS28-CRP at Week 12 at alpha = 0.05 (1-sided)

Sample Size	Delta	SD	Power (%)
85 (34:51)	0.6	1.1	79
	0.6	1.2	72
	0.6	1.3	66
	0.67	1.1	86
	0.67	**1.2**	**80**
	0.67	1.3	75
	0.7	1.1	89
	0.7	1.2	83
	0.7	1.3	79
	0.8	1.1	95
	0.8	1.2	91
	0.8	1.3	87

Note: The T-test was used for intervention comparison

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

Table 3 defines the analysis sets for purposes of analysis.

Table 3: Description of Analysis Sets used to Analyze the Data in the Study

Analysis Sets	Description
Enrolled	All participants who sign the ICF.
Randomized	All participants who were randomized in the study
Full Analysis Set (FAS)	The Full Analysis Set (FAS) includes all participants who were randomized in the study and received at least one (partial or complete) administration of study intervention.
Safety Analysis Set (SAS)	The Safety Analysis Set includes all participants who were randomized in the study and received at least one (partial or complete) administration of study intervention.
Immunogenicity Analysis Set (IAS)	The immunogenicity analysis set is defined as all participants who received at least one dose (partial or complete) administration of nipocalimab and have appropriate serum samples for ADA detection.
Pharmacokinetics (PK) Analysis Set	The PK analysis set is defined as all participants randomized who received at least one complete administration of nipocalimab and/or certolizumab and have at least one valid post-dose blood sample collection.
Pharmacodynamic (PD) Analysis Set	The PD analysis set is defined as participants who received at least 1 dose (partial or complete) of study intervention and have at least 1 valid post-dose blood sample drawn for PD analysis.

5. STATISTICAL ANALYSES

5.1. General Considerations

For efficacy analyses, the FAS will be analyzed according to the participants' assigned treatment to which they were randomized, regardless of the actual study treatment received.

For efficacy analyses within a subset (i.e. baseline ACPA high, baseline csDMARD use and baseline MTX use), the analysis set will consist of the individual subset of the FAS. The subset of the FAS will be analysed according to the participants' assigned treatment to which they were randomized, regardless of the actual study treatment received.

For safety analyses, the safety analysis set will be analysed based on the treatment they actually received, regardless of the intervention groups to which they were assigned.

Unless otherwise indicated, continuous variables will be summarized with the following descriptive statistics: n (number of observations), mean, standard deviation (SD), interquartile (IQ) range, minimum, median, and maximum. Categorical data will be summarized with frequencies and percentages. Percentages will be based on the number of participants included in the analysis set under consideration unless otherwise specified. Graphical data displays may also be used to summarize the data.

Demographic and baseline variables will be summarized by intervention group, and overall, for the FAS, as described in Appendix 2. Participants with major protocol deviations will be identified prior to database lock and will be summarized by category as described in Appendix 3. Summaries of concomitant medications will be presented by anatomic and therapeutic class (ATC) term, and intervention group, as described in Appendix 5.

For continuous endpoints, Analysis of Covariance (ANCOVA) will be used. The ANCOVA model will include treatment (combination therapy or monotherapy), baseline endpoint value and randomization stratification factors [baseline csDMARDs use (yes or no), screening ACPA level (high ≥400 U/mL; low <400 U/mL), country and continuous baseline DAS28-CRP, instead of using baseline DAS28-CRP categories (high ≥5.8; low <5.8), where appropriate]. The Least Squares (LS) mean estimates and their corresponding 90% confidence intervals (CI), and the estimates of LS mean differences between the combination therapy and monotherapy groups and their corresponding 90% CI will be provided. In addition, LS mean estimates, LS mean difference and their corresponding 95% CI will be reported for interpretability of results in the context of other studies.

For binary endpoints, treatment comparisons will be performed using a logistic regression model adjusting for randomization stratification factors [baseline csDMARDs use (yes or no), screening ACPA level (high ≥400 U/mL; low <400 U/mL), country and continuous baseline DAS28-CRP, where appropriate]. The odds ratio and associated 90% CI Wald, asymptotic method confidence intervals and statistical significance of the difference in treatment effect will be reported. The number and proportion of participants who achieve binary composite response at Week 12 will be summarized for combination therapy or monotherapy groups with associated 90% and 95% CI. In

addition, the difference in proportions and its associated 90% and 95% CI along with 90% and 95% CI associated with odds ratio will be reported.

Note, randomization stratification factor might be excluded from subgroup analysis, for example if the subgroup analysis is performed by baseline csDMARDs, then randomization stratification baseline csDMARDs use (yes or no) will not be included in ANCOVA or logistic regression model.

Specific details for Primary and Key Secondary endpoints are provided below.

Statistical comparison between combination therapy and monotherapy will be performed at primary endpoint and key secondary endpoints at Week 12. In addition, the primary, secondary, and exploratory/tertiary endpoints will be performed by visit through Week 24. No treatment comparison will be performed after Week 24.

The primary hypothesis will be tested for statistical significance at the α=0.05 (1-sided) level. For all endpoints and comparisons no multiplicity control is planned, and nominal p-values will be reported.

5.2. Strategy for Intercurrent Events

5.2.1. Composite Strategy

The Composite Strategy assesses the treatment effects not only based on the variable measurements, but also based on intercurrent events defined for an estimand. Any participants who have intercurrent events that are handled with the Composite Strategy will be a non-responder for response type variables (binary endpoints) and will have a score of no improvement/no change for continuous variables (continuous endpoints) for timepoints after the occurrence of the intercurrent event. This strategy acknowledges that these intercurrent events are unfavourable.

5.2.2. Hypothetical Strategy

The Hypothetical Strategy assesses the treatment effect in a hypothetical scenario in which the intercurrent event would not occur, i.e, a hypothetical scenario is considered in which participants would take treatment as directed. For those participants who have intercurrent events that are handled with the Hypothetical Strategy, the value that the variable of interest would have taken in the hypothetical scenario defined for that estimand, will be used for analysis.

5.2.3. Treatment Policy Strategy

The Treatment Policy strategy is to use all observed data collected for the endpoint. The occurrence of the intercurrent event that is handled with the Treatment Policy strategy is irrelevant: the observed value for the variable of interest is used regardless of whether the intercurrent event occurs.

5.2.4. Data Handling Rules

5.2.4.1. Missing Data

Missing data will be imputed using multiple imputation (MI) for continuous endpoints, see details in Appendix 14. Data from participants who experience an intercurrent event, i.e, outcome values collected after ICE will not be used in the imputation model. Under the assumption of missing at random (MAR), MI will be used to impute the missing data for the continuous/ordinal measurements at components level. The missing data will be imputed using the predicted value from an imputation model using the Full Conditional Specification (FCS) regression method for any missing pattern. Each variable will be restricted to only imputed within its possible range of values (e.g., HAQ-DI Score may only be imputed to a value within the range of 0-3). The explanatory variables in the imputation model include imputation variables and ancillary variables with all data from baseline through Week 12 or through Week 24. Refer to Appendix 14. for more details regarding the imputation models, the number of imputations (N) and the starting seeds.

For the composite continuous endpoints, the above imputation will be performed at each component with missing data and then the composite score will be derived based on the imputed components.

For supplementary estimand that are applied to the continuous endpoints, all observed data is used regardless of the intercurrent event, thus, MI will be implemented using observed data and component with missing data that is due to any reason will be imputed.

For the composite and supplementary binary endpoints (i.e. ACR 20), missing response status that is not due to an intercurrent event, will conservatively be imputed as non-responders at a scheduled visit through Week 12 or through Week 24.

5.2.5. Reference Date, Study Day and Relative Day

The reference date is the date of the first study intervention administration. If the date of the first study intervention administration is missing or the first study intervention is not done, then the reference date equals the corresponding visit date (eg, Week 0 visit date). If the corresponding visit date is also missing, then the reference date equals the randomization date.
Study day 1 or day 1 refers to the reference date (there is no study day 0). All efficacy and safety assessments on all visits will be assigned a day relative to this date.

5.2.6. Visit Windows

Unless otherwise specified, nominal visits will be used for all by visit analyses. The study visits scheduled after randomization are expected to occur at the time delineated in the Schedule of Activities (SoA).

All post-baseline visits from Baseline through Week 24 will have a visit window of ± 3 days. The final safety follow-up at Week 30 will have a visit window of ± 1 week (7 days).

For PK analyses, if a participant has an administration outside the visit window at a visit, the concentration data collected at and after that visit will be excluded from the by-visit data analyses.

5.2.7. Pooling Algorithm for Analysis Centers

Unless otherwise specified, data from all investigational centers/sites will be pooled for analyses.

5.2.8. Rescue Therapy Criteria

Rescue therapy criteria is applied during the double-blind portion of the study, starting from Week 14, see Appendix 7 for details. Participants who achieve a CDAI low disease activity >10 score at Week 14 are eligible to adjust their current baseline SoC treatment starting from Week 14 (investigator's choice of SoC treatment appropriate for the participant).

Participants who meet 1 or more of the following criteria (due to CDAI LDA >10) at Week 14 will be defined as treatment failures (TF) from the time the TF occurs (Week 14) onward through end of double-blind period (Week 24):

- If a participant initiates an immunosuppressants or immunomodulators therapies for RA, listed in Appendix 6, then the participant will be defined as TF from Week 14 onwards through end of double-blind period.
- If participant exceed the baseline dose or initiate a new csDMARDs medication between Week 14 and Week 24 visits. A "new" csDMARDs medication is one that was not present at randomization then the participant will be defined as TF from Week 14 onwards through end of double-blind period.

Note: Participants who meet the rescue criteria i.e CDAI LDA >10 and do **not** change their current baseline SoC treatment will not be defined as TF. Substitution of an agent with a similar mechanism of action at an equivalent or lower dose will not be considered a TF.

5.3. Participant Disposition

The number of participants screened, the number of participants screen failed, primary reason for screen failure and the number of participants re-screened will be presented.

The number of participants in the following disposition categories will be summarized by intervention group and overall.

- Participants randomized
- Participants who received study intervention
- Participants who completed the study
- Participants who discontinued study intervention
- Reasons for discontinuation of study intervention
- Participants who terminated the study prematurely
- Reasons for terminating the study prematurely

The above categories will include summaries through Week 30 (including any final safety visit that occurs prior to the Week 12 DBL).

A listing of participants will be provided for the following categories:

- Participants who discontinued study intervention
- Participants who terminated study prematurely
- Participants who were unblinded during the study period
- Participants who were randomized yet did not receive study intervention
- Participants who experienced intercurrent events
- Participants who met rescue therapy (RT) criteria at Week 14, defined in Appendix 7

In addition, time from randomization/Week 0 to study termination/discontinuation of study intervention will be displayed with Kaplan-Meier curves, as described in Appendix 1.

5.4. Primary Endpoint

5.4.1. Definition of Primary Endpoint

The primary endpoint is change from baseline in DAS28-CRP at Week 12.

The Disease Activity Index Score 28 using CRP [DAS28-CRP] is a validated and derived score combining tender joints (28 joints) (TJC28), swollen joints (28 joints) (SJC28), CRP, and Patient's Global Assessment of Disease Activity (PtGA) (van Riel 2000).

The set of 28 joint count is based on evaluation of the shoulder, elbow, wrist, MCP1, MCP2, MCP3, MCP4, MCP5, IP1, PIP2, PIP3, PIP4, PIP5 joints of both the upper right extremity and the upper left extremity as well as the knee joints of lower right and lower left extremities.

The DAS28-CRP is a continuous parameter and is defined as follows:

$$\text{DAS28-CRP} = [0.56*\text{SQRT}(\text{TJC28}) + 0.28*\text{SQRT}(\text{SJC28}) + 0.36*\text{Ln}(\text{CRP}_{\text{mg/L}}+1) + 0.014*\text{PtGA}_{\text{mm}} + 0.96]$$

1. TJC28: a total number of tender joints among the 28 joints evaluated for tenderness. Each of the 28 joints will be evaluated for tenderness, categorized as tender or not tender. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of tender joints.
2. SJC28: a total number of swollen joints among the 28 joints evaluated for swelling. Each of the 28 joints will be evaluated for swelling, categorized as swollen or not swollen. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of swollen joints.

3. $CRP_{mg/L}$: C-reactive protein (CRP) in mg/L. In the calculation of DAS28-CRPvalue, the natural logarithm of ($CRP_{mg/L}$ + 1) is used. LLOQ rule specified in Section 6.4.3 will be applied to values <LLOQ. Note, Labcorp CRP values are used.
4. $PtGA_{mm}$: Patient's Global Assessment of Disease Activity scaled from 0 (very well) to 100 (very poor) on a 100-unit VAS scale for the calculation of DAS28-CRPvalue.

If the DAS28-CRP score cannot be calculated (i.e., 3 components available) at a visit, the DAS28-CRP score will be considered missing.

Change from baseline in DAS28-CRP measures the change in disease activity, where a negative change indicates an improvement, and a positive change indicates a worsening.

5.4.2. Estimands

5.4.2.1. Primary Estimand

Primary Trial Objective: To evaluate the superiority of combination therapy (Nipocalimab 30 mg/kg IV q2w and Certolizumab 400 mg at Week 0, 2, 4; then 200 mg Q2W) versus monotherapy (Certolizumab 400 mg at Week 0, 2, 4; then 200 mg Q2W and Placebo IV q2w) in patients with moderately to severely active RA despite ≥1 advanced therapy (bDMARD or tsDMARD).

Estimand Scientific Question of Interest:

For a patient with moderately to severely active RA despite taking at least one or more advanced therapy with bDMARD or tsDMARD, what is the expected effect at Week 12 on change from baseline in DAS28-CRP of being assigned to a Nipocalimab + Certolizumab combination therapy vs Certolizumab monotherapy in addition to SoC, where patients with treatment discontinuation, a change of standard-of-care (SoC) dosing or initiation of an immunosuppressants or immunomodulators therapies for RA are considered to experience no effect?

The primary estimand, aligned with the above primary objective and clinical question of interest, provide precise descriptions of the treatment effect of interest that are defined by the following 5 attributes:

- **Treatment condition of interest vs Alternative treatment condition:**
  - Combination therapy: (Nipocalimab 30 mg/kg IV q2w and Certolizumab 400 mg at Week 0, 2, 4; then 200 mg Q2W) up to Week 12 in addition to SoC (see Table 1)

    vs

  - Monotherapy: (Certolizumab 400 mg at Week 0, 2, 4; then 200 mg Q2W and Placebo IV q2w) up to Week 12 in addition to SoC
- **Population:** Adult patients with moderately to severely active RA despite ≥1 advanced therapy (bDMARD or tsDMARD).
- **Variable:** Change from baseline in DAS28-CRP at Week 12, where patients with ICEs 1-3 are considered to have a zero change.

- **Population-level summary:** Differences in means between treatment conditions
- **Intercurrent events (ICEs) and their corresponding strategies:**

ICEs	Strategy for Addressing ICEs
5. Initiated an immunosuppressants or immunomodulators therapies for RA (see Appendix 6)	**Composite Strategy:** A patient experiencing ICE 1-3 is considered to have no treatment effect, as reflected in the Variable definition.
1. Initiated or increased the dose of csDMARDs (such as MTX, SSZ, HCQ, CQ or leflunomide) or oral corticosteroid therapy above the baseline dose for RA	
2. Discontinued study intervention due to any reason **except** ICE4	
3. Discontinued study intervention due to Natural Disaster or Major Disruption	A hypothetical scenario is considered in which patients would continue treatment as assigned (rather than discontinuing treatment due to Natural Disaster or Major Disruption)

5.4.2.2. Supplementary Estimand

The supplementary estimand has the same components as the primary estimand for primary endpoint, except for the strategies used for ICEs and not considering no treatment effect (zero change) in the variable definition.

Variable: Change from baseline in DAS28-CRP at Week 12

Treatment Policy Strategy: Strategy targeting the effect of treatment assignment, regardless of the occurrence of ICEs.

5.4.3. Analysis Methods

5.4.3.1. Primary Analysis (Main Estimator)

Analyses of the primary efficacy endpoint will include data from all participants in the FAS (Section 4), which includes all randomized participants who received at least 1 dose (partial or complete) of study intervention. Participants will be analysed according to the intervention group to which they were randomized regardless of the study intervention they actually received.

For participants with ICE 4, observed data after this ICE will not be used in the analysis, i.e. values of outcome measure collected after ICE 4 will not be used in the analysis nor in the imputation model. For participants with ICE 1-3, observed data after ICE will not be used in the imputation model, i.e. values collected after ICE1-3 are not contributing to the imputed components.

Missing data due to missed visits or missed data collections not related to study withdrawal, or data not used will be assumed as MAR and imputed using MI by FCS on individual components for continuous endpoints After the missing data in the components are imputed, then the ICE rule will be applied and the primary endpoint, change from baseline in DAS28-CRP, will be determined. Steps for performing multiple imputation are provided in Appendix 14.

The primary analysis will be aligned with the Primary Estimand defined in Section 5.4.2.1. The Treatment comparisons will be performed using an Analysis of Covariance (ANCOVA) model on each of the imputation datasets. The treatment difference between certolizumab monotherapy and the combination therapy (nipocalimab and certolizumab) will be tested for each imputation dataset and then the results across all imputation datasets will be combined according to Rubin's Rules (Rubin 1987). The least squares means (LS means) and its associated 90% CI, and treatment difference and the associated 90% CI for the differences in LS means and the p-value for the treatment difference will be obtained based on the ANCOVA model. The ANCOVA model will be based on the original scale and will include intervention groups, continuous baseline DAS28 (CRP), and randomization stratification factor levels [baseline csDMARDs use (yes or no), screening ACPA level (high ≥400 U/mL; low <400 U/mL) and country] as explanatory variables. In addition, LS means, and LS means difference, and associated 95% CI will be reported.

5.4.3.2. Sensitivity Analysis

A sensitivity analysis will be performed using a re-randomization test to evaluate the impact of the dynamic randomization in the case that primary endpoint analysis result shows superiority in combination therapy over monotherapy targeting the primary estimand. Steps for performing the re-randomization test, (EMA, 2015), (Han, et al., 2013) and (Ge, et al., 2011).

1. Obtain the test statistics $\boldsymbol{T}^*$ for the superiority of combination therapy against monotherapy in primary endpoint analysis (Section 5.4.3.1)
2. Estimate the distribution of the test statistics with repetitions by following steps:
    a. Re-randomize the treatment assignment with the dynamic randomization by maintaining the same biased coin probability, but rather with a new seed to generate new serious of random number for treatment assignment. Based on the new treatment assignment, obtain the new test statistics $\boldsymbol{T}$
    b. Repeat the re-randomize step for 5000 times
3. Derive the p-value associated with the observed test statistics $\boldsymbol{T}^*$ based on the empirical distribution in step 2. Here the p-value can be calculated as $(n + 1)/(5000 + 1)$, where the n is the number of times that $\boldsymbol{T} > \boldsymbol{T}^*$ among the 5,000 repetitions
    a. If this p-value is greater than 0.05 (i.e., study pre-specified one-sided type-I error), then it indicates the superiority observed in original test (Section 5.4.3.1) might have been due to chance, and therefore no superiority result should be claimed.
    b. If this p-value is equal to or less than 0.05 (i.e., study pre-specified one-sided type-I error), then we claim the re-randomization test supports the main study results, i.e., the superiority of combination therapy over monotherapy is demonstrated.

5.4.3.3. Subgroup Analysis

To evaluate the consistency of the primary endpoint (change from baseline in DAS28-CRP at Week 12), across subgroups defined by baseline demographics and disease characteristics, and prior and baseline medication specified in Section 5.8.5 will be performed. Subgroup analyses may be performed when the number of participants in the subgroup level permits (at least ≥5 participants for each of the intervention group within a subgroup). If the numbers of participants in a stratum is too small (eg. <5), then strata may be pooled if appropriate, i.e. if the subgroup has more than two levels. Note that, for subgroup analyses, the analysis sets are the individual subgroups of the FAS.

For each subgroup analysis, the main estimator for the primary estimand will be used and same analysis described in Section 5.4.3.1 will be followed.

The treatment difference in change from baseline in DAS28-CRP at Week 12 by each of subgroup factors listed in Section 5.8.5, will be analysed using an ANCOVA model. The ANCOVA model includes intervention groups, baseline DAS28-CRP score and stratification factors (if appropriate) as explanatory variables. A forest plot of the LS means and 90% CI, LS mean of the difference between treatment groups in change from baseline in DAS28-CRP at Week 12 and the associated 90% CI estimated by the ANCOVA model will be plotted for each of the subgroup factors defined Section 5.8.5.

In addition, the treatment*subgroup interaction p-value will be obtained separately for each subgroup that has at least two categories.

Moreover, a forest plot will be presented for LS means, and LS means difference and the associated 95% CI for each of the subgroup factors defined Section 5.8.5.

5.4.3.4. Supplementary Analysis (Treatment Policy Estimand)

The primary endpoint will also be analysed utilizing the supplementary estimand, treatment policy estimand. For participants who experience an ICE through Week 12, the analysis will be performed using observed data regardless of intercurrent events, i.e. values collected after ICE will be included in the imputation model and analysis model. Missing data will be imputed using MI by FCS, steps for performing multiple imputation are provided in Appendix 14. The same ANCOVA model as the primary analysis described in Section 5.4.3.1 will be performed.

5.5. Secondary Endpoint(s) Analysis

5.5.1. Key Secondary Endpoint(s)

There are eight secondary endpoints:

- ACR20 responses at Week 12
- ACR50 responses at Week 12
- ACR70 responses at Week 12
- ACR90 responses at Week 12
- DAS28-CRP remission at Week 12
- DAS28-CRP LDA at Week 12
- Change from baseline in HAQ-DI score at Week 12
- Change from baseline in CDAI at Week 12

5.5.2. Definition of Endpoint(s)

5.5.2.1. American College of Rheumatology Response

ACR response is a composite measurement of change in RA signs and symptoms and is presented as the numerical measurement of improvement in multiple disease assessment criteria. An ACR response (Anderson 2011, Felson 1995) is defined as:

1. ≥20% improvement from baseline in both swollen joint count (66 joints) and tender joint count (68 joints)

AND

2. ≥20% improvement from baseline in at least 3 of the following 5 assessments:
   - Patient's Global Assessment of Disease Activity (VAS)
   - Patient's assessment of pain (VAS)
   - Patient's assessment of physical function as measured by HAQ-DI
   - Physician's Global Assessment of Disease Activity (VAS)
   - CRP

The following are the definitions of each of the forgoing disease assessment criteria (components) that are used in the determination of ACR 20 response.

3. Tender Joint Count 68 (TJC68): a total number of tender joints among the 68 joints evaluated for tenderness. Each of the 68 joints will be evaluated for tenderness, categorized as tender or not tender joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 68-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of tender joints.

4. Swollen Joint Count 66 (SJC66): a total number of swollen joints among the 66 joints evaluated for swelling. (Note: The 2 hip joints are excluded from swelling assessment.) Each of the 66 joints will be evaluated for swelling, categorized as swollen or not swollen. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 66-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of tender joints.
5. Patient's Assessment of Pain (PAIN): a measure from 0 (no pain) to 10 cm (the worst possible pain) on a VAS scale
6. Patient's Global Assessment of Disease Activity (PtGA): a measure from 0 (very well) to 10 cm (very poor) on a VAS scale
7. Physician's Global Assessment of Disease Activity (PGA): a measure from 0 (no arthritis activity) to 10 cm (extremely active arthritis) on a VAS scale
8. HAQ-DI: a measure of difficulty a participant may have in accomplishing tasks in 8 functional areas. For additional details, please refer to the definition of HAQ-DI, see Section 5.5.2.3.
9. C-reactive protein (CRP): a lab parameter measured in mg/dL. LLOQ rule specified in Section 6.4.3 will be applied to values <LLOQ. Note, Labcorp CRP values are used.

If a participant's baseline value for a component is zero (i.e., no disease activity as measured by that component), the participant should be considered as not achieving 20% improvement from baseline for that component since there is no room for improvement.

ACR50, ACR70, and ACR90 are similarly defined except improvement threshold from baseline is 50%, 70%, and 90%, respectively.

5.5.2.2. DAS28 Remission and Low Disease Activity

Refer to Section 5.4.1, for the definition based on DAS28-CRP.

- DAS28 remission is defined as a DAS28-CRP value of <2.6 at a visit
- DAS28 LDA is defined as a DAS28-CRP value of ≤ 3.2 at a visit

5.5.2.3. Change from Baseline in HAQ-DI Score

HAQ disability index (HAQ-DI) score is an evaluation of the functional status for a participant (Fries 1980). The 20-question instrument assesses the degree of difficulty a person has in accomplishing tasks in 8 functional areas (dressing, arising, eating, walking, hygiene, reaching, gripping, and activities of daily living). Responses in each functional area are scored from 0, indicating no difficulty, to 3, indicating inability to perform a task in that area (i.e., lower scores are indicative of better functioning).

The HAQ-DI score is the sum of computed category scores divided by the number of categories answered. The HAQ-DI score will not be computed if the participant does not have scores for at least 6 of the 8 categories. The scoring algorithm (including adjusting the use of aids or devices) is provided by http://patienteducation.stanford.edu/research/haq20.html.

Change from baseline in HAQ-DI score is a measure of the change in the functional status, where a negative change reflects an improvement, and a positive change reflects a worsening.

Number of HAQ-DI responders (Decrease of ≥0.22 points from baseline in HAQ-DI), defined as participants who have a change of < -0.22 from baseline in HAQ-DI score.

5.5.2.4. Clinical Disease Activity Index Score

The Clinical Disease Activity Index (CDAI) score is a derived score combining 4 disease assessments: tender joint counts (28 joints), swollen joint counts (28 joints), Patient's Global Assessment of Disease Activity (PtGA), and Physician's Global Assessment of Disease Activity (PGA) (Aletaha 2005). Note that CDAI is modified SDAI by excluding CRP component. The 28 joints evaluated for swelling and tenderness are the same set of the 28 joints as used in DAS28 and include shoulder, elbow, wrist, MCP1, MCP2, MCP3, MCP4, MCP5, IP1, PIP2, PIP3, PIP4, PIP5 joints of the upper right and upper left extremities and knee joints of the lower right and lower left extremities.

The CDAI is a continuous parameter and is defined as followings:

$$CDAI = TJC28 + SJC28 + PtGA + PGA$$

1. TJC28: a total number of tender joints among the 28 joints evaluated for tenderness. Each of the 28 joints will be evaluated for tenderness, categorized as tender or not tender. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of tender joints.
2. SJC28: a total number of swollen joints among the 28 joints evaluated for swelling. Each of the 28 joints will be evaluated for swelling, categorized as swollen or not swollen. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of swollen joints.
3. PtGA: Patient's Global Assessment of Disease Activity (VAS). In the calculation of CDAI score, PtGA is scaled from 0 (very well) to 10 cm (very poor) on a VAS scale.
4. PGA: Physician's Global Assessment of Disease Activity (VAS). In the calculation of CDAI score, PGA is scaled from 0 (no arthritis activity) to 10 cm (extremely active arthritis) on a VAS scale.

If any of the components required for computing the CDAI score is missing, the CDAI score will be set to missing.

Change from baseline in CDAI score measures the change in disease activity, where a negative change indicates an improvement, and a positive change indicates a worsening.

CDAI Low Disease Activity

- CDAI LDA is defined as a CDAI score of ≤10 at a visit.

CDAI Remission

- CDAI remission is defined as a CDAI score of ≤2.8 at a visit.

5.5.3. Estimands

5.5.3.1. Main Estimand for Key Secondary Endpoints

The main estimands for key secondary endpoints at Week 12 have same attributes as the primary estimand, except for the attributes of variable and population level summary defined in Table 4.

Table 4: Variables and Population-level Summary

	Variable (Endpoint)	Population-level summary
Secondary Endpoint #1-4	Achieving an ACR (ACR20, ACR50, ACR70 & ACR90) response at Week 12, where patients are considered to have achieved an ACR response if they fulfil the ACR responder criteria (defined in Section 5.5.2.1) at Week 12 and not experiencing ICEs in categories 1-3	Difference in proportions between treatment conditions
Secondary Endpoint #5	Achieving clinical remission at Week 12 (yes/no), where patients are considered to have achieved clinical remission if they fulfil the clinical remission criteria based on DAS28-CRP remission (defined in Section 5.5.2.2) at Week 12, and do not experience intercurrent events in categories 1-3	
Secondary Endpoint #6	Achieving low disease activity (LDA) at Week 12 (yes/no), where patients are considered to have achieved low disease if they fulfil the clinical low disease activity criteria based on DAS28-CRP LDA (defined Section 5.5.2.2) at Week 12, and do not experience intercurrent events in categories 1-3	
Secondary Endpoint #7	Change from baseline in HAQ-DI at Week 12, where patients with ICEs in categories 1-3 are considered to have a zero change	Differences in means between treatment conditions
Secondary Endpoint #8	Change from baseline in CDAI at Week 12, where patients with ICEs in categories 1-3 are considered to have a zero change	

5.5.3.2. Supplementary Estimand for Key Secondary Endpoints

The supplementary estimand has the same components as the main estimands for key secondary endpoints, except for the strategies used for ICEs and not counting ICEs as non-responders/no treatment effect in the variable definition. Treatment Policy strategy will be used to address all ICEs in the supplementary estimand for each key secondary endpoint. For binary secondary missing endpoint, non-responder imputation will be applied. While for continuous secondary endpoints, missing data will be handled by MI.

5.5.4. Analysis Methods

5.5.4.1. Analytical Approaches for Main Estimands

All secondary endpoints will be analysed at Week 12 DBL based on the FAS (Section 4), that is all randomized participants who have received at least one administration of study intervention will be included in analysis according to the intervention they were randomized into, regardless of the intervention participant received.

Binary Endpoint Analyses

The binary response efficacy endpoints will be analysed based on the main estimands defined in Table 4. After accounting for intercurrent events defined in Section 5.4.2.1, any remaining missing response at Week 12 will be considered as non-responders.

The treatment comparison between combination therapy and monotherapy will be tested using logistic regression on binary endpoint at Week 12. The logistic regression model will include intervention group, randomization stratification factors [continuous baseline DAS28 (CRP), baseline csDMARDs use (yes or no), screening ACPA level (high ≥400 U/mL; low <400 U/mL) and country].

The odds ratio and associated 90% Wald, asymptotic method confidence intervals and statistical significance of the difference in treatment effect will be reported at Week 12.

In addition, the number and proportion of participants who achieve a binary response at Week 12 will be summarized for each intervention group and the difference in proportion of response and associated 90% CI at Week 12 will be provided for each binary endpoint. Note that this analysis will not adjust any randomization stratification factors.Same parameter results with 95% CI will also be reported.

Continuous Endpoint Analyses

For the continuous secondary endpoints, change from baseline in HAQ-DI and change from baseline in CDAI score will be analysed based on the main estimands defined in Table 4. The same analytical method for the primary endpoint described in Section 5.4.3.1 will be used to analyze continuous secondary endpoints.

5.5.4.2. Supplementary Analysis (Treatment Policy Estimand)

Similar to the main estimand for key secondary endpoints, the methods specified in Section 5.5.4.1 will be applied to the key secondary binary and continuous endpoints using Treatment Policy Estimand (Section 5.5.3.2). Any missing data for binary secondary endpoints will be imputed as non-responder, while for continuous it will be handled by MI.

5.6. Tertiary/Exploratory Endpoint(s) Analysis

In addition to the primary and secondary endpoints, exploratory endpoints related to disease status and Patient-reported outcomes (PROs) through Week 24 will be performed. This section lists these endpoints, followed by their definitions and analysis methods. These endpoints will be summarized

and compared between combination therapy and monotherapy groups. For safety analysis, inflammatory biomarkers, PK, immunogenicity, and PD will be analysed through Week 30 and covered in Section 5.7, Section 5.8, respectively. Additionally, graphical data displays may also be used to summarize the data over time, if applicable.Analyses of the following endpoints are applicable to visits from Week 0 through Week 24

- ACR20, ACR50, ACR70, and ACR90 responses through Week 24
- DAS28-CRP remission through Week 24
- DAS28-CRP LDA through Week 24
- DAS28 (ESR) remission through Week 24
- DAS28 (ESR) LDA through Week 24
- CDAI remission through Week 24
- CDAI LDA through Week 24
- SDAI-based ACR/EULAR remission through Week 24
- Boolean-based ACR/EULAR remission through Week 24
- SDAI LDA through Week 24
- Decrease of ≥0.22 points from baseline in HAQ-DI score through Week 24
- ACPA seropositive through Week 24
- RF seropositive through Week 24
- Percent improvement from baseline in ACR components through Week 24
- Change from baseline in ACR components through Week 24
- Change from baseline in DAS28-CRP through Week 24
- Change from baseline in DAS28 (ESR) through Week 24
- Change from baseline in CDAI through Week 24
- Change from baseline in SDAI through Week 24
- Change from baseline in HAQ-DI through Week 24
- Change from baseline in Joint Pain Severity score through Week 24
- Change from baseline in ESR through Week 24

5.6.1. Definition of Exploratory Endpoints

Definitions for DAS28-CRPand clinical responses based on DAS28 (CRP), ACR responses, HAQ-DI and CDAI are provided in Section 5.4.1, Section 5.5.2.1, Section 5.5.2.3, and Section 5.5.2.4, respectively.

5.6.1.1. Simplified Disease Activity Index Score

The Simplified Disease Activity Index (SDAI) for RA score is a derived score combining 5 disease assessments: tender joint counts (28 joints), swollen joint counts (28 joints), Patient's Global Assessment of Disease Activity (PtGA), Physician's Global Assessments of Disease Activity (PGA), and CRP (Aletaha 2006). The 28 joints evaluated for swelling and tenderness are the same set of the 28 joints as used in DAS28 and include shoulder, elbow, wrist, MCP1, MCP2, MCP3, MCP4, MCP5, IP1, PIP2, PIP3, PIP4, PIP5 joints of the upper right and upper left extremities and knee joints of the lower right and lower left extremities.

The SDAI is a continuous parameter and is defined as follows:

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

1. TJC28: a total number of tender joints among the 28 joints evaluated for tenderness. Each of the 28 joints will be evaluated for tenderness, categorized as tender or not tender. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of tender joints.
2. SJC28: a total number of swollen joints among the 28 joints evaluated for swelling. Each of the 28 joints will be evaluated for swelling, categorized as swollen or not swollen. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of swollen joints.
3. PtGA: Patient's Global Assessment of Disease Activity (VAS). In the calculation of SDAI score, PtGA is scaled from 0 (very well) to 10 cm (very poor) on a VAS scale.
4. PGA: Physician's Global Assessment of Disease Activity (VAS). In the calculation of SDAI score, PGA is scaled from 0 (no arthritis activity) to 10 cm (extremely active arthritis) on a VAS scale.
5. C-reactive protein (CRP): a lab parameter measured in (mg/dL). In the calculation of SDAI score, the CRP value will be in mg/dL. LLOQ rule specified in Section 6.4.3 will be applied to values <LLOQ. Note, Labcorp CRP values are used.

If any of the components required for computing the SDAI score is missing, the SDAI score will be set to missing.

Change from baseline in SDAI score measures the change in disease activity, where a negative change indicates an improvement, and a positive change indicates a worsening.

SDAI Low Disease Activity

SDAI LDA is defined as a SDAI score of ≤11 at a visit.

5.6.1.2. ACR/EULAR Remission

5.6.1.2.1. SDAI-based ACR/EULAR Remission

For the definition of SDAI, refer to Section 5.6.1.1.

A participant is considered to have achieved SDAI-based ACR/EULAR remission at a visit if SDAI score is ≤ 3.3 (Felson 2011).

5.6.1.2.2. Boolean-Based ACR/EULAR Remission

A participant is considered as having achieved the Boolean-based ACR/EULAR remission at a visit if he/she meets all of the following 4 criteria at that visit (Felson 2011):

1. Tender joint count (68 joints) ≤1
2. Swollen joint count (66 joints) ≤1
3. CRP ≤1 mg/dL
4. Patient's Global Assessment of Disease Activity (PtGA) ≤1 on a 0 (very well) to 10 cm (very poor) VAS scale

The definition of these 4 disease assessment criteria (components) can be found in the definition of ACR response in Section 5.4.2.1.

If any of the above 4 disease assessment criteria (components) required for remission determination is missing, the Boolean-based ACR/EULAR remission will be set to missing. Note, Labcorp CRP values are used.

5.6.1.3. Pain Assessment

This is a patient reported outcome to provide a self-assessment of pain by marking on a Visual Analog Scale (VAS) that ranges from 0-10 units. A value of 0 corresponds to "no pain" and a value of 10 corresponds to "the worst possible pain" on the VAS scale (Felson 1993; Hawley 1992).

5.6.1.4. Joint Pain Severity NRS

Participant' joint pain will be assessed using a single item that asks the participant to report the worst severity of their joint pain over the past 7 days on a 0 to 10 NRS. Responses range from 'No joint pain' (0) to 'Severe joint pain' (10).

5.6.1.5. DAS28 Using Erythrocyte Sedimentation Rate

The Disease Activity Index Score 28 using erythrocyte sedimentation rate (DAS28 (ESR)) is a statistically derived index combining tender joints (28 joints) (TJC28), swollen joints (28 joints) (SJC28), erythrocyte sedimentation rate (ESR), and Patient's Global Assessment of Disease Activity (PtGA) (Prevoo 1995).

It is a continuous parameter and is defined as follows:

DAS28 (ESR) = 0.56*SQRT(TJC28) + 0.28*SQRT(SJC28) + 0.70*Ln (ESR) + 0.014 × $PtGA_{mm}$

1. TJC28: a total number of tender joints among the 28 joints evaluated for tenderness. Each of the 28 joints will be evaluated for tenderness, categorized as tender or not tender. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of tender joints.
2. SJC28: a total number of swollen joints among the 28 joints evaluated for swelling. Each of the 28 joints will be evaluated for swelling, categorized as swollen or not swollen. Joint evaluability rules specified in Section 6.4.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For participants with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.4.2, will be applied in determining the ultimate count of swollen joints.
3. $PtGA_{mm}$: Patient's Global Assessment of Disease Activity (PtGA) scaled from 0 (very well) to 100 (very poor) on a 100-unit VAS for the calculation of DAS28 (ESR) value.
4. ESR: a lab parameter measured in mm/hr. In the calculation of DAS28 (ESR) value, the ESR value is in mm/hr and natural logarithm of ESR is used.

5.6.1.6. DAS28 Remission and Low Disease Activity

Refer to Section 5.6.1.5, for the definition based on DAS28 (ESR) score.

- DAS28 remission is defined as a DAS28-ESR value of <2.6 at a visit.
- DAS28 LDA is defined as a DAS28-ESR value of ≤3.2 at a visit.

5.6.1.7. ACPA and RF seropositive

- Anti-CCP seropositive is defined as Anti-CCP value of ≥17 U/ml at a visit
- RF seropositive is defined as a RF value of ≥15 U/ml at a visit.

Note, Labcorp Anti-CCP values are used.

5.6.2. Estimands

5.6.2.1. Main Estimand for Tertiary/Exploratory Endpoints

The estimand for the tertiary/exploratory endpoints through Week 24 have the same attributes as the primary estimand in treatment, population and ICEs. The attributes for variable, and population level summary for the tertiary/exploratory endpoints are listed in Table 5.

Table 5: Variables and Population-level Summary for Each Tertiary/Exploratory Endpoint

	Variable (Endpoint)	Population-level summary
Tertiary Endpoint #1-4	Achieving an ACR (ACR20, ACR50, ACR70 & ACR90) response through Week 24 (yes/no), where patients are considered to have achieved an ACR response if they fulfil the ACR responder criteria (defined in Section 5.5.2.1) at each scheduled visit and not experiencing ICEs in categories 1-3	
Tertiary Endpoint #5	Achieving clinical remission through Week 24 (yes/no), where patients are considered to have achieved clinical remission if they fulfil the clinical remission criteria based on DAS28-CRP remission (defined in Section 5.5.2.2) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	

Table 5: Variables and Population-level Summary for Each Tertiary/Exploratory Endpoint

	Variable (Endpoint)	Population-level summary
Tertiary Endpoint #6	Achieving clinical low disease activity through Week 24 (yes/no), where patients are considered to have achieved clinical low disease activity if they fulfil the clinical low disease activity criteria based on DAS28-CRP LDA (defined in Section 5.5.2.2) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	Difference in proportions between treatment conditions
Tertiary Endpoint #7	Achieving clinical remission through Week 24 (yes/no), where patients are considered to have achieved clinical remission if they fulfil the clinical remission criteria based on DAS28-ESR remission (defined in Section 5.6.1.6) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	
Tertiary Endpoint #8	Achieving clinical low disease activity through Week 24 (yes/no), where patients are considered to have achieved clinical low disease activity if they fulfil the clinical low disease activity criteria based on DAS28-ESR LDA (defined in Section 5.6.1.6) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	
Tertiary Endpoint #9	Achieving clinical remission through Week 24 (yes/no), where patients are considered to have achieved clinical remission if they fulfil the clinical remission criteria based on CDAI remission (defined in Section 5.5.2.4) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	
Tertiary Endpoint #10	Achieving clinical low disease activity through Week 24 (yes/no), where patients are considered to have achieved if they fulfil the clinical low disease activity clinical low disease activity criteria based on CDAI LDA (defined in Section 5.5.2.4) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	
Tertiary Endpoint #11	Achieving clinical remission through Week 24 (yes/no), where patients are considered to have achieved clinical remission if they fulfil the clinical remission criteria based on SDAI-based ACR/EULAR remission (defined in Section 5.6.1.2.1) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	
Tertiary Endpoint #12	Achieving clinical remission through Week 24 (yes/no), where patients are considered to have achieved clinical remission if they fulfil the clinical remission criteria based on Boolean-based ACR/EULAR remission (defined in Section 5.6.1.2.2) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	
Tertiary Endpoint #13	Achieving clinical low disease activity through Week 24 (yes/no), where patients are considered to have achieved if they fulfil the clinical low disease activity clinical low disease activity criteria based on SDAI LDA (defined in Section 5.6.1.1) at each scheduled visit, and do not experience intercurrent events in categories 1-3.	
Tertiary Endpoint #14	Achieving a decrease of ≥0.22 points from baseline in HAQ-DI response through Week 24 (yes/no), where patients are considered to have achieved a decrease of ≥0.22 from baseline in HAQ-DI if they fulfil the decrease of ≥0.22 from baseline in HAQ-DI criteria (defined in Section 5.5.2.3) at each scheduled visit and not experiencing ICEs in categories 1-3	
Tertiary Endpoint #15	Achieving an ACPA seropositive response through Week 24 (yes/no), where patients are considered to have achieved ACPA response if they fulfil the ACPA seropositive response criteria (defined in Section 5.6.1.7) at each scheduled visit and not experiencing ICEs in categories 1-3	
Tertiary Endpoint #16	Achieving an RF seropositive response through Week 24 (yes/no), where patients are considered to have achieved RF response if they fulfil the RF seropositive response criteria (defined in Section 5.6.1.7) at each scheduled visit and not experiencing ICEs in categories 1-3	
Tertiary Endpoint #17	Percent improvement from baseline in ACR components at all the study scheduled visit (through Week 24), where patients with ICEs 1-3 are considered to have a zero change	Differences in means between treatment conditions
Tertiary Endpoint #18	Change from baseline in ACR component score at all the study scheduled visit (through Week 24), where patients with ICEs1-3 are considered to have a zero change	
Tertiary Endpoint #19	Change from baseline in DAS28-CRP at all the study scheduled visit (through Week 24), where patients with ICEs 1-3 are considered to have a zero change	
Tertiary Endpoint #20	Change from baseline in DAS28-ESR at all the study scheduled visit (through Week 24), where patients with ICEs 1-3 are considered to have a zero change	
Tertiary Endpoint #21	Change from baseline in CDAI score at all the study scheduled visit (through Week 24), where patients with ICEs 1-3 are considered to have a zero change	
Tertiary Endpoint #22	Change from baseline in SDAI score at all the study scheduled visit (through Week 24), where patients with ICEs1-3 are considered to have a zero change	

Table 5: Variables and Population-level Summary for Each Tertiary/Exploratory Endpoint

	Variable (Endpoint)	Population-level summary
Tertiary Endpoint #23	Change from baseline in HAQ-DI score at all the study scheduled visit (through Week 24), where patients with ICEs1-3 are considered to have a zero change	
Tertiary Endpoint #24	Change from baseline in Joint Pain Severity score at all the study scheduled visit (through Week 24), where patients with ICEs 1-3 are considered to have a zero change	
Tertiary Endpoint #25	Change from baseline in ESR score at all the study scheduled visit (through Week 24), where patients with ICEs 1-3 are considered to have a zero change	

5.6.2.2. Supplementary Estimand for Tertiary/Exploratory Endpoints

The supplementary estimand has the same components as the main estimands for tertiary/exploratory endpoints, except for the strategies used for ICEs and not counting ICEs as non-responders/no treatment effect in the variable definition. Treatment Policy strategy will be used to address all ICEs in the supplementary estimand for each key tertiary/exploratory endpoint.

5.6.3. Analysis Methods

5.6.3.1. Analytical Approaches for Main Estimands

All exploratory endpoints including subgroup analysis of baseline csDMARDs use, baseline MTX use and baseline ACPA high subpopulation analysis, will be analysed through Week 24 based on the FAS (Section 4), that is all randomized participants who have received at least one administration of study intervention will be included in analysis according to the intervention they were randomized into, regardless of the intervention participant received.

Binary Endpoint Analyses

The binary exploratory response efficacy endpoints through Week 24 will be analysed based on the main estimand defined in Table 5. After accounting for intercurrent events defined in Section 5.4.2.1, any remaining missing response will be considered as non-responders at a visit through Week 24.

The treatment comparison between combination therapy and monotherapy will be tested using logistic regression on binary endpoint through Week 24. The logistic regression model will include intervention group, randomization stratification factors [continuous baseline DAS28 (CRP), and baseline csDMARDs use (yes or no), screening ACPA level (high ≥400 U/mL; low <400 U/mL) and country] as explanatory variables.

The odds ratio and associated 90% Wald, asymptotic method confidence intervals and statistical significance of the difference in treatment effect will be reported by the scheduled visits, through Week 24.

In addition, the number and proportion of participants who achieve a binary response through Week 24 will be summarized for each intervention group and the difference in proportion of response and associated 90% CI will also be provided for each binary endpoint.

Additionally, the odds ratio, proportion and difference in proportions of participants who achieve a binary response and associated 95% CI will be reported.

Continuous Endpoint Analyses

For the continuous exploratory endpoints through Week 24 will be analysed based on the main estimand defined in Table 5.

Missing data will be assumed as MAR and imputed using MI by FCS. After the missing data in the components are imputed, and then each continuous endpoint will be determined and then ICE rule applied Steps for performing multiple imputation are provided in Appendix 14.

The Treatment comparisons will be performed using an ANCOVA model on each of the imputation datasets. The treatment difference between certolizumab monotherapy and the combination therapy will be tested for each imputation dataset and then the analysis results across all imputation datasets will be combined according to Rubin's Rules (Rubin 1987), and the treatment difference in the least squares means (LS means) and the 90% CI for the differences in LS means and the p-value for the treatment difference will be obtained based on the ANCOVA model, by the scheduled visits, through Week 24. The ANCOVA model will be based on the original scale and will include intervention group, baseline score for the endpoint, and randomization stratification factors as the explanatory variables. In addition, LS means, and LS mean difference, and associated 95% CI will also be reported.

5.6.3.2. Supplementary Analysis (Treatment Policy Estimand)

Similar to the main estimand for tertiary/exploratory endpoints, the methods specified in Section 5.6.3.1 will be applied to the tertiary/exploratory binary and continuous endpoints using Treatment Policy Estimand (Section 5.6.2.2).

5.6.3.3. Subgroup

5.6.3.3.1. Baseline csDMARDs use

Overtime subgroup analyses comparing baseline csDMARDs use (yes, no) for binary endpoints [ACR (20,50,70, and 90) responses, DAS28-CRP remission, and DAS28-CRP LDA, CDAI LDA and CDAI remission]; and continuous endpoints (change from baseline in DAS28-CRP, change from baseline in HAQ-DI and change from baseline in CDAI score) through Week 24 between the combination therapy and monotherapy will be performed.

Analyses for binary and continuous endpoints will be based on the main estimand and supplementary estimand described in Section 5.6.2. The same analytical approach as the exploratory endpoints analyses described above Section 5.6.3 will be performed.

Note, in this subgroup analysis, randomization stratification factor- baseline csDMARD use (Yes/No) will be excluded from the ANCOVA and logistic regression models. If the number of participants in a stratum level is too small (eg. <5) then analysis for that stratum will not be performed.

5.6.3.3.2. Baseline ACPA High

Overtime analyses of baseline ACPA high (defined as baseline ACPA value ≥400 U/mL) subpopulation for binary endpoints [ACR (20,50,70, and 90) responses, DAS28-CRP remission, and DAS28-CRP LDA, CDAI LDA and CDAI remission]; and continuous endpoints (change from baseline in DAS28-CRP, change from baseline in HAQ-DI and change from baseline in CDAI score) through Week 24 will be performed. Note, Labcorp baseline ACPA values will be used to perform this analysis.

Analyses for binary and continuous endpoints will be based on the main estimand and supplementary estimand described in Section 5.6.2. The same analytical approach as the exploratory endpoints analyses described above Section 5.6.3 will be performed.

Note, in this subgroup analysis, randomization stratification factor - Screening ACPA level (high ≥400 U/mL; low <400 U/mL) will be excluded from the ANCOVA and logistic regression models If the number of participants in a stratum level is too small (eg. <5) then analysis for that stratum will not be performed.

Sensitivity analysis of baseline ACPA high (defined as baseline ACPA value ≥ median) will also be performed. Note, PBL baseline ACPA values will be used to perform the sensitivity analysis.

5.6.3.3.3. Baseline MTX use

Overtime analyses of baseline MTX use (yes, no) for binary endpoints [ACR (20,50,70, and 90) responses, DAS28-CRP remission, and DAS28-CRP LDA, CDAI LDA and CDAI remission]; and continuous endpoints (change from baseline in DAS28-CRP, change from baseline in HAQ-DI and change from baseline in CDAI score) through Week 24 will be performed.

Analyses for binary and continuous endpoints will be based on the main estimand and supplementary estimand described in Section 5.6.2. The same analytical approach as the exploratory endpoints analyses described above Section 5.6.3 will be performed.

Note, in this subgroup analysis, randomization stratification factor- baseline csDMARD use (Yes/No) will be excluded from the ANCOVA and logistic regression model. If the number of participants in a stratum level is too small (eg. <5) then analysis for that stratum will not be performed.

5.6.3.4. Sensitivity Analysis (ESR data)

To evaluate the robustness of the ESR data from participants using the old kit only, new kit only, and a combination of the new and old kits, the following sensitivity analysis might be performed.

The main estimand and supplementary estimand approaches and corresponding analysis methods specified for tertiary/exploratory continuous endpoints in Section 5.6.3.1 will be applied to the ESR data from participants using the old kit only, new kit only, and a combination of the new and old kits.

5.7. Safety Analyses

All safety analyses will be based on the safety analysis set based on actual intervention received, unless otherwise specified.

For all continuous safety variables, descriptive statistics by intervention group will include N, mean, standard deviation (SD), median, and range (minimum, maximum). Categorical variables will be summarized by the intervention group using frequency counts and percentages.

5.7.1. Extent of Exposure

The number and percentage of participants who receive study intervention (combination therapy, monotherapy) will be summarized.

Descriptive statistics for duration of a study intervention and duration of follow-up in weeks (N, mean, SD, median, and range (minimum, maximum) will be summarized. Participant-years of study intervention duration are calculated as days of intervention/365.25. Participant-years will be presented by the intervention group.

Study intervention duration is defined as (date of last dose of study intervention – date of first dose of study intervention) +1.

Follow-up duration is defined as [date of last study visits or last study contact (whichever is later)-date of first dose of study intervention] +1.

Descriptive statistics will be presented by intervention group for the following parameters:

- Number of administrations
- Cumulative total dose: Cumulative total dose is the total dose in mg a participant received over all infusions in the study. The dose in mg for a given infusion is the dose level administered in mg/kg multiplied by the participant's weight.

A listing of participants with an infusion interruption or injection or change in infusion flow rate will be provided. Study intervention compliance will be summarized descriptively, see Appendix 8 for further details.

5.7.2. Adverse Events

The verbatim terms used in the CRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any AE occurring at or after the initial administration of study intervention is considered to be a treatment emergent. If the event occurs on the day of the initial administration of study intervention, and either event time or time of administration are missing, then the event will be assumed to be treatment emergent. If the AE

onset time is missing and the AE onset date is the same as the infusion date, the missing time will be imputed with the time of the infusion. If the event date is recorded as partial or completely missing, then the event will considered to be treatment emergent unless it is known to be prior to the first administration of study intervention based on partial onset date or resolution date. All reported treatment-emergent adverse events will be included in the analysis. For each adverse event, the number and percentage of participants who experience at least 1 occurrence of the given event will be summarized by intervention group.

Summary tables will be provided for treatment-emergent (TE) adverse events for:

- All AEs
- Serious AEs (SAEs)
- AEs leading to discontinuation of study intervention/termination of study participation
- AEs by severity
- AEs by relationship to study intervention
- AEs of special interest (AESI)
  - Infections that are severe or require IV anti-infective or operative/invasive intervention
  - Hypoalbuminemia with albumin < 20 g/L
- AEs of clinical interest
  - Infusion reactions
  - Infusion site reaction
  - Opportunistic infections
  - Potential Hypersensitivity reactions
  - Anaphylactic reactions or serum sickness reactions
  - Potentially associated with glucocorticoid toxicity

In addition to the summary tables, listings will be provided for participants who had:

- SAEs
- AEs leading to discontinuation of study intervention/termination of study participation
- Deaths
- AEs of special interest or clinical interest
- Major adverse cardiovascular event (MACE) identified through independent adjudication committee
- Serious infections
- Malignancies

- Pregnancies

See Appendix 10 for the definitions of adverse events in each special or clinical interest category.

Since safety should be assessed relative to exposure and follow-up, most AE summary tables will include the average number of study agent administrations and average weeks of follow-up for each intervention group.

5.7.3. AEs, SAEs and AESI by IgG

Treatment-emergent AEs, SAEs and AESIs will be summarized by minimum IgG levels (<1.0,1.0-<3.0, 3.0-6.0, >6.0 g/L). Events that will be included in these summaries are those that occurred since previous dosing prior to the date of minimum IgG level until the end of the phase. Additionally, IgG values through Week 24 will be plotted for those participants who experience an AESI.

5.7.4. Supplementary Safety Analysis

To adjust for unequal lengths of study treatment duration among participants, and potentially between intervention groups, exposure adjusted incidence rate (EAIR) based on **while on treatment estimand strategy** will be performed for these specific treatment emergent AEs: infection, SAE, and AE leading to treatment discontinuation that is more than 10%. This estimand strategy only accounts for safety events (specific TEAE of interest, i.e infection) prior to the occurrence of an intercurrent event, i.e., before a patient receives immunosuppressants or immunomodulators therapies for their RA.

5.7.4.1. While on treatment estimand

The while on treatment estimand have the same attributes as the primary estimand in population. The attributes for treatment, variable, and population level summary are listed below.

- **Treatment condition of interest vs Alternative treatment condition:** Assignment to Combination therapy – (Nipocalimab 30 mg/kg IV q2w and Certolizumab 400 mg at Week 0, 2, 4; then 200 mg Q2W) up to Week 24 in addition to SoC **vs** Monotherapy – (*Certolizumab 400 mg at Week 0, 2, 4; then 200 mg Q2W and Placebo IV q2w) up to Week 24* in addition to SoC
- **Variable/Endpoint:** Occurrence of TEAE event (i.e infection), while treatment is being taken and other RA therapies are not initiated.
- **Population-level summary:** Difference in Incidence rate for each treatment conditions
- **Intercurrent events (ICEs) and their corresponding strategies:**

ICEs	Strategy for Addressing ICEs
1. Initiation of an immunosuppressants or immunomodulators therapies for RA 2. Treatment discontinuation	**While on treatment:** Strategy targeting a treatment effect captured while on treatment

5.7.4.2. Analysis Method

The analysis will be censored at the occurrence of the first event per patient plus 30 days and will ignores the existence of later (multiple) events, as these cannot be assumed to occur independent of previous events (e.g., patient suffering from infections may have in general a higher risk of having other complications and may even have a higher risk of getting other infections). For this reason, the EAIR should be interpreted as 'rate until the first event occurs'.

Analysis for these specific TEAEs will be aligned on while on treatment estimand. Exposure-adjusted incidence rates (EAIR per 100 patient-month) along with their 90% confidence intervals will be computed using the Exact Poisson method. These incidence rates estimate the number of individual patients experiencing events, such as infections, within the period from the initial to the final dose of the study treatment plus an additional 30 days (equivalent to 5 half-lives). This estimation is derived by dividing the count of unique patients experiencing events within this period by the accrued time during the risk period, defined as the time between the initial and final study treatment doses plus 30 days, or the time leading up to the first event, whichever occurs earlier.

A graphical representation of Kaplan-Meier curve will be displayed for these specific TEAE by the intervention groups.

5.7.5. Additional Safety Assessments

5.7.5.1. Clinical Laboratory Tests

Clinical laboratory tests will be displayed for the participants included in the safety analysis set.

Hematology: hemoglobin, hematocrit, lymphocytes, mean corpuscular hemoglobin (MCH)%, mean corpuscular volume (MCV), platelet count, red blood cell (RBC) count, white blood cell (WBC) count with differential (basophils, eosinophils, lymphocytes, monocytes, neutrophils).

Clinical Chemistry: alanine aminotransferase (ALT)/serum glutamic-pyruvic transaminase (SGPT), albumin, alkaline phosphatase (ALP), aspartate aminotransferase (AST)/serum glutamic-oxaloacetic transaminase (SGOT), bicarbonate, blood urea nitrogen (BUN), calcium, chloride, cholesterol, creatinine, creatine phosphokinase (CPK), gamma-glutamyltransferase (GGT), glucose, lactic acid dehydrogenase (LDH), magnesium, phosphate, potassium, sodium, total bilirubin, total protein.

Lipid Panel: high density lipoprotein [HDL], low density lipoprotein [LDL], LDL subfractions, total cholesterol, triglycerides, and ratio of total cholesterol to HDL. Only fasting lipid samples will be included in summaries.

Urinalysis: dipstick (bilirubin, blood, glucose, ketones, leukocyte esterase, nitrite, pH, specific gravity, and urobilinogen), sediment (bacteria, casts, crystals, epithelial cells, RBCs, WBCs).

Descriptive statistics will be presented for chemistry, hematology and urinalysis (pH and specific gravity) laboratory tests at scheduled time points. Change from baseline to Week 30 will be

summarized for chemistry, hematology and urinalysis (pH and specific gravity) tests and displayed by intervention group. Descriptive statistics will be provided for percent change from baseline to each scheduled time point for albumin and total cholesterol by intervention group. Boxplots and plots of mean (±SE) value, change, and percent change from baseline to each scheduled time point will be provided for albumin and total cholesterol. Plots of mean (±SE) value and change to each scheduled time point will be provided for ALT, AST, and total bilirubin.

Shift tables will be provided summarizing the shift in select laboratory values from baseline to the maximum post-baseline value and to the minimum post-baseline value with respect to the normal range criteria (low, normal, high).

Markedly abnormal criteria will be applied to baseline and postbaseline values and are provided in Appendix 12. The number and percentage of participants with treatment-emergent markedly abnormal values will be presented by intervention group through Week 24. A listing of markedly abnormal laboratory values will be provided. Additionally, abnormal laboratory findings to be reported for albumin and liver enzyme tests are described below. For criteria that do not include an increase or decrease from baseline, post baseline abnormalities will be compared with their corresponding baseline result. Specifically, if the postbaseline value is above the upper limit and the baseline value is below the upper limit (e.g., Normal or Low), then the postbaseline abnormality will be considered TE. The same applies to the postbaseline value being below the lower limit with the baseline value being above the lower limit (e.g., Normal or High). If the baseline value is missing, a postbaseline abnormality will always be considered as TE.

- Albumin <20 g/L
- ALT, AST: >1xULN; ≥3xULN; ≥5xULN; ≥10xULN; ≥20xULN
- ALP: >1xULN; ≥3xULN; ≥5xULN; ≥10xULN; ≥20xULN
- Bilirubin: ≥2xULN

A listing of patients meeting biochemical Hy's law criteria will be provided:

- ALT or AST ≥3xULN and
- ALP <2xULN and
- Total bilirubin ≥2xULN or INR >1.5 (if measured)

5.7.5.2. Vital Signs and Physical Examination Findings

Continuous vital sign parameters including temperature, weight, pulse, blood pressure (systolic and diastolic) will be summarized at each assessment time point. Change from baseline will be summarized through Week 12 or through Week 24. Descriptive statistics (mean, SD, median, and range) will be presented.

Abnormality criteria (based on criteria in Table 6) will be applied to baseline and postbaseline values. For baseline values, increase or decrease criteria are not applied.

Postbaseline values will be considered TE if they meet both value and change criteria in the Table 6.

For criteria that do not include an increase or decrease from baseline:

- TE will be concluded if the postbaseline value is above the upper limit and the baseline value is below the upper limit (e.g., Normal or Low). The same applies to the postbaseline value being below the lower limit with the baseline value being above the lower limit (e.g., Normal or High).
- If the baseline value is missing, a postbaseline abnormality will always be considered as TE.

Incidence of TE markedly abnormal vital signs during intervention will be summarized. A listing of participants with markedly abnormal vital signs will be presented. Markedly abnormal criteria are defined in Table 6 below.

Table 6: Markedly Abnormal Vital Signs Criteria

Vital Sign	Criteria
Pulse	≥120 bpm and with ≥15 bpm increase from baseline
	≤50 bpm and with ≥15 bpm decrease from baseline
Systolic blood pressure	≥160 mmHg and with ≥20 mmHg increase from baseline
	≤90 mmHg and with ≥20 mmHg decrease from baseline
Diastolic blood pressure	≥100 mmHg and with ≥15 mmHg increase from baseline
	≤50 mmHg and with ≥15 mmHg decrease from baseline
Temperature	>38°C
	<36°C

5.8. Other Analyses

5.8.1. Pharmacokinetics

Pharmacokinetics (PK) samples for measuring serum nipocalimab or certolizumab concentrations will be collected from all participants at the specified visits as shown in the Schedule of Activities. All PK evaluations will be performed on the PK analysis set for the main study, defined as participants who have received at least 1 administration of nipocalimab or certolizumab and have at least one post-dose sample collection.

Descriptive statistics (N, mean, SD, median, range, CV (%) and IQ range) will be used to summarize nipocalimab or certolizumab serum concentrations at each sampling time point. PK data may be displayed graphically, such as median +/- IQ range PK concentrations over time by intervention group. The following analyses will be performed as appropriate:

- Summary of serum nipocalimab or certolizumab concentrations at each visit
- Summary of serum nipocalimab or certolizumab concentrations at each visit by baseline body weight quartiles

- Summary of serum nipocalimab or certolizumab concentrations at each visit by baseline age categories (< 65 years, ≥65 years)
- Summary of serum nipocalimab or certolizumab concentrations at each visit by baseline CRP level (< 15 mg/L, ≥15 mg/L at Week 0)
- Summary of nipocalimab or certolizumab concentrations at each visit by baseline MTX use (yes, no)
- Summary of nipocalimab or certolizumab concentrations at each visit by baseline Rheumatoid factor (positive, negative)
- Summary of nipocalimab or certolizumab concentrations at each visit by baseline Anti-CCP (positive, negative)
- Summary of nipocalimab or certolizumab concentrations at each visit by baseline Rheumatoid factor and Anti-CCP (positive for both, otherwise)
- Summary of nipocalimab or certolizumab concentrations at each visit by number of prior anti-TNF therapies used (1, ≥2)
- Summary of nipocalimab or certolizumab concentrations at each visit by prior failed anti-TNF therapy (yes/no)
- Proportion of participants without detectable serum nipocalimab or certolizumab concentrations (below the lower limit of quantification x minimum required dilution (LLOQxMRD)) at each visit
- Plot of median +/- IQ range serum nipocalimab or certolizumab concentrations over time
- Change from screening visit in serum ACPA and RF levels over time by treatment group, in participants seropositive at screening visit for ACPA and RF
- In addition, the relationship between serum nipocalimab or certolizumab concentrations and antibodies to nipocalimab status, safety and efficacy may be explored using graphical displays. Median serum nipocalimab or certolizumab concentrations will be plotted. Box plots of serum nipocalimab or certolizumab concentrations will be plotted by DAS28-CRPat Week 12.

PK analyses will be summarized through Week 12 and End of Study.

If sufficient data are available, then population PK analysis using serum nipocalimab or certolizumab concentration time data will be performed using nonlinear mixed-effects modeling to estimate total systemic clearance and volume of distribution. Details will be given in a population PK analysis plan and the results of the analysis will be presented in a separate report.

5.8.1.1. Data Handling Guidelines

Unless otherwise specified, the following data handling guideline will apply to PK analyses:

- All serum concentration summaries for a particular timepoint will include data obtained from treated participants at the timepoint of interest without imputing any missing data
- A concentration not quantifiable (below LLOQ) will be treated as 0 in the summary statistics and shown as the lower limit of quantification (<LLOQ) in the data listings

- The data from a participant who meets one of the following dosing deviation criteria will be excluded from the by-visit data analyses and from that point onwards:
  - Discontinue nipocalimab or certolizumab administrations
  - Skipped a nipocalimab or certolizumab administration
  - Received incomplete/incorrect dose
  - Received incorrect study agent
  - Received additional dose

Of note, serum nipocalimab or certolizumab concentrations prior to the first of such events will be included in the summaries. In addition, if a participant has an administration outside of dosing window, the concentration data collected at and after that will be excluded from the by-visit data analyses. Additional exclusions for PK data to be implemented based on TV-GDL-00362. All participants and samples excluded from analysis will be documented in the Clinical Study Report.

5.8.2. Immunogenicity

Blood samples will be collected to examine the formation of antibodies to nipocalimab at the specified visits as shown in the Schedule of Activities of the protocol.

"Sample ADA status" and sample titer as well as the cumulative "participant ADA status" and peak titer through the visit will be coded and provided by the bioanalytical group.

5.8.2.1. Participant ADA Classifications

Participants evaluable for immunogenicity are defined as having at least one postdose ADA time point collected for antibodies to nipocalimab detection.

1. Participants with treatment-emergent antibodies to nipocalimab include participants with treatment-induced antibodies to nipocalimab and treatment-boosted antibodies to nipocalimab.
2. Participants with treatment-induced antibodies to nipocalimab have an ADA negative sample prior to nipocalimab administration and at least one ADA positive sample after nipocalimab.
3. Participants with treatment-boosted antibodies to nipocalimab have an ADA positive sample prior to nipocalimab administration and at least one ADA positive sample after nipocalimab with a 2-fold increase in titer over baseline (the fold difference, referred to as the baseline multiplier, could be greater than 2 for some assays).

If titer remains the same after intervention or if ADA titer reduces or ADA disappears, the participant is classified as "treatment-emergent ADA negative". Participants that are unavailable for treatment-emergent ADA following intervention will be classified as "participants with baseline samples only", i.e., no appropriate sample is available after intervention.

5.8.2.2. Immunogenicity Analyses

The summary and evaluation of antibodies to nipocalimab will be based on the observed data; therefore, no imputation of missing data will be performed. Note: participant status is through each

visit, thus, participant status and peak titers may change as the study progresses over time. Therefore, the 'participant ADA status' at a visit represents the cumulative ADA status through that visit. For example, if a study has a database lock at Week 24, datasets through Week 24 will have participant level status (e.g., negative) but at Week 58, they may have developed ADA and the participant status becomes "treatment-emergent ADA positive" from the interim to the final DBL. Peak titers can also change (increase) if a higher titer occurs after an initial DBL.

The summary of participants with baseline positive samples is taken from the sample status at baseline. There is no participant level status at baseline.

The data analysis of antibodies to nipocalimab includes the following:

- Incidence of antibody (evaluable, treatment-emergent ADA positive, treatment-emergent ADA negative) status and neutralizing antibodies (NAb) to nipocalimab will be summarized.
- Descriptive statistics (N, mean, SD, median, range, and IQ range) and incidence (N, %) of the relationship between treatment-emergent antibodies to nipocalimab status (positive or negative) and PK concentration will be assessed.
- Participants in response (N, %) for treatment-emergent antibodies to nipocalimab status (positive or negative) and efficacy endpoints will be assessed:
  - Participants evaluable for immunogenicity
  - Efficacy endpoints repeated for different levels of response (e.g., ACR 20/ 50/70/ 90, DAS28 (CRP))
    - Number of participants (N)
    - Participants in response (N, %)
- Incidence (N, %) between treatment-emergent antibodies to nipocalimab status (positive or negative) and infusion-related reactions will be assessed:
  - Participants evaluable for immunogenicity
  - Participants with infusion-related reaction
  - Participants with severe infusion-related reaction
  - Participants with serious infusion-related reaction
  - Participants with infusion-related reaction leading to discontinuation
  - Nipocalimab infusions with infusion-related reactions (out of total number of nipocalimab infusions)
  - Placebo infusions with infusion-related reactions (out of total number of placebo infusions).

In addition, listings of participants with baseline positive ADA samples, participants who are classified as positive for treatment-emergent antibodies to nipocalimab and participants who discontinue the study by antibodies to nipocalimab status as well as graphical representation of median serum concentration by antibody status may be presented.

Additional immunogenicity analyses by peak titer groups may be performed if the number of participants who are positive of antibodies to nipocalimab are sufficient to permit subgroup analyses (eg > 30 to 40 participants).

5.8.3. Biomarkers

Biomarker assessments will be made to examine the biologic response to treatment and to identify biomarkers that are relevant to certolizumab or certolizumab/nipocalimab combination therapy and RA, where local regulations permit. Assessments (detailed below) will include the evaluation of relevant biomarkers in serum, plasma, whole, and urine collected as specified in the SoA where local regulations permit.

Data collected from these samples will be used for exploratory research that will include the following objectives:

1. To understand the molecular effects of certolizumab monotherapy and nipocalimab/certolizumab combination therapy
2. To understand RA pathogenesis
3. To understand why individual participants may respond differently to certolizumab monotherapy and nipocalimab/certolizumab combination therapy
4. To understand the impact of certolizumab monotherapy and nipocalimab/certolizumab combination therapy on RA or systemic inflammation
5. To develop diagnostic tests to identify RA populations that may be responsive or nonresponsive to treatment with certolizumab monotherapy and nipocalimab/certolizumab combination therapy

Biomarker analyses are dependent upon the availability of appropriate biomarker assays and clinical response rates. Biomarker analysis may be deferred or not performed, if during or at the end of the study, it becomes clear that the analysis will not have sufficient scientific value for biomarker evaluation, or if there are not enough samples or responders to allow for adequate biomarker evaluation. In the event the study is terminated early or shows poor clinical efficacy, completion of biomarker assessments is based on justification and intended utility of the data.

5.8.3.1. Pharmacodynamics

Venous blood samples will be collected at time points shown in the SoA.

Serum concentrations of total IgG, IgM, IgE, and IgA will be measured to assess the PD effect. IgG subclasses (IgG1, IgG2, IgG3, IgG4), albumin, CICs, and inflammatory markers (e.g., CRP) may also be evaluated using separate assays to assess the PD effect of nipocalimab and certolizumab. The relationship between PD effects and clinical responses may be assessed.

5.8.3.2. Pharmacokinetic/Pharmacodynamic Relationships

Exploratory PK-PD analyses, including graphical exploration of PK-PD data, may be performed such as PK/IgG or PK/pathogenic antibodies, PK/RF, PK/albumin, PK/lipids.

For efficacy, percent responders within each quartile of PK concentration may be presented, if data permits.

If deemed feasible and necessary, exposure-response analyses may be performed. The analysis methods may be summarized in a separate analysis plan. Results of such analyses may be presented in a separate technical report.

5.8.3.3. Serum and Plasma Biomarkers

Blood samples for serum and plasma biomarker analyses will be collected from all participants, where local regulations permit. Serum and plasma may be analyzed for levels of circulating proteins, autoantibodies (e.g., ACPA, RF), other inflammation-related molecules, and disease-associated serologies relevant to RA and treatment and response to nipocalimab and certolizumab.

5.8.3.4. Whole Blood Gene Expression Profile

Whole blood will be collected by venipuncture from participants for RNA expression analysis, where local regulations permit. Total RNA will be isolated and used for differential gene expression analyses to identify gene expression patterns that are relevant to nipocalimab and certolizumab treatment or RA and to evaluate markers that can predict clinical response. Transcriptomic studies may be conducted using microarray or alternative equivalent technologies, which facilitate the simultaneous measurement of the relative abundance of multiple RNA species for each blood sample. The samples may also be used for targeted assessment of genes relevant to RA and nipocalimab/certolizumab treatment. These analyses may be used to evaluate the changes in gene expression profiles that may correlate with biologic response relating to RA and the action of nipocalimab and certolizumab and may be used to identify population subtypes that respond differently to an intervention.

5.8.3.5. Peripheral Blood Mononuclear Cells

If operationally feasible, whole blood may also be collected and processed for PBMC isolation and cryopreserved for later analysis. Analysis may include but is not limited to flow cytometric assessment of cell populations, single cell transcriptomics, or functional assessment of cells. The samples may also be used to evaluate cellular and molecular changes in response to nipocalimab/ certolizumab treatment or related to RA pathogenesis. These analyses may not be performed if cryopreserved PBMC samples do not meet the quality or quantity standards.

5.8.4. Additional PK, PD, Immunogenicity and Pharmacogenomic Analysis

An additional planned analysis for Pharmacodynamic biomarker for total IgG, IgG subclasses, disease, anti-CCP2, RF-IgM, RF-IgG and Serum Molecular Disease Profile.

The list of endpoints:

- Change and % change from baseline in total IgG, IgM, IgE, and IgA over time
- Change and % change from baseline in IgG subclasses (IgG1, IgG2, IgG3, IgG4) over time

- Change from baseline in anti-CCP2
- Change from baseline in RF IgM and RF IgG
- Change from baseline in Serum Molecular Disease Profile (M-DP 4 and M-DP 9 analyte)

5.8.4.1. Analysis Methods

These analyses will be based on PD analysis set; however, if a participant misses a planned dose of study intervention at any visit, their data is excluded from all subsequent visits after the first occurrence of a missed dose due to any reasons (which includes patients who discontinue study treatment but continue to the study). For example, if a participant had all planned doses through Week 12 and then missed the Week 14 dose, they would only be included in the summaries through Week 14 (if available) and excluded from all subsequent visit summaries.

Descriptive statistics summarizing each endpoint at scheduled visits will be provided, encompassing the change from baseline through Week 30 by intervention group. This summary will detail changes and percent changes from baseline at each scheduled time point, and for each endpoint. For each endpoint, it will display the geometric mean (95% CI), arithmetic mean (SD), median, range and IQ range for both scheduled visit and by intervention groups.

In addition, Median (IQR) plots for measured value, and percent change from baseline over time will be provided for each endpoint. Note, see additional DPS developed for these PK, PD, Immunogenicity and Pharmacogenomic analysis.

5.8.4.2. Data Handling Rule

Unless otherwise specified, the following data handling rule will be applied to Pharmacodynamic biomarker analyses only:

- A concentration not quantifiable (below LLOQ) will be replaced by half of the value of the lower limit of quantification (LLOQ) for numerical calculations in PD analysis, i.e. if Anti-CPP LLOQ <25, then value will be replaced by 12.5 U/ml (25*0.5).

5.8.5. Subgroup Analysis

To evaluate the consistency in the primary efficacy endpoints, change from baseline in DAS28-CRP at Week 12 over demographics, baseline disease characteristics, prior and baseline medication use, subgroup analyses will be performed. The subgroups for subgroup analyses include, but are not limited to, the following:

5.8.5.1. Definition of Subgroups

The subgroups for subgroup analysis may include, but are not limited to, the following:

- Demographics

Subgroup	Definition
Region	Define based on UN guidance as per the M49 standard

Subgroup	Definition
	• North America • Western Europe • Eastern Europe • Latin America
Age Group	• <65 • ≥65
BMI	• normal <25 kg/m^2 • overweight 25-<30 kg/m^2 • obese ≥30 kg/m^2
Body Weight Group	• 1st Quartile • 2nd Quartile • 3rd Quartile • 4th Quartile
Body Weight Group	• ≤55kg • > 55 and ≤85 kg • > 85 kg
Race	• White • Black • Asian • All other categories
Ethnicity	• Hispanic or Latino • Not Hispanic or Latino
Gender	• Male • Female
Disease duration	• < 1 year • ≥1 year to < 3 years, • ≥3 years

- Baseline Disease Characteristics subgroups

Subgroup	Definition
Rheumatoid factor	• Positive • Negative
Anti-CCP (from Labcorp)	• Positive • Negative
Rheumatoid factor and anti-CCP	• Positive • Negative
Anti-CCP (from PBL)	• ≥ Median • < Median
Rheumatoid factor, IgM, and IgG both considered separately	• ≥ Median • < Median
Number of swollen 66 joints	• <10 • ≥10
Number of tender 68 joints	• <10 • ≥10
HAQ	• <2 • ≥2
CRP (from Labcorp)	• <15 mg/L • ≥15 mg/L
ESR	• <30 • ≥30

- Prior and Baseline Medicine Subgroups

Subgroup	Definition
Baseline MTX use	• Yes • No
Baseline NSAIDs	• Yes • No
Baseline oral corticosteroids	• Yes • No
Number of failed anti TNF α medications prior to study entry (1, ≥2)	• Yes • No
Failed (due to Lack of Efficacy) any prior anti TNF therapy	• Yes • No
Intolerant and safety to any prior anti TNF therapy	• Yes • No
Number of failed Advance Therapies prior to study entry	• 0 • 1 • 2 • ≥3

5.9. Interim Analysis

No Interim analysis is planned for this study.

5.10. Data Monitoring Committee or Other Review Board

An independent Data Monitoring Committee (DMC) will be established to monitor data on an ongoing basis. This committee will consist of at least 2 medical experts in the relevant therapeutic area and at least one statistician; committee membership responsibilities, authorities, and procedures will be documented in its charter. An early safety review of the first 15 participants randomized to any treatment arm and treated will be performed by DMC. Detailed guidance for the DMC regarding these reviews will be provided in the DMC charter, see (EDMS-RIM-587693). Once the 15th participant is randomized, the DMC will perform a review of unblinded safety tables. If no new safety concerns are identified during this initial review period, then the subsequent DMC reviews will include monthly reports of all SAEs in enrolled participants, and the quarterly scheduled safety reviews, the DMC will make a recommendation to the sponsor whether the study should continue, be modified, or stopped for safety concerns. Further details of the safety reviews will be in the DMC SAP (EDMS-RIM-520688).

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1: Participant Dispositions

The distribution of the time to study termination/discontinuation of study intervention will be displayed with Kaplan-Meier curves. Participants who terminate study participation prematurely/discontinue study intervention at any time will be considered an 'Event' and their date of study termination (premature or permanent)/study intervention discontinuation will be used in the time to event calculation. Participants who complete the study/study intervention will be censored and the date of study completion/last dose of study intervention will serve as the time of censoring.

6.2. Appendix 2: Baseline Characteristics and Demographics

Table 7 presents a list of the demographic and some baseline variables that will be summarized by intervention group, and an overall for the following analysis sets: FAS, all randomized (should it differ from FAS.

Table 7: Demographic Variables and clinical characteristics

Continuous Variables:	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum], and IQ range).
Weight (kg)	
Height (cm)	
Body Mass Index (BMI) (kg/m^2)	
RA disease duration (years)	
Number of swollen joints (0-66)	
Number of tender joints (0-68)	
Patient's assessment of pain (VAS; 0-10)	
Patient's global assessment of disease activity (VAS; 0-10)	
Physician's global assessment of disease activity (VAS; 0-10)	
HAQ disability index (0-3)	
CRP (mg/dL)	
Number of swollen joints (0-28)	
Number of tender joints (0-28)	
DAS28 (CRP)	
DAS28 (ESR)	
CDAI Score	
SDAI Score	
Joint Pain Severity NRS	
ESR (mm/hr)	
Categorical Variables	**Summary Type**
Age (<65 Years; >=65 years])	Frequency distribution with the number and percentage of participants in each category.
Sex (male, female, undifferentiated)	
Race[a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Multiple)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino)	
BMI (<25 kg/m^2, >=25 to <30 kg/m^2, obese >=30 kg/m^2])	
ACR/EULAR Remission (Participants achieved remission/ N)	
Anti-CCP antibody (Participants with positive anti-CCP antibody/N)	
Rheumatoid factor (Participants with positive rheumatoid factor/N)	

[a] If multiple race categories are indicated, the Race is recorded as 'Multiple'

6.3. Appendix 3: Protocol Deviations

In general, the following list of major protocol deviations may have the potential to impact participants' rights, safety or well-being, or the integrity and/or result of the clinical study. Participants with major protocol deviations will be identified prior to database lock.

Protocol deviations will be summarized overall for the FAS analysis set through Week 0-30, and by treatment group and overall including the final safety visit occurring prior to the Week 30 DBL.

In addition to the summary tables, the following listings will be provided from Week 0 through Week 24, including the final safety visit occurring prior to the Week 30 DBL:

- List of participants with major protocol deviations
- List of participants who did not meet study selection criteria by category
- List of participants who had a protocol deviation in study intervention administration

6.3.1. Major Protocol Deviations

Participants with major protocol deviations will be identified prior to the Week 24 database lock and the participants with major protocol deviations will be summarized by category.

- Developed withdrawal criteria but not withdrawn
- Entered but did not satisfy criteria
- Received a disallowed concomitant treatment
- Received wrong medication or incorrect dose
- Other

In these summaries a participant can be included in more than 1 deviation category.

6.3.2. Inclusion/Exclusion Criteria Deviations

Participants who entered the study but did not meet the inclusion/exclusion selection will be grouped into the following 5 categories: RA disease criteria, medication criteria, laboratory criteria, medical history criteria, and other.

In these summaries a participant can be included in more than 1 deviation category.

6.3.3. Study Administration Deviations

Protocol deviations in study agent administrations includes missing doses, incorrect doses, and treatments administrated out of dosing windows.

In these summaries a participant can be included in more than 1 deviation category.

6.4. Appendix 4: Rules Applied in definitions of endpoints

6.4.1. Joint Evaluability Rules for Sign and Symptom Data

Joints should only be designated as "non-evaluable" by the joint assessor on the Joint Assessment Worksheet if it is physically impossible to assess the joint (ie, joint inaccessible due to a cast, joint was replaced, joint not present due to an amputation, joint deformed so as to make it impossible to assess). In all other cases, the joint assessor should assess each joint for tenderness and swelling (hips are excluded for swelling) and complete the worksheet with their assessments. This should be completed regardless of any visual indications of prior surgical procedures (eg, scars) or knowledge they may have of a participant's prior joint procedures/injections.

For participants having a joint injection(s)/surgical joint procedure(s) prior to the date of study entry (e.g., randomization) or during the study, the affected joint(s) will be valued according to the following rules:

- For participants having a joint injection and/or surgical joint procedure prior to the date of randomization, the affected joints will be analyzed according to the impact of the joint injection and/or surgical joint procedure on the evaluability of the involved joints.
- If a joint is considered un-evaluable at baseline due to certain procedure/injection performed prior to the date of randomization, the joint will be considered un-evaluable throughout the study.
- For participants undergoing surgical joint procedures for the treatment of RA during the study, the affected joints will be considered as swollen and tender from the date of procedure onwards.
- For participants undergoing joint injections during the study, the affected joints will be considered as swollen and tender from the date of injection for the next 90 days.

6.4.2. Joint Count Adjustment Rule

For participants who have an incomplete set of evaluable joints the joint count/score will be adjusted to the total number joints of interest (e.g., 68 joints for tenderness and 66 joints for swelling) by dividing the number of affected joints by the number of evaluable joints and multiplying by the total number joints of interest.

6.4.3. LLOQ Rule

Any CRP value < LLOQ will be replaced by half of the value of the lower limit of quantification (LLOQ) for numerical calculations of the efficacy endpoints.

6.5. Appendix 5: Prior and Concomitant Medications

Prior and Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). Prior medications are defined as any therapy used before the day of first dose (partial or complete) of study intervention. Concomitant medications are defined as any therapy used on or after the same day as the first dose of study intervention, including those that started before and continue after the first dose of study intervention.

Summaries of RA specific concomitant medications will be presented by anatomic-therapeutic-chemical classification [ATC] term and intervention group. The proportion of participants who receive each concomitant medication will be summarized as well as the proportion of participants who receive at least one RA concomitant medication.

Background medication use for active RA (csDMARDs (MTX: ≤25 mg/week, SSZ: 1000-2000 mg/day, HCQ or CQ: 200-400 mg/day, Leflunomide: 10-20 mg/day), oral corticosteroids including prednisone and NSAIDs) will be summarized by intervention group and schedule visits through Week 24.

In addition, summaries of RA related medication history (history of inadequate response to or tolerance of TNF and other biologics/biosimilar will be summarized by intervention group and ATC term.

6.6. Appendix 6: Prohibited changes in concomitant medications

Had prohibited changes in RA medication as defined below:

Oral Corticosteroids:

Increase above baseline in the prednisone equivalent dosage of oral corticosteroids due to worsening RA.

csDMARDs

- Increase above baseline in the dosage of csDMARDs: (MTX: ≤25 mg/week (IM, SC, or PO), SSZ: 1000-2000 mg/day, HCQ or CQ: 200-400 mg/day and Leflunomide: 10-20 mg/day) due to worsening of RA disease.
- No csDMARDs at baseline and initiation of csDMARDs after baseline due to worsening RA disease.
- Switch between one csDMARDs to another DMARD due to worsening RA disease

Protocol-prohibited medications:

Initiation of any of the following immunosuppressants or immunomodulators agents after baseline due to worsening of RA disease:

- Systemic immunosuppressives (other than MTX, SSZ, HCQ, CQ, and leflunomide) such as azathioprine, oral cyclosporine A, tacrolimus, mycophenolate mofetil, oral or parenteral gold
- Cytotoxic drugs such as cyclophosphamide, chlorambucil, nitrogen mustard, or other alkylating agents
- Anti-TNFα therapy such as: infliximab, golimumab, etanercept, adalimumab, and biosimilars to those anti-TNFs
- Anti-IL-6 receptor mAb (eg, tocilizumab or sarilumab)
- Anti-IL-1 receptor mAb (eg, anakinara)
- B-cell depleting biologic therapy (eg, rituximab)
- Co-stimulatory inhibitors (eg, abatacept)
- JAKi (eg, tofacitinib, baricitinib, upadacitinib)
- Any other targeted biologic therapy
- Any investigational intervention or use of an invasive investigational medical device
- Use of complementary therapies or traditional medicine (eg, Chinese, acupuncture, ayurvedic) through Week 24
- Live virus or live bacterial vaccination during the study, unless approved by the sponsor (80202135ARA2002 Protocol Section 6.8)

As these lists cannot be exhaustive, please consult the Medical Monitor to discuss prior to starting any biologic or other advanced therapies

6.7. Appendix 7: Rescue Medication

The following rescue medications may be used:

- csDMARDs such as MTX, SSZ, HCQ, CQ, and LEF
- oral corticosteroids

At Week 14, participants who have not achieved low disease activity (defined as CDAI ≤10) will receive the current SoC treatment starting from Week 14 (investigator's choice of SoC treatment appropriate for the participant).

If participants initiate or increase their dose of csDMARDs, oral corticosteroids and/or NSAIDs, there will be no need to discontinue study intervention, and data collection will continue through Week 30. The study intervention must be discontinued if rescue medications are the prohibited medications listed in Appendix 6. The date of rescue medication administration as well as the name and dosage regimen of the rescue medication must be recorded.

6.8. Appendix 8: Intervention Compliance

Compliance will be summarized descriptively for each study agent within a study intervention. Compliance to randomized intervention versus actual intervention will be presented in a summary table. Since the study has a combination intervention, therefore the 'study agent' refer to the individual components included in the combination intervention.

Compliance (%) will be calculated as (actual number of injection and infusion received/total number of injections and infusion planned) x 100.

6.9. Appendix 9: Medications History

Medical history will be listed by intervention groups and participants.

6.10. Appendix 10: Adverse Events of Interest

Adverse events of special interest are indicated on the eCRF via a checkbox and are reconciled with the definition below:

AE Special Interest Category	SOC	Additional condition
Infections	Infection and Infestations	Checkbox for AESI that indicates it was severe or required IV anti-infective or operative/invasive intervention

AE Special Interest Category	Preferred term	Additional condition
Hypoalbuminaemia	Hypoalbuminemia	Checkbox for AESI that indicates Albumin <20 g/L

Other adverse events of clinical interest are defined as follows:

AE of Clinical Interest Category	Preferred term	Additional condition
Infusion reaction	Any	Indicated as infusion reaction by investigator on eCRF and relationship to study intervention='Related'. Exclude infusion site reactions.
Infusion site reaction	HLT of Infusion site reaction	
Serum sickness reaction	Serum sickness	Requires clinical review
	Serum sickness-like reaction	Requires clinical review
Anaphylactic reaction	Anaphylactic reaction; Anaphylactic shock; Anaphylactoid reaction; Anaphylactoid shock; Type I hypersensitivity; Kounis syndrome	Determined through SMQ algorithmic approach/Sampson's criteria
Potentially Associated with Glucocorticoid Toxicity	**Cardiovascular:** Acute myocardial infarction*; Angina pectoris; Arteriosclerosis; Blood cholesterol increased; Blood pressure increased; Cardiac failure; Cardiovascular insufficiency; Congestive cardiomyopathy; Dyslipidaemia; Fluid retention; Hypercholesterolaemia; Hyperlipidaemia; Hypertension; Hypertensive emergency; Hypertriglyceridaemia; Hypervolaemia; Low density lipoprotein increased; Myocardial infarction*; Myocardial ischaemia; Oedema; Oedema peripheral; Peripheral swelling	*Also a MACE preferred term
	Infections: all serious infections	
	Gastrointestinal: Duodenal ulcer; Gastritis; Gastritis erosive; Gastrointestinal disorder; Pancreatitis; Pancreatitis acute	
	Psychological: Affective disorder; Agitation; Anxiety; Confusional state; Depressed mood; Depression; Insomnia; Irritability; Libido decreased; Major depression; Mania; Mental status changes; Mood altered; Nervousness; Poor quality sleep	

AE of Clinical Interest Category	Preferred term	Additional condition
	Endocrine/metabolic: Adrenal insufficiency; Blood glucose increased; Blood potassium decreased; Central obesity; Cushingoid; Cushing's syndrome; Diabetes mellitus; Diabetes mellitus inadequate control; Glucose tolerance impaired; Gynaecomastia; Hyperglycaemia; Hypokalaemia; Influenza like illness; Polymenorrhagia; Menometrorrhagia; Systemic inflammatory response syndrome; Type 2 diabetes mellitus; Waist circumference increased; Weight increased	
	Dermatological: Acne; Dermatitis acneiform; Ecchymosis; Hirsutism; Increased tendency to bruise; Skin atrophy; Skin striae	
	Musculoskeletal: Hip fracture; Humerus fracture; Lower limb fracture; Lumbar vertebral fracture; Muscle atrophy; Muscular weakness; Myopathy; Osteonecrosis; Osteopenia; Osteoporosis; Spinal compression fracture; Tendon rupture; Wrist fracture	
	Ophthalmological: Cataract; Cataract nuclear; Glaucoma; Intraocular pressure increased; Open angle glaucoma; Retinopathy hypertensive	
Potentially abuse-related	Aggression; Confusional state; Decreased activity; Dependence; Disorientation; Dissociation; Dissociative disorder; Dizziness; Drug abuse; Drug abuser; Drug dependence; Drug detoxification; Drug diversion; Drug rehabilitation; Drug tolerance; Drug tolerance increased; Drug use disorder; Drug withdrawal convulsions; Drug withdrawal headache; Drug withdrawal syndrome; Euphoric mood; Feeling abnormal; Feeling drunk; Feeling of relaxation; Hallucination; Hallucination, auditory; Hallucination, gustatory; Hallucination, olfactory; Hallucination, synaesthetic; Hallucination, tactile; Hallucination, visual; Hallucinations, mixed; Inappropriate affect; Mental impairment; Product tampering; Psychomotor hyperactivity; Psychotic disorder; Rebound effect; Somatic hallucination; Somnolence; Substance abuser; Substance dependence; Substance use; Substance use disorder; Substance-induced mood disorder; Substance-induced psychotic	

AE of Clinical Interest Category	Preferred term	Additional condition
	disorder; Thinking abnormal; Withdrawal arrhythmia; Withdrawal syndrome	
Hyperlipidemia	Hyperlipidaemia; Lipids increased Lipids abnormal; Dyslipidaemia Hypertriglyceridaemia; Blood triglycerides increased; Blood triglycerides abnormal; Hypercholesterolaemia; Blood cholesterol increased; Blood cholesterol abnormal; Low density lipoprotein increased; Low density lipoprotein abnormal; Very low density lipoprotein abnormal; Very low density lipoprotein increased; Non-high-density lipoprotein cholesterol increased	
Activation of latent virus	Epstein-Barr Virus Infection reactivation; Epstein-Barr Virus Infection; Epstein-Barr Viraemia; Cytomegalovirus infection reactivation; Cytomegalovirus infection; Cytomegalovirus Viraemia; Herpes simplex reactivation; Herpes zoster reactivation; Herpes virus infection; Cytomegalovirus urinary tract infection; Cytomegalovirus gastrointestinal infection; Gastritis herpes; Genital herpes; Genital herpes simplex; Genital herpes zoster; Herpes dermatitis; Herpes oesophagitis; Herpes ophthalmic Herpes pharyngitis; Herpes sepsis Herpes simplex; Herpes simplex bronchitis; Herpes simplex cervicitis; Herpes simplex colitis; Herpes simplex encephalitis; Herpes simplex gastritis; Herpes simplex hepatitis; Herpes simplex meningitis; Herpes simplex meningoencephalitis; Herpes simplex meningomyelitis; Herpes simplex necrotising retinopathy; Herpes simplex oesophagitis; Herpes simplex otitis externa; Herpes simplex pharyngitis; Herpes simplex pneumonia; Herpes simplex sepsis; Herpes simplex viraemia; Herpes simplex virus urethritis; Herpes simplex visceral; Herpes zoster; Herpes zoster cutaneous disseminated; Herpes zoster disseminated; Herpes zoster infection	

AE of Clinical Interest Category	Preferred term	Additional condition
	neurological; Herpes zoster meningitis; Herpes zoster meningoencephalitis; Herpes zoster meningomyelitis; Herpes zoster meningoradiculitis; Herpes zoster necrotising retinopathy; Herpes zoster oticus; Herpes zoster pharyngitis; Herpetic radiculopathy; Lower respiratory tract herpes infection; Meningitis herpes; Meningoencephalitis herpetic; Meningomyelitis herpes; Nasal herpes; Necrotising herpetic retinopathy; Ophthalmic herpes simplex; Ophthalmic herpes zoster; Oral herpes; Oral herpes zoster; Pneumonia herpes viral; Proctitis herpes; Cytomegalovirus chorioretinitis; Cytomegalovirus colitis; Cytomegalovirus duodenitis; Cytomegalovirus enteritis; Cytomegalovirus enterocolitis; Cytomegalovirus gastritis; Cytomegalovirus gastroenteritis; Cytomegalovirus gastrointestinal ulcer; Cytomegalovirus hepatitis; Cytomegalovirus mononucleosis; Cytomegalovirus mucocutaneous ulcer; Cytomegalovirus myelomeningoradiculitis; Cytomegalovirus myocarditis; Cytomegalovirus nephritis; Cytomegalovirus oesophagitis; Cytomegalovirus pancreatitis; Cytomegalovirus pericarditis; Cytomegalovirus syndrome; Disseminated cytomegaloviral infection; Encephalitis cytomegalovirus; Pneumonia cytomegaloviral; Epstein Barr virus positive mucocutaneous ulcer	

6.11. Appendix 11: Quality Tolerance Limits

The 80202135ARA2002 trial will be supported with ARBM Light model. Thus, Quality Tolerance Limit (QTL) will not be defined for this study.

6.12. Appendix 12: Markedly Abnormal Laboratory Values

Category	Lab test	Sex	Unit	Markedly Abnormal Lows*	Markedly Abnormal High*
CHEMISTRY	Albumin	BOTH	g/L	Decrease >10 and value <20	Increase >10 and value > 60
CHEMISTRY	Alkaline phosphatase	BOTH	U/L	N/A	Increase >100 and value >250
CHEMISTRY	Alanine transaminase (ALT)(SGPT)	BOTH	U/L	N/A	≥3xULN
CHEMISTRY	Aspartate transaminase (AST) (SGOT)	BOTH	U/L	N/A	≥3xULN
CHEMISTRY	Bicarbonate	BOTH	mmol/L	Decrease >20% and value <15.1	Increase >20% and value >34.9
CHEMISTRY	Blood urea nitrogen	BOTH	mmol/L	N/A	Increase >20% and value >17.9
CHEMISTRY	Calcium	BOTH	mmol/L	Decrease >20% and value <1.5	Increase >20% and value >3
CHEMISTRY	Chloride	BOTH	mmol/L	Decrease >5 and value <85	Increase >5 and value >120
CHEMISTRY	Creatinine	BOTH	umol/L	N/A	Increase >20% and value >250
CHEMISTRY	Gamma glutamyl transferase	BOTH	U/L	N/A	Increase >100 and value >300
CHEMISTRY	Glucose	BOTH	mmol/L	Decrease >20% and value <2.2	Increase >30% and value >16.7
CHEMISTRY	Phosphate	BOTH	mmol/L	Decrease >10% and value <0.6	Increase >10% and value >2.6
CHEMISTRY	Potassium	BOTH	mmol/L	Decrease >10% and value <3	Increase >20% and value >6.0
CHEMISTRY	Sodium	BOTH	mmol/L	Decrease >10% and value <125	Increase >10% and value >155
CHEMISTRY	Total bilirubin	BOTH	umol/L	N/A	Increase >20% and value >45
CHEMISTRY	Total protein	BOTH	g/L	Decrease >20% and value <50	N/A
CHEMISTRY	Creatine Kinase	BOTH	U/L	N/A	Increase >20% and value >960
CHEMISTRY	Total Cholesterol	BOTH	mmol/L	N/A	≥6.2
CHEMISTRY	HDL	BOTH	mmol/L	<1.0	≥2.1
CHEMISTRY	LDL	BOTH	mmol/L	N/A	≥4.1
CHEMISTRY	Triglycerides	BOTH	mmol/L	N/A	>5.6
HEMATOLOGY	Hematocrit female	F	fraction	Decrease >15% and value <0.28	Increase >15% and value >0.50

Category	Lab test	Sex	Unit	Markedly Abnormal Lows*	Markedly Abnormal High*
HEMATOLOGY	Hematocrit male	M	fraction	Decrease >15% and value <0.28	Increase >15% and value >0.55
HEMATOLOGY	Hemoglobin	BOTH	g/L	Decrease >10% and value <80	Increase >10% and value >190
HEMATOLOGY	Neutrophils/Leukocytes	BOTH	fraction	Decrease> 30% and value <0.30	Increase >30% and value >0.90
HEMATOLOGY	Monocytes/Leukocytes	BOTH	fraction	N/A	Increase >20% and value >0.20
HEMATOLOGY	Eosinophils/Leukocytes	BOTH	fraction	N/A	Increase >20% and value >0.10
HEMATOLOGY	Basophils/Leukocytes	BOTH	fraction	N/A	Increase >20% and value >0.06
HEMATOLOGY	Lymphocytes/Leukocytes	BOTH	fraction	Decrease > 20% and value <0.08	Increase >20% and value >0.60
HEMATOLOGY	Platelet count	BOTH	x10E9/L	Decrease >20% and value <100	Increase >20% and value >600
HEMATOLOGY	Red blood cell count Female	F	x10E12/L	Decrease >20% and value <3	Increase >20% and value >6.1
HEMATOLOGY	Red blood cell count Male	M	x10E12/L	Decrease >20% and value <3	Increase >20% and value >6.4
HEMATOLOGY	White blood cell count	BOTH	x10E9/L	Decrease >10% and value <2.5	Increase >20% and value >15
URINALYSIS	Urine pH	BOTH		N/A	>8

*Increases and decreases are calculated from baseline values.

6.13. Appendix 13: Changes to Protocol-Planned Analyses

Not Applicable.

6.14. Appendix 14: Summary of Analyses Based on Multiple Imputation

A multiple imputation method will be utilized to impute missing DAS28/CDAI/SDAI and Joint Pain severity composite scores through Week 12 or through Week 24, after which the ICE rules will be applied, and then change from baseline in endpoint will be determined. Following steps should be followed to implement MI with ICE rules (applicable to composite estimand):

1. Data after ICE 1-4 are not used in the imputation model, i.e., data collected at or after ICEs are not contributing to the MI model, for example, if participant has an ICE, then from the date of the ICE, their data will be treated as missing instead and imputed with MI.
2. Perform MI as specified in Table 8. For each of the multiply imputed datasets, determine the endpoint value. For example, the composite endpoint (e.g., DAS28-CRP), calculate the composite endpoint value then drive change from baseline for continuous endpoint at a visit (e.g, change from baseline in DAS28-CRP at Week 12)

3. After the imputation, for those with an ICE 1-3, apply the ICE rule, i.e., composite estimand
4. Each of the imputed dataset will be analysed using an ANCOVA model for continuous endpoints). The result from each of the imputed dataset will then be combined to produce inferential results based on Rubin's rule.

Following steps should be followed to implement MI without ICE rules (applicable to supplementary estimand):

1. Perform MI as specified in Table 8. For each of the multiply imputed datasets, determine the endpoint value. For example, the composite endpoint (e.g., DAS28-CRP), calculate the composite endpoint value then drive change from baseline for continuous endpoint at a visit (e.g, change from baseline in DAS28-CRP at Week 12)
2. Each of the imputed dataset will be analysed using an ANCOVA model for continuous endpoints). The result from each of the imputed dataset will then be combined to produce inferential results based on Rubin's rule.

Continuous endpoint

For continuous endpoint imputation, imputation will be performed on continuous scale, and it will be restricted to only impute within its possible range of values (e.g., TJC68 score may only be imputed to be between 0 and 68). The missing data will be imputed using the predicted value from an imputation model using the full conditional specification regression method for any missing pattern.

Table 8 below summarizes the imputation models and seeds for different endpoints.

Table 9: Summary of Multiple Imputation Method

Imputation Variable	**Imputed dataset** [a,b]	Change from baseline/percent change from baseline in ACR components[d,e]	Change from baseline in DAS28-CRP [d,e]	Change from baseline in CDAI [d,e]	Change from baseline in SDAI [e]	Change from baseline in HAQ [d,e]	Change from baseline in DAS28 (ESR) [e]	Change from baseline in Joint Pain severity score[e,f]
TJC28	MIdat_TJC28 (N=100, Seed[c]=**13471**)		✓	✓	✓		✓	
SJC28	MIdat_SJC28 (N=100, Seed[c]=**14594**)		✓	✓	✓		✓	
TJC68	MIdat_TJC68 (N=100, Seed[c]=**11235**)	✓						
SJC66	MIdat_SJC66 (N=100, Seed[c]=**12358**)	✓						
CRP (mg/dL)	MIdat_CRP (N=100, Seed[c]=**15617**)	✓	✓		✓			
GDPTRA (cm)	MIdat_GDPTRA (N=100, Seed[c]=**15627**)	✓	✓	✓	✓		✓	
GDEV(cm)	MIdat_GDEV (N=100, Seed[c]=**21347**)	✓		✓	✓			
HAQ-DI	MIdat_HAQ (N=100, Seed[c]=**16730**)	✓				✓		
Pain (cm)	MIdat_PAIN (N=100, Seed[c]=**18976**)	✓						
ESR	MIdat_PAIN (N=100, Seed[c]=**18986**)						✓	

[a]Ancillary variables: treatment group, randomization strata factors (continuous baseline DAS28-CRP and randomization stratification levels (country, csDMARDs usage at baseline (yes or no), and Screening ACPA level (high ≥400 U/mL; low <400 U/mL)

[b] For composite continuous endpoints, the imputation will be performed on each component with missing values and then the continuous scale of the composite will be constructed based on imputed component result, then change from baseline in the endpoint is calculated.

[c]The starting seed for FCS regression MI is used to generate a series of imputation seeds using the algorithm: *INT((2**31-2)*RANUNI(starting seed))*, where each imputation seed will be used for a single imputation. To account for the possibility that some imputations may fail to complete due to out-of-range issues, 100+ initial imputation seeds will be prepared, and the first 100 successful imputations will be used for analysis.

[d]These endpoints are between Week 0 to Week 12 visit. Additional MI datasets will be created for Supplemantary estimand using all observed and impute the missing data

[e]These endpoints are between Week 0 to Week 24 visit. Additional MI datasets will be created for Supplemantary estimand using all observed and impute the missing data

[f]MIdat_JPS (N=100, Seed 16730[c])

6.15. Appendix 15: Description of Statistical Models

1. Analysis of Covariance (ANCOVA)

For most continuous endpoints where any portion of the missing data is imputed using MI (see Appendix 14), treatment comparisons will be performed using an ANCOVA model on each of the imputation sets. The estimate of the mean change from baseline is the average of the mean change taken over all the MI data sets. The estimate of the variance of the mean change from baseline is the weighted sum of the average within-imputation variance and the between-imputation variance. The confidence interval for the mean change from baseline uses critical values from the t-distribution. The treatment difference between each combination therapy group versus the monotherapy group will be tested for each imputation dataset and then the analysis results across all imputation datasets will be combined by Rubin's rule (Rubin 1987). The treatment difference in the change from baseline is estimated by the average of the treatment differences over the MI data sets. The estimate of the variance of the treatment difference in the change from baseline is the weighted sum of the average within-imputation variance and the between-imputation variance, under the assumption of homogeneity of variance between intervention groups for performing ANCOVA within each imputation dataset. The confidence interval is based on the critical values from the t-distribution.

The ANCOVA model will be based on the original scale and will include intervention group, baseline score, and randomization stratification factors as the explanatory factors. The treatment difference between a combination therapy group and the monotherapy group will be estimated by the difference in the least squares means (LSmeans). The 90% confidence interval (CI) for the LS means, and differences in LSmeans and p-values will be calculated based on ANCOVA. In addition, 95% confidence interval (CI) for the LS means, and differences in LSmeans and p-values will be reported based on ANCOVA.

2. Logistic Regression

For binary analyses, a logistic regression model will be used to calculate odds ratios their corresponding 90% or 95% CI. If there are any remaining missing data after the ICE rule is applied, then a non-response imputation will be applied to the binary endpoint. The logistic model will include treatment group and randomization strata factors (continuous baseline DAS28-CRP, country, csDMARDs usage at baseline (yes or no), and Screening ACPA level (high ≥400 U/mL; low <400 U/mL)) where appropriate.

7. REFERENCES

Aletaha D (2006), Smolen JS. The definition and measurement of disease modification in inflammatory rheumatic diseases. Rheum Dis Clin North Am. 2006;32(1):9-44, vii

Anderson JK (2011), Zimmerman L, Caplan L, Michaud K. Measures of rheumatoid arthritis disease activity: Patient (PtGA) and Provider (PrGA) Global Assessment of Disease Activity, Disease Activity Score (DAS) and Disease Activity Score With 28-Joint Counts (DAS28), Simplified Disease Activity Index (SDAI), Cl. Arthritis Care Res. 2011;63(S11): S14–36

EMA, 2015. Guideline on adjustment for baseline covariates in clinical trials, London: Committee for Medicinal Products for Human Use.

Felson DT (1993), Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum. 1993;36(6):729-740

Felson DT (1995), Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum. 1995;38(6):727-735

Felson DT (2011), Smolen JS, Wells G, et al. American College of Rheumatology; European League Against Rheumatism. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum. 2011;63(3):573-586

Fries JF (1980), Spitz P, Kraines RG, Holman HR. Measurement of patient outcomes in arthritis. Arthritis Rheum. 1980;23:137-145

Ge, Murray, Meyer, et al. Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences. Drug Information Journal 2011;45:481-93

Han, B., Yu, M. & McEntegart, D., 2013. Weighted re-randomization tests for minimization with unbalanced allocation. Pharmaceutical Statistics, 12(5), pp. 243-253.

Hawley DJ (1992), Wolfe F. Sensitivity to change in the Health Assessment Questionnaire (HAQ-DI) and other clinical and health status measures in rheumatoid arthritis. Arthritis Care Res. 1992;5(3):130-136

Prevoo ML (1995), van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995;38(1):44–48

Keystone E, Genovese MC, Klareskog L, et al. Golimumab in patients with active rheumatoid arthritis despite methotrexate therapy: 52-week results of the GO-FORWARD study. Annals of the Rheumatic Diseases. 2010;69:1129-1135.

Ratitch, Bohdana et al. Combining analysis results from multiply imputed cartegorical data. PharmaSUG 2013-Paper SP03.

Rubin DB. Multiple Imputations for Nonresponse in Surveys. New York: John Wiley & Sons. 1987.

van Riel PL (2000), van Gestel AM, Scott DL. EULAR Handbook of Clinical Assessments in Rheumatoid Arthritis.Alphen Aan Den Rijn, The Netherlands: Van Zuiden Communications B.V.; 2000:40.

Weinblatt ME, Fleischmann R, Huizinga TW, et al. Efficacy and safety of certolizumab pegol in a broad population of patients with active rheumatoid arthritis: results from the REALISTIC phase IIIb study. Rheumatology (Oxford). 2012;51(12):2204-2214.

Yeonhee K, Seunghyun W. Adjusted proportion difference and confidence interval in stratified randomized trials. PharmaSUG 2013 – Paper SP04.