Janssen Research & Development

Statistical Analysis Plan

A Multicenter, Randomized, Double-blind, Parallel-group, Placebo-controlled, Proof of concept Study Evaluating the Efficacy and Safety of Nipocalimab Administered Intravenously in Participants with Active Rheumatoid Arthritis Despite Standard Therapy

Protocol JNJ-80202135ARA2001; Phase 2a
AMENDMENT 1
JNJ-80202135 (Nipocalimab)

Status:	Approved
Date:	26 July 2022
Prepared by:	Janssen Research & Development, LLC; Janssen Research & Development, a division of Janssen Pharmaceutica NV
Document No.:	EDMS-RIM-554178, 3.0

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

TABLE OF CONTENTS

VERSION HISTORY

SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	8-November-2021	Not Applicable	Initial release
1.1		Section 5.7.6 Updated to include additional biomarker assessments (M-DP scores)	SAP is updated to align with Protocol Amendment 3 (January 2022)
		Section 5.6 Safety updated to include additional AEs of clinical interest, and changes in markedly Abnormal Laboratory Values in Appendix 9	updated Program Safety Analysis Plan (PSAP) for Nipocalimab (Feb 2022)
		Section 5.1, 5.3.3, 5.4.3 and 5.5.3 • For primary endpoint, Change from baseline in DAS28 (CRP) • Secondary endpoint, change from baseline in HAQ-DI • Tertiary endpoint, change from baseline in DAS28 (ESR) The analysis of endpoints will be performed using an ANCOVA model. If missing data is not a problem, otherwise missing data will be handled through Mixed Effect Repeated Measures (MMRM)	Changed main analysis for continuous endpoints to ANCOVA if there is no missing data problem
		Section 5.1, 5.4.3 and 5.5.3 changed analysis method of binary endpoints from GLMM model to CMH test	GLMM convergence problem
		Throughout the SAP, a non-content related wording or formatting changes were made	To enhance clarity

1. INTRODUCTION

This Statistical Analysis Plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for all planned analyses of efficacy, safety, tolerability, pharmacokinetics (PK), pharmacodynamics (PD) and immunogenicity of Nipocalimab. This SAP incorporates all analyses planned through the Week 18 final database lock (DBL) for the JNJ-80202135ARA2001 study.

1.1. Objectives

Primary Objective

- To evaluate the efficacy of nipocalimab vs placebo in participants with moderate to severe active RA

Secondary Objectives

- To evaluate the safety and tolerability of nipocalimab vs placebo in participants with moderate to severe active RA
- To evaluate the PK and immunogenicity of IV nipocalimab in participants with moderate to severe active RA

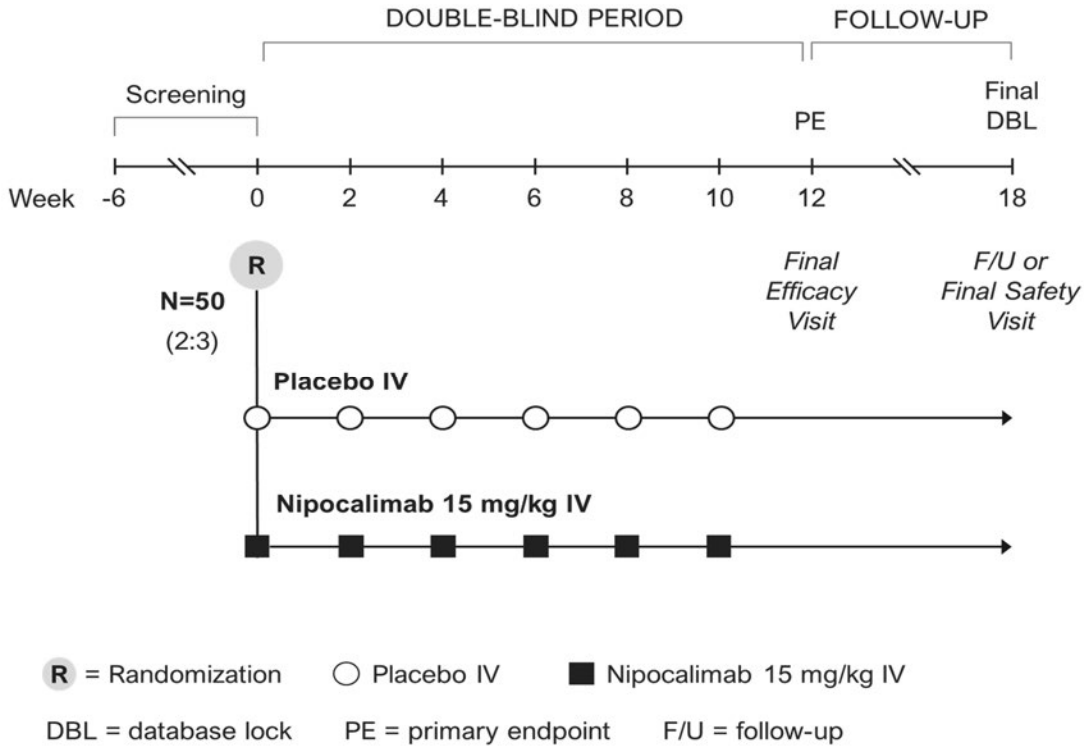
Exploratory Objectives

- To evaluate the impact of nipocalimab vs placebo on disease activity in participants with moderate to severe active RA
- To evaluate the impact of nipocalimab vs placebo on PROs in participants with moderate to severe active RA
- To evaluate the effect of nipocalimab vs placebo on PD and disease biomarkers in participants with moderate to severe active RA
- To evaluate changes in serological profiles in response to nipocalimab vs placebo treatment in participants with moderate to severe active RA
- CCI [redacted]

1.2. Study Design

This is a Phase 2a, randomized, double-blind, placebo-controlled, parallel-group, proof-of-concept, multicenter clinical study designed to evaluate the efficacy and safety of Nipocalimab 15 mg/kg IV every 2 weeks (q2w) in participants with moderate to severe active RA despite standard therapy, including anti-TNF agents. An overview of the study design is provided in Figure 1.

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



The total duration of the study is up to 24 weeks, consisting of 3 study periods: a ≤6-week screening period (rescreening is permitted once per subject), a 12-week double-blind treatment period, and a 6-week safety follow-up period (8 weeks after the last study intervention administration).

Through 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. Subjects should not initiate any new treatment for RA through Week 18.

Table 1: Permitted Concomitant Medications for RA, and the Maximum Allowed Doses at Study

Permitted Concomitant Medications for RA[a]	**Maximum Allowable Dosage**
Conventional synthetic DMARDs	
Methotrexate (MTX)	25 mg/week
Sulfasalazine (SSZ)	1000-2000 mg/day
Hydroxychloroquine (HCQ) or CQ	200-400 mg/day
Leflunomide (LEF)	10-20 mg/day
Oral corticosteroids	Equivalent to 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

[a]Permitted concomitant medication are not supplied by the Sponsor

Efficacy data will be collected through Week 12 and safety data will be collected through Week 18. All assessments will be performed according to the schedule of activities (SoA) detailed in Section 1.3 in the protocol. Subjects who discontinue study agent administration prior to Week 12 will continue to be evaluated for efficacy and safety through Week 18 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), and vital signs.

An interim analysis (IA) is planned when 60% (30 participants) of the planned number of participants complete the Week 12 assessment.

A database lock (DBL) for primary efficacy analysis is planned at Week 12, when 100% of participants complete their Week 12 assessments or terminate study prior to Week 12 – refer to as Week 12 DBL hereafter. A final DBL is planned at Week 18, when all subjects complete the study or have terminated the study prior to Week 18- refer to as final safety DBL. The end of the study is defined as the last follow up assessment (8 weeks after the last infusion of study intervention) for the last participant. Additional DBLs may be added as necessary and will be specified in the SAP prior to the additional DBLs.

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

1.2.1. Randomization and Blinding

Randomization at Week 0 will be used to minimize bias in the assignment of subjects to treatment groups and to increase the likelihood that known and unknown subject attributes (eg, demographic and baseline characteristics) are evenly balanced across treatment groups. The maintenance blinded treatment will be used to reduce potential bias during data collection and evaluation of clinical endpoints. At the Week 12 DBL, the data will be unblinded for analysis to some Sponsor personnel while subjects are still participating in the study. Identification of Sponsor personnel who will have access to the unblinded subject-level-data will be documented prior to unblinding the data. Investigative study sites and subjects will remain blinded to initial treatment assignment until after the final database is locked at Week 18.

1.2.1.1. 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 subject 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 subject details to uniquely identify that subject. Based on a computer-generated randomization

schedule prepared before the study under the supervision of the Sponsor, the IWRS will then assign a unique treatment code, which will dictate the treatment assignment and matching study agent kit for that subject.

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

- Baseline MTX use (Not currently taking Methotrexate, >0 to <12.5 mg/week, ≥12.5 mg/week),
- Baseline swollen and tender joint counts level (this factor is combined baseline tender and swollen 28 joint counts using this formula [0.56*√(TJC28) + 0.28*√(SJC28)], which is part of DAS28(CRP) calculation), can be categorised as four levels:
  - <2.690 =1
  - ≥2.690 to <3.160 =2
  - ≥3.160 to <3.656 =3
  - ≥3.656 =4
- Anti-TNF IR/intolerance (Yes, No)

1.2.1.2. Blinding

To maintain the study blind, the study intervention (placebo and nipocalimab 15 mg/kg) container will have a label containing the study name, study intervention number, and reference number. The label will not identify the study intervention in the container. However, if it is necessary for a participant's safety, the study blind may be broken, and the identity of the study intervention ascertained. The study intervention number will be entered in the electronic case report form (eCRF) when the study intervention is administered. The study interventions will be identical in appearance and will be packaged in identical containers.

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-nipocalimab antibodies, intervention allocation, or other specific laboratory data, [eg, IgG levels, serum albumin levels]) will be handled with special care (eg, masked) to ensure that the integrity of the blind is maintained and the potential for bias is minimize. 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 DBL and unblinding.

Under normal circumstances, the blind should not be broken until all participants have completed the study and the database is finalized (Week 18). 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 IWRS. While the responsibility to break the intervention code in emergency situations resides solely with the investigator, 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 and reason for the unblinding must be documented in the appropriate section of the eCRF and in the source document. The documentation received from the 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 will not be eligible to receive further study intervention but should complete evaluations specified in the Schedule of Activities for participants who discontinue study intervention in the protocol Section 1.3.

Additionally, a given participant's treatment assignment may be unblinded to the sponsor, the Independent Ethics Committee/Institutional Review Board (IEC/IRB), and site personnel to fulfill regulatory reporting requirements for suspected unexpected serious adverse reactions (SUSARs).

A separate code break procedure will be available for use by the sponsor's Global Medical Safety group to allow for unblinding of individual participants to comply with specific requests from regulatory or health authorities.

An independent Data Monitoring Committee (DMC) will monitor the safety of the study in an unblinded fashion on a regular basis and whenever deemed necessary. The contents of the unblinded data to which the DMC and Statistical Support Group (SSG) have access should not be divulged, in any way, to members of the study team or to any members of the Sponsor Committee, unless specifically requested by the Sponsor Committee Chairperson, until the study has completed, and clinical database is locked. Additional details related to DMC are provided in Section 5.8.1

2. STATISTICAL HYPOTHESES

The primary hypothesis is that treatment with nipocalimab 15 mg/kg intravenously (IV) for every 2 weeks (q2w) is superior to placebo in participants with moderate to severe active rheumatoid arthritis (RA) as assessed by the mean change from baseline in the Disease Activity Index Score 28 using C-reactive Protein (DAS28-CRP) at Week 12.

3. SAMPLE SIZE DETERMINATION

CCI

[redacted]

CCI

Table 2: Statistical Power for Treatment Difference in Change from Baseline in DAS28 (CRP) at Week 12

CCI

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

CCI

5. STATISTICAL ANALYSES

The statistical analyses will include all analyses from Week 0 through Week 18 (the final safety follow-up) visit.

5.1. General Considerations

Unless otherwise specified, efficacy data summaries will be provided by intervention group for the FAS. Data primarily will be summarized using descriptive statistics. Continuous variables will be summarized using the number of observations, mean, standard deviation (SD), median, interquartile range (IQ), minimum and maximum, as appropriate. Categorical values will be summarized using the number of observations and percentages as appropriate. Graphical data displays (i.e, line plots) and participant listings may also be used to summarize/present the data.

Continuous Efficacy Endpoints

For all continuous endpoints of change from baseline score by scheduled visits, treatment comparison will be performed using an Analysis of Covariance (ANCOVA) model, if missing data is not a problem. After intercurrent event rules are applied, the remaining missing data will not be imputed for analysis that is using an ANCOVA model assuming Missing Complete at Random (MCAR) assumption. Otherwise, any remaining missing data after ICEs rules applied, it will be accounted for under MAR assumptions using Mixed-Effect Model Repeated Measures (MMRM) model (for more details see Section 6.13).

The ANCOVA model will include treatment group, baseline score, and stratification factor (baseline MTX use, (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week) as the explanatory factors. The treatment difference between a nipocalimab group and the placebo group will be estimated by the difference in the least squares means (LSmeans). The 95% confidence interval (CI) for the differences in LSmeans and p-values will be calculated based on the ANCOVA model.

Binary Efficacy Endpoints

Binary response efficacy endpoints, treatment comparisons will be performed using a Cochran-Mantel-Haenszel (CMH) chi-square stratified by stratification factor (baseline MTX use, (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week). After intercurrent event rules are applied, subjects with missing response status will be considered a non-responder (i.e. non-responder imputation) assuming Missing Complete at Random (MCAR) assumption. The magnitude of the treatment difference will be estimated by the difference in response rates between the nipocalimab and placebo groups with a 95% confidence interval (CI) calculated based on Wald statistics. The Mantel Fleiss criterion will be used to determine the appropriateness of using the CMH test. If the Mantel Fleiss criterion is not satisfied the Fisher's exact test will be used instead of the CMH test to compare the two intervention groups.

Unless otherwise specified, all statistical tests will be performed at a 2-sided level of significance α=0.05 and the p-values will be considered nominal.

5.1.1. Baseline Windows

In general, unless otherwise stated, baseline is defined as the last observation prior to the start of the first study intervention administration.

5.1.2. 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).

5.1.3. Pooling Algorithm for Analysis Centers

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

5.1.4. 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 at all visits will be assigned a day relative to this date.

Note, all post-baseline visits through Week 12 will have a visit window of ± 3 days from scheduled visit day. All the visits after Week 12 through Week 18 will have a visit window of ± 7 days from scheduled visit day.

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. Participant Dispositions

Screened participants and reason for screen failures will be summarized overall.

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

- Participants randomized
- Participants who received study intervention
- Participants who completed the study
- Participants who discontinued study intervention and reason for discontinuation
- Participants who terminated study prematurely and reason for termination
- Participants who experience an intercurrent events (defined under Section 5.3.2.1) and reasons

The above categories will include summaries through Week 18 (including any final safety visit that occurs prior to the Week 18 DBL).

Listings 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 experience an intercurrent events

5.3. Primary Endpoint Analysis

The primary endpoint is the change from baseline in DAS28 (CRP) at Week 12.

5.3.1. Definition of Endpoint

5.3.1.1. Change from baseline in DAS28(CRP)

Change from baseline in DAS28 (CRP) 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, PIP1, 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:

DAS28 (CRP) = 0.56*SQRT(TJC28) +0.28*SQRT(SJC28) + 0.36*Ln($CRP_{mg/L}$+1) + 0.014*$PtGA_{mm}$ + 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 6.12.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.12.1, 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 6.12.2, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.12.2, 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 (CRP) value, the natural logarithm of ($CRP_{mg/L}$ + 1) is used. LLOQ rule specified in Section 6.12.3, will be applied to values <LLOQ.

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 (CRP) value.

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.

5.3.2. Primary Endpoint Estimands

Primary Trial Objective: To evaluate the efficacy of Nipocalimab plus standard of care therapy vs. placebo plus standard of care therapy in participants with moderate to active RA.

Estimand Scientific Question of Interest: What level of mean change in DAS28 (CRP) is considered to have benefited from nipocalimab 15mg/kg vs placebo for the pre-specified duration (12 weeks), administered together with the protocol allowed background standard of care medication?

5.3.2.1. Primary Estimand

Primary Estimand, based on the primary objective and scientific question of interest above, is defined by the following 5 components:

a. **Study intervention:**
   - Nipocalimab (experimental treatment/intervention) 15 mg/kg IV q2w in addition to standard of care therapy
   - Placebo IV q2w in addition to standard of care therapy

b. **Population:** Participants between the ages of 18 and 75, inclusive, with moderate to severe active RA who have an inadequate response anti-TNF agent despite standard therapy

c. **Variable (Endpoint)**: Change from baseline in DAS28 (CRP) at Week 12. A participant who initiates or adjusts medication and experience an intercurrent event (ICEs) in categories 1, 2 and 3 (defined below) prior to the Week 12 visit will be considered a treatment failure and will be assigned a zero change from baseline in the DAS28 (CRP) at Week 12 regardless of the observed change. For a subject who experience an ICE in category 4, the observed change from baseline in the DAS28 (CRP) at Week 12 will not be used, it will be assumed missing.

d. **Summary measure (Population-level summary)**: Difference in mean change in DAS28 (CRP) as defined in the variable between nipocalimab group and placebo group

e. **Intercurrent events (ICEs) and their corresponding strategies:**

ICEs	**Strategy for Addressing ICEs and Its Description**
1. Initiated protocol-prohibited medications/therapies for RA	**Composite Strategy:** A participant with this intercurrent event is considered as to have no change from baseline after this event; the occurrence of this ICE has been captured in the variable definition.
2. 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	
3. Discontinued study intervention due to any reasons, including COVID-19 infection but excluding other COVID-19 reasons	
4. Treatment discontinuation due to study conduct affected by COVID-19 reasons i.e site closure/restricted access	**Hypothetical Strategy:** A participant with this intercurrent event is addressed with a hypothetical strategy, as if the intercurrent event would not have occurred. * *Note: Data after this ICE will not be used in the analysis.

For participants experiencing multiple ICEs simultaneously, ICEs in categories 1, 2 and 3 will take priority over ICE in category 4. If a participant experiences more than one ICE in ICEs 1-3 then the overall strategy will be based on the first ICE to occur in ICEs 1-3 categories.

5.3.2.2. Treatment Policy Estimand (Supplementary Estimand)

Estimand Scientific Question of Interest: What level of mean change in DAS28 (CRP) is considered to have benefited from nipocalimab vs placebo for the pre-specified duration (12 weeks), administered together with the protocol allowed background standard therapy, regardless of occurrence of any intercurrent events?

Treatment policy estimand is defined to support the primary estimand.

Supplementary estimand are defined as

a. **Study intervention:** Same as primary estimand
b. **Population:** Same as primary estimand
c. **Variable**: Change from baseline in DAS28 (CRP) at Week 12, participants with any ICEs in categories 1-4 prior to Week 12 will still use their observed data

d. **Summary measure (Population-level summary)**: Difference in mean change in DAS28 (CRP) score as defined in the variable between nipocalimab group and placebo group

e. **ICEs and their corresponding strategies:**

The definition of the four categories of ICEs are the same as the Primary Estimand.

ICE strategy Difference from Primary Estimand: In the primary estimand, ICEs are treated with composite variable and hypothetical, that means subjects with any of the ICEs will be considered as either no change from baseline in DAS28 (CRP) or data is not used at Week 12. The treatment policy strategy is to use all observed data collected for the endpoint. The occurrence of the intercurrent event is irrelevant and the value for the variable of interest is used regardless of whether or not the intercurrent event occurs.

5.3.3. Analysis Methods

5.3.3.1. Analysis (estimators) for the primary estimand

5.3.3.1.1. Main (primary) estimator

The primary endpoint, defined in Section 5.3.1, will be analyzed based on the Primary Estimand (Section 5.3.2.1). After accounting for the ICEs for the primary estimand, participants who are missing change from baseline in DAS28 (CRP) at Week 12 will not be imputed, it will remain missing for analysis that is using an ANCOVA model under MCAR assumption, if missing data is not a problem. Otherwise remaining missing data will be handled through MMRM model (for more details see Section 6.13)

In the primary analysis, data from all participants in the FAS (Section 4) will be analyzed according to the randomized study intervention regardless of the study intervention they actually received. The treatment difference between nipocalimab group and placebo group at Week 12 will be tested using an ANCOVA model. The ANCOVA model will include treatment group, DAS28 (CRP) baseline score, and stratification factor (baseline MTX use, (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week) as the explanatory factors. The treatment difference between a nipocalimab group and the placebo group will be estimated by the difference in the least squares means (LSmeans). The 95% confidence interval (CI) for the differences in LSmeans and p-values will be calculated based on the ANCOVA model.

5.3.3.1.2. Sensitivity Analyses

To evaluate the robustness of dynamic randomization algorithm, a re-randomization test will be performed targeting the primary estimand as a sensitivity analysis. In this analysis, remaining missing data after ICE rules are applied will remain as missing.

Participants will be re-randomization to new treatment assignment by changing the seed (#) to generate a new sequence random number, while keeping the same biased coin probability and weight of each factor. The same analysis as the primary analysis will be performed to obtain a test statistic, this will be repeated over 5000 times. Details of the re-randomization test are provided in Section 6.14.

5.3.3.1.3. Subgroups

To evaluate the consistency of the primary endpoint (change from baseline in DAS28 (CRP) at Week12), across subgroups defined by baseline demographics and disease characteristics, and prior and baseline medication specified in Section 5.7.8. subgroup analyses may be performed when the number of subject in the subgroup permits (at least ≥5 subjects for each of the intervention group within a subgroup), If the number of subject in a stratum are too small (eg. <5), strata may be pooled. Note that, for subgroup analyses, the analysis sets are the individual subgroups of the FAS.

For subgroup analysis, the primary estimand (see Section 5.3.2.1) will be used. After accounting for intercurrent events specified in Section 5.3.2.1, the remaining missing data will not be imputed. The treatment difference in change from baseline in DAS28 (CRP) at Week 12 by each of subgroup factors listed in Section 5.7.8, will be performed using an ANCOVA model. The ANCOVA model include treatment group, baseline DAS28 (CRP) score, stratification factor (baseline MTX use) as explanatory variables. A forest plot of the LS means, LS mean of the difference between treatment groups in change from baseline in DAS28 (CRP) at Week 12 and the associated 95% CI estimated by the ANCOVA model will be plotted for each of the subgroup factors defined Section 5.7.8, In addition, the p-values for interaction of the treatment groups and the subgroups will also be provided when a subgroup has at least 2 categories.

5.3.3.2. Analysis (estimators) for the Supplementary Estimand

Statistical analysis for both primary Estimand and supplementary estimand (treatment policy) will be the same. The only difference is the strategy used to handle the ICEs.

- For the primary Estimand: the occurrence of ICE 1-3, subjects will be considered as not having no change from baseline in DAS28 (CRP) at or after the ICE occurrence. In the event that subjects only had ICE 4, subjects will be treated as if the ICE would not have occurred. Therefore, the data that collected after ICE 4 are not used in the analysis. If subjects experience multiple ICEs, then an ICE 1-3 will supersede ICE 4
- For the supplementary Estimand (Treatment Policy): Observed value for primary endpoint at Week 12 will be used regardless of intercurrent events occurring, that is the subjects experiencing an ICE will not be treated differently from subjects without an ICE.

Therefore, in the analyses based on the supplementary estimand, the same analysis as described for the primary Estimand in Section 5.3.3.1.1, the analysis will be performed using observed data, missing data will not be imputed.

5.3.3.3. Summary of Analyses Related to Primary Endpoint

Table 3 below provides an overview on all the analyses related to the primary endpoint of change from baseline in DAS28 (CRP) at Week 12, the estimands, the data handling rules to be used, and the analysis methods and summary statistics.

Table 3: Primary Endpoint Analysis Summary

Continuous endpoints	Missing data	Additional notes

	Primary Estimand (Composite) (Section 5.3.2.1)	After accounting the ICEs 1-3 specified in specified in Section 5.3.2.1, any remaining missing data at Week 12 will not be imputed for ANCOVA model. If missing data is a problem, any remaining missing data after ICE rules are applied will be handled through MMRM model.	• Summarized descriptively • ANCOVA/MMRM model ○ LS mean (95%CI) ○ Treatment difference in LS mean (95% CI) ○ P-value of comparing LS mean
	Primary Estimand (Hypothetical) (Section 5.3.2.1)	ICE 4 related to COVID-19 reasons i.e site closure/restricted access, observed data at Week 12 for subjects with an ICE 4 will be set to missing	
	Supplementary Estimand (Treatment Policy) (Section 5.3.2.2)	Observed data at Week 12 will be used regardless the occurrence of ICEs and any remaining missing data will not be imputed	
	Sensitivity Analysis (Section 5.3.3.1.2) Primary Estimand	After accounting the ICEs 1-3 specified in specified in Section 5.3.2.1, any remaining missing data at Week 12 will not be imputed for ANCOVA model. If missing data is a problem, any remaining missing data after ICE rules are applied will be handled through MMRM model.	
	Subgroup Analysis (Section 5.3.3.1.3) Primary Estimand	After accounting the ICEs 1-3 specified in specified in Section 5.3.2.1, any remaining missing data at Week 12 will not be imputed.	○ ANCOVA model ○ LS mean (95% CI) ○ Treatment difference in LS mean (95% CI) ○ P-value from ANCOVA for the interaction of treatment group and subgroup variable ○ Graphical: forest plots

5.4. Secondary Efficacy Endpoints Analysis

Objective: To characterize additional assessments of efficacy for nipocalimab vs placebo in subject with moderate to active RA who have an inadequate response to anti-TNF agent despite standard therapy

There are seven secondary endpoints at Week 12 to address the objective:

- Proportion of participants achieving an ACR20 responses at Week 12
- Proportion of participants achieving an ACR50 responses at Week 12
- Proportion of participants achieving an ACR70 responses at Week 12
- Proportion of participants achieving an ACR90 responses at Week 12
- Proportion of participants achieving DAS28-CRP remission at Week 12
- Proportion of participants achieving DAS28-CRP LDA at Week 12
- Change from baseline in HAQ-DI score at Week 12

5.4.1. Definition of Endpoint(s)

5.4.1.1. 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 (Felson 1995) is defined as:

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

AND

≥20% improvement from baseline in 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

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

1. 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 6.12.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 68-joint set, joint count adjustment rules described in Section 6.12.1, will be applied in determining the ultimate count of tender joints
2. 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 6.12.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 66-joint set, joint count adjustment rules described in Section 6.12.1, will be applied in determining the ultimate count of tender joints
3. Patient's Assessment of Pain (PAIN): a measure from 0 (no pain) to 10 cm (the worst possible pain) on a VAS scale
4. Patient's Global Assessment of Disease Activity (PtGA): a measure from 0 (very well) to 10 cm (very poor) on a VAS scale
5. 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
6. HAQ-DI: a measure of difficulty a subject may have in accomplishing tasks in 8 functional areas. For additional details, please refer to the definition of HAQ-DI, see Section 5.4.1.3

7. C-reactive protein (CRP): a lab parameter measured in mg/dL. LLOQ rule specified in Section 6.12.3, will be applied to values <LLOQ.

If a subject's baseline value for a component is zero (i.e., no disease activity as measured by that component), the subject 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.

5.4.1.2. DAS28 Remission and Low Disease Activity

Refer to Section 5.3.1.1, for the definition based on DAS28 (CRP) and change from baseline in DAS28 (CRP) score.

DAS28 remission is defined as a DAS28 (CRP) value of <2.6 at Week 12

DAS28 LDA is defined as a DAS28 (CRP) value ≥2.6 and <3.2 at Week 12

5.4.1.3. Change from Baseline in HAQ-DI Score

HAQ disability index (HAQ-DI) score is an evaluation of the functional status for a subject (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 subject 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.

5.4.2. Secondary Endpoint Estimand

5.4.2.1. Main Estimand

The following describes the attributes of the estimands for the secondary endpoint's binary endpoints (#1, 2, 3) and continuous endpoint (#4) in Section 5.4.1.

a. **Study intervention:** Same attributes as primary estimand for primary endpoint

b. **Population:** Same attributes as primary estimand for primary endpoint

c. **Variable/Endpoint and Population-level summary:** see Table 4

Table 4: Variables and Summary measure (Population-level Summary) for Each Secondary Estimand

Estimand	Variable (Endpoint)	Summary measure (Population-level summary)
Secondary Endpoint Group #1	Response binary variable, where a response is defined as the proportion of participants who achieve an ACR response (ACR20, ACR50, ACR70 & ACR90) response at Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of participants who achieved a response as defined in the variable between nipocalimab and placebo.
Secondary Endpoint #2	Response binary variable, where a response is proportion of participants who achieve a DAS28 (CRP) binary endpoint (remission or LDA) response at Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of participants who achieved a binary response as defined in the variable between nipocalimab and placebo.
Secondary Endpoint Group #3	Continuous variable, defined as change from baseline in HAQ-DI score at Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change in HAQ-DI score as defined in the variable between the nipocalimab and placebo group

d. **Intercurrent events (ICEs) and their corresponding strategies:**

ICEs	Analysis Strategy for Addressing Intercurrent Events
1. Initiated protocol-prohibited medications/therapies for RA	**Composite Strategy:** A subject with this intercurrent event is considered as not achieving the binary endpoint and having no change from baseline for the continuous endpoint after this event, the occurrence of this intercurrent event is captured in the variable definition.
2. Initiation or increased the dose of csDMARDs (such as MTX, SSZ, HCQ, CQ or leflunomide) or oral corticosteroid therapy above the baseline dose for RA	
3. Discontinued study intervention infusion due to any reasons including COVID-19 infection but excluding other COVID-19 reasons	

4. Treatment discontinuation due to study conduct affected by COVID-19 reasons i,e site closure/restricted access	**Hypothetical Strategy:** A participant with this ICE is considered to have missing data at and after the event occurred* *Note: Data after this ICE will not be used in the analysis

For participants experiencing multiple ICEs simultaneously, ICEs in categories 1, 2 and 3 will take priority over ICE in category 4. If a participant experiences more than one ICE in ICEs 1-3 then the overall strategy will be based on the first ICE to occur in ICEs 1-3 categories.

5.4.2.2. Supplementary Estimand (Treatment Policy)

For each of the secondary endpoints' supplementary estimand, it shares the same population, study intervention, ICEs, variable, and summary measures as the main estimand, see Section 5.4.2.1. The only difference is the ICE strategy, which will be handled by Treatment Policy Strategy. Under the treatment policy strategy, observed data collected at or after intercurrent events will still be used in the analysis.

5.4.3. Analysis Methods

5.4.3.1. Main (Secondary) estimator for main estimands

All secondary endpoints will be analyzed at Week 12 DBL based on the FAS (Section 4), that is all randomized subjects 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 treatment subjects actually received.

5.4.3.1.1. Binary Endpoints

In the binary secondary endpoint responses (ACR20, ACR50, ACR70, ACR90, DAS28 (CRP) remission and DAS28 (CRP) LDA) at Week 12 analysis will be based on main estimands defined in Table 4. After accounting for intercurrent events specified in Section 5.4.2.1, any remaining missing response status at Week 12 will be considered as non-responders.

The treatment comparison between a nipocalimab group and the placebo group will be performed using a CMH chi-square test stratified by stratification factor (baseline MTX use (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week) at Week 12. The magnitude of the treatment difference will be estimated by the difference in binary response rates between the nipocalimab and placebo groups with a 95% CI calculated based on Wald statistics. If the Mantel Fleiss criterion is not satisfied the Fisher's exact test will be used instead of the CMH test to compare the two treatment groups.

5.4.3.1.2. Continuous Endpoint

In the continuous secondary endpoint, change from baseline in HAQ-DI will be analyzed based on the main estimands defined in Table 4. After accounting for the ICEs for the main estimand, participants who are missing change from baseline in HAQ-DI at Week 12 will not be imputed, it will remain missing. The treatment difference between nipocalimab group and placebo group at

Week 12 will be tested using an ANCOVA model, if missing data is not a problem, otherwise remaining missing data will be handled through MMRM model (for more details see Section 6.13).

The ANCOVA model will include treatment group, HAQ-DI baseline score, and stratification factor (baseline MTX use, (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week) as the explanatory factors. The treatment difference between a nipocalimab group and the placebo group will be estimated by the difference in the least squares means (LSmeans). The 95% confidence interval (CI) for the differences in LSmeans and p-values will be calculated based on the ANCOVA model.

5.4.3.1.3. Subgroups

A subgroup analyses comparing baseline csDMARDs use (yes, no) for ACR (20,50,70, and 90) responses, DAS28 (CRP) remission and DAS28 (CRP) LDA at Week 12 may be performed when the number of subjects in the subgroup permits (at least ≥5 subjects for each of the intervention group within a subgroup), If the number of subjects in a stratum are too small (eg. <5), strata may be pooled.

The treatment difference in proportion of subjects achieving an ACR20, ACR50, ACR70 and ACR90 response, DAS28 (CRP) remission and DAS28 (CRP) LDA at Week 12 by baseline csDMARDs (yes, no) will be performed using a logistics regression model including treatment group, stratification factor (baseline MTX use, (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week) and an interaction between treatment group and csDMARDs use (yes, no). For this analysis, the main estimand defined in Table 4 will be used. After accounting for intercurrent events specified in Section 5.4.2.1, any remaining missing response status at Week 12 will be considered as non-responders. A forest plot will be produced for each of the responses ACR20, ACR50, ACR70, ACR90, DAS28 (CRP) remission and DAS28 (CRP) LDA. Odds ratios and the corresponding 95% CIs will also be provided for each of subgroups. In addition, the p-values for interaction of the treatment groups and the subgroup will also be provided when a subgroup has at least 2 categories.

5.4.3.2. Analysis (estimators) for the Supplementary Estimand

In the analyses based on the supplementary estimand, observed data will used and any missing data will not be imputed. Analysis methods based on the supplementary estimand (treatment policy) will be the same as that based on the main estimands defined in Section 5.4.3.1.

5.4.3.3. Summary of Analyses Related to Secondary Endpoints

Table 5 below provides an overview on all the analyses related to the secondary endpoint binary response and continuous endpoint at Week 12, the estimands, the data handling rules to be used, and the analysis methods and summary statistics.

Table 5: Secondary Endpoint Analysis Summary			
Binary endpoints		**Missing data**	**Additional notes**
	Main Estimand ((Section 5.4.2.1)	After accounting the ICEs specified in Section 5.4.2.1, any remaining missing data at Week 12 will be imputed as non-responder	• Summarized descriptively • Response • Treatment difference 95% CI for treatment comparison • P-value from CMH test
	Supplementary Estimand (Treatment Policy) (Section 5.4.2.2)	Observed value at Week 12 will be used regardless of intercurrent events occurring. Remaining missing data at Week 12 will not be imputed	
	Subgroup analysis (Section 5.4.3.1.1) (main estimand)	After accounting the ICEs specified in Section 5.4.2.1, any remaining missing data at Week 12 will be imputed as non-responder	• Odds ratio and 95% CI for treatment comparison • P-value from logistic regression for the interaction of treatment group and subgroup variable • Graphical: forest plots
Continuous endpoints			
	Main Estimand ((Section 5.4.2.1)	After accounting the ICEs specified in Section 5.4.2.1, any remaining missing data at Week 12 will not be imputed for the ANCOVA model. If missing data is a problem, any remaining missing data after ICE rules are applied will be handled through MMRM model.	• Summarized descriptively • ANCOVA/MMRM model for ○ LS mean (SD) ○ Treatment difference in LS mean (95% CI) ○ P-value of comparing LS mean
	Supplementary Estimand (Treatment Policy) (Section 5.4.2.2)	Observed data at Week 12 will be used regardless of intercurrent events occurring. Remaining missing data at Week 12 will not be imputed	

5.5. Tertiary/Exploratory Endpoint(s) Analysis

In addition to the primary and secondary endpoints, exploratory endpoints related to disease status over time, Patient-reported outcomes (PROs), inflammatory biomarkers, PK and Actigraph will be analyzed. This section lists these endpoints, followed by their definitions and analysis methods. These endpoints will be summarized and compared between nipocalimab and placebo groups. PK analysis is covered in Section 5.7.1.

Analyses of the following are at all applicable visits from Week 0 through Week 12

1. Proportion of participants achieving an ACR20, ACR50, ACR70, and ACR90 responses over time
2. Proportion of participants achieving DAS28 (CRP) remission over time
3. Proportion of participants achieving DAS28 (CRP) LDA over time
4. Proportion of participants achieving DAS28 (ESR) remission over time
5. Proportion of subject achieving DAS28 (ESR) LDA over time

6. Proportion of participants achieving CDAI remission over time
7. Proportion of participants achieving CDAI LDA over time
8. Proportion of participants achieving SDAI-based ACR/EULAR remission over time
9. Proportion of participants achieving Boolean-based ACR/EULAR remission over time
10. Proportion of participants achieving SDAI LDA over time
11. Proportion of participants achieving a decrease of ≥0.22 points from baseline in HAQ-DI score over time
12. Proportion of subject who are ACPA seropositive over time
13. Proportion of subject who are RF seropositive over time
14. Percent improvement from baseline in ACR components over time
15. Change from baseline in ACR components over time
16. Change from baseline in DAS28 (CRP) over time
17. Change from baseline in DAS28 (ESR) over time
18. Change from baseline in CDAI over time
19. Change from baseline in SDAI over time
20. Change from baseline in the Tender/Painful and Swollen Joint Counts over time
21. Change from baseline in duration of morning stiffness over time
22. Change from baseline in FACIT-Fatigue score over time
23. Change from baseline in Joint Pain Severity score over time
24. Change from baseline in SF-36 domain scores, PCS and MCS summary scores over time
25. Change from baseline in Serum Molecular Disease Profile (M-DP 4 and M-DP 9 analyte) over time
26. Change from baseline in ESR over time

5.5.1. Definition of Exploratory Endpoints

Definitions for DAS28 (CRP) and clinical responses based on DAS28 (CRP), ACR responses and HAQ-DI are provided in Section 5.3.1, Section 5.4.1, and Section 5.4.1, respectively.

All other definitions are listed below.

5.5.1.1. 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, PIP1, 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: is 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 6.12.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.12.1, 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 6.12.2, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.12.2, 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.

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 Week 12.

CDAI Remission

- CDAI remission is defined as a CDAI score of ≤2.8 at Week 12

5.5.1.2. 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, PIP1, 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: is 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 6.12.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.12.1, 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 6.12.2, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.12.2, 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.
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 6.12.3, will be applied to values <LLOQ.

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 Week 12.

5.5.1.3. ACR/EULAR Remission

5.5.1.3.1. SDAI-based ACR/EULAR Remission

For the definition of SDAI, refer to Section 5.5.1.2.

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

5.5.1.3.2. Boolean-Based ACR/EULAR Remission

A subject 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 5.4.1.1.

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.

5.5.1.4. 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).

5.5.1.5. Functional Assessment of Chronic Illness Therapy-Fatigue

The FACIT-Fatigue is a questionnaire that assess self -reported tiredness, weakness, and difficulty conducting usual activities due to fatigue. The questionaries consist of 13-item questions that assess subject's level of fatigue and tiredness over the last 7 days. Subject will be asked to answer each question using a 5-point Likert-type scale (0=Not at all; 1=A little bit; 2=Somewhat; 3=Quite a bit; and 4=Very Much); and accordingly, the total FACIT-Fatigue scores can range from 0 to 52, with lower score reflecting more fatigue and higher score scores reflecting less fatigue (Cella 2002; Yellen 1997).

Note that, when at least 7 questions are answered, the score can be calculated and should be adjusted by the number of available questions.

FACIT-Fatigue has been used in clinical trials of subject with RA and has demonstrated sensitivity to change in these subject (Smolen 2008; Yount 2007).

5.5.1.6. Joint Pain Severity NRS

Subject' joint pain will be assessed using a single item that asks the subject 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).

5.5.1.7. Duration of Morning Stiffness

Duration of morning stiffness is the average daily duration of morning stiffness during the previous week in minutes (0 – 1440 minutes).

5.5.1.8. 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:

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

1. TJC28: is 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 6.12.1, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.12.1, 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 6.12.2, for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 28-joint set, joint count adjustment rules described in Section 6.12.2, 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.

5.5.1.9. SF36-item Short Form Health Survey (Standard)

The SF-36 version 2 Standard health survey is a self-administered, 36-item questionnaire measuring health-related quality of life, with a recall period of the past 4 weeks. It includes 8 domains that measure physical functioning, role limitations due to physical health problems, bodily pain, general health, vitality, social functioning, role limitations due to emotional problems and mental health. Each of these 8 (domains) is scored from 0 to 100 with higher scores indicating better health. Based on the scale scores, the summary scores that reflect physical and mental health: a physical component scale (PCS) and a mental component score (MCS) will be derived. These summary scores are also scaled with higher scores indicating better health.

Note that, the SF-36 will be derived based on the algorithm and software provided by developer (Quality Metric 2011). The norm-based score will be used for data analysis and derived by using the 2009 general US population with missing value estimated based on Advanced Data Recovery with Missing Score Estimator (MSE) provided in the user manual (Quality Metric 2011).

Change from baseline in MCS/PCS score measures the change in health-related quality of life, where a positive change indicates an improvement, and a negative change indicates a worsening.

5.5.1.10. Serum Molecular Disease Profile (M-DP) Score

A molecular disease profile (M-DP) score is a composite value derived from a set of serum biomarkers (4 or 9 analytes) that are known to be dysregulated in rheumatoid arthritis patients.

Retrospective analysis of serum samples collected from previous studies in RA patients treated with advanced biologic therapies demonstrated reduction in M-DP scores from their baseline values with treatments achieving clinical efficacy endpoints at Week 12.

For calculating M-DP scores for this study, concentrations of 4 or 9 serum analytes will be measured. Changes in the serum levels of each analyte/subject at different visits will be calculated by normalizing to baseline values followed by log2 transformation of the normalized value. M-DP scores for samples collected from each patient will be calculated by taking the median of log2 transformed (fold-change/baseline) values for 4 or 9 analytes, panels used in the calculation of M-DP 4 or 9 analytes are defined as:

a. M-DP 4 analyte panel = [C-X-C motif chemokine ligand 10 (CXCL10), C-X-C motif chemokine ligand 13 (CXCL13), C-reactive protein (CRP) and Serum Amyloid A (SAA)]
b. M-DP 9 analyte panel = [M-DP 4 panel + Interlukin-6, Annexin-1, Hepcidin, Matrix Metalloproteinase-1 (MMP-1) and Matrix Metalloproteinase-3 (MMP-3)]

5.5.2. Tertiary Endpoint Estimand

5.5.2.1. Main Estimand

a. **Study intervention:** Same as main estimand for secondary endpoint
b. **Population:** Same as main estimand for secondary endpoint
c. **Intercurrent events (ICE) and the corresponding strategies:** Same as main estimand for secondary endpoints (Section 5.4.2.1)
d. **Variable/Endpoint and Population-level summary:** see Table 6

Table 6: Variables and Summary measure (Population-level Summary) for Each Tertiary Endpoint

Estimand	Variable (Endpoint)	Summary measure (Population-level summary)
Tertiary Endpoint #1	Response binary variable, where a response is defined as the proportion of subjects who achieve an ACR response (ACR20, ACR50, ACR70 & ACR90) through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #2	Response binary variable, where a response is proportion of subjects who achieve a DAS28 (CRP) remission through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #3	Response binary variable, where a response is proportion of subjects who achieve DAS28 (CRP) LDA through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #4	Response binary variable, where a response is proportion of subjects who achieve a DAS28 (ESR) remission through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #5	Response binary variable, where a response is proportion of subjects who achieve DAS28 (ESR) LDA through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #6	Response binary variable, where a response is proportion of subjects who achieve a CDAI remission through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.

Table 6: Variables and Summary measure (Population-level Summary) for Each Tertiary Endpoint

Estimand	Variable (Endpoint)	Summary measure (Population-level summary)
Tertiary Endpoint #7	Response binary variable, where a response is proportion of subjects who achieve CDAI LDA through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #8	Response binary variable, where a response is proportion of subjects who achieve a SDAI-based ACR/EULAR remission through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #9	Response binary variable, where a response is proportion of subjects who achieve a Boolean-based ACR/EULAR remission through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #10	Response binary variable, where a response is proportion of subjects who achieve SDAI LDA through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #11	Response binary variable, where a response is proportion of subjects who achieve a decrease of ≥0.22 points from baseline in HAQ-DI through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #12	Response binary variable, where a response is proportion of subjects who achieve ACPA seropositive through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #13	Response binary variable, where a response is proportion of subjects who achieve RF seropositive through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in proportion of subjects who achieved a response as (defined in the variable) between nipocalimab and placebo.
Tertiary Endpoint #14	Continuous variable, defined as percent improvement from baseline in ACR components by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in percent improvement as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #15	Continuous variable, defined as change from baseline in ACR components by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #16	Continuous variable, defined as change from baseline in DAS28 (CRP) by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #17	Continuous variable, defined as change from baseline in DAS28 (ESR) by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #18	Continuous variable, defined as change from baseline in CDAI by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #19	Continuous variable, defined as change from baseline in SDAI by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #20	Continuous variable, defined as change from baseline in Tender/Painful and Swollen Joint Counts by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #21	Continuous variable, defined as change from baseline in duration of morning stiffness by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #22	Continuous variable, defined as change from baseline in FACIT-Fatigue score by visit through Week 12.	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group

Table 6: Variables and Summary measure (Population-level Summary) for Each Tertiary Endpoint		
Estimand	**Variable (Endpoint)**	**Summary measure (Population-level summary)**
Tertiary Endpoint #23	Continuous variable, defined as change from baseline in Joint Pain Severity score by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #24	Continuous variable, defined as change from baseline in SF-36 domain, PCS and MCS score by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #25	Continuous variable, defined as change from baseline in M-DP score, M-DP4 and M-DP9 score by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group
Tertiary Endpoint #26	Continuous variable, defined as change from baseline in ESR by visit through Week 12 and not experiencing an ICE 1-3 prior to Week 12	Difference in mean change as (defined in the variable) between the nipocalimab and placebo group

5.5.2.2. Supplementary Estimand (Treatment Policy)

For tertiary endpoints' supplementary estimand share the same population, study intervention, ICEs, variable, and summary measures as the main estimand, see Section 5.5.2.1. The only difference is the ICE strategy, which will be handled by Treatment Policy Strategy, i.e., a subject with an ICE will not be treated differently from subjects without an ICE. Under the treatment policy strategy, observed data collected at or after intercurrent events will still be used in the analysis.

5.5.3. Analysis Methods

5.5.3.1. Main (Tertiary) estimator for main Estimand

Unless otherwise specified, the analysis population will be the FAS defined in Section 4. All statistical testing will be performed at the 2-sided 0.05 significance level and nominal p-values will be presented.

Analyses for all tertiary endpoints will be based on the main estimands defined in Table 6.

5.5.3.1.1. Binary Endpoints

In the binary tertiary endpoint responses analysis over time will be based on main estimands defined in Table 6 After accounting for intercurrent events specified in Section 5.4.2.1, any remaining missing response status will be considered as non-responders.

The treatment comparison between a nipocalimab group and the placebo group will be performed using a CMH chi-square test stratified by stratification factor (baseline MTX use (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week) by the scheduled visits. The magnitude of the treatment difference will be estimated by the difference in binary response rates between the nipocalimab and placebo groups with a 95% CI calculated based on Wald statistics. If the Mantel Fleiss criterion is not satisfied the Fisher's exact test will be used instead of the CMH test to compare the two treatment groups.

5.5.3.1.2. Continuous Endpoints

For tertiary continuous endpoints analysis over time will be based on main estimands defined in Table 6. After accounting for the ICEs for the main estimand, participants who are missing change from baseline in an endpoint will not be imputed, it will remain missing. The treatment difference between nipocalimab group and placebo group will be tested using an ANCOVA model by the scheduled visits, if missing data is not a problem, otherwise remaining missing data will be handled through MMRM model. The ANCOVA model will include treatment group, baseline score, and stratification factor (baseline MTX use, (0 mg/week, >0 to <12.5 mg/week, or ≥ 12.5 mg/week) as the explanatory factors. The treatment difference between a nipocalimab group and the placebo group will be estimated by the difference in the least squares means (LSmeans). The 95% confidence interval (CI) for the differences in LSmeans and p-values will be calculated based on the ANCOVA model. Note, an ANCOVA analysis will not be performed for these endpoints, percent improvement from baseline in ACR components and change from baseline in ACR components, only descriptive summary statistics will be presented.

5.5.3.2. Analysis (estimators) for the Supplementary Estimand

Analysis methods based on the supplementary estimand (treatment policy) will be the same as that based on the main estimands defined in Section 5.5.3.1 and observed data will be used regardless of the inter current event occurring, missing data will not be imputed.

5.5.3.3. Summary of Analyses Related to Tertiary Endpoints

Table 7 below provides an overview on all the analyses related to the tertiary endpoint binary response and continuous endpoint over time, the estimands, the data handling rules to be used, and the analysis methods and summary statistics.

Table 7: Tertiary Endpoint Analysis Summary

Binary endpoints		**Missing data**	**Additional notes**
	Main Estimand ((Section 5.5.2.1)	After accounting the ICEs specified in Section 5.4.2.1, any remaining missing data will be imputed as non-responder	• Summarized descriptively • Response • Treatment difference for treatment comparison and 95% CI • P-value from CMH test
	Supplementary Estimand (Treatment Policy) (Section 5.5.2.2)	Observed values over time will be used regardless of intercurrent events occurring. Remaining missing data will not be imputed	
Continuous endpoints			

	Main Estimand ((Section 5.5.2.1)	After accounting the ICEs specified in Section 5.4.2.1, any remaining missing data will not be imputed. If missing data is a problem, any remaining missing data after ICE rules are applied, it will be handled through MMRM model	• Summarized descriptively
			• ANCOVA/MMRM model
			  ○ LS mean (95 %CI)
			  ○ Treatment difference in LS mean (95% CI)
			  ○ P-value of comparing LS mean

	Supplementary Estimand (Treatment Policy) (Section 5.5.2.2)	Observed values over time will be used regardless of intercurrent events occurring. Remaining missing data will not be imputed	

5.6. 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 the N, mean, standard deviation (SD), median, and range (minimum, maximum). Categorical variables will be summarized by intervention group using frequency counts and percentages.

5.6.1. Extent of Exposure

The number and percentage of participants who receive study intervention (placebo, nipocalimab 15 mg/kg IV q2w) will be summarized. The number and percentage of participants will also be summarized by visit.

Descriptive statistics for duration of a study intervention and duration of follow-up (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 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 visit 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 change in infusion flow rate will be provided.

Study intervention compliance will be summarized descriptively. See Section 6.7, for further details

5.6.2. 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 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 be 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
  - CCI [redacted]
- AEs of clinical interest
  - Infusion reactions
  - Infusion site reaction
  - Opportunistic infections
  - 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 Section 6.8, 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 average number of study agent administrations and average weeks of follow-up for each intervention group.

5.6.3. AEs, SAEs and AESI by IgG

Treatment-emergent AEs, SAEs and AESIs will be summarized by minimum IgG levels (<=2.0, 2.0-4.0, >4.0 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 over time will be plotted for those subjects who experience an AESI.

5.6.4. Additional Safety Assessments

5.6.4.1. Clinical Laboratory Tests

Clinical laboratory tests will be displayed for the participants included in the Safety Analysis Set.

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 HDL to total cholesterol. Only fasting lipid samples will be included in summaries. In addition, all samples with fasting and non-fasting at baseline and post baseline will be summarised.

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).

Coagulation:

Prothrombin Intl. Normalized Ratio (INR), Prothrombin Time, Activated Partial Thromboplastin Time (aPTT)

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

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 XX 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 Section 6.9. The number and percentage of participants with treatment-emergent markedly abnormal values will be presented by intervention group over time. 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 <20g/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)

5.6.4.2. 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. Descriptive statistics (mean, SD, median, and range) will be presented.

Abnormality criteria (based on criteria in Table 8) 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 8

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 8 below.

Table 8: 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

5.7. Other Analyses

5.7.1. Pharmacokinetics

Pharmacokinetics (PK) samples for measuring serum nipocalimab concentrations will be collected from all subjects 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 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 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 concentrations at each visit
- Summary of serum nipocalimab concentrations at each visit by baseline body weight quartiles
- Summary of serum nipocalimab concentrations at each visit by baseline age categories (< 65 years, ≥ 65 years)
- Summary of serum nipocalimab concentrations at each visit by baseline CRP level (< 15 mg/L, ≥ 15 mg/L at Week 0)
- Summary of nipocalimab concentrations at each visit by baseline MTX use (yes, no)
- Summary of nipocalimab concentrations at each visit by baseline Rheumatoid factor (positive, negative)

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

PK analyses will be summarized through Week 12

If sufficient data are available, then population PK analysis using serum nipocalimab 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.

5.7.1.1. 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 subject 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 administrations
  - Skipped a nipocalimab administration
  - Received incomplete/incorrect dose
  - Received incorrect study agent
  - Received additional dose

Of note, serum nipocalimab concentrations prior to the first of such events will be included in the summaries. In addition, if a subject 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.

5.7.2. 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 "subject ADA status" and peak titer through the visit will be coded and provided by the bioanalytical group.

5.7.2.1. 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.

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.

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", ie, no appropriate sample is available after intervention.

5.7.2.2. 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 'subject 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 (eg, 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 (eg, 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 subjects who are positive of antibodies to nipocalimab are sufficient to permit subgroup analyses (eg > 30 to 40 subjects).

5.7.3. Pharmacodynamics

Summary statistics of total IgG, IgG subclasses, ACPA and RF pathogenic antibodies, percent change from baseline, and change from baseline will be evaluated. LLOQ rule specified in Section 6.12.3 will be applied to values <LLOQ if LLOQ value is available.

PD Analyses will be summarized through Week 12 DBL and Week 18 final safety DBL.

5.7.4. 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.

5.7.5. Pharmacogenomic Analyses

CCI [redacted]

5.7.6. Biomarkers

CCI [redacted]

[redacted]

[redacted]

5.7.7. Digital Health-Actigraphy Measurements

CCI [redacted]

CCI [redacted]

5.7.8. Definition of Subgroups

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:

1. Subgroups defined by demographics:

- Gender (Male, Female)
- Region (North America Europe).
- Age (< 65 years, ≥ 65 years)
- Weight (quartiles)
- BMI(kg/m^2): <25; >=25 to <30; >=30

Subgroups defined by baseline disease characteristics:

- Disease duration (< 1 year, ≥1 year to < 3 years, ≥3 years)
- Rheumatoid factor (Positive, Negative)
- Anti-CCP (positive, negative)
- Rheumatoid factor and anti-CCP (positive for both, otherwise)
- Anti-CCP (above median, below median)
- Rheumatoid factor (above median, below median)
- Number of swollen 66 joints (<10, ≥ 10)
- Number of tender 68 joints (<10, ≥ 10)
- HAQ (<2, ≥ 2)
- CRP (< 15 mg/L, ≥ 15 mg/L at Week 0)
- ESR (< 30, ≥30)

Subgroups defined by prior and baseline medicine use:

- Baseline MTX (0 mg/week, > 0 to < 12.5 mg/week, ≥ 12.5 mg/week)
- Baseline NSAIDs (yes, no)
- Baseline oral corticosteroids (yes, no)

- Baseline cDMARDs (yes, no)
- Number of anti TNF α medications used prior to study entry (1, ≥2)
- Failed (due to Lack of Efficacy) any prior anti TNF therapy (yes, no)
- Intolerant to any prior anti TNF therapy (yes, no)
- Prior Medications other than anti-TNF, Biosimilar and csDMARD (1, 2, ≥ 3)

5.8. Interim Analyses

CCI

[redacted]

[redacted]

5.8.1. Data Monitoring Committee (DMC) or Other Review Board

A 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. The committee will meet periodically to review interim data through unblinding after the Week 18 final database lock (DBL). For the 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/Charter.

Interim Analysis Committee

An internal unblinded interim analysis committee (IAC) will be established to review interim data. This committee will consist of physicians, statisticians, clinical pharmacologists, and pharmacometricians unaffiliated with the study; committee membership responsibilities, authorities, and procedures will be documented in its charter. The committee will meet periodically to review the interim analysis data (Week 12, as specified above).

Further details of the interim analysis will be in the IAC SAP/Charter.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1: 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
CMH	cochran-mantel-haenszel
CL	total systemic clearance
Cmax	maximum concentration
CRF	case report form
CRP	C-reactive protein
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DBL	database lock
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
GCP	good clinical practice
ICE	Inter current event
IEC	independent ethics committee
IQ	interquartile
IVRS	interactive voice response system
IWRS	interactive web response system
LLOQ	lower limit of quantification
LSmeans	least squares means
M-DP	Molecular Disease Profile
MCAR	missing completely at random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-Effect Model Repeated Measure
MTX	Methotrexate
NAb	neutralizing antibodies
NONMEM	nonlinear mixed-effects modeling
PD	pharmacodynamic(s)
PI	principal investigator
PK	pharmacokinetic(s)
PP	per protocol
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation
SoA	schedule of activities
SMQs	standardised MedDRA queries
TEAE	treatment-emergent adverse event
TF	treatment failure
TNF	tumor necrosis factor
Tmax	time to maximum concentration

ULN	upper limit of normal
US	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
WBC	white blood cell

6.2. Appendix 2: Changes to Protocol-Planned Analyses

N/A

6.3. Appendix 3: Demographics and Baseline Characteristics

Table 9 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 9: Demographic Variables

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)	
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])	

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

6.4. Appendix 4: Protocol Deviations

Participants with major protocol deviations will be identified prior to database lock.

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 0-18, and by treatment group and overall including the final safety visit occurring prior to the Week 18 DBL.

In addition to the summary tables, the following listings will be provided from Week 0 through Week 12, including the final safety visit occurring prior to the Week 18 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

6.4.1. Major Protocol Deviations

Participants with major protocol deviations will be identified prior to the Week 12 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
  - Missed visit/assessment due to COVID-19

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

6.4.2. 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.

6.4.3. 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.

6.5. 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 on 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 12.

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: Medical History

Medical history will be listed by intervention groups and participants.

6.7. Appendix 7: Intervention Compliance

Compliance will be summarized descriptively for each study intervention. Compliance to randomized intervention versus actual intervention will be presented in a summary table.

6.8. Appendix 8: 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	Hypoalbuminaemia	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, insomnia, irritability, libido decreased, major depression, mania, mental status changes, mood altered, nervousness, poor quality sleep	
	Endocrine/metabolic: adrenal insufficiency, blood glucose increased, blood potassium decreased, central obesity,	

	Cushingold, 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; myopathy; osteonecrosis; osteopenia; osteoporosis; spinal compression fracture; tendon rupture; weakness; wrist fracture	
	Ophthalmological: Cataract; cataract nuclear; glaucoma; intraocular pressure increased; open angle glaucoma; retinopathy hypertensive.	

AE of Clinical Interest Category	**SMQ**	**Additional condition**
Opportunistic infections	Opportunistic infections	Narrow scope
Hypersensitivity reaction	Hypersensitivity reaction	Narrow scope
Suicidal ideation/behavior	Suicide/self-injury	Narrow and broad scope
MACE	Conditions associated with central nervous system haemorrhages and cerebrovascular accidents	All potential cases of MACE to be independently adjudicated Narrow and broad scope
	Haemorrhagic central nervous system vascular conditions	Narrow scope
	Ischaemic central nervous system vascular conditions	Narrow scope
	Myocardial infarction	Narrow and broad scope
	Preferred terms of Troponin abnormal, Troponin I abnormal	Preferred terms
	All deaths	

6.9. Appendix 9: 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	≥~~240~~6.2
CHEMISTRY	HDL	BOTH	mmol/L	<~~40~~1.0	≥~~80~~2.1
CHEMISTRY	LDL	BOTH	mmol/L	N/A	≥~~160~~ 4.1
CHEMISTRY	Triglycerides	BOTH	mmol/L	N/A	~~>500~~ 5.6
HEMATOLOGY	Hematocrit female	F	fraction	Decrease >15% and value <0.28	Increase >15% and value >0.50
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.

6.10. Appendix 10: COVID-19 Related Summaries

Participant disposition as related to COVID-19 will be summarized by intervention group. This includes the following COVID-19 related disposition events:

- Termination of study due to COVID-19 and reason
- Discontinuation of stud agent due to COVID-19 and reason
- Death related to COVID-19

Participants discontinuing treatment or terminating study participation due to COVID-19 and reason(s) will be listed.

Assessment of study compliance as related to COVID-19 will be summarized, including number of missed visits, and number of remote visits conducted. Participants whose visit compliance was impacted by COVID-19 will be listed.

Concomitant medications used for COVID-19 will be summarized. Participants receiving concomitant medications related to COVID-19 will be listed.

Protocol deviations (major and minor) as related to COVID-19 will be summarized and listed.

Study agent modifications due to COVID-19 will be summarized and listed.

Adverse events related to COVID-19 will be identified and coded using the MedDRA coding guidance for COVID-19, and summary tables will be provided for COVID-19 related TEAEs. The following will be listed by participant:

- TEAEs related to COVID-19
- Serious TEAEs related to COVID-19
- Deaths related to COVID-19

6.11. Appendix 11: 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

1. 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 or DMARDs (other than MTX, SSZ, HCQ, CQ, and leflunomide) including azathioprine, oral cyclosporine A, tacrolimus, mycophenolate mofetil, oral or parenteral gold
- Anti-TNF therapy such as: infliximab, golimumab, etanercept, adalimumab, certolizumab, and biosimilars to those anti-TNFs
- Anti-IL-6 receptor mAb (eg, tocilizumab or sarilumab)
- IL-1 inhibitors (eg, anakinra)
- B-cell depleting biologic therapy (eg, rituximab)
- Co-stimulatory inhibitors (eg, abatacept)
- JAKi (eg, tofacitinib, baricitinib, upadacitinib)
- Cytotoxic drugs such as cyclophosphamide, chlorambucil, nitrogen mustard, or other alkylating agents
- Any other targeted biologic therapy
- Any investigational intervention (including investigational vaccines) or use of an invasive investigational medical device
- The use of complementary therapies (eg, herbs, ointments, traditional Chinese medicine, acupuncture)
- A live virus or bacterial vaccine, unless approved by the sponsor (JNJ-80202135ARA2001 Protocol Section 6.8)

6.12. Appendix 12: Rules Applied in definitions of endpoints

6.12.1. 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 subjects 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 subjects 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 subjects 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 subjects 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.

6.12.2. Joint Count Adjustment Rule

For subjects 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.

6.12.3. LLOQ Rule

Any value < LLOQ is considered equal to half of the value of LLOQ for numerical calculations

6.13. Appendix 13: Description of Statistical Models

1. Mixed-effect Model Repeated Measures (MMRM)

To account for the missing data for continuous endpoints, an MMRM model will be used on the change from baseline, under the assumption of MAR, to test the difference between a nipocalimab group and the placebo group. The model will include treatment group, visit, an interaction terms of visit with treatment group, stratification factor (baseline MTX use (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week), baseline score, an interaction term between baseline score and visit, and an interaction term between stratification factor and visit as explanatory variables. An unstructured covariance matrix for repeated measure within a subject will be used. The F-test will use Kenward-Roger's approximating for degree of freedom. In case of lack of convergence, empirical structured covariances will be used in the following order until convergence is reached: 1) first order Autoregressive Moving Average 2) Toeplitz. The treatment difference between a nipocalimab group and the placebo group will be estimated by the difference in the least squares means (LSmeans). The 95% confidence interval (CI) for the differences in LSmeans and p-values will be calculated based on the MMRM model.

2. Analysis of Covariance (ANCOVA) Model

The ANCOVA model will be used on the change from baseline, under the assumption of Missing Completely at Random (MCAR), to test the difference between a nipocalimab group and the placebo group, if missing data is not a problem. The model will include treatment group, stratification factor (baseline MTX use (0 mg/week, > 0 to < 12.5 mg/week, or ≥ 12.5 mg/week)), and baseline score as explanatory variables. For analyses through Week 12 the model will include data from all 2 treatment groups at Week 12. The treatment difference between a nipocalimab group and the placebo group will be estimated by the difference in the least squares means (LSmeans). The 95% confidence interval (CI) for the differences in LSmeans and p-values will be calculated based on ANCOVA.

6.14. Appendix 14: Re-randomization Test

The original treatment assignment algorithm is used to re-randomize subject while keeping the biased coin probability and weights of each of the three stratification factors the same, changing only the seed (#). Test statistic is used to evaluate the observed data (original randomization) and the re-randomized test data. The procedure for implementing re- randomization test has the following steps:

1. Obtain the test statistics **T*** for the superiority of nipocalimab versus Placebo in primary endpoint analysis

For each participant j in the study (j = 1;50), re-randomize their treatment allocation assignment using the same randomization algorithm with a new "seed #" as a starting point. Under the new "seed" random numbers will be different from the original random number, this will produce a new set of treatment assignments for the 50 participants in the study. Based on the new treatment assignment:

i. Obtain the new test statistics **T**
ii. Repeat the re-randomize step for 5000 times.

Derive the p-value associated with the observed test statistics **T*** based on the empirical distribution in step 2. The p-value can be estimated as $(n + 1)/(5000 + 1)$, where the **n** is the number of times that $\mathbf{T} > \mathbf{T}^*$ among the 5000 repetitions. If p-value is greater than one-sided study type one error (p-value=0.025) then the randomization algorithm used in this study is a valid randomization method under the null hypothesis.

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

Cella D (2002), Lai J, Chang C, et al. Fatigue in cancer patients compared with fatigue in the general United States population. Cancer. 2002; 94(2):528-538

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

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

Quality Metric Inc. (2011) User's Manual for the SF-36v2 Health Survey. 3rd edition;2011: http://www.qualitymetric.com/WhatWeDo/ManualsUserGuides/UsersManusalfortheSF36v2HealthSurvey/tabid/328/Default.aspx

Smolen JS (2008), Beaulieu A, Rubbert-Roth A, et al, for the OPTION Investigators. Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial. Lancet. 2008;371(9617):987-997

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.

Yellen SB (1997), Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Sympt Manage. 1997; 13;63-74.

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

Yount S (2007), Sorensen MV, Cella D, Sengupta N, Grober J, Chartash EK. Adalimumab plus methotrexate or standard therapy is more effective than methotrexate or standard therapies alone in the treatment of fatigue in patients with active, inadequately treated rheumatoid arthritis. Clin Exp Rheum. 2007; 25(6):838-846.