Janssen Research & Development

Statistical Analysis Plan

A Phase 3b, Multicenter, Randomized, Double-blind, Placebo-controlled Study Evaluating the Efficacy and Safety of Subcutaneously Administered Guselkumab in Improving the Signs and Symptoms and Inhibiting Radiographic Progression in Participants with Active Psoriatic Arthritis

Protocol CNTO1959PSA3004; Phase 3b
Amendment 4
CNTO1959 (guselkumab)

Status:	Approved
Date:	21 May 2025
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-RIM-541958, 5.0

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

Confidentiality Statement

The information provided herein is considered trade secrets, commercial or financial, or privileged or confidential information that we customarily hold close and treat as confidential. The information is being provided under the assurance that the recipient will maintain the confidentiality of the information under the applicable statutes, regulations, rules, case law, contractual provisions, protective orders, or as otherwise necessary to protect the information from disclosure. The recipient agrees and acknowledges that release of information considered trade secrets, commercial or financial, or privileged or confidential will result in competitive harm and, as such, the information provided herein is exempt from disclosure under the Freedom of Information Act ("FOIA"), or other similar law or regulation. Particularly, the information shall not be utilized in any AI solutions, including GenAI, without Company's prior written approval.

TABLE OF CONTENTS

LIST OF IN TEXT TABLES AND FIGURES

Tables

Figures

VERSION HISTORY

Table 1: SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	2021-10-01	Not Applicable	Initial release
2	2022-09-08	• Replaced COVID-19 related wording for ICEs and missing data handling to refer only to site closure, site access restrictions, or lockdowns caused by COVID-19 (Natural Disaster). • FAS analysis set changed to equal all randomized participants • Analysis window for radiographic assessments changed from +/-8 weeks, into Study Week -6 to Study Day 1 for bl, +/-2 for Week 24; sensitivity analysis added for subgroup with narrower baseline window of Study Week -4 to Study Day 1; a sensitivity analysis added using 2-Step MI to utilize out-of-window data; a sensitivity analysis added using original window of +/- 8 weeks. • Added tipping point sensitivity analysis for Primary Endpoint based on Adjusted Composite Estimand with MI for missing data; removed the similar tipping point supplementary analysis using the Treatment Policy Estimand • Clarified use of re-reads for Xray is solely to assess the intra-reader and inter-reader variability and calculate the intra-class correlation • Added thromboembolic AE events for safety	Received regulatory agency feedback for guselkumab programs
		• Added mFAS analysis set as the main efficacy analysis set. • Added mFAS-UKR analysis set for sensitivity/supplementary analysis for Primary and Key Confirmatory Secondary endpoints. Additionally, supportive analyses using Treatment Policy Estimand for Other Efficacy endpoints were changed from using mFAS to mFAS-UKR, as those analyses were meant to assess the de facto treatment effect in a typical real world setting, not one under which entire countries are under conflict or disruption. • Added Per-Protocol analysis set for sensitivity analysis • Changed main estimand for radiographic endpoints from Treatment Policy to Adjusted Treatment Policy, which incorporates Major Disruption related ICEs • Updated the Adjusted Composite Estimand to incorporate Major Disruption ICEs • Removed the Composite Estimand and the associated sensitivity analysis for the Primary Endpoint	Development of Major Disruptions involving Ukraine and neighboring countries/territories beginning February 24, 2022

SAP Version	Approval Date	Change	Rationale
		• Added handling for missing data due to Major Disruption • Added several sensitivity and supplementary analyses to assess impact of Major Disruption on the Primary and Key Confirmatory Secondary endpoints • Adjusted tipping point analyses to distinguish which values are to be varied and which are part of the estimand and not to be varied • Added summary of missed doses due to Major Disruption or Natural Disaster • Added laboratory analysis distinction for handling local labs • Added a sentence stating more participants may be enrolled to account for the impact of the Major Disruption	
		• Adjusted the LFT analyses	For consistency with other studies for the compound
		• Various typographical or internal consistency corrections	Other
3	2023-06-20	• Updated the definition of the mFAS analysis set so that site exclusion criteria are based off of participant randomization date, site closure dates, and site study intervention supply interruption dates. • Changed the missing data assumption for the Key Confirmatory Secondary Endpoint Supplementary Analysis 1 from MCAR to MAR; changed the missing data imputation for that analysis from no imputation to FCS MI. • Added tipping point supplementary analyses for evaluating deviations in missing data assumption for the Primary and Key Confirmatory Secondary Endpoints under the Treatment Policy Estimand.	Received regulatory agency feedback
		• Removed Randomized Analysis Set, as Amendment 1 redefined the FAS to be identical to the Randomized Analysis Set, rendering the latter redundant. Replaced analyses where Randomized Analysis Set is used with FAS (eg. demographic and baseline characteristics summary). • Various typographical or internal consistency corrections	Other
4	2024-11-15	• Added clinically important hepatic disorder event summary table and listing; added MACE listing; added opportunistic infections listing	For consistency with other studies for the compound

SAP Version	Approval Date	Change	Rationale
		• Removed language involving possible Hy's Law, and replaced with 2 combined criteria in Section 5.4.4.1.	
		• Added exploratory analyses for mvdH-S using AI models • For subgroups involving JSN, combined the two lower categories into a single category; defined the categories in the Participating Countries/Territories subgroup; added Asian category to Race subgroup • In Appendix 7, added a table on how AEs of interest are categorized • Various typographical or internal consistency corrections	Other
5	2025-05-21	• In Appendix 4, replaced the mention of using WHO-DD for prior/concomitant medications to stating they will be coded using an appropriate drug dictionary to be defined in the study metadata. This administrative update is only to accommodate an update in drug dictionary name, no change to analysis is intended.	Administrative amendment: to comply with updated nomenclature in FDA data standards catalog, where WHO-DD was recently renamed to UMC-DD.
		• Updated the Confidentiality Statement	Administrative amendment: to comply with updated wording in Janssen guidelines for confidentiality

1. INTRODUCTION

This statistical analysis plan contains definitions of analysis sets, derived variables, and statistical methods for the analysis of efficacy, safety, pharmacokinetics (PK), and pharmacodynamics (PD), and Immunogenicity in the CNTO1959PSA3004 study.

1.1. Objectives and Endpoints

Primary Objective

The primary objective of this study is to evaluate the efficacy of guselkumab treatment in participants with active psoriatic arthritis (PsA) by assessing the reduction in signs and symptoms of PsA. The proportion of participants with American College of Rheumatology (ACR) 20 response at Week 24 will be used for this assessment.

Major Secondary Objective

The secondary objective of this study is to evaluate the inhibition of progression of structural damage in participants with active PsA. The change from baseline in PsA modified van der Heijde-Sharp (vdH-S) score at Week 24 will be used for this assessment.

Other Secondary Objectives

Other secondary objectives of this study are to evaluate the safety in participants with active PsA, as well as their PK and immunogenicity. These objectives will be assessed by:

- Frequency and type of adverse events (AEs), serious adverse events (SAEs), reasonably related AEs, AEs leading to discontinuation of study intervention, infections, infusion reactions, and injection-site reactions.
- Laboratory abnormalities (chemistry, hematology), maximum toxicity (Common Terminology Criteria for Adverse Events [CTCAE 5.0]) grades.
- Serum guselkumab concentration.
- Incidence of antibodies to guselkumab

1.2. Study Design

This is a Phase 3b, multicenter, randomized, double-blind, placebo-controlled, 3-arm study in participants with active PsA who are biologic naïve and have had inadequate response to current standard therapies (eg, DMARDs/apremilast, corticosteroids, NSAIDs).

A target of approximately 950 participants will be randomly assigned in this study with 350 participants planned in each of the placebo and guselkumab 100 mg q8w groups, and 250 participants in the guselkumab 100 mg q4w group. Stable doses of concomitant NSAIDs, oral corticosteroids (≤ 10 mg/day prednisone equivalent), selected non-biologic DMARDs (MTX, SSZ, HCQ, LEF) will be allowed but are not required (see Table 2 below).

Table 2: Permitted Concomitant Medications for PsA and the Maximum Allowed Doses During the Study

Permitted Concomitant Medications for Psoriatic Arthritis (PsA)[a]	**Maximum Allowed Dose**
NSAIDs and other analgesics	Maximum marketed dose approved in in the country where the study is being conducted
Oral corticosteroids	Equivalent to 10 mg/day of prednisone
Methotrexate (MTX)[b]	25 mg/week
Sulfasalazine (SSZ)	3 g/day
Hydroxychloroquine (HCQ)	400 mg/day
Leflunomide (LEF)	20 mg/day

[a] Permitted concomitant medications are not supplied by the Sponsor.
[b] It is recommended that all participants taking MTX in this study receive at least 5 mg oral folate or 5 mg folinic acid weekly. Guidelines for dose adjustment in the event of MTX toxicity are included in the Trial Center File.
Abbreviations: HCQ = Hydroxychloroquine; LEF = Leflunomide; MTX = Methotrexate; PsA = psoriatic arthritis; SSZ = Sulfasalazine

Participants who satisfy all inclusion and exclusion criteria will be randomly assigned to one of the following 3 treatment groups in a 7:5:7 ratio using permuted block randomization and will be stratified by a combined factor of baseline radiographic variability, corticosteroid use, number of joints with erosion, and the most recent available CRP value prior to randomization into 4 strata levels (high radiographic variability [HRV], no progression [NP], low to moderate progression [LMP], and rapid progression [RP]). All participants with HRV will be assigned to HRV stratum, and the rest of participants will be assigned to the other strata based on probability of NP, LMP, and RP.

The stratification levels are derived as follows:

- **HRV**: R_DIFF_ERN = Yes
- **RP:** [(CRP>=2.1 and ERN>=16) OR (CRP>=5.7)] AND (R_DIFF_ERN = No)
- **NP:** [(CRP<5.7 and ERN<=5) OR (CRP<5.7 and 5<ERN<16 and COR=Yes)] AND (R_DIFF_ERN = No)
- **LMP:** [Other] AND (R_DIFF_ERN = No)

With the abbreviations standing for:

COR = the participant's baseline oral corticosteroid use
ERN = the participant's baseline number of joints with erosion
R_DIFF_ERN = (Yes/No) if the absolute reader difference for number of joints with erosions at baseline was >8
CRP = the participant's most recent screening CRP Level in mg/dL

Group I (n=350)

Participants will receive SC guselkumab 100 mg at Weeks 0, 4, 12, 20, 28, 36 and 44 and placebo at Weeks 8, 16, 24, 32, 40 and 48 to maintain the blind.

Group II (n = 250)

Participants will receive SC guselkumab 100 mg at Weeks 0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44 and 48.

Group III (n=350)

Participants will receive SC placebo at Weeks 0, 4, 8, 12, 16 and 20, and will cross over at Week 24 to receive SC guselkumab 100 mg at Weeks 24, 28, 32, 36, 40, 44 and 48.

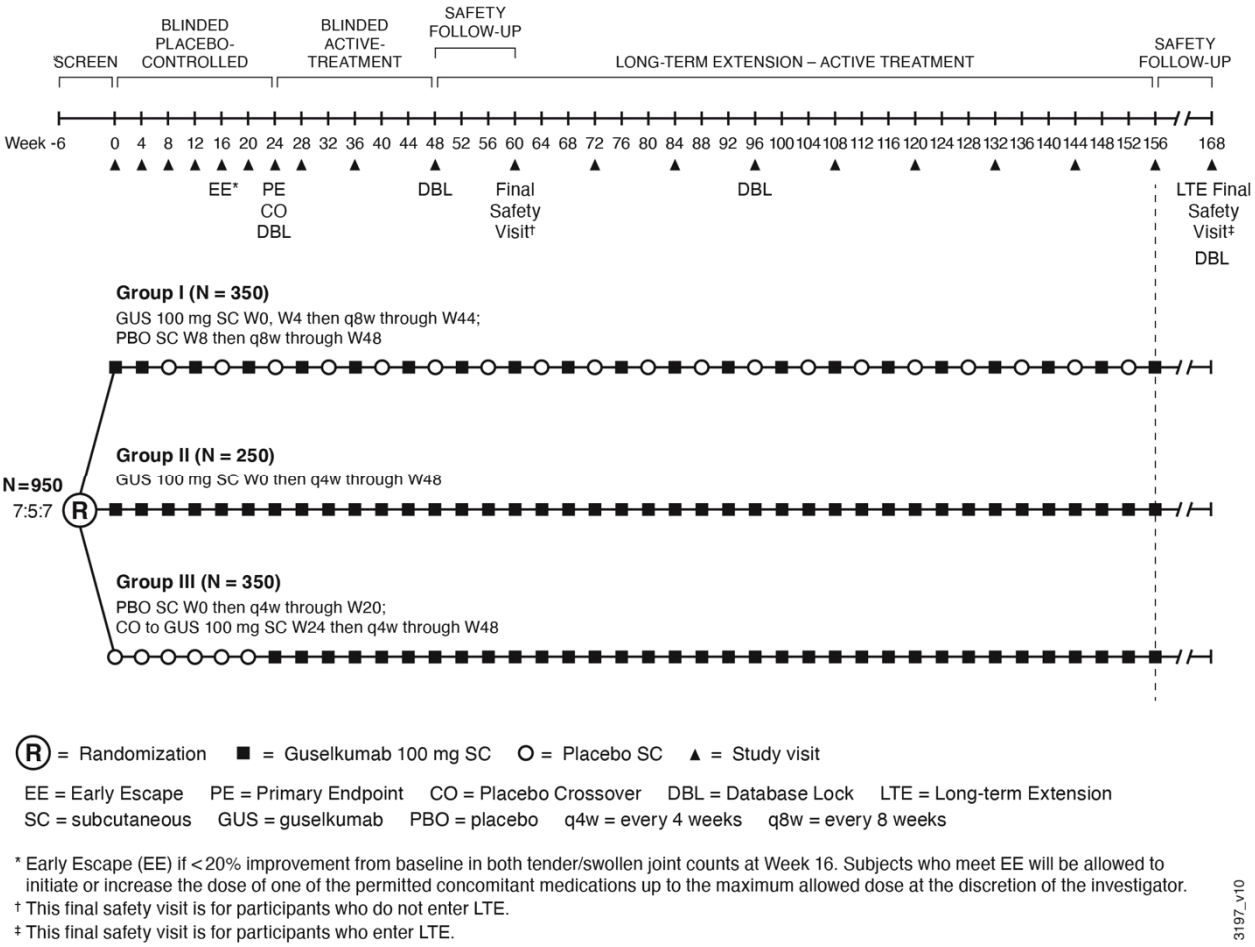
At Week 16, all participants in Groups I II and III with < 20% improvement from baseline in both tender and swollen joint counts will qualify for early escape (EE) and will be allowed to initiate or increase the dose of one of the permitted concomitant medications up to the maximum allowed dose as specified in Table 2, as selected by the investigator.

The core study lasts from Screening to Week 48. At Week 48, participants who have not discontinued will be eligible to enter a long-term extension (LTE) for a period of up to approximately two years (ie, Week 48 through Week 156) where they will continue on the same treatment regimen.

Database locks (DBLs) are scheduled at Weeks 24, 48, 96 and end of study (Week 168). The first DBL will occur when all randomized participants have either completed the Week 24 assessments or terminated study participation prior to the Week 24 visit (referred to as Week-24 DBL). The second DBL will occur when all randomized participants have either completed the Week 48 assessments or terminated study participation prior to the Week 48 visit (referred to as Week-48 DBL). The third DBL will occur when all randomized participants have either completed the Week 96 assessments or terminated study participation prior to the Week 96 visit (referred to as Week-96 DBL). The fourth and final DBL will occur when all randomized participants have either completed their final safety visit or have terminated study participation.

A diagram of the study design is provided in Figure 1, Schema.

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



Blinding

To maintain the study blind, the study intervention container will have a label containing the study name, study intervention number, reference number and storage instructions. 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 interactive web response system (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-guselkumab antibodies) will be handled with special care to ensure that the integrity of the blind is maintained and the potential for bias is minimized. This can include making

special provisions, such as segregating the data in question from view by the investigators, clinical team, or others as appropriate until the time of DBL and unblinding.

Under normal circumstances, the blind should not be broken until all participants have completed Week 48 or discontinued prior to Week 48 and the Week-48 DBL has occurred. 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 IWRS 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 should continue to return for scheduled evaluations, and may not be eligible for further treatment.

Once the Week-48 DBL has occurred, and the study is unblinded to the investigative sites, participants receiving guselkumab q8w will no longer be required to dose with placebo to maintain the blind.

2. STATISTICAL HYPOTHESES

The primary efficacy endpoint of this study is the proportion of participants who achieved an ACR 20 response at Week 24 (refer to Section 5.3.4 for endpoint definition and analyses). This endpoint was chosen because it is well-accepted by regulatory authorities and the clinical PsA community.

The hypotheses related to the primary endpoint are that:

1. (**H1**) treatment with guselkumab 100 mg SC q4w is superior to treatment with placebo SC with respect to reduction of PsA signs and symptoms as measured by proportion of participants who achieved an ACR 20 response at Week 24 (***primary hypothesis***); and
2. (**H2**) treatment with guselkumab 100 mg SC at Week 0, Week 4 and then q8w is superior to treatment with placebo SC with respect to reduction of PsA signs and symptoms as measured by proportion of participants who achieved an ACR 20 response at Week 24 (***key secondary hypotheses*****).**

The first hypothesis is the primary hypothesis for this study. If the first hypothesis achieves the statistical significance at a 2-sided α-level of 0.05, the study will be considered positive.

In addition to the primary endpoint, there is one key confirmatory secondary endpoint in this study (refer to Section 5.3.5 for endpoint definitions and analyses). The hypotheses related to the key confirmatory secondary endpoint (all are key secondary hypotheses) are as follows:

1. (**H3**) treatment with guselkumab 100 mg SC q4w is superior to treatment with placebo SC with respect to inhibition of progression of structural damage as measured by change from baseline in modified vdH-S score at Week 24
2. (**H4**) treatment with guselkumab 100 mg SC at Week 0, Week 4 and then q8w is superior to treatment with placebo SC with respect to inhibition of progression of structural damage as measured by change from baseline in modified vdH-S score at Week 24

For hypothesis testing order and multiplicity control, refer to Section 5.3.3.2.

3. SAMPLE SIZE DETERMINATION

The planned enrollment in the study is approximately 950 participants. The sample size selection was determined based on the primary endpoint of proportion of participants who achieve an ACR 20 response at Week 24 and the key confirmatory secondary endpoint of change from baseline in modified vdH-S score at Week 24 by considering power for each comparison individually. The assumptions are based on the PSA3002 study.

Additionally, more participants may be enrolled to account for those impacted by the Major Disruption in Ukraine and neighboring countries/territories beginning February 24, 2022.

3.1.1. Primary Endpoint – ACR 20 Response at Week 24

In the PSA3002 study, the ACR 20 response rates at Week 24 were 33.1%, 64.6%, and 63.7%, respectively, for the placebo, guselkumab 100 mg SC at Weeks 0, 4, then q8w, and guselkumab 100 mg SC q4w treatment groups.

For this study, assuming a 60% ACR 20 response rate in the guselkumab group and a 35% ACR 20 response rate in the placebo group, a sample size of 250 or 350 participants in the guselkumab group and 350 participants in the placebo group will provide a power of approximately >99% to detect a significant treatment difference at a significance level of α=0.05 using a 2-sided Chi-square test. Table 3 shows the power to detect a difference in the proportion of participants achieving ACR20 response between guselkumab groups and placebo group with various assumptions.

Table 3: Statistical Power for Treatment Difference in ACR 20 Response at Week 24

Sample size per arm Guselkumab/Placebo	ACR 20 Response Rate			Power
	Placebo Group	Guselkumab Group	Difference (Δ)	
250/350	35%	55%	20%	>99%
250/350	35%	60%	25%	>99%
250/350	35%	65%	30%	>99%
350/350	35%	55%	20%	>99%
350/350	35%	60%	25%	>99%
350/350	35%	65%	30%	>99%

3.1.2. Key Confirmatory Secondary Endpoint – Change from Baseline in Modified vdH-S Score at Week 24

For change from baseline in modified vdH-S score, participants in each treatment group can be considered a mixture of two subpopulations: one subpopulation (Spop 1) with a change score of 0 regardless of treatment and another subpopulation (Spop2) with a change score sampled from a normal distribution. Therefore, the distribution of the modified vdH-S change scores is determined by 3 parameters: the probability that a participant has a change score of 0, the mean of the normal distribution, and the standard deviation (SD) of the normal distribution. The overall mean (ie, crude mean) of the change scores for a treatment group is the overall average of the change scores among all participants (including both Spops 1 and 2) in that treatment group.

In the PSA3002 study, the following statistics were observed for change from baseline in modified vdH-S score at Week 24 for each treatment group:

The overall mean (SD) of change from baseline in modified vdH-S score at Week 24 was 0.90 (3.14), 0.25 (2.52), and 0.45 (2.38) respectively, for the placebo, guselkumab 100 mg q4w, and guselkumab 100 mg q8w treatment groups. The assumptions for power calculations in this study are based in part on these data, adjusted for the difference in enrichment criteria between studies.

For this study, assuming an overall mean (SD) of change from baseline in modified vdH-S score as 1.13 (3.2), 0.25 (3.1), and 0.45 (3.1) respectively in the placebo, guselkumab 100 mg q4w, and guselkumab 100 mg q8w groups, a sample size of 350/250/350 participants (ie, 7:5:7 ratio, 950 in total) will provide a power of at least 90% and 80% to detect a significant treatment difference at a 2-sided significance level of α=0.05 for guselkumab q4w vs placebo and guselkumab q8w vs placebo comparisons respectively.

Table 4 provides the statistical power for guselkumab 100mg q4w vs placebo under various assumptions, where the sample size is 250 in the guselkumab group and 350 in the placebo group.

Table 4: Statistical Power for Treatment Difference in Modified vdH-S Change from Baseline at Week24 for Guselkumab 100mg q4w vs Placebo (N=250, 350)

Percent Participant with extra 0 (Spop 1)	Normal Placebo (Spop 2)		Normal Guselkumab (Spop 2)		Overall Placebo		Overall Guselkumab		Overall Mean Difference	Power
	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
15%	**1.33**	**3.4**	**0.29**	**3.4**	**1.13**	**3.2**	**0.25**	**3.1**	**-0.88**	**92**
15%	1.33	3.4	0.34	3.4	1.13	3.2	0.29	3.1	-0.84	89
15%	1.33	3.0	0.29	3.0	1.13	2.8	0.25	2.8	-0.88	96
15%	1.33	3.0	0.34	3.0	1.13	2.8	0.29	2.8	-0.84	95
25%	1.33	3.4	0.29	3.4	1.00	3.0	0.22	2.9	-0.78	87
25%	1.33	3.4	0.34	3.4	1.00	3.0	0.26	2.9	-0.74	83
25%	1.33	3.0	0.29	3.0	1.00	2.7	0.22	2.6	-0.78	94
25%	1.33	3.0	0.34	3.0	1.00	2.7	0.26	2.6	-0.74	91

Table 5 provides the statistical power for guselkumab 100mg q8w vs placebo under various assumptions, where the sample size is 350 in each group.

Table 5: Statistical Power for Treatment Difference in Modified vdH-S Change from Baseline at Week 24 for Guselkumab 100mg q8w vs Placebo (N=350, 350)

Percent Participant with extra 0 (Spop 1)	Normal Placebo (Spop 2)		Normal Guselkumab (Spop 2)		Overall Placebo		Overall Guselkumab		Overall Mean Difference	Power
	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
15%	**1.33**	**3.4**	**0.53**	**3.4**	**1.13**	**3.2**	**0.45**	**3.1**	**-0.68**	**80**
15%	1.33	3.4	0.58	3.4	1.13	3.2	0.49	3.1	-0.64	75
15%	1.33	3.0	0.53	3.0	1.13	2.8	0.45	2.8	-0.68	89
15%	1.33	3.0	0.58	3.0	1.13	2.8	0.49	2.8	-0.64	84
25%	1.33	3.4	0.53	3.4	1.00	3.0	0.40	3.0	-0.60	74
25%	1.33	3.4	0.58	3.4	1.00	3.0	0.44	3.0	-0.56	69
25%	1.33	3.0	0.53	3.0	1.00	2.7	0.40	2.6	-0.60	84
25%	1.33	3.0	0.58	3.0	1.00	2.7	0.44	2.6	-0.56	79

The statistical power for each comparison was estimated based on 10000 simulations with treatment comparison performed at each simulation using an analysis of variance (ANOVA) test on the van der Waerden normal score. Under these assumptions, the power ranges approximately from 83% to 96% for the guselkumab q4w vs placebo comparison, and from 69% to 89% for the guselkumab q8w vs placebo comparison.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

4.1. Enrolled Participants

All participants who signed the informed consent form (ICF).

4.2. Full Analysis Set (FAS)

All participants who were randomized in the study. This analysis set will be used for efficacy analyses.

In the analyses for this set, participants will be analyzed according to the randomized study intervention they were **assigned to**, regardless of the study intervention they actually received.

4.3. Modified Full Analysis Set (mFAS)

All participants who were randomized, excluding participants from Ukrainian sites rendered unable to support key study operations due to Major Disruption. This analysis set will be used as the main efficacy analysis set.

Due to the crisis in Ukraine and neighboring countries/territories beginning February 24, 2022, several Ukrainian sites were temporarily closed, and additionally, all Ukrainian sites with randomized participants experienced an interruption of study intervention supply. A site will be deemed unable to support key study operations and will not be included in the mFAS if, based on the randomization date, **every** participant at a site is either projected to miss the Primary and Key Confirmatory Secondary efficacy endpoints at Week 24, OR projected to miss ≥2 doses of study intervention prior to Week 24. The detailed definition of these two criteria are as follows:

1. If 2 or more projected dosing dates up to and including Week 20 fall within site closure or site study intervention supply interruption periods, where projected dosing dates are extrapolated based on participants' randomization date (a window of ±4 days for dosing is allowed, as per the protocol specified window).
2. If the projected Week 24 date falls within the site closure period, where projected Week 24 date is extrapolated based on participants' randomization date (a window of ±14 days is allowed, as per the protocol specified window for collecting radiographs).

Post-baseline dosing and visit dates are all projected using the randomization date, and the actual observed post-baseline dates, missing status, or early discontinuations are NOT used. Site closures and study intervention supply interruption dates are applied consistently at the site level for all participants at that site.

In the analyses for this set, participants will be analyzed according to the randomized study intervention they were **assigned to**, regardless of the study intervention they actually received.

4.4. Modified Full Analysis Set Excluding Ukraine (mFAS-UKR)

All participants in the mFAS analysis set, excluding those participants from sites in Ukraine. This analysis set will be used for efficacy analyses.

No participant from Ukrainian sites reached Week 24 prior to February 24, 2022. Thus while the mFAS excludes the sites most impacted by Major Disruption, all Ukrainian sites were potentially impacted, ranging from study intervention supply interruptions (supply depot to the entire country cut off for a period of time), to temporary site closures in some sites, and other difficulties involved in operation within an area of ongoing conflict. This analysis set will be used for sensitivity and supplementary analyses of the Primary and Key Confirmatory Secondary endpoints, as well as supportive analyses for the Other Efficacy endpoints based on the Treatment Policy Estimand.

In the analyses for this set, participants will be analyzed according to the randomized study intervention they were assigned to, regardless of the study intervention they actually received.

4.5. Per-Protocol Analysis Set

The per-protocol analysis set (PPAS) includes all participants in mFAS who met all inclusion and exclusion criteria and had no major protocol deviations prior to Week 24 that could have impacted efficacy assessment per clinical judgement. This analysis set will be used for the analyses of selected efficacy endpoints through Week 24. Participants to be excluded from this analysis will be identified prior to the Week-24 DBL and un-blinding.

In the analyses for this set, participants will be analyzed according to the randomized study intervention they were assigned to, regardless of the study intervention they actually received.

4.6. Safety Analysis Set

All participants who received at least one (complete or partial) administration of any study intervention, ie, the treated population. This analysis set will be used for the safety analyses.

In the analyses for this set, participants will be analyzed per the study intervention they actually received, regardless of the study intervention they are randomized to.

4.7. Pharmacokinetics (PK) Analysis Set

All participants who received at least one complete administration of guselkumab and had at least one valid blood sample drawn for PK analysis.

4.8. Immunogenicity Analysis Set

All participants who received at least one (complete or partial) administration of guselkumab and who had at least one sample obtained after their first administration of guselkumab.

4.9. Pharmacodynamic (PD) Analysis Set

The PD analysis set will be defined in a separate analysis plan for biomarkers and PD analyses.

5. STATISTICAL ANALYSES

5.1. General Considerations

5.1.1. Visit Windows

In general for safety and efficacy analyses, the nominal visit will be used for by-visit analyses. Exceptions exist for the following.

5.1.1.1. Visit Windows for Dosing and PK Analysis

All post-baseline visits from Baseline through Week 24 will have a visit window of ± 4 days, and from Week 28 through Week 156 will have a visit window of ± 1 week (7 days). The final safety follow-up at Week 168 will have a visit window of ± 2 weeks (14 days). This information will be used to identify out-of-window dosing or visits (non-radiographic).

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.

5.1.1.2. Visit Windows for Radiographic Assessments and Analyses

The windows for taking radiographs of hands and feet at respective scheduled visits are specified as follows:

All eligible participants should have radiographs taken approximately 2 weeks but not greater than 4 weeks prior to randomization.

Week 24 radiographs should be taken within ±2 weeks of the Week 24 visit. For participants who discontinue study intervention prior to Week 24, radiographs of the hands and feet should be performed at the Week 24 visit.

Week 48, 96, and 156 radiographs should be taken within ±2 weeks of the Week 48, 96 and 156 visits. For participants who permanently discontinue study intervention after Week 24 but prior to Week 156, and participation at any time during the study radiographs of hands and feet should be performed at the time of discontinuation of study intervention or as soon as possible unless another set of radiographs has been obtained within the past 6 weeks.

The analytical window (inclusive) to be used in radiographic data analyses will be as follows:

- Screening (baseline) analysis visit: Study week -6 to study day 1
- Week 24 analysis visit: ±2 weeks around study week 24
- Weeks 48, 96, and 156 analysis visit: ±8 weeks around their respective study weeks

Radiographs taken at end of study intervention will be slotted based the analysis window to the appropriate visits to be included in the data analysis.

5.1.2. Pooling Algorithm for Analysis

Data from all investigational centers/sites will be pooled for analyses.

5.2. Participant Dispositions

The number of participants screened, randomized and treated will be summarized by geographic region, country, and investigational site.

The number of participants in the following disposition categories will be summarized throughout the study by intervention group and overall. Note that some categories will only be relevant for later DBLs (e.g., participants who decided to enter the LTE would only be relevant from the Week-48 DBL onwards).

- Participants who received study intervention
- Participants who decided to enter the LTE
- Participants who received study intervention during the LTE
- Participants who completed the study: core, LTE
- Participants who discontinued study intervention: core, LTE
- Reasons for discontinuation of study intervention: core, LTE
- Participants who terminated study prematurely: core, LTE
- Reasons for termination of study: core, LTE

Listings of participants will be provided for the following categories:

- Participants who discontinued study intervention
- Participants who terminated study prematurely
- Participants who were unblinded prior to the Week-48 DBL.
- Participants who were randomized yet did not receive study intervention

Summaries of participant demographic and baseline characteristics are in Section 6.2 (Appendix 2).

5.2.1. Intercurrent Events (ICEs) and Early Escape (EE)

Tabulations by randomized intervention group will also be provided for participants who met EE criteria at Week 16 and for participants who met 1 or more ICEs for the Adjusted Composite Estimand prior to Week 24 as defined in Section 5.3.4.2.

Listings of participants who met any ICE criteria will be presented.

5.3. Efficacy Analyses

5.3.1. General Method of Analysis

In general, descriptive statistics, such as mean, standard deviation (SD), median, inter quartile (IQ) range, minimum, and maximum for continuous variables, and counts and percentages for discrete variables will be used to summarize most data.

Statistical comparison between a guselkumab group (100 mg q4w or 100 mg at Weeks 0, 4 and then q8w) and the placebo group will be performed by visit through Week 24. No treatment comparison will be performed after Week 24.

Binary Response Efficacy Endpoints

For binary response efficacy endpoints where any portion of the missing data is imputed using Multiple Imputation (MI) (see Section 6.11 for technical details), treatment comparisons will be performed using a Cochran-Mantel-Haenszel (CMH) on each of the imputation sets, and the inferences from the analysis of each imputed data are pooled. The within imputation variance and between imputation variance are combined to estimate the total variance of the stratum adjusted difference of proportions. This estimate of the variance and critical values from the t-distribution are used to calculate the confidence interval for the stratum adjusted difference of proportions. The SAS procedure PROC MIANALYZE is used where the critical value is based on the t-distribution which is different from the analysis not based on MI where normal distribution is used. The large number of observations in our data imply that the critical values from the t-distribution are almost identical to the critical values from the standard normal distribution. The Wilson-Hilferty (Ratitch 2013) transformation is used to pool the CMH statistics from each imputed data set to calculate the p-value.

For analyses where all missing data is imputed only a single time (e.g., non-responder imputation for all missings) or not imputed with the missing assumed Missing Completely at Random (MCAR), treatment comparisons will generally be performed using a CMH test. The magnitude of the treatment difference will be estimated by the difference in response rates between the guselkumab and placebo groups with a 95% confidence interval (CI) calculated based on Wald statistics (Yeonhee 2013). 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.

For endpoints where any portion of the missing data is not imputed, but still needs to be accounted for under Missing at Random (MAR) assumptions, the **Generalized Linear Mixed Model (GLMM)** will be used (for more details see Section 6.13).

For subgroup analyses, a **logistic regression model** using data from all 3 intervention group will be used to obtain odds ratios and associated 95% CI. If any missing data were multiply imputed, the logistic regression model will be run on each of the imputation sets, and the odds ratios from each imputation set will have the log transformation applied before pooling across imputation sets. Once combined, the log transformed odds ratio and 95% CI will be back transformed to their original scale (Ratitch 2013) and presented. The treatment*subgroup interaction p-value will be obtained separately, based on logistic regression on the observed data and with only the data from the 2 intervention groups currently being analyzed fed into the model.

Continuous Efficacy Endpoints

For most continuous endpoints where any portion of the missing data is imputed using MI (see Section 6.11 for technical details), treatment comparisons will be performed using an **Analysis of Covariance (ANCOVA)** model on each of the imputation sets. The estimate of the mean change from baseline is the average of the mean change taken over all the MI data sets. The estimate of the variance of the mean change from baseline is the weighted sum of the average within-imputation variance and the between-imputation variance. The confidence interval for the mean change from baseline uses critical values from the t-distribution. The treatment difference between each guselkumab group versus the placebo group will be tested for each imputation dataset and then the analysis results across all imputation datasets will be combined. The treatment difference in the change from baseline is estimated by the average of the treatment differences over the MI data sets. The estimate of the variance of the treatment difference in the change from baseline is the weighted sum of the average within-imputation variance and the between-imputation variance, under the assumption of homogeneity of variance between intervention groups for performing ANCOVA within each imputation dataset. The confidence interval is based on the critical values from the t-distribution. The large number of observations in our data imply that the critical values from the t-distribution are almost identical to the critical values from the standard normal distribution.

The ANCOVA model will be based on the original scale and will include intervention group, baseline score, and randomization strata levels as the explanatory factors. The model will include data from all the 3 intervention groups.

For analyses where all missing data is imputed only a single time (e.g., change=0 for all missings) or not imputed with the missing assumed MCAR, treatment comparisons will generally be performed using ANCOVA.

For endpoints where any portion of the missing data is not imputed, but still needs to be accounted for under MAR assumptions, the **Mixed-Effect Model Repeated Measures (MMRM)** model will be used (for more details see Section 6.13).

5.3.2. General Data Handling Rules

Missing data will be handled depending on the assumed mechanism behind the missingness, whether the endpoint is a key endpoint, and the estimand being used. Handling rules are summarized in Table 6 below:

Table 6: Data Handling Rules for Missing Data

Estimand	Assumption	Continuous endpoints	Binary response endpoints
Adjusted Composite	MAR due to Natural Disaster or Major Disruption	**Clinical Other Efficacy endpoints**: not imputed but accounted for in MMRM model	**Primary endpoint**: imputed using Full Conditional Specifications (FCS) MI on continuous component scale then dichotomized **Clinical Other Efficacy endpoints**: not imputed but accounted for in GLMM
	MAR due to not entering LTE for reasons OTHER than Lack of Efficacy or Worsening of PsA	**Clinical Other Efficacy endpoints:** not imputed but accounted for in MMRM model	**Clinical Other Efficacy endpoints**: not imputed but accounted for in GLMM
	MAR NOT due to above reasons	**Clinical Other Efficacy endpoints:** not imputed but accounted for in MMRM model	**Primary endpoint:** conservatively imputed as non-response for consistency with historical studies **Clinical Other Efficacy endpoints:** conservatively imputed as non-response for consistency with historical studies
	Missing Not at Random (MNAR)	**N/A**	***For tipping point analyses specifically:*** Sensitivity analysis with exhaustive assessment of all possible combinations of response status for missing data; Sensitivity analysis systematically assessing scenarios which deviate from MAR or MCAR assumption
Adjusted Treatment Policy	MAR	**Radiographic endpoints**: imputed using FCS MI; **Radiographic sensitivity analysis**: 2-step MI	**Radiographic endpoints**: imputed using FCS MI on continuous component scale then dichotomized
	MNAR	***For tipping point analyses specifically:*** Sensitivity analysis systematically assessing scenarios which deviate from MAR or MCAR assumption	**N/A**
	MCAR	**Radiographic endpoints:** not imputed	**Radiographic endpoints:** not imputed
	MAR	**Radiographic supplementary analyses:** imputed using FCS MI	**Clinical supplementary analyses:** imputed using FCS MI on

Table 6: Data Handling Rules for Missing Data

Estimand	Assumption	Continuous endpoints	Binary response endpoints
Treatment Policy			continuous component scale then dichotomized
	MNAR	***For tipping point analyses specifically:*** Supplementary analysis systematically assessing scenarios which deviate from MAR or MCAR assumption	***For tipping point analyses specifically:*** Supplementary analysis systematically assessing scenarios which deviate from MAR or MCAR assumption
	MCAR	**Clinical Other Efficacy endpoints:** not imputed	**Clinical Other Efficacy endpoints:** not imputed

5.3.2.1. Imputing Missing Data Evaluating Deviation from Assumptions (Tipping Point Analyses)

Primary Endpoint

For the **Primary endpoint** (ACR 20 response at Week 24) using the Adjusted Composite Estimand (see Section 5.3.4.2 for definition of estimand), the exhaustive scenario tipping point analyses will be performed to evaluate any deviation from the imputation of: missings NOT due to Natural Disaster or Major Disruption (see Section 5.3.4.2) being imputed as non-responder.

Note that as part of the ICE strategies for the Adjusted Composite Estimand (see Section 5.3.4.3), participants who meet ICEs 1-3 prior to Week 24 are set to non-responder regardless of observed data, and participants who meet ICEs 4 or 5 prior to Week 24 are imputed using FCS MI regardless of observed data. These participants, combined with participants with observed data at Week 24 who did not meet any ICEs prior that timepoint, comprise the not-to-be-varied group for the Adjusted Composite Estimand, and their response rate remains static through the tipping point analysis.

Let T_A be the total number of imputed values to-be-varied in the Active arm, where i of them will be set to 'Yes' response and (T_A-i) of them set to 'No' response. In the same vein, let T_P be the total number of imputed values to-be-varied in the Placebo arm, where j of them will be set to 'Yes' response and (T_P-j) of them set to 'No' response. The range of i is from 0 to T_A, and a range of j is from 0 to T_P, which is an 'exhaustive approach'.

Also for the **Primary endpoint** using the Adjusted Composite Estimand, a second tipping point analyses based on imputed data by MI will be performed to evaluate the impact of imputed data when deviating from the MAR assumption.

- A pair of deltas (e.g., Dg =-0.1, Dp =0.2) will be added to the predicted response rates of each missing data from the MI method depending on guselkumab or placebo group.

- With the new response rate, the missing response will be imputed for N (e.g., N=200) times to generate N multiple imputations based on a Bernoulli distribution. Treatment comparisons will then be performed same as treatment comparison with MI .
- The range of delta values include the scenarios where participants on guselkumab have worse outcomes than participants on placebo.

A third tipping point analysis for the **Primary endpoint** evaluating the impact of imputed data when deviating from the MAR assumption using the Treatment Policy Estimand will be also performed using similar methodology as the second tipping point analysis. The Treatment Policy Estimand considers the ICEs irrelevant to the endpoint, thus all the missing data in the specified analysis population will be imputed and will comprise the to-be-varied group.

Key Confirmatory Secondary Endpoint

For the **Key Confirmatory Secondary endpoint** (change from baseline to Week 24 in modified vdH-S score) using the Adjusted Treatment Policy Estimand (see Section 5.3.5.1.2 for definition of estimand), the tipping point analyses based on imputed data by FCS MI will be performed to evaluate the impact of imputed data when deviating from the MAR assumption.

Note that as part of the ICE strategies for the Adjusted Treatment Policy Estimand (see Section 5.3.5.1.3), participants who meet ICEs 4 or 5 prior to Week 24 are imputed using FCS MI regardless of observed data. These participants, combined with participants with observed data at Week 24 who did not meet ICEs 4 or 5 prior that timepoint, comprise the not-to-be-varied group for the Adjusted Treatment Policy Estimand, and their change scores remain static through the tipping point analysis.

- A delta (e.g., Dg =0.2, Dp =0.1) will be added to the imputed value for each participant with missing value from the MI depending on whether the participant is in the guselkumab or placebo group.
- With the new datasets, treatment comparisons will be performed similar to treatment comparisons with MI data.
- The analysis will be repeated for a range of Dg and Dp by varying Dg and Dp independently, including the scenarios where participants on guselkumab have worse outcomes than participants on placebo.

A second tipping point analysis for the **Key Confirmatory Secondary endpoint** evaluating the impact of imputed data when deviating from the MAR assumption using the Treatment Policy Estimand will be also performed using similar methodology as the first tipping point analysis. The Treatment Policy Estimand considers the ICEs irrelevant to the endpoint, thus all the missing data in the specified analysis population will be imputed and part of the to-be-varied group.

5.3.3. Analysis Specifications

5.3.3.1. Level of Significance

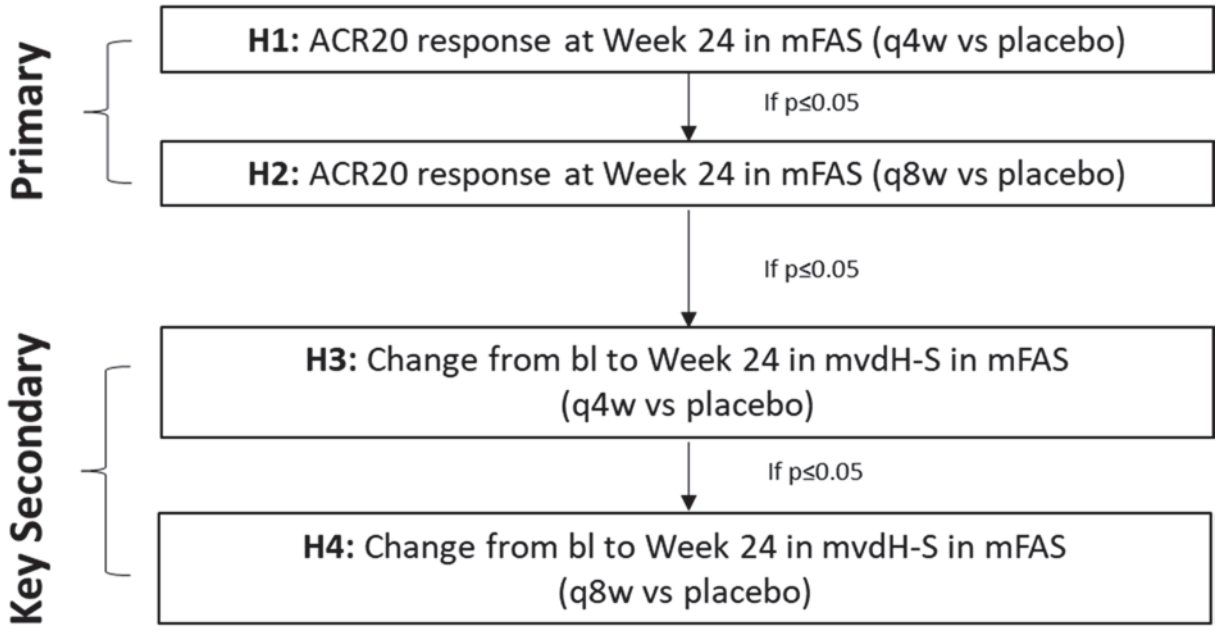
The overall type I error will be controlled among the primary and key confirmatory secondary endpoints at 5% as specified in Section 5.3.3.2.

5.3.3.2. Multiplicity Adjustment for Testing Procedures

This study has 1 primary endpoint (proportion of participants who achieved an ACR 20 response at Week 24) and 1 key confirmatory secondary endpoint (change from baseline in modified vdH-S score at Week 24). With 2 treatment comparisons each, there are a total of 4 hypotheses to be tested. These hypotheses are explicitly listed in Section 2.

The overall Type I error of these 4 hypotheses will be controlled at a significance level of ≤ 0.05. The testing procedure tests the primary and key confirmatory secondary endpoints, for the two regimens of guselkumab vs placebo, in a fixed sequence and each endpoint is tested at the two sided 0.05 level of significance. The fixed sequence testing method tests an endpoint only if the null hypotheses of no difference between the guselkumab regimen and placebo was rejected at the 0.05 level for all the endpoints above it in the sequence. This is shown visually in Figure 2 below.

Figure 2: Multiplicity Control



5.3.4. Primary Endpoint Analysis

The primary endpoint of this study is the proportion of participants who achieved an ACR 20 response at Week 24. This section outlines the definitions and analyses of this primary endpoint.

5.3.4.1. Definition of Endpoint

ACR response is a composite measurement of change in PsA signs and symptoms and is presented as the numerical measurement of improvement in multiple disease assessment criteria (Felson 1993) (Felson 1995). An ACR20 response is defined as:

[redacted]

5.3.4.2. Estimand

Adjusted Composite Estimand (*Primary*)

The ***primary*** analysis for the primary endpoint will be based on the Adjusted Composite Estimand. This estimand is defined by the 5 components:

- **Population**: Participants with active PsA who are biologic naïve
- **Treatment**:
  - Placebo
  - Guselkumab
- **Variable**: ACR20 composite binary response variable at Week 24, where a responder is defined as a participant who achieves ACR20 response at Week 24 and does not experience ICE categories 1 to 3 prior to that time, in the hypothetical situation where Natural Disaster or Major Disruption and associated ICE categories 4 and 5 did not occur.
- **Intercurrent Events**:

1. Discontinued study intervention injections due to any reason **except** due to Natural Disaster or Major Disruption.
2. Initiated or increased the dose of non-biologic DMARDs (MTX, SSZ, HCQ, LEF) or oral corticosteroids over baseline for PsA.
3. Initiated protocol prohibited medications/therapies for PsA.
4. Discontinued study intervention injections due to Natural Disaster or Major Disruption.
5. Severe treatment non-compliance due to Natural Disaster or Major Disruption, defined as ≥2 doses of study intervention missed due to Natural Disaster or Major Disruption.

- **Population level summary**: difference in proportion of responders (as per definition of **Variable** above) between guselkumab group and placebo group.

*Note that in the context of ICEs and missing data, the following are defined:

1. **Natural Disaster**: site closure, site access restrictions, or lockdowns caused by COVID-19.
2. **Major Disruption**: the disruption involving Ukraine and neighboring countries/territories beginning February 24, 2022

This estimand examines the difference in proportion of participants who achieve ACR20 response after 24 weeks without increasing/initiating select background PsA medications or discontinuing study intervention (for reasons not due to Natural Disaster or Major Disruption) prior to that point and in the hypothetical scenario where the Natural Disaster or Major Disruptions did not occur,

between each guselkumab group individually versus placebo, amongst participants with active PsA who are biologic naïve.

Additional supplementary analyses will be conducted on the primary endpoint using the Treatment Policy Estimand, aiming to achieve a robust treatment effect for regulatory decision making:

Treatment Policy Estimand (*Supplementary*)

This estimand is defined by the 5 components:

- **Population**: same as adjusted composite estimand
- **Treatment**: same as adjusted composite estimand
- **Variable**: ACR20 composite binary response variable at Week 24, where a responder is defined as a participant who achieves ACR20 response at Week 24 irrespective of background PsA medication or adherence to study intervention.
- **Intercurrent Events**: the definition of the 5 categories of ICE are the same as the Adjusted Composite Estimand
- **Population level summary**: difference in proportion of responders (as per definition of **Variable** above) between guselkumab group and placebo group

This estimand examines the difference in proportion of participants who achieve ACR20 response after 24 weeks irrespective of background PsA medication or adherence to study intervention, between each guselkumab group individually versus placebo, amongst participants with active PsA who are biologic naïve.

5.3.4.3. Analysis Methods

The primary efficacy analysis of the primary endpoint will be analyzed at Week-24 DBL based on the **Adjusted Composite Estimand**.

In the primary efficacy analysis, data from all participants in mFAS (Section 4.2) will be analyzed according to randomized intervention group regardless of the treatment actually received.

ICE Strategies

For the Adjusted Composite Estimand, ICEs 1-3 are incorporated as part of the endpoint using the composite strategy. This estimand acknowledges that a participant increasing their background PsA medications or electing to discontinue study intervention for reasons other than Natural Disaster or Major Disruption prior to the assessment timepoint is an unfavorable outcome, and thus for the purpose of analysis they are considered treatment failures (TF) and set to ACR20 non-responders at Week 24.

This estimand also possesses a hypothetical component, regarding significant unplanned changes to study conduct as a result of Natural Disaster or Major Disruption. This estimand seeks to estimate the treatment effect of study intervention as if the above, as well as ICEs directly resulting from it, did not occur. Thus, data observed after the occurrence of ICEs 4 and 5 will not be used, and will be considered MAR and imputed using FCS MI on the individual ACR components.

Should there be a participant who meets ICEs from both the composite strategy and the hypothetical strategy, the composite strategy has precedence and they will be set to an ACR20 non-responder at Week 24.

Handling Rules for Missing Data

After ICE strategies have been implemented, remaining missing data will be handled as follows:

1. **Missing Week 24 ACR 20 response due to Natural Disaster or Major Disruption,** will be assumed to be MAR and imputed using FCS MI on the individual ACR components.
2. **Missing Week 24 ACR 20 response NOT due to Natural Disaster or Major Disruption,** will be considered MAR but imputed conservatively as an ACR 20 non-responder (NRI rule) at Week 24 to be consistent with historical studies.

Analysis Testing

The treatment difference between each guselkumab group versus the placebo group will be tested using a CMH test stratified by the randomization strata levels for each imputation set, and the Wilson-Hilferty transformation will be applied to the CMH statistics across the imputation sets. The transformed CMH statistics will be combined to calculate the p-values according to Rubin (Rubin 1987). The magnitude of the treatment difference will be estimated by the difference in ACR 20 response rates between the guselkumab and placebo groups with a 95% CI calculated based on Wald statistics.

In order to control the overall Type 1 error rate, the primary endpoint will be tested in a fixed sequence.

1. Guselkumab 100 mg q4w versus placebo in ACR 20 response at Week 24, among mFAS participants
2. Guselkumab 100 mg at Weeks 0, 4, and then q8w versus placebo in ACR 20 response at Week 24, among mFAS participants

If the first test is significant at a 2-sided α-level of 0.05, the study will be considered positive and the second test can then be performed.

5.3.4.4. Sensitivity and Supplementary Analyses

1. (**Sensitivity Analysis 1**) To evaluate the robustness of the **Adjusted Composite Estimand** regarding the assumptions for missing data, sensitivity analyses with the exhaustive two-dimensional scenario tipping point analyses will be performed. The analysis will be conducted for an 'exhaustive approach' testing all combinations of missing data imputation as responder and NR (Section 5.3.2.1). Note that data imputed as part of the ICE strategy (i.e, non-responder due to meeting ICEs 1-3, FCS MI due to meeting ICEs 4 and 5) will be performed prior to the tipping point analysis. The chi-square test will be used to compare each guselkumab group versus the placebo group. This will avoid the complication of having to incorporate baseline stratification in the mix when generating all combinations of responders and NR for the missing data for CMH test. As all combinations will be presented, both the points where tipping occurs, as well as the proportion of non-tipping combinations, are of interest.

2. (**Sensitivity Analysis 2**) An additional tipping point analysis for the **Adjusted Composite Estimand** will be performed. Note that data imputed as part of the ICE strategy (i.e, non-responder due to meeting ICEs 1-3, FCS MI due to meeting ICEs 4 and 5) will be performed prior to the tipping point analysis. In this analysis, a pair of deltas will be added to the predicted response rates from MI method depending on guselkumab or placebo group to new MI datasets (Section 5.3.2.1). The same analysis method as in the primary analysis will be applied for the pairs of deltas. The analysis will be done for pairs of delta values include the scenarios where participants on guselkumab have worse outcomes than participants on placebo.
3. (**Sensitivity Analysis 3**) To assess the effect of including all randomized participants for the **Adjusted Composite Estimand**, even those from sites rendered unable to support key study operations due to Natural Disaster or Major Disruption, the same analysis of ACR 20 response at Week 24 as the primary analysis will be performed on the FAS analysis set.
4. (**Sensitivity Analysis 4**) To assess the effect of excluding all Ukrainian participants from the analysis for the **Adjusted Composite Estimand**, as all sites in Ukraine have been impacted by Major Disruption (e.g., study intervention interruptions, temporary site closures, etc) to a larger or smaller extent. The same analysis of ACR 20 response at Week 24 as the primary analysis will be performed on the mFAS-UKR analysis set.
5. (**Sensitivity Analysis 5**) To assess the effect of protocol deviations which may affect efficacy, a sensitivity analysis similar to the main analysis will be performed, but based on the PPAS.
6. (**Sensitivity Analysis 6**) As the Major Disruption has the potential to prevent source data verification (SDV) from taking place (e.g, due to site closure, site inaccessibility, etc), an analysis will be conducted to assess the impact by excluding affected participants. The same analysis as the main analysis will be performed on the mFAS analysis set, excluding participants for which expected SDV was not completed for data on or prior to Week 24.
7. (**Supplementary Analysis 1**) To support regulatory decision making, the **Treatment Policy Estimand** (Section 5.3.4.2) will also be evaluated as a supplementary analysis. In this analysis, the observed ACR 20 response for all participants will be used regardless of whether or not ICE criteria are met prior to Week 24, and the missing ACR 20 response will be imputed by FCS MI (Section 5.3.1) on the component level under the assumption that data are MAR. Treatment comparisons for each imputation data set will be based on a CMH test stratified by randomization strata levels, as per Section 5.3.1. This analysis will be conducted based on the mFAS-UKR analysis set.
8. (**Supplementary Analysis 2**) An additional tipping point analysis for the **Treatment Policy Estimand** will be performed. The observed ACR 20 response for participants will be used regardless of whether or not ICE criteria are met prior to Week 24, and the missing ACR 20 response will be imputed by FCS MI (Section 5.3.1) on the component level under the assumption that data are MAR. In this analysis, a pair of deltas will be added to the predicted response rates from MI method depending on guselkumab or placebo group to new MI datasets (Section 5.3.2.1) to evaluate deviation from the assumption of MAR for missing data. The same analysis method as in the primary analysis will be applied for the pairs of deltas. The analysis will be done for pairs of delta values including scenarios where participants on guselkumab have worse outcomes than participants on placebo. This analysis will be conducted based on the mFAS-UKR analysis set.

5.3.4.5. Subgroup Analyses

Subgroup analyses will be performed using the Adjusted Composite Estimand using a logistic regression model on the multiply imputed data to evaluate treatment consistency in proportion of participants who achieve an ACR 20 response at Week 24 over baseline demographics, baseline disease characteristics, and prior and baseline medication use. A forest plot will be produced for all subgroups listed in Section 5.5.6. The odds ratios and the corresponding 95% Cis will also be provided for each of subgroups (Section 5.3.1). In addition, the p-values for interaction of the intervention groups and the subgroups will also be provided when a subgroup has at least 2 categories.

If the number of participants in a subgroup is too small (eg., < 10), subgroups may be pooled for analyses.

5.3.4.6. Summary of Analyses Related to the Primary Endpoint of ACR 20 Response at Week 24

Table 7 below provides an overview on all the analyses related to the primary endpoint of ACR 20 response at Week 24, the estimands, the data handling rules to be used, and the analysis methods and summary statistics. Section 5.3.2 provides a summary of the Multiple Imputation method.

Table 7: Summary of Analyses Related to the Primary Endpoint of ACR 20 Response at Week 24

Analysis (Estimand) – Analysis Set	ICE/Missing data imputation	Additional notes
Primary Analysis (based on Adjusted Composite Estimand) – mFAS	ICEs 1-3 considered as TF; ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data due to Natural Disaster or Major Disruption assumed MAR, imputed via FCS MI Missing data not due to Natural Disaster or Major Disruption assumed MAR but conservatively imputed as non-response	• Summarized descriptively • Pooled response rates, treatment difference in response rates and 95% CI across multiply imputed data sets • P-value from CMH statistic with Wilson-Hilferty transformation across multiply imputed data sets
Sensitivity Analysis 1 (based on Adjusted Composite Estimand) – mFAS	ICEs 1-3 considered as TF; ICEs 4 and 5, data not used, but rather imputed using FCS MI, then averaged across MI datasets. Missing data rules not applied. Response rate of missing data varied with all possible combinations of response status for missing data. Single imputation.	• Exhaustive 2-dimensional tipping point analysis; graphical • The chi-squared test is used to compare intervention groups, for each coordinate on the graph separately.
Sensitivity Analysis 2 (based on Adjusted Composite Estimand) - mFAS	ICEs 1-3 considered as TF; ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data imputed via FCS MI, then imputed data response rate	• 2-dimensional tipping point analysis to assess robustness of analysis results should there be deviation from MAR

Table 7: Summary of Analyses Related to the Primary Endpoint of ACR 20 Response at Week 24

Analysis (Estimand) – Analysis Set	ICE/Missing data imputation	Additional notes
	varied and generated using Bernoulli distribution	assumption of missing data; graphical • P-value from CMH statistic with Wilson-Hilferty transformation across multiply imputed data sets, for each coordinate on the graph separately.
Sensitivity Analysis 3 (based on Adjusted Composite Estimand) – FAS	ICEs 1-3 considered as TF; ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data due to Natural Disaster or Major Disruption assumed MAR, imputed via MI Missing data not due to Natural Disaster or Major Disruption assumed MAR but conservatively imputed as non-response	• Summarized descriptively • Pooled response rates, treatment difference in response rates and 95% CI across multiply imputed data sets • P-value from CMH statistic with Wilson-Hilferty transformation across multiply imputed data sets
Sensitivity Analysis 4 (based on Adjusted Composite Estimand) – mFAS-UKR	ICEs 1-3 considered as TF; ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data due to Natural Disaster or Major Disruption assumed MAR, imputed via FCS MI Missing data not due to Natural Disaster or Major Disruption assumed MAR but conservatively imputed as non-response	• Summarized descriptively • Pooled response rates, treatment difference in response rates and 95% CI across multiply imputed data sets • P-value from CMH statistic with Wilson-Hilferty transformation across multiply imputed data sets
Sensitivity Analysis 5 (based on Adjusted Composite Estimand) – PPAS	ICEs 1-3 considered as TF; ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data due to Natural Disaster or Major Disruption assumed MAR, imputed via FCS MI Missing data not due to Natural Disaster or Major Disruption assumed MAR but conservatively imputed as non-response	• Summarized descriptively • Pooled response rates, treatment difference in response rates and 95% CI across multiply imputed data sets • P-value from CMH statistic with Wilson-Hilferty transformation across multiply imputed data sets
Sensitivity Analysis 6 (based on Adjusted Composite Estimand) – mFAS, Excluding Participants with Non-SDV'd Data Through Week 24	ICEs 1-3 considered as TF; ICEs 4 and 5, data not used, but rather imputed using FCS MI	• Summarized descriptively • Pooled response rates, treatment difference in response rates and

Table 7: Summary of Analyses Related to the Primary Endpoint of ACR 20 Response at Week 24

Analysis (Estimand) – Analysis Set	ICE/Missing data imputation	Additional notes
	Missing data due to Natural Disaster or Major Disruption assumed MAR, imputed via FCS MI Missing data not due to Natural Disaster or Major Disruption assumed MAR but conservatively imputed as non-response	95% CI across multiply imputed data sets • P-value from CMH statistic with Wilson-Hilferty transformation across multiply imputed data sets
Supplementary Analysis 1 (based on Treatment Policy Estimand) – mFAS-UKR	No ICEs considered as TF; Missing data assumed MAR, imputed via FCS MI	• Summarized descriptively • Pooled response rates, treatment difference in response rates and 95% CI across multiply imputed data sets • P-value from CMH statistic with Wilson-Hilferty transformation across multiply imputed data sets
Supplementary Analysis 2 (based on Treatment Policy Estimand) – mFAS-UKR	No ICEs considered as TF; Missing data imputed via FCS MI, then imputed data response rate varied and generated using Bernoulli distribution	• 2-dimensional tipping point analysis to assess robustness of analysis results should there be deviation from MAR assumption of missing data; graphical • P-value from CMH statistic with Wilson-Hilferty transformation across multiply imputed data sets, for each coordinate on the graph separately.
Subgroup Analyses (based on Adjusted Composite Estimand) – mFAS subgroups	ICEs 1-3 considered as TF; ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data due to Natural Disaster or Major Disruption assumed MAR, imputed via FCS MI Missing data not due to Natural Disaster or Major Disruption assumed MAR but conservatively imputed as non-response	• Odds ratio and 95% CI for treatment comparison • P-value from logistic regression for the interaction of intervention group and subgroup variable • Graphical: forest plots

5.3.5. Secondary Endpoint Analysis

5.3.5.1. Key Confirmatory Secondary Endpoint

The key confirmatory secondary endpoint in this study is the change from baseline in modified van der Heijde-Sharp (vdH-S) score at Week 24, under the subgroup of participants without high baseline radiographic variability, as well as over the entire mFAS.

5.3.5.1.1. Definition of Endpoint

The vdH-S score is an original vdH-S score (van der Heijde 1992) modified for PsA. The modification for PsA includes addition of distal inter-phalangeal (DIP) joints of both hands scored for erosions and joint space narrowing (JSN), and assessments of radiographic features known as "pencil in cup" (PIC) and "gross osteolysis" (GO) that are specific to PsA. The vdH-S score is a measurement of progression in structural damage. It is the sum of joint erosion score and JSN score. The erosion score and JSN score, respectively, is a measurement of 2 types of structural damage.

The **joint erosion score** is a summary of erosion severity in 40 joints of the hands (20 joints per hand) and 12 joints in the feet (6 in each foot). Each joint is scored according to the surface area involved, from 0 to 5, with 0 indicating no erosion and 5 indicating complete collapse of bone. To identify the presence of PIC and GO in the hands, a modification of erosions scores of 6 and 7 are applied by IRC radiologists. For joints with one of these abnormalities, the maximum score of 5 will be applied. To identify the presence of PIC and GO in the feet, a modification of erosions scores of 11 and 12 are applied by IRC radiologists. For joints with one of these abnormalities the maximum score of 10 will be applied. Therefore, the maximum erosion score for a hand joint is 5 and the maximum erosion score for hands is 200. Because each side of a foot joint is graded on the scale of 0 to 5, the maximum erosion score for a foot joint is 10 and the maximum erosion score for feet is 120. Thus, the maximal **erosion score (i.e., hand erosion score + foot erosion score)** is 320.

The **joint space narrowing score** summarizes the severity of JSN in 40 joints in the hands and 12 joints of the feet. Assessment of JSN is scored from 0 to 4, with 0 indicating no JSN and with 4 indicating absence of a joint space, presumptive evidence of ankylosis or complete luxation. Therefore, the maximum JSN score for a hand joint is 4 and the maximum JSN score for hands is 160. The maximum JSN score for a foot joint is 4 and the maximum JSN score for feet is 48. Thus, the maximal **JSN score (i.e., hand JSN score + foot JSN score)** is 208.

The maximal erosion score of 320 combined with the maximal JSN score of 208 gives worst possible **modified vdH-S score (ie, erosion score + JSN score)** of 528.

Joint Evaluability Rules specified in Section 6.10 for joint evaluation will be applied to those joints with surgery/joint replacement or with radiographically insufficient data for reading. For

participants with incomplete set of evaluable joints, Erosion and JSN Score Adjustment Rules described in Section 6.10 will be applied to determine the ultimate **sub-scores (i.e., scores of hand erosion, hand JSN, foot erosion, and foot JSN)** for each reader. A composite score [including erosion, JSN, **hand (i.e., hand erosion + hand JSN), foot (i.e., foot erosion + food JSN)**, and modified vdH-S scores)] will be set to missing if any of its corresponding sub-scores is missing.

Confirmation will occur when the 2 primary readers do not agree with each other with respect to change from baseline in modified vdH-S score at any post-baseline visit. The criteria for triggering confirmation can be found in Section 6.10. ***Note*** that the criteria for triggering confirmation should be determined based on the observed data (i.e., data without application of those data handling rules specified in Section 5.3.2). Section 6.10 outlines the rules to select which 2 readers' scores to be used in the analysis for participants with confirmation.

The final scores or sub-scores at each visit are the average of corresponding scores from the 2 primary readers for a participant without confirmation, the 2 primary readers for a participant with confirmation but both primary readers equally distant from confirmation reader, and the confirmation reader and the primary reader closest to the confirmation reader otherwise for a participant with confirmation .

Change from baseline in modified vdH-S score measures the change in progression of structural damage, where a negative change indicates an improvement and a positive change indicates a worsening.

5.3.5.1.2. Estimand

Adjusted Treatment Policy Estimand *(Primary)*

This estimand is defined by the 5 components:

- **Population**: Participants with active PsA who are biologic naïve
- **Treatment**:
  - Placebo
  - Guselkumab
- **Variable**: change from baseline in modified vdH-S score at Week 24 irrespective of background PsA medication or adherence to study intervention, in the hypothetical situation where Natural Disaster or Major Disruption and associated ICE categories 4 and 5 did not occur.
- **Intercurrent Events**:

1. Discontinued study intervention injections due to any reason **except** due to Natural Disaster or Major Disruption.
2. Initiated or increased the dose of non-biologic DMARDs (MTX, SSZ, HCQ, LEF) or oral corticosteroids over baseline for PsA.
3. Initiated protocol prohibited medications/therapies for PsA.

4. Discontinued study intervention injections due to Natural Disaster or Major Disruption.
5. Severe treatment non-compliance due to Natural Disaster or Major Disruption, defined as ≥2 doses of study intervention missed due to Natural Disaster or Major Disruption.

- **Population level summary**: difference in mean changes (as per definition of **Variable** above)between guselkumab group and placebo group.

*Note that in the context of ICEs and missing data, the following are defined:

1. **Natural Disaster: site closure**, site access restrictions, or lockdowns caused by COVID-19.
2. **Major Disruption**: the disruption involving Ukraine and neighboring countries/territories beginning February 24, 2022

This estimand examines the difference in mean change in modified vdH-S score after 24 weeks irrespective of background PsA medication or adherence to study intervention, in the hypothetical scenario where Natural Disaster or Major Disruptions did not occur, between each guselkumab group individually versus placebo, amongst participants with active PsA who are biologic naïve.

Treatment Policy Estimand *(Supplementary)*

This estimand is defined by the 5 components:

- **Population**: same as adjusted treatment policy estimand
- **Treatment**: same as adjusted treatment policy estimand
- **Variable**: change from baseline in modified vdH-S score at Week 24 irrespective of background PsA medication or adherence to study intervention.
- **Intercurrent Events**: the definition of the 5 categories of ICE, and the definition of Natural Disaster and Major Disruption, are the same as for the Adjusted Treatment Policy Estimand.
- **Population level summary**: difference in mean changes (as per definition of **Variable** above) between guselkumab group and placebo group.

This estimand examines the difference in mean change in modified vdH-S score after 24 weeks irrespective of background PsA medication or adherence to study intervention, between each guselkumab group individually versus placebo, amongst participants with active PsA who are biologic naïve.

5.3.5.1.3. Analysis Methods

The main analysis of the key confirmatory secondary endpoint will be analyzed at Week-24 DBL based on the Adjusted Treatment Policy Estimand.

In the primary efficacy analysis, data from all participants in mFAS(Section 4.2) will be analyzed according to randomized intervention group regardless of the treatment actually received.

ICE Strategies

For the Adjusted Treatment Policy Estimand, the treatment policy strategy will be used for ICEs 1 to 3, wherein all observed data collected for the endpoint is used. The occurrence of the intercurrent event is irrelevant: the value for the variable of interest is used regardless of whether or not the intercurrent event occurs.

The hypothetical strategy will be used for ICEs 4 and 5, wherein observed data collected after the ICE will not be used and will be assumed to be MAR, then imputed using FCS MI.

For participants experiencing multiple ICEs, an ICE in categories 4 and 5 will supersede an ICE in categories 1 to 3.

Handling Rules for Missing Data for Adjusted Treatment Policy Estimand

After ICE strategies have been implemented, remaining missing data will be handled as follows:

- **Missing data for any reason** will be imputed using FCS MI as described in Section 5.3.2.

Analysis Testing

Data from unscheduled visits which do not fit into the analysis window (Section 5.1.1.2) will not be used. Treatment comparisons for each imputation data set will be based on an ANCOVA model adjusted for baseline score and randomization strata levels. The analysis results from the N imputation datasets will be combined, according to Rubin (Rubin 1987), and the p-value for testing the treatment difference will be obtained.

In order to control the overall Type 1 error rate, the key confirmatory secondary endpoint will be tested in a fixed sequence if both hypotheses of the primary endpoint are statistically significant.

3. Guselkumab 100 mg q4w SC is superior to treatment with placebo SC as measured by change from baseline in modified vdH-S score at Week 24
4. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to treatment with placebo SC as measured by change from baseline in modified vdH-S score at Week 24.

5.3.5.2. Sensitivity and Supplementary Analyses

1. (**Sensitivity Analysis 1**) To evaluate the robustness of the **Adjusted Treatment Policy Estimand** regarding the assumptions for missing data, two-dimensional tipping point analyses based on FCS MI imputed data will be performed to evaluate the deviation from the assumption of MAR for missing data (Section 5.3.2.1) using the same MI imputed dataset as that used for the main analysis. Note that data imputed as part of the ICE strategy (MI due to meeting ICEs 4 and 5) will be performed prior to the tipping point analysis. The same analyses method as that for the main analysis, will be fitted for each combination of deltas. The analysis results from the N imputation datasets will be combined, according to Rubin (Rubin 1987) and the p-value for testing the treatment difference will be obtained.

2. (**Sensitivity Analysis 2**) To assess the effect of including all randomized participants, even those from sites rendered unable to support key study operations due to Natural Disaster or Major Disruption, the same analysis as the main analysis will be performed on the FAS analysis set.
3. (**Sensitivity Analysis 3**) To assess the influence of outliers and extreme observations, a trimmed analysis will be performed. The same analysis method as for the main analysis, ANCOVA on multiply imputed data, will be performed on a subset of data where k% (k ranging from 1 to 10) of data are removed from the highest and lowest change scores of modified vdH-S in each intervention group.
4. (**Sensitivity Analysis 4**) To assess the influence of difficult to read radiographs. The same analysis method as for the main analysis, ANCOVA on multiply imputed data, will be performed on a subset of data where participants randomized to the HRV randomization strata are excluded.
5. (**Sensitivity Analysis 5**) An analysis similar to the main analysis, except replacing the randomization stratification factor in the ANCOVA analysis model with the baseline covariates: number of joints with erosion (numeric), log-transformed CRP (mg/dL, numeric), and oral corticosteroid use (Y/N).
6. (**Sensitivity Analysis 6**) To assess the impact of a narrower baseline window, the same analysis as the main analysis will be performed, but only participants with baseline and Week 24 observations within +/-2 weeks of the expected assessment time, ie, study week -2 for baseline and study week 24 for Week 24 respectively, are included in the analysis.
7. (**Sensitivity Analysis 7**) To assess the impact of using the original analysis window, the same analysis as the main analysis will be performed, but both baseline and Week 24 observations within +/- 8 weeks of the expected assessment time, ie, study week -2 for baseline and study week 24 for Week 24 respectively, are included in the analysis.
8. (**Sensitivity Analysis 8**) To use all collected post-bl data even outside of the 2 week analysis window, a 2-step MI will be used for imputation: 1) Missing Week 24 data from participants with an x-ray measurement at baseline and a measurement post-baseline during the placebo controlled period will be imputed through a mixed effect linear growth curve (MLGC) model (see Section 6.13); 2) Other missing data will be imputed using the FCS regression method based on the imputed dataset from the 1st step. The 2-step MI will be applied only to the modified vdH-S score and not the erosion or JSN scores.
9. (**Sensitivity Analysis 9**) To assess the effect of excluding all Ukrainian participants from the analysis, as all sites in Ukraine have been impacted by Major Disruption (e.g., study intervention interruptions, temporary site closures, etc) to a larger or smaller extent. The same analysis as the main analysis will be performed on the mFAS-UKR analysis set.
10. (**Sensitivity Analysis 10**) As the Major Disruption has the potential to prevent SDV from taking place (e.g, due to site closure, site inaccessibility, etc), an analysis will be conducted to assess the impact by excluding affected participants. The same analysis as the main analysis will be performed on the mFAS analysis set, excluding participants for which expected SDV was not completed for data on or prior to Week 24.
11. (**Supplementary Analysis 1**) To support regulatory decision making, an analysis based on the **Treatment Policy Estimand** will be performed. The observed modified vdH-S score for participants will be used regardless of whether or not ICE criteria are met prior to Week 24,

and missing modified vdH-S score will be imputed using FCS MI under the assumption that data are MAR. The same analysis method as the main analysis will be performed. The analysis will be performed based on the mFAS-UKR analysis set.

12. (**Supplementary Analysis 2**) A two-dimensional tipping point analysis for the **Treatment Policy Estimand** will be performed. The observed modified vdH-S score for participants will be used regardless of whether or not ICE criteria are met prior to Week 24, and missing modified vdH-S score will be imputed using FCS MI under the assumption that data are MAR. Varying pairs of deltas will then be added to the imputed values (Section 5.3.2.1) to evaluate deviation from the assumption of MAR for missing data. The same analysis method as in the main analysis of this endpoint will be applied for each pair of deltas. The analysis will be done for pairs of delta values including scenarios where participants on guselkumab have worse outcomes than participants on placebo. This analysis will be performed based on the mFAS-UKR analysis set.

5.3.5.3. Subgroup Analyses

Subgroup analyses will be performed using the Adjusted Treatment Policy Estimand using the same ANCOVA model on the multiply imputed data as the main analysis to evaluate treatment consistency over baseline demographics, baseline disease characteristics, baseline modified vdH-S score, and prior and baseline medication use. A forest plot will be produced for all subgroups listed in Section 5.5.6. The LSMean for each intervention group, LSMean difference between intervention groups, and the corresponding 95% CIs will also be provided for each of subgroups (Section 5.3.1). In addition, the p-values for interaction of the intervention groups and the subgroups will also be provided when a subgroup has at least 2 categories, and will be calculated based on the **observed data**.

If the number of participants in a subgroup is too small (eg., < 10), subgroups may be pooled for analyses.

5.3.5.4. Exploratory Analyses

An exploratory analysis may be conducted for the change from baseline to Week 24 in modified vdH-S, similar to the main analysis, except that data observed after missing any active (ie, guselkumab) dose will not be used and instead imputed using FCS MI assuming MAR. Additional analyses of radiographic data will be explored using computer vision based AI models.

5.3.5.5. Summary of Analyses Related to the Key Confirmatory Secondary Endpoint of Change form Baseline to Week 24 in Modified vdH-S Score

Table 8 below provides an overview on all the analyses related to the key confirmatory secondary endpoint of change from baseline to Week 24 in modified vdH-S score, the data handling rules to be used, and the analysis methods and summary statistics. For subgroup analyses the analysis sets are further subset to the individual subgroups. Section 5.3.2 provides a summary of the Multiple Imputation method.

Table 8: Summary of Analyses Related to the Key Confirmatory Secondary Endpoint of Change from Baseline in Modified vdH-S Score at Week 24

Analysis/Population/Estimand	ICE/Missing data imputation	Additional notes
Main Analysis/ mFAS/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data assumed MAR, imputed via FCS MI	• Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets
Sensitivity Analysis 1/ mFAS/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data imputed using FCS MI, then imputed data varied	• 2-dimensional tipping point analysis to assess robustness of analysis results should there be deviation from MAR assumption of missing data; graphical • P-value from combining ANCOVA model treatment difference across multiply imputed datasets, for each coordinate on the graph separately.
Sensitivity Analysis 2/ FAS/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data assumed MAR, imputed via FCS MI	• Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets
Sensitivity Analysis 3/ mFAS/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data assumed MAR, imputed via FCS MI	• Trimmed analysis: k% (k ranging from 1 to 10) of data are removed from the highest and lowest change scores of in each intervention group. • Summarized descriptively at each level of k • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets, separately at each level of k

Table 8: Summary of Analyses Related to the Key Confirmatory Secondary Endpoint of Change from Baseline in Modified vdH-S Score at Week 24

Analysis/Population/Estimand	ICE/Missing data imputation	Additional notes
Sensitivity Analysis 4/ mFAS excluding HRV participants/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data assumed MAR, imputed via FCS MI	• Analysis for non-HRV participants • Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets
Sensitivity Analysis 5/ mFAS/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data assumed MAR, imputed via FCS MI	• Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets. ANCOVA model, randomization stratification factor replaced with bl #joints with erosions, CRP, oral cort use
Sensitivity Analysis 6/ mFAS excluding participants with baseline assessment outside +/- 2 weeks of Week -2/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI	• Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets
Sensitivity Analysis 7/ mFAS/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI	• Analysis window using +/- 8 weeks for both bl and Week 24 • Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets
Sensitivity Analysis 8/ mFAS/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using MI 2-step MI	• Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets

Table 8: Summary of Analyses Related to the Key Confirmatory Secondary Endpoint of Change from Baseline in Modified vdH-S Score at Week 24

Analysis/Population/Estimand	ICE/Missing data imputation	Additional notes
Sensitivity Analysis 9/ mFAS-UKR/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data assumed MAR, imputed via FCS MI	• Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets
Sensitivity Analysis 10/ mFAS, Excluding Participants with Non-SDV'd Data Through Week 24/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data assumed MAR, imputed via FCS MI	• Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets
Supplementary Analysis 1/ mFAS-UKR/ Treatment Policy	ICEs considered irrelevant Missing data assumed MAR, imputed via FCS MI	• Summarized descriptively • Treatment comparisons including p-value from combining ANCOVA model treatment difference across multiply imputed datasets
Supplementary Analysis 2/ mFAS-UKR/ Treatment Policy	ICEs considered irrelevant Missing data imputed using FCS MI, then imputed data varied	• 2-dimensional tipping point analysis to assess robustness of analysis results should there be deviation from MAR assumption of missing data; graphical • P-value from combining ANCOVA model treatment difference across multiply imputed datasets, for each coordinate on the graph separately.
Subgroup Analyses/ mFAS into subgroup categories/ Adjusted Treatment Policy	ICEs 4 and 5, data not used, but rather imputed using FCS MI Missing data assumed MAR, imputed via FCS MI	• LSMean difference and 95% CI from combining ANCOVA model estimates across multiply imputed data • P-value from ANCOVA model for the interaction of intervention group and subgroup variable, using

Table 8: Summary of Analyses Related to the Key Confirmatory Secondary Endpoint of Change from Baseline in Modified vdH-S Score at Week 24

Analysis/Population/Estimand	ICE/Missing data imputation	Additional notes
		observed data and NOT MI imputed data • Graphical: forest plots

5.3.6. Other Efficacy Endpoints

In addition to the primary and key confirmatory secondary endpoints, other efficacy analyses related to reduction of signs and symptoms and physical function, skin disease, nail psoriasis, joint structural damage, and health related quality of life will be analyzed. These endpoints are NOT adjusted for multiplicity, and any p-values calculated will be considered nominal.

Treatment comparisons will ONLY be performed up to the Week 24 visit. Subsequent visits through Week 156 will be limited to descriptive summaries or model based estimates by study intervention.

5.3.6.1. Estimands

The estimand is composed of 5 components: Population, treatment, variable, intercurrent events, and population level summary. Of these, population, variable, and population level summary vary across the different endpoints, while treatment and intercurrent events are constant. For all endpoints, the treatments are placebo and the 2 guselkumab dose groups. Additionally, the definition of intercurrent events are defined below.

Adjusted Composite Estimand (for non-radiographic endpoints)

This estimand was previously defined for the primary analysis of ACR 20 response at Week 24 in Section 5.3.4.2. The ICEs are generalized across endpoints and for all efficacy visits below:

- **Intercurrent Events**:

1. Discontinued study intervention injections due to any reason **except** due to Natural Disaster or Major Disruption.
2. Initiated or increased the dose of non-biologic DMARDs (MTX, SSZ, HCQ, LEF) or oral corticosteroids over baseline for PsA.
3. Initiated protocol prohibited medications/therapies for PsA.
4. Discontinued study intervention injections due to Natural Disaster or Major Disruption.
5. Severe treatment non-compliance due to Natural Disaster or Major Disruption. This is defined for a given visit, when the total number of doses of study intervention missed due to Natural Disaster or Major Disruption exceeds 30% of the total protocol defined doses from Week 0 up to and including that visit. For Week 20, this amounts to ≥2 dose missed:

6. Decided to NOT enter the LTE due to lack of efficacy OR adverse event of worsening of PsA (only relevant after Week 48).

*Note that in the context of ICEs and missing data, the following are defined:

1. **Natural Disaster**: site closure, site access restrictions, or lockdowns caused by COVID-19.
2. **Major Disruption**: the disruption involving Ukraine and neighboring countries/territories beginning February 24, 2022

Treatment Policy Estimand (for non-radiographic endpoints)

This estimand was previously defined for the primary endpoint of ACR 20 response at Week 24 in Section 5.3.4.2 and for the change from baseline in modified vdH-S score at Week 24 in Section 5.3.5.1.2. The ICEs and their handling are generalized across endpoints and for all efficacy visits below:

- **Intercurrent Events**: the definition of the 6 categories of ICE are the same as the generalized Adjusted Composite Estimand in this Section.

Adjusted Treatment Policy Estimand (for radiographic endpoints only)

This estimand was previously defined for the main analysis of change from baseline to Week 24 in modified vdH-S score in Section 5.3.5.1.2. The ICEs are generalized across endpoints and for all efficacy visits below:

- **Intercurrent Events**: the definition of the 6 categories of ICE are the same as the generalized Adjusted Composite Estimand in this Section.

5.3.6.2. Endpoints Related to Reduction of Signs and Symptoms and Physical Function

In this study, Other Efficacy endpoints for reduction of signs and symptoms and physical function include those related to ACR responses, HAQ-DI, DAS28 (CRP), modified PsARC, enthesitis (LEI), dactylitis, mCPDAI, DAPSA, MDA, and VLDA.

All Other Efficacy endpoints related to reduction of signs and symptoms and physical function will be conducted under the mFAS population for Adjusted Composite Estimand, and mFAS-UKR for Treatment Policy Estimand, unless explicitly stated otherwise.

5.3.6.2.1. ACR Related Endpoints

ACR 20 response was previously defined in Section 5.3.4.1. ACR 50 and ACR 70 responses are defined similarly to ACR 20 response, except that the improvement threshold of 20% from baseline in ACR 20 response is replaced by 50% and 70%, respectively.

The ACR related Other Efficacy endpoints include:

- Proportions of participants who achieve ACR 20, ACR 50, and ACR 70 responses by visit over time through Week 156.

- Proportion of participants who maintain an ACR 20 response at Week 48, Week 96, and Week 156 among the participants who achieved an ACR 20 response at Week 24.
- Proportion of participants who maintain an ACR 50 response at Week 48, Week 96, and Week 156 among the participants who achieved an ACR 50 response at Week 24.
- Proportion of participants who maintain an ACR 70 response at Week 48, Week 96, and Week 156 among the participants who achieved an ACR 70 response at Week 24.
- Value, change, and percent change from baseline in ACR components by visit over time through Week 156.

5.3.6.2.2. HAQ-DI Related Endpoints

HAQ (Fries 1980) disability index (HAQ-DI) score is an evaluation of the functional status for a participant. CCI [redacted]

- [redacted]
- [redacted]
- [redacted]

5.3.6.2.3. DAS28 (CRP) Related Endpoints

CCI [redacted]

CCI [redacted]

9.

CCI [redacted]

The DAS28 (CRP) related Other Efficacy endpoints include:

- Change from baseline in DAS28 (CRP) by visit over time through Week 156.

- Proportion of participants who achieve a DAS28 (CRP) response by visit over time through Week 156.
- Proportion of participants who achieve a DAS28 (CRP) remission by visit over time through Week 156.

5.3.6.2.4. modified PsARC Related Endpoints

CCI [redacted]

The modified PsARC related Other Efficacy endpoint is:

- Proportion of participants who achieve a response based on modified PsARC by visit over time through Week 156.

5.3.6.2.5. Enthesitis Related Endpoints

Enthesitis is an important feature of psoriatic arthritis and other spondyloarthropathies. In this study, enthesitis will be assessed by an independent joint assessor using the Leeds Enthesitis Index (LEI) (Healy 2008). CCI [redacted]

CCI [redacted]

The enthesitis (LEI) related Other Efficacy endpoints include:

- Change from baseline in enthesitis score (LEI) by visit over time through Week 156 among the participants with enthesitis at baseline.
- Proportion of participants with resolution of enthesitis (LEI) by visit over time through Week 156 among the participants with enthesitis at baseline.

5.3.6.2.6. Dactylitis Related Endpoints

CCI [redacted]

The dactylitis related Other Efficacy endpoints include:

- Change from baseline in dactylitis score by visit over time through Week 156 among the participants with dactylitis at baseline.
- Proportion of participants with resolution of dactylitis by visit over time through Week 156 among the participants with dactylitis at baseline.

5.3.6.2.7. mCPDAI Related Endpoints

CCI [redacted]

CCI [redacted]

CCI [redacted]

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

Participants with low disease activity based on the mCPDAI score are those participants who have a mCPDAI score less than or equal to 3.2.

The mCPDAI related Other Efficacy endpoints include:

- Change from baseline in mCPDAI score by visit over time through Week 156.
- Proportion of participants with low disease activity based on mCPDAI by visit over time through Week 156

5.3.6.2.8. DAPSA Related Endpoints

CCI [redacted]

CCI [redacted]

The DAPSA related Other Efficacy endpoints include:

- Change from baseline in Disease Activity Index for Psoriatic Arthritis (DAPSA) score by visit over time through Week 156.
- Proportion of participants who achieve DAPSA low disease activity by visit over time through Week 156.
- Proportion of participants who achieve DAPSA remission by visit over time through Week 156.

5.3.6.2.9. MDA and VLDA

CCI [redacted]

The MDA and VLDA related Other Efficacy endpoints include:

- Proportion of participants who achieve minimal disease activity (MDA) by visit over time through Week 156.
- Proportion of participants who achieve very low disease activity (VLDA) by visit over time through Week 156.

5.3.6.2.10. Method of Analysis

In general, descriptive statistics, such as mean, standard deviation (SD), median, inter quartile (IQ) range, minimum, and maximum for continuous variables, and counts and percentages for discrete variables will be used to summarize most data.

The **main analysis** for Other Efficacy endpoints related to reduction of signs and symptoms and physical function will be conducted using the Adjusted Composite Estimand. The ICE Strategy handling rules for the estimand will be applied first, after which the data handling rules for missing data will be applied. Specifically:

ICE Strategies for Adjusted Composite Estimand

For the Adjusted Composite Estimand, ICE categories 1, 2, 3, and 6 are incorporated as part of the endpoint using the composite strategy. This estimand acknowledges that a participant increasing their background PsA medications, electing to discontinue study intervention for reasons other than Natural Disaster or Major Disruption, or not entering the LTE due to lack of efficacy or worsening of PsA is an unfavorable outcome where participants who meet them prior to the visit will be considered as treatment failures at that visit and subsequently through the final efficacy visit. Participants meeting TF are considered non-responders for binary response endpoints, and to have no improvement (change from baseline = 0) for continuous endpoints, regardless of observed data.

This estimand also employs the hypothetical strategy. Primarily, regarding significant unplanned changes to study conduct as a result of Natural Disaster or Major Disruption. This estimand seeks to estimate the treatment effect of study intervention as if Natural Disaster or Major Disruption, as well as ICEs directly resulting from them, did not occur. Thus, for ICE category 4 all observed data after meeting the ICE through end of study will not be used and will be assumed to be MAR. For ICE category 5, observed data at the visit immediately subsequent to meeting the ICE will not be used and will be assumed to be MAR.

For participants experiencing multiple ICEs, an ICE in categories 1, 2, 3, and 6 (ie, using the composite strategy) will supersede an ICE in categories 4 or 5 (ie, using the hypothetical strategy).

These ICE strategies are also summarized in tabular form in Section 6.14.

Handling Rules for Missing Continuous Data for Adjusted Composite Estimand

1. **Missing data for any reason** will be assumed to be MAR. The data is not explicitly imputed, but is accounted for in the analysis model.

Handling Rules for Missing Binary Data for Adjusted Composite Estimand

1. Missing data due to Natural Disaster (caused site closure, site access restrictions, or lockdowns) or Major Disruption will be assumed to be MAR. The data is not explicitly imputed, but is accounted for in the analysis model.
2. Missing data due to not entering LTE for reasons OTHER THAN lack of efficacy, worsening of PSA will be assumed to be MAR. The data is not explicitly imputed, but is accounted for in the analysis model.
3. Missing data for any other reason will be assumed to be MAR but conservatively imputed as NR

Analysis Testing for Adjusted Composite Estimand

Statistical comparison between a guselkumab group (100 mg q4w or 100 mg q8w) and the placebo group will be performed by visit through Week 24 using the Mixed-Effect Model Repeated Measures (MMRM) model for continuous endpoints and the Generalized Linear Mixed Model (GLMM) for binary response endpoints (see Section 6.13). No treatment comparison will be performed after Week 24. For continuous endpoints, only descriptive summaries and LSmeans by study intervention and visit will be presented after Week 24 through Week 156; for binary endpoints, only descriptive summaries and model based response rates by study intervention and visit will be presented after Week 24 through Week 156.

ICE Strategies for Treatment Policy Estimand

For the Treatment Policy Estimand, the occurrence of ICEs is considered irrelevant. This estimand looks at the effect of assignment to intervention group irrespective of changes to background PsA medications, study intervention adherence, or study retention.

Handling Rules for Missing Data for Treatment Policy Estimand

1. **Missing data for any reason** will not be imputed, assumed to be MCAR.

Analysis Testing for Treatment Policy Estimand

A corresponding supportive analysis will be conducted using the Treatment Policy Estimand. For these analyses, the mFAS-UKR analysis set or further subset (if specified) will be used. For visits through Week 24, an ANCOVA model will be used for continuous endpoints, and the CMH test for treatment difference and 95% CI calculated based on Wald statistics will be used for the binary response endpoints. No treatment comparison will be performed after Week 24. For continuous endpoints, only descriptive summaries and LSmeans by study intervention and visit will be presented after Week 24 through Week 156; for binary endpoints, only descriptive summaries of response rates by study intervention and visit will be presented after Week 24 through Week 156.

Table 10 summarizes the analyses for supportive efficacy endpoints related to reduction of signs and symptoms and physical function, the methods for analyses, and the data handling rules used.

Table 10: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL, ADJUSTED COMPOSITE ESTIMAND				
1	Proportions of participants with ACR 20, ACR 50, and ACR 70 response	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM

Table 10: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
2	ACR components	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group
3	Change from baseline in ACR components	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group
4	Percent change from baseline in ACR components	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group
5	Change from baseline in HAQ-DI score	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
6	Proportion of participants with HAQ-DI response	mFAS whose baseline HAQ-DI score ≥ 0.35	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
7	Change from baseline in DAS28 (CRP)	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
8	Proportion of participants with DAS28 (CRP) response	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
9	Proportion of participants with DAS28 remission	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
10	Proportion of participants with modified PsARC response	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM

Table 10: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
11	Change from baseline in enthesitis score	mFAS with enthesitis at baseline	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
12	Proportion of participants with resolution of enthesitis	mFAS with enthesitis at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
13	Change from baseline in dactylitis score	mFAS with dactylitis at baseline	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
14	Proportion of participants with resolution of dactylitis	mFAS with dactylitis at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
15	Change from baseline in mCPDAI	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
16	Proportion of participants with mCPDAI low disease activity	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
17	Change from baseline in DAPSA score	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM

Table 10: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
18	Proportion of participants with low disease activity or remission based on DAPSA	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
19	Proportion of participants with MDA	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
20	Proportion of participants with VLDA	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
'-' indicates no missing data rules to be applied				
ENDPOINTS BY VISIT AFTER WEEK 24 THROUGH WEEK 48 AT WEEK-48 DBL, ADJUSTED COMPOSITE ESTIMAND				
1, 2, 3 4, 6, 8, 9, 10, 12, 14, 16, 18, 19, 20	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • Response rates by intervention group based on GLMM
5, 7 11, 13 15, 17	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model
21	Proportion of participants who maintained an ACR 20 response at Week 48	mFAS who achieved an ACR 20 response at Week 24	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates by intervention group based on GLMM
22	Proportion of participants who maintained an ACR 50 response at Week 48	mFAS who achieved an ACR 50 response at Week 24	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates by intervention group based on GLMM

Table 10: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
23	Proportion of participants who maintained an ACR 70 response at Week 48	mFAS who achieved an ACR 70 response at Week 24	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates by intervention group based on GLMM
24	Proportion of participants who maintained HAQ-DI response at Week 48	mFAS who achieved HAQ-DI response at Week 24	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates by intervention group based on GLMM
ENDPOINTS BY VISIT AFTER WEEK 48 THROUGH WEEK 96 AT WEEK-96 DBL, ADJUSTED COMPOSITE ESTIMAND				
1, 2, 3 4, 6, 8, 9, 10, 12, 14, 16, 18, 19, 20	Same as through Week 24	Same as through Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
5, 7 11, 13 15, 17	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model
21	Proportion of participants who maintained an ACR 20 response at Week 96	mFAS who achieved an ACR 20 response at Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
22	Proportion of participants who maintained an ACR 50 response at Week 96	mFAS who achieved an ACR 50 response at Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM

Table 10: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
23	Proportion of participants who maintained an ACR 70 response at Week 96	mFAS who achieved an ACR 70 response at Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
24	Proportion of participants who maintained HAQ-DI response at Week 96	mFAS who achieved HAQ-DI response at Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
ENDPOINTS BY VISIT AFTER WEEK 96 THROUGH WEEK 156 AT FINAL DBL, ADJUSTED COMPOSITE ESTIMAND				
1, 2, 3 4, 6, 8, 9, 10, 12, 14, 16, 18, 19, 20	Same as through Week 24	Same as through Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
5, 7 11, 13 15, 17	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model
21	Proportion of participants who maintained an ACR 20 response at Week 156	mFAS who achieved an ACR 20 response at Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
22	Proportion of participants who maintained an ACR 50 response at Week 156	mFAS who achieved an ACR 50 response	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of	• Summarized descriptively • Response rates by intervention group based on GLMM

Table 10: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
		at Week 24	Efficacy/Worsening PsA	
23	Proportion of participants who maintained an ACR 70 response at Week 156	mFAS who achieved an ACR 70 response at Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
24	Proportion of participants who maintained HAQ-DI response at Week 156	mFAS who achieved HAQ-DI response at Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL, AFTER WEEK 24 THROUGH WEEK 48 AT WEEK-48 DBL, AFTER WEEK 48 THROUGH WEEK 96 AT WEEK-96 DBL, AND AFTER WEEK 96 THROUGH WEEK 156 AT FINAL DBL, USING TREATMENT POLICY ESTIMAND AND THE mFAS-UKR ANALYSIS SET INSTEAD OF mFAS				
• The endpoints analyzed using the Adjusted Composite Estimand for corresponding visits/DBL are repeated here, except: No imputation will be applied and all analyses will be based on observed data Continuous endpoints will use ANCOVA instead of MMRM; binary response endpoints will use CMH test with CI based on Wald statistics instead of GLMM for visits up to and including Week 24				

5.3.6.3. Endpoints Related to Skin Disease

In this study, Other Efficacy endpoints for skin disease include those related to PASI, IGA, and DLQI.

All Other Efficacy endpoints related to skin disease will be conducted under the mFAS population among participants with ≥3% body surface area (BSA) psoriatic involvement and an Investigator's Global Assessment (IGA) score of ≥2 (mild) at baseline for Adjusted Composite Estimand, and mFAS-UKR among participants with ≥3% body surface area (BSA) psoriatic involvement and an Investigator's Global Assessment (IGA) score of ≥2 (mild) at baseline for Treatment Policy Estimand, ***unless*** explicitly stated otherwise.

5.3.6.3.1. PASI Related Endpoints

CCI

-
-
-
  -
-

The PASI related Other Efficacy endpoints include:

- Change and percent change from baseline in PASI score by visit over time through Week 156 among the participants with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.
- Proportions of participants who achieve ≥75%, ≥90%, and 100% improvement in PASI score from baseline by visit over time through Week 156 among the participants with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.
- Proportion of participants who achieve both PASI 75 and ACR 20 responses by visit over time through Week 156 among the participants with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.
- Proportion of participants who achieve both PASI 75 and modified PsARC response by visit over time through Week 156 among the participants with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.

5.3.6.3.2. IGA Related Endpoints

CCI [redacted]

The IGA related Other Efficacy endpoints include:

- Proportion of participants with IGA response by visit over time through Week 156 among the participants with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.
- Proportion of participants with an IGA score of 0 (cleared) by visit over time through Week 156 among the participants with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.

5.3.6.3.3. DLQI Related Endpoints

CCI [redacted]

The DLQI related Other Efficacy endpoints include:

- Change from baseline in DLQI score by visit through Week 156 among the participants with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.
- Proportion of participants who achieve a DLQI score of 0 or 1 by visit through Week 156 among the participants with baseline DLQI score >1 and with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.
- Proportion of participants who achieve ≥5-point improvement from baseline in DLQI score by visit through Week 156 among the participants with baseline DLQI score ≥5 and with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.

5.3.6.3.4. Method of Analysis

In general, the same methods of analysis for the Other Efficacy endpoints related to reduction of signs and symptoms and physical function in Section 5.3.6.2.10 will also be used for those related to skin disease.

Table 11 summarizes the analyses for supportive efficacy endpoints related to skin disease, the methods for analyses, and the data handling rules used.

Table 11: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Skin Disease

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL, ADJUSTED COMPOSITE ESTIMAND				
1	Change from baseline in PASI score	mFAS with ≥3% BSA and IGA≥2 at baseline	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
2	Percent change from baseline in PASI score	mFAS with ≥3% BSA and IGA≥2 at baseline	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
3	Proportions of participants with PASI ≥75%, ≥90%, ≥100% improvement	mFAS with ≥3% BSA and IGA≥2 at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
4	Proportion of participants with both PASI ≥75% improvement and ACR 20 response	mFAS with ≥3% BSA and IGA≥2 at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM

Table 11: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Skin Disease

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
5	Proportion of participants with both PASI ≥75% improvement and modified PsARC response	mFAS with ≥3% BSA and IGA≥2 at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
6	Proportion of participants with IGA response	mFAS with ≥3% BSA and IGA≥2 at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
7	Proportions of participants with IGA score of 0	mFAS with ≥3% BSA and IGA≥2 at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
8	Change from baseline in DLQI score	mFAS with ≥3% BSA and IGA≥2 at baseline	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
9	Proportions of participants with DLQI score of 0 or 1	mFAS with **DLQI score>1,** ≥3% BSA and IGA≥2 at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM.
10	Proportions of participants with ≥5-point improvement in DLQI score	mFAS with **DLQI score≥5,** ≥3% BSA and IGA≥2 at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM.

'-' indicates no missing data rules to be applied

ENDPOINTS BY VISIT AFTER WEEK 24 THROUGH WEEK 48 AT WEEK-48 DBL, USING ADJUSTED COMPOSITE ESTIMAND

1, 2, 8	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model

Table 11: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Skin Disease

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
3, 4, 5, 6, 7, 9, 10	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • Response rates by intervention group based on GLMM
ENDPOINTS BY VISIT AFTER WEEK 48 THROUGH WEEK 96 AT WEEK-96 DBL, AND AFTER WEEK 96 THROUGH WEEK 156 AT FINAL DBL, USING ADJUSTED COMPOSITE ESTIMAND				
1, 2, 8	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model
3, 4, 5, 6, 7, 9, 10	Same as through Week 24	Same as through Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL, AFTER WEEK 24 THROUGH WEEK 48 AT WEEK-48 DBL, AFTER WEEK 48 THROUGH WEEK 96 AT WEEK-96 DBL, AND AFTER WEEK 96 THROUGH WEEK 156 AT FINAL DBL, USING TREATMENT POLICY ESTIMAND AND THE mFAS-UKR ANALYSIS SET INSTEAD OF THE mFAS				
• The endpoints analyzed using the Adjusted Composite Estimand for corresponding visits/DBL are repeated here, except: No imputation will be applied and all analyses will be based on observed data Continuous endpoints will use ANCOVA instead of MMRM; binary response endpoints will use CMH test with CI based on Wald statistics instead of GLMM for visits up to and including Week 24				

5.3.6.4. Endpoints Related to Psoriasis of the Nails

In this study, Other Efficacy endpoints for psoriasis of the nails include those related to mNAPSI and PGA-F.

All Other Efficacy endpoints related to psoriasis of the nails will be conducted under the mFAS population among participants with presence of psoriatic nail disease as measured by the instrument in question for the Adjusted Composite Estimand, and mFAS-UKR population among participants with presence of psoriatic nail disease as measured by the instrument in question for the Treatment Policy Estimand, ***unless*** explicitly stated otherwise.

5.3.6.4.1. mNAPSI Related Endpoints

CCI

-
-
-

CCI [redacted]

The mNAPSI related Other Efficacy endpoints include:

- Percent change from baseline in total fingernail mNAPSI score by visit over time through Week 156 among the participants with total fingernail mNAPSI score >0 at baseline.
- Proportions of participants who achieve total fingernail mNAPSI 50/75/100 response by visit over time through Week 156 among the participants with total fingernail mNAPSI score >0 at baseline.

5.3.6.4.2. PGA-F Related Endpoints

CCI [redacted]

The PGA-F related Other Efficacy endpoints include:

- Proportion of participants with PGA-F response by visit over time through Week 156 among the participants with PGA-F score ≥2 (mild) at baseline.

5.3.6.4.3. Method of Analysis

In general, the same methods of analysis for the Other Efficacy endpoints related to reduction of signs and symptoms and physical function in Section 5.3.6.2.10 will also be used for those related to nail psoriasis.

Table 12 summarizes the analyses for supportive efficacy endpoints related to nail psoriasis, the methods for analyses, and the data handling rules used.

Table 12: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Nail Psoriasis

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL, ADJUSTED COMPOSITE ESTIMAND				
1	Percent change from baseline in total fingernail mNAPSI score	mFAS with total fingernail mNAPSI score>0 at baseline	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM •
2	Proportions of participants with total fingernail mNAPSI 50/75/100 response	mFAS with total fingernail mNAPSI score>0 at baseline	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
3	Proportion of participants with PGA-F response	mFAS with PGA-F score≥2	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively • Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
'-' indicates no missing data rules to be applied				
ENDPOINTS BY VISIT AFTER WEEK 24 THROUGH WEEK 48 AT WEEK-48 DBL, USING ADJUSTED COMPOSITE ESTIMAND				
1	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model
2, 3	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • Response rates by intervention group based on GLMM
ENDPOINTS BY VISIT AFTER WEEK 48 THROUGH WEEK 96 AT WEEK-96 DBL, AND AFTER WEEK 96 THROUGH WEEK 156 AT FINAL DBL, USING ADJUSTED COMPOSITE ESTIMAND				
1	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model
2, 3	Same as through Week 24	Same as through Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of	• Summarized descriptively • Response rates by intervention group based on GLMM

Table 12: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Nail Psoriasis

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
			Efficacy/Worsening PsA	
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL, AFTER WEEK 24 THROUGH WEEK 48 AT WEEK-48 DBL, AFTER WEEK 48 THROUGH WEEK 96 AT WEEK-96 DBL, AND AFTER WEEK 96 THROUGH WEEK 156 AT FINAL DBL, USING TREATMENT POLICY ESTIMAND AND THE mFAS-UKR ANALYSIS SET INSTEAD OF THE mFAS				
• The endpoints analyzed using the Adjusted Composite Estimand for corresponding visits/DBL are repeated here, except: No imputation will be applied and all analyses will be based on observed data Continuous endpoints will use ANCOVA instead of MMRM; binary response endpoints will use CMH test with CI based on Wald statistics instead of GLMM for visits up to and including Week 24				

5.3.6.5. Endpoints Related to HRQOL

In this study, Other Efficacy endpoints for health related quality of life (HRQOL) measures include questionnaires of FACIT-Fatigue and PsAID-12.

All Other Efficacy endpoints related to HRQOL will be conducted under the mFAS population for the Adjusted Composite Estimand, and mFAS-UKR population of the Treatment Policy Estimand, ***unless*** explicitly stated otherwise.

5.3.6.5.1. FACIT-Fatigue Related Endpoints

CCI [redacted]

The FACIT-Fatigue related Other Efficacy endpoints include:

- Change from baseline in FACIT-F by visit over time through Week 156.
- Proportion of participants who achieve ≥4-point improvement from baseline in FACIT-F score by visit over time through Week 156.

5.3.6.5.2. PsAID-12 Related Endpoints

CCI [redacted]

The PsAID-12 related Other Efficacy endpoints include:

- Change from baseline in PsAID-12 by visit over time through Week 156.

5.3.6.5.3. Method of Analysis

In general, the same methods of analysis for the Other Efficacy endpoints related to reduction of signs and symptoms and physical function in Section 5.3.6.2.10 will also be used for those related to HRQOL.

Table 13 summarizes the analyses for supportive efficacy endpoints related to HRQOL, the methods for analyses, and the data handling rules used.

Table 13: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of HRQOL

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL, ADJUSTED COMPOSITE ESTIMAND				
1	Change from baseline in FACIT-F score	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM •
2	Proportions of participants with≥4-point improvement from baseline in FACIT-F score	mFAS	NRI for MAR not due to Natural Disaster or Major Disruption	• Summarized descriptively Response rates, and treatment difference in response rates and 95% CI and pvalue, based on GLMM
3	Change from baseline in PsAID-12 score	mFAS	-	• Summarized descriptively • LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups based on MMRM
'-' indicates no missing data rules to be applied				
ENDPOINTS BY VISIT AFTER WEEK 24 THROUGH WEEK 48 AT WEEK-48 DBL, USING ADJUSTED COMPOSITE ESTIMAND				
1, 3	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model
2	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • Response rates by intervention group based on GLMM
ENDPOINTS BY VISIT AFTER WEEK 48 THROUGH WEEK 96 AT WEEK-96 DBL, AND AFTER WEEK 96 THROUGH WEEK 156 AT FINAL DBL, USING ADJUSTED COMPOSITE ESTIMAND				
1, 3	Same as through Week 24	Same as through Week 24	Same as through Week 24	• Summarized descriptively • LSMeans and 95% CI by intervention group based on MMRM model

Table 13: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of HRQOL

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
2	Same as through Week 24	Same as through Week 24	NRI for MAR not either: due to Natural Disaster or Major Disruption, OR not enter LTE due reason ***other*** than Lack of Efficacy/Worsening PsA	• Summarized descriptively • Response rates by intervention group based on GLMM
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL, AFTER WEEK 24 THROUGH WEEK 48 AT WEEK-48 DBL, AFTER WEEK 48 THROUGH WEEK 96 AT WEEK-96 DBL, AND AFTER WEEK 96 THROUGH WEEK 156 AT FINAL DBL, USING TREATMENT POLICY ESTIMAND AND THE mFAS-UKR ANALYSIS SET INSTEAD OF mFAS				
• The endpoints analyzed using the Adjusted Composite Estimand for corresponding visits/DBL are repeated here, except: No imputation will be applied and all analyses will be based on observed data Continuous endpoints will use ANCOVA instead of MMRM; binary response endpoints will use CMH test with CI based on Wald statistics instead of GLMM for visits up to and including Week 24				

5.3.6.6. Endpoints Related to Joint Structural Damage

In this study, Other Efficacy endpoints for joint structural damage include modified vdH-S scores and its components.

All Other Efficacy endpoints related to joint structural damage will be conducted under the mFAS population.

5.3.6.6.1. Modified vdH-S Score Related Endpoints

CCI [redacted]

[redacted]

- [redacted]
- [redacted]

CCI [redacted]

The modified vdH-S related Other Efficacy endpoints include:

- Change from baseline in modified vdH-S score by visit over time through Week 156.
- Change in modified vdH-S score from Week 24 to Week 48; from Week 48 to Weeks 96 and 156; and from Week 96 to Week 156.
- Change from baseline in modified vdH-S erosion score and joint space narrowing (JSN) score by visit over time through Week 156.
- Change in modified vdH-S erosion score and JSN score from Week 24 to Week 48; from Week 48 to Weeks 96 and 156; and from Week 96 to Week 156.
- Proportion of participants with a change of ≤0 from baseline, and proportion of participants with a change of ≤0.5 from baseline in modified vdH-S score by visit over time through Week 156.
- Proportion of participants with a change of ≤0 from baseline, and proportion of participants with a change of ≤0.5 from baseline in modified vdH-S erosion score and JSN score by visit over time through Week 156.
- Proportion of participants with radiographic progression (based on the smallest detectable change [SDC]) from baseline by visit over time through Week 156. ***Alternatively***, this endpoint may be presented as the proportion of participants without radiographic progression.
- Proportion of participants with radiographic joint erosion progression and radiographic JSN progression (based on SDC) from baseline by visit over time through Week 156. ***Alternatively***, these endpoints may be presented as the proportion of participants without radiographic joint erosion progression and without radiographic JSN progression.
- Change from baseline in modified vdH-S score by region and type of damage (ie, hand erosion, hand JSN, foot erosion, foot JSN subscores) by visit over time through Week 156.
- Proportion of participants with pencil in cup or gross osteolysis deformities by visit over time through Week 156.

5.3.6.6.2. Method of Analysis

Radiographic images will be read in 4 read campaigns. The same 2 primary readers and 1 confirmation reader from Read Campaign 1 are planned to reprise their roles in each subsequent read campaign. From Read Campaign 2 onwards, only participants with new images taken since the prior read campaign will be read, however for these chosen participants all the relevant images for the read campaign will be read. The intended images to be read at each read campaign are as follows:

- Read Campaign 1: Baseline and Week 24
- Read Campaign 2: Baseline, Week 24, and Week 48/ED2
- Read Campaign 3: Baseline, Week 48, and Week 96/ED3
- Read Campaign 4: Baseline, Week 48, Week 96, and Week 156/ED4

  ED2/3/4 stand for early discontinuation of a participant (should it occur) associated with the last visit of the specific read campaign

Endpoints at Week 24 (i.e., placebo controlled period) based on data generated from Read Campaign 1 will be analyzed at Week-24 DBL. Statistical analysis of treatment comparisons is based on this data. All Other Efficacy endpoints related joint structural damage will be conducted under the mFAS population.

Treatment comparison is not planned on reads generated from Read Campaigns 2, 3, or 4. All Other Efficacy endpoints related to joint structural damage from these Read Campaigns will be conducted under the mFAS population. Generally, data generated from Read Campaign 2 will be analyzed at the Week-48 DBL; data generated from Read Campaign 3 will be analyzed at the Week-96 DBL; data generated from Read Campaign 4 will be analyzed at the Final DBL. ***Exception*** exists for the following:

Since the same readers are planned to be used throughout this study for all read campaigns, participants who have no new images to read for the current read campaign, may have their radiographic reads from the latest read campaign with data used for the analysis of the current read campaign. An example with 2 participants is shown below:

> Participant 1 is randomized and treated, then early discontinues the study at Study Week 22. They have 2 images taken, which slot into the analysis windows for Analysis baseline and Week 24 respectively. Their Analysis baseline and Week 24 images are read by the central readers in *Read Campaign 1*, but not re-read in *Read Campaign 2*.
>
> Participant 2 is randomized and treated, and are still on study at Study Week 48. They have 3 images taken, which slot into the analysis windows for Analysis baseline and Weeks 24 and 48 respectively. Their Analysis baseline and Week 24 images are read by the central readers in *Read Campaign 1*, and their Analysis baseline and Weeks 24 and 48 images read in *Read Campaign 2*.
>
> At the **Week-24 DBL**, the Analysis baseline and Week 24 readings from *Read Campaign 1* for both participants will be used for the analysis. At the **Week-48 DBL**, for Participant 1 the Analysis baseline and Week 24 readings from *Read Campaign 1* will be used for the analysis, while for Participant 2 the Analysis baseline and Weeks 24 and 48 from *Read Campaign 2* will be used for the analysis.

Should an unexpected situation arise where readers are changed for whatever reason before all 4 read campaigns are completed, it is planned that all participants, ***even those who have no new***

images taken since the last read campaign, will be read for the read campaign with change in readers.

In general, descriptive statistics, such as mean, standard deviation (SD), median, inter quartile (IQ) range, minimum, and maximum for continuous variables, and counts and percentages for discrete variables will be used to summarize most data. ***For multiply imputed data***, the descriptive statistics also include the mean across imputations and the standard error (SE) of this mean (Rubin 1987).

The **main analysis** for Other Efficacy endpoints related to joint structural damage will be conducted using the Adjusted Treatment Policy Estimand.

ICE Strategies for Adjusted Treatment Policy Estimand

The ICEs for this estimand were defined in Section 5.3.6.1. For the Adjusted Treatment Policy Estimand, the treatment policy strategy will be used for ICE categories 1, 2, 3, and 6, wherein all observed data collected for the endpoint is used. This estimand also employs the hypothetical strategy. For ICE category 4 all observed data after meeting the ICE through end of study will not be used and will be assumed to be MAR and imputed using FCS MI. For ICE category 5, observed data at the visit immediately subsequent to meeting the ICE will not be used, it will be assumed to be MAR and imputed using FCS MI.

For participants experiencing multiple ICEs, an ICE in categories 4 or 5 (ie, using the hypothetical strategy) will supersede an ICE in categories 1, 2, 3, and 6 (ie, using the treatment policy strategy) will supersede an ICE in categories.

These ICE strategies are also summarized in tabular form in Section 6.14.

Handling Rules for Missing Continuous Data for Adjusted Treatment Policy Estimand

1. **Missing data for any reason** will be assumed to be MAR. The FCS MI method is used to impute the missing component data, and the total score is derived from the sum of components.

Handling Rules for Missing Binary Response Data for Adjusted Treatment Policy Estimand

1. **Missing data for any reason** will be assumed to be MAR. The FCS MI method is used to impute the missing component data, and the total score is derived from the sum of components. The result is dichotomized into the binary response.

Analysis Testing for Adjusted Treatment Policy Estimand

Statistical comparison between a guselkumab group (100 mg q4w or 100 mg at Weeks 0, 4 and then q8w) and the placebo group will be performed at Week 24 using Read Campaign 1 data. No treatment comparison will be performed after Week 24, or using the other read campaign data; only descriptive summaries by study intervention and visit will be presented.

Table 14 summarizes the analyses for supportive efficacy endpoints related to joint structural damage, the methods for analyses, and the data handling rules used.

Table 14: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Joint Structural Damage

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
ENDPOINTS AT WEEK-24 DBL USING READ CAMPAIGN 1, ADJUSTED TREATMENT POLICY ESTIMAND				
1	Change from baseline in modified vdH-S erosion score, JSN score at Week 24	mFAS	FCS MI for MAR	• Summarized descriptively • Combined ANCOVA results across imputation sets, including: LS mean (95% CI) for each intervention group, LS mean difference (95% CI) and p-values for differences between groups
2	Proportion of participants with change ≤0 from baseline in modified vdH-S score at Week 24	mFAS	FCS MI for MAR	• Summarized descriptively • Combined CMH statistic[a] across imputation sets for p-value • Response rates, and treatment difference in response rates and 95% CI, based on combined results across imputation sets.
3	Proportion of participants with change ≤0 from baseline in modified vdH-S erosion score, JSN score at Week 24	mFAS	FCS MI for MAR	• Summarized descriptively • Combined CMH statistic[a] across imputation sets for p-value • Response rates, and treatment difference in response rates and 95% CI, based on combined results across imputation sets.
4	Proportion of participants with change ≤0.5 from baseline in modified vdH-S score at Week 24	mFAS	FCS MI for MAR	• Summarized descriptively • Combined CMH statistic[a] across imputation sets for p-value • Response rates, and treatment difference in response rates and 95% CI, based on combined results across imputation sets.
5	Proportion of participants with change ≤0.5 from baseline in modified vdH-S erosion score, JSN score at Week 24	mFAS	FCS MI for MAR	• Summarized descriptively • Combined CMH statistic[a] across imputation sets for p-value • Response rates, and treatment difference in response rates and 95% CI, based on combined results across imputation sets.
6	Proportion of participants without radiographic progression (based on SDC) at Week 24	mFAS	FCS MI for MAR	• Summarized descriptively • Combined CMH statistic[a] across imputation sets for p-value • Response rates, and treatment difference in response rates and 95% CI, based on combined results across imputation sets.

Table 14: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Joint Structural Damage

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
7	Proportion of participants without radiographic erosion progression, radiographic JSN progression (based on SDCs) at Week 24	mFAS	FCS MI for MAR	• Summarized descriptively • Combined CMH statistic[a] across imputation sets for p-value • Response rates, and treatment difference in response rates and 95% CI, based on combined results across imputation sets.
8	Change from baseline in modified vdH-S by region and type of damage at Week 24	mFAS	-	• Summarized descriptively
9	Proportion of participants with pencil in cup or gross osteolysis deformities at Baseline and Week 24	mFAS	-	• Summarized descriptively
ENDPOINTS AT WEEK-48 DBL USING PRIMARILY READ CAMPAIGN 2 (and supplemented by Read Campaign 1), ADJUSTED TREATMENT POLICY ESTIMAND				
1	Changes from: bl to Wk24, bl to Wk48, Wk24 to Wk48, in modified vdH-S score	mFAS	FCS MI for MAR	• Summarized descriptively
2	Changes from: bl to Wk24, bl to Wk48, Wk24 to Wk48, in modified vdH-S erosion score and JSN score	mFAS	FCS MI for MAR	• Summarized descriptively
3	Proportion of participants with change ≤0 from: bl to Wk24, bl to Wk48, Wk24 to Wk48, in modified vdH-S score	mFAS	FCS MI for MAR	• Summarized descriptively
4	Proportion of participants with change ≤0 from : bl to Wk24, bl to Wk48, Wk24 to Wk48, in modified vdH-S erosion score and JSN score	mFAS	FCS MI for MAR	• Summarized descriptively

[a] When combining analysis results for the CMH test, the Wilson-Hilferty transformation will be applied to the test statistics to achieve an approximate normal distribution.

'-' indicates no missing data rules to be applied

Table 14: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Joint Structural Damage

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
5	Proportion of participants with change ≤0.5 from: bl to Wk24, bl to Wk48, Wk24 to Wk48, in modified vdH-S score	mFAS	FCS MI for MAR	• Summarized descriptively
6	Proportion of participants with change ≤0.5 from : bl to Wk24, bl to Wk48, Wk24 to Wk48, in modified vdH-S erosion score and JSN score	mFAS	FCS MI for MAR	• Summarized descriptively
7	Proportion of participants without radiographic progression (based on SDC) from: bl to Wk24, bl to Wk48, Wk24 to Wk48	mFAS	FCS MI for MAR	• Summarized descriptively
8	Proportion of participants without radiographic erosion progression, radiographic JSN progression (based on SDCs) from: bl to Wk24, bl to Wk48, Wk24 to Wk48	mFAS	FCS MI for MAR	• Summarized descriptively
9	Change from baseline in modified vdH-S by region and type of damage at Week 48	mFAS	-	• Summarized descriptively
10	Proportion of participants with pencil in cup or gross osteolysis deformities at Baseline, Week 24, and Week 48	mFAS	-	• Summarized descriptively
'-' indicates no missing data rules to be applied				
ENDPOINTS AT WEEK-96 DBL USING PRIMARILY READ CAMPAIGN 3 (and supplemented by Read Campaigns 1 and 2), ADJUSTED TREATMENT POLICY ESTIMAND				
1-8	Similar to analyses at Week-48 DBL, except for time periods: bl to Wk48, bl to Wk96, Wk48 to Wk96	mFAS	FCS MI for MAR	• Summarized descriptively

Table 14: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Joint Structural Damage

	Endpoint	Analysis Set	Missing Data Rules	Analysis Methods
9	Change from baseline in modified vdH-S by region and type of damage at Week 96	mFAS	-	• Summarized descriptively
10	Proportion of participants with pencil in cup or gross osteolysis deformities at Baseline, Week 24, Week 48, and Week 96	mFAS	-	• Summarized descriptively
'-' indicates no missing data rules to be applied				
ENDPOINTS AT FINAL DBL USING PRIMARILY READ CAMPAIGN 4 (and supplemented by Read Campaigns 1, 2, and 3), ADJUSTED TREATMENT POLICY ESTIMAND				
1-8	Similar to analyses at Week-48 DBL, except for time periods: bl to Wk48, bl to Wk96, bl to Wk156, Wk48 to Wk96, Wk48 to Wk156, Wk96 to Wk156	mFAS	FCS MI for MAR	• Summarized descriptively
9	Change from baseline in modified vdH-S by region and type of damage at Week 156	mFAS	-	• Summarized descriptively
10	Proportion of participants with pencil in cup or gross osteolysis deformities at Baseline, Week 24, Week 48, Week 96, and Week 156	mFAS	-	• Summarized descriptively
'-' indicates no missing data rules to be applied				

5.3.6.7. Radiographic Readers' Agreement

The agreement between the 2 primary reader scores will be assessed at treatment group level and subject level.

In order to assess intra-reader variability, images of 30 subjects will be randomly selected and re-read by each of the 2 primary readers (Read Campaigns 1, 2, 3, and 4). The scores from the re-read will be used for intra-class correlation analysis.

The readers' agreement at treatment group level will be evaluated by descriptive summarization of each primary reader's score by treatment group overtime.

The readers' agreement at subject level will be evaluated using the methods of Bland and Altman, as applied by and referred to as the "limits of agreement" method by plots of the differences between the 2 primary readers' modified vdH-S scores versus the mean of the 2 primary readers' vdH-S scores (Lassere 1999).

In addition, intra-reader and inter-reader variability will be assessed. The scores from the re-read will be used for intra-class correlation analysis. The intra-class correlation for intra-reader and inter-reader variability will be calculated on modified vdH-S score at baseline, Week 24, Week 48, Week 96, and Week 156, and modified vdH-S score change from baseline at Weeks 24, 48, 96, and 156.

The purpose of the re-reads within each Read Campaign is solely for the purpose of assessing inter-reader variability, intra-reader variability, and intra-class correlation. They are not meant to be used in the formal analyses assessing treatment effect and treatment difference.

No data handling rules will be applied.

5.4. Safety Analyses

All safety analyses will be based on the safety analysis set based on **actual** intervention received.

Safety will be assessed by summarizing the occurrences and type of AEs, vital signs (pulse, blood pressure, and weight) and examining the changes in the laboratory parameters. No formal statistical comparison is planned.

5.4.1. Safety Tables Presentation

There are 4 DBLs in this study, respectively, at Week 24, Week 48, Week 96, and End of Study (Week 168). Depending on the safety data categories, the cumulative safety data will be analyzed through different study periods which include, but are not limited to, through Week 24, through Week 48, through Week 96, and through end of study periods. Tabular summaries of safety events for key study periods are in general presented as follows:

5.4.1.1. Summaries Through Week 24

Safety data through Week 24 will be analyzed according to the following intervention groups:

1. **Placebo**: Participants who received placebo only and no guselkumab prior to Week 24.
2. **Guselkumab 100 mg at Weeks 0, 4, and then q8w**: Participants who received guselkumab 100 mg q8w prior to Week 24 with an additional dose at Week 4.
3. **Guselkumab 100 mg q4w**: Participants who received guselkumab 100 mg q4w prior to Week 24.
4. **Guselkumab Combined:** Participants in Groups 2 and 3.

The above intervention groups 1-3 are **mutually exclusive**. This allows between-group comparisons of safety between a guselkumab group and the placebo group based on similar follow-up period in each group. The safety tables will have the column headings below:

		Guselkumab		
	Placebo	100 mg q8w	100 mg q4w	Combined
Analysis set: Safety Analysis Set	###	###	###	###

For participants who started treatment with placebo only but later received any amount of guselkumab prior to Week 24 inadvertently, the safety events/measurements on and after the first dose of guselkumab, will be excluded from the data summaries through Week 24. Only the safety events/measurements that occurred while the participants had been receiving placebo only will be included in the data summaries through Week 24.

5.4.1.2. Summaries Through Week 48

Safety data through Week 48 will be analyzed according to the following intervention groups:

1. **Placebo:** Participants who received placebo only. Follow-up will be based on the period that the participant was on placebo from the first dose up to Week 48.
    a. For participants who started treatment with placebo and later received treatment with guselkumab (due to CO or inadvertently), follow-up will end at the first dose of guselkumab, and only the safety events/measurements that occurred prior to the first dose of guselkumab will be included in this group
2. **Placebo → Guselkumab 100 mg q4w:** Participants who started treatment with placebo and later received treatment with guselkumab (due to CO or inadvertently). Follow-up will start from the first dose of guselkumab up to Week 48. All the safety events/measurements that occurred on and after the first dose of guselkumab up to Week 48 will be included in this group.
3. **Guselkumab 100 mg at Weeks 0, 4, and then q8w**: Participants who received guselkumab 100 mg q8w prior to Week 24 with an additional dose at Week 4. Follow-up will be from the first dose up to Week 48.

4. **Guselkumab 100 mg q4w**: Participants who received guselkumab 100 mg q4w prior to Week 24. Follow-up will be from the first dose up to Week 48.
5. **Guselkumab 100 mg q4w Combined:** Participants in Groups 2 and 4.
6. **All Guselkumab Combined:** Participants in Groups 2, 3, and 4.

The above intervention groups 1-2 are **not mutually exclusive**. The safety tables will have the column headings below:

		Guselkumab				
	Placebo	Placebo → 100 mg q4w	100 mg q8w	100 mg q4w	100 mg q4w Combined	All Combined
Analysis set: Safety Analysis Set	###	###	###	###	###	###

5.4.1.3. Summaries Through Week 96

Safety data through Week 96 will be analyzed similarly to safety data through 48, replacing Week 48 with Week 96.

5.4.1.4. Summaries Through End of Study

Safety data through End of Study (Week 168) will be analyzed similarly to safety data through 48, replacing Week 48 with End of Study.

5.4.2. Extent of Exposure and Study Follow-up

The number and percentage of participants who receive study intervention will be summarized. Descriptive statistics for duration study intervention (N, mean, SD, median, and range (minimum, maximum)) will be summarized.

Study intervention duration is defined as (date of last dose of study intervention – date of first dose of study intervention) +1. For the placebo intervention group which has planned crossover at Week 24, the study intervention duration prior to first guselkumab dose will be summarized separately to the study intervention duration on/after first guselkumab dose.

Study follow-up duration is defined in Section 5.4.1.

Study intervention compliance will be summarized descriptively. See Section 6.6 for further details.

5.4.3. Adverse Events

The verbatim terms used in the eCRF 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 and those AEs that were present at baseline but worsened in severity after the start of initial study intervention are 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 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 adverse events:

- Treatment emergent adverse events (TEAEs)
- Treatment emergent serious AEs (SAEs)
- TEAEs with severe intensity
- TEAEs leading to permanent discontinuation of study intervention
- TEAEs related to study intervention
- SAEs related to study intervention
- TEAEs leading to dose interruption of study intervention.
- Treatment emergent infections
- Treatment emergent serious infections
- Treatment emergent infections requiring oral or parenteral anti-microbial treatment
- Injection-site reactions
- Venous thromboembolism (VTE) events
- Clinically important hepatic disorder events
- Anaphylaxis(, hypersensitivity, and serum sickness reactions
- TEAEs leading to death

All AE summary tables will include average weeks of follow-up and average number of study intervention for each intervention group.

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

- SAEs
- TEAEs leading to permanent discontinuation of study intervention
- Anaphylactic reactions or serum sickness reactions
- Malignancies
- Serious infections including TB
- TEAEs leading to death
- VTE events
- Major adverse cardiovascular events (MACE) events

- Clinically important hepatic disorder events
- Opportunistic infections

Section 6.7 contains the methods of identification for selected AEs of interest.

A listing of participants who died will be provided, including cause of death, date of death, relationship to study intervention, and study day of death in relation to reference date.

5.4.4. Additional Safety Assessments

5.4.4.1. Clinical Laboratory Tests

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

The clinical laboratory parameters to be evaluated by the central laboratory include but are not limited to:

- **Hematology:** basophils, eosinophils, hemoglobin, hematocrit, lymphocytes, monocytes, neutrophils, platelets, red blood cell (RBC) count and white blood cell (WBC) count
- **Clinical chemistry:** albumin, alkaline phosphatase (ALP), alanine aminotransferase (serum glutamate pyruvate transaminase) [ALT (SGPT)], aspartate aminotransferase (serum glutamic oxaloacetic transaminase) [AST (SGOT)], bicarbonate, blood urea nitrogen (BUN), calcium, chloride, creatinine, glucose, potassium, sodium, total bilirubin, and total protein

Due to the Major Disruption resulting from the conflict involving Ukraine and neighboring countries/territories beginning February 24, 2022, central labs were unavailable for some sites over a period of time. In these instances, local labs, limited to the following parameters, may have been entered into the eCRFs: sodium, potassium, chloride, bicarbonate, BUN, creatinine, total bilirubin, direct bilirubin (conditionally), indirect bilirubin (conditionally), AST, ALT, GGT, ALP, and LDH. Certain analyses will use central lab data only, while other analyses may use a combination of central lab data and local lab data.

Descriptive statistics and graphical displays of observed values and changes from baseline will be presented for selected chemistry and hematology laboratory tests at scheduled time points. Only central lab data will be used for this analysis.

Shift tables from baseline to post-baseline will be produced for select laboratory parameters. Both central and local lab data will be used for this analysis.

Abnormality criteria based on toxicity grade will be applied to baseline and postbaseline values using National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI-CTCAE) for parameters with NCI-CTCAE criteria defined. Applicable laboratory results will be graded according to NCI-CTCAE version 5.0. Both central and local lab data will be used for this analysis. Abnormality for selected chemistry and hematology laboratory tests will be summarized by study intervention for:

- Number and percent of participants with post-baseline values by maximum toxicity grade

- Listings of participants with any post-baseline lab value of NCI-CTCAE toxicity Grade 3 or higher

Number and percent of participants with post-baseline elevated liver chemistry tests will also be produced using both central and local lab data, for

ALT categories:

- >1x to <3x Upper limit of normal (ULN)
- ≥3x to <5x ULN
- ≥5x to <8x ULN
- ≥8x ULN

AST categories

- >1x to <3x ULN
- ≥3x to <5x ULN
- ≥5x to <8x ULN
- ≥8x ULN

Total Bilirubin categories:

- >1x to <2x ULN
- ≥2x ULN

ALP categories:

- >1x to <2x ULN
- ≥2x to <4x ULN
- ≥4x ULN

Both central and local lab data will be used for this analysis.

A listing of participants with ANY of post-baseline ALT≥3x ULN, AST≥3x ULN, ALP≥2x ULN, OR total bilirubin≥2x ULN will be created. Additionally, they will be assessed for the following two combined criteria:

1) Total bilirubin ≥2xULN within 5 days after either ALT or AST ≥3xULN

2) International normalized ratio (INR) >1.5 within 5 days after either ALT or AST ≥3xULN

Number and percent of participants who met each of the following 5 liver function criteria individually, as well as overall (i.e, met any of the 5), as determined by the investigator, will be summarized:

- **ALT or AST absolute**:

  ALT or AST ≥8xULN
- **ALT or AST increase**:

  ALT or AST ≥5xULN but <8xULN persists for ≥2 weeks, OR

  ALT or AST ≥3xULN but <5xULN persists for ≥4 weeks
- **Bilirubin or INR**:

  ALT or AST ≥3xULN **and** total bilirubin ≥2xULN, OR

  ALT or AST ≥3xULN **and** INR >1.5
- **Cannot monitor**:

  ALT or AST ≥5xULN but <8xULN and cannot be monitored weekly for ≥2 weeks, OR

  ALT or AST ≥3xULN but <5xULN and cannot be monitored weekly for ≥4 weeks
- **Symptomatic**:

  ALT or AST ≥3xULN associated with symptoms (new or worsening) believed to be related to liver injury or hypersensitivity

.

5.4.4.2. Vital Signs and Physical Examination Findings

Continuous vital sign parameters including pulse, blood pressure (systolic and diastolic), and weight will be summarized at each assessment time point. The observed value and change from baseline will be summarized by intervention group. Descriptive statistics (mean, standard deviation, median, minimum and maximum) will be presented.

Incidence of markedly abnormal vital signs during intervention, as defined in Table 15, will be summarized for participants who had a baseline assessment and at least 1 postbaseline assessment for that vital sign. A listing of participants with markedly abnormal vital signs will be presented.

Table 15: Markedly Abnormal Vital Signs

Vital Sign	Criteria
Pulse	>120 bpm and with >30 bpm increase from baseline
	<50 bpm and with >20 bpm decrease from baseline
Systolic blood pressure	>180 mm Hg and with >40 mm Hg increase from baseline
	<90 mm Hg and with >30 mm Hg decrease from baseline

Table 15: Markedly Abnormal Vital Signs

Vital Sign	Criteria
Diastolic blood pressure	>105 mm Hg and with >30 mm Hg increase from baseline
	<50 mm Hg and with >20 mm Hg decrease from baseline

5.4.4.3. Electrocardiogram

No analysis is planned.

5.4.4.4. Other Safety Parameters

5.4.4.4.1. Suicidal Ideation and Behavior

The electronic Columbia-Suicide Severity Rating Scale (eC-SSRS) will be used as a screening tool to prospectively evaluate the potential of guselkumab to induce suicidal ideation and behavior. The eC-SSRS defines five subtypes of suicidal ideation and behavior in addition to self-injurious behavior with no suicidal intent, and is a fully-structured participant self-report questionnaire, including standardized questions, follow-up prompts, error handling routines, and scoring conventions (Mundt 2013) (Posner 2011). Two versions of the eC-SSRS will be used in this study, the Lifetime version and the Since Last Contact version. The Lifetime version will be conducted during the screening visit and the Since Last Contact version will be conducted at all other visits through Week 168.

Participants will complete the eC-SSRS questionnaire using the Sponsor-provided electronic tablets (or through an Interactive Voice Response System, if available). Study site personnel will train the participants on how to use the electronic device and/or a telephone system. The eC-SSRS will be provided in the local languages in accordance with local guidelines.

The eC-SSRS will be performed during each evaluation visit according to the Time and Events schedule. The eC-SSRS should be performed after the joint assessment at the screening visit (after signing informed consent). At Week 0/baseline and at all post-baseline visits, the eC-SSRS will be the first assessment/questionnaire that the participant completes prior to study intervention administration.

CCI [redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

Negative suicidality indication reports are generated from the eC-SSRS vendor when there are NO indications of the above.

Any eC-SSRS findings, which in the opinion of the investigator are new or considered to be a worsening and clinically significant, should be reported on the AE eCRF.

CCI

5.5. Other Analyses

5.5.1. Pharmacokinetics

PK analyses will be performed on the PK analysis set, defined as participants who have received at least 1 complete dose of guselkumab and have at least 1 valid blood sample drawn for PK analysis (Section 4.7). Subjects will be analyzed according to the treatment groups that they actually receive. No imputation for missing concentration data will be performed.

Descriptive statistics (N, mean, SD, median, range, CV (%) and IQ range) will be used to summarize serum guselkumab concentrations at each sampling time point by treatment group. PK data may be displayed graphically. The following analyses will be performed by treatment group as appropriate:

- Summary of serum guselkumab concentrations at each visit by treatment group

- Proportion of subjects without detectable serum guselkumab concentration at each visit by treatment group
- Summary of serum guselkumab concentrations at each visit by treatment group and body weight
- Summary of serum guselkumab concentrations at each visit by treatment group and baseline MTX use (Yes, No)
- Summary of serum guselkumab concentrations by baseline CRP levels
- Plot of median (IQ) serum guselkumab concentrations over time by treatment group

In addition, the relationship between serum guselkumab concentrations and safety or efficacy may be explored.

For summary statistics of serum guselkumab concentrations, concentration values below the lower limit of quantification will be treated as zero. Once a subject meets one of the following dosing deviation criteria, the subject's data will be excluded from the by-visit data analyses from that point onwards.

Dosing deviation criteria:

- Discontinue SC guselkumab administrations.
- Skipped an SC guselkumab administration.
- Received an incomplete/ incorrect SC dose.
- Received an incorrect SC study agent.
- Received an additional SC guselkumab dose.

In addition, if a subject has an administration outside of visit windows (Section 5.1.1), the concentration data collected at and after that visit will be excluded from the by-visit data analyses. Additional exclusions for incongruous PK data to be implemented based on Janssen SOP-07948.

5.5.2. Immunogenicity

The antibodies to guselkumab will be summarized based on all participants who received at least one (complete or partial) administration of guselkumab and who had at least 1 sample obtained after their first administration of guselkumab (Section 4.7). Subjects will be analyzed according to the treatment groups that they actually receive. No imputation for missing concentration data will be performed.

The following analysis of antibodies to guselkumab will be performed by treatment group:

- Summary of antibodies to guselkumab status
- Summary of neutralizing antibodies to guselkumab status
- List of subjects positive for antibodies to guselkumab

In addition, to explore the relationship between antibodies to guselkumab status and serum guselkumab concentrations, efficacy and safety, the following analysis may be performed as appropriate:

- Summary of clinical response (e.g., ACR 20 and ACR50, IGA) by antibody to guselkumab Status
- Summary of injection-site reactions by antibody to guselkumab status
- Summary of serum guselkumab concentrations by antibody to guselkumab status
- Plots of median (IQ) trough serum guselkumab concentrations over time by antibody to guselkumab status.

5.5.3. Biomarker/Pharmacodynamic Analysis

Methods and results for biomarker/pharmacodynamic analyses will be presented in a separate technical report.

5.5.4. Pharmacokinetic/Pharmacodynamic Relationships

If data permit, the relationships between serum guselkumab concentration and efficacy may be analyzed graphically. If any visual trend is observed, a suitable population PK/PD model may be developed to describe the exposure-response relationship. Details will be given in a population PK/PD analysis plan and results of the population PK/PD analysis will be presented in a separate technical report.

5.5.5. Health Economics

In this study, Other Efficacy endpoints for health economics include the Work Productivity and Activity Impairment Questionnaire - Specific Health Problem (**WPAI-SHP**), a validated instrument that has been used to study the impact of various diseases on patients' ability to work and perform daily activities (http://www.reillyassociates.net/WPAI_General.html). The WPAI-PsA assesses the impact of PsA on work and other daily activities during the past 7 days. The WPAI-PsA consists of the following 6 questions:

Q1: currently employed (working for pay)? (yes, no) ***If No, skip to Q6.***

Q2: hours missed from work in the past 7 days due to PsA? (hours)

Q3: hours missed from work in the past 7 days due to other reasons? (hours)

Q4: hours actually worked in the past 7 days? (hours)

Q5: degree to which PsA affected work productivity while at work in the past 7 days? [0 (no effect) to 10 (completely prevented from working)]

Q6: degree to which PsA affected regular activities outside of work in the past 7 days? [0 (no effect) to 10 (completely prevented from daily activities)]

Based on the answers to the above 6 questions, 4 types of scores (in percentage) are calculated, with higher scores indicating greater impairment and less productivity, i.e., worse outcomes, as follows. ***Note*** that for participants with answer='No' to Q1, only the 4th score (ie., percent activity impairment outside work due to PsA) can be calculated.

1. **Percent work time missed** due to PsA (absenteeism): 100*Q2/(Q2+Q4)
2. **Percent impairment while working** due to PsA (presenteeism): 100*Q5/10
3. **Percent overall work impairment** due to PsA (combining absenteeism and presenteeism): 100*{Q2/(Q2+Q4)+[(1-Q2/(Q2+Q4)) * (Q5/10)]}
4. **Percent activity impairment outside work** due to PsA: 100* Q6/10

Change from baseline in WPAI scores measures the change in work productivity and/or activity impairment, where a positive change indicates a worsening and a negative change indicates an improvement.

The WPAI related Other Efficacy endpoints include:

- Change from baseline in WPAI **Percent Activity Impairment Outside of Work** scores by visit over time through Week 156.
- Change from baseline in WPAI **Percent Work Time Missed**, **Percent Impairment While Working**, and **Percent Overall Work Impairment** scores by visit over time through Week 156 among participants who were employed at baseline

Estimands

The same estimands defined in Section 5.3.6.1 will be used for WPAI.

Method of Analysis

The same methods of analysis for the Other Efficacy endpoints related to reduction of signs and symptoms and physical function in Section 5.3.6.2.10 will also be used for WPAI.

5.5.6. Definition of Subgroups

To evaluate the consistency in the primary efficacy endpoint (proportion of participants who achieve ACR 20 at Week 24) and the key confirmatory secondary endpoint of structural damage (change from baseline in modified vdH-S score at Week 24) over demographics, baseline characteristics, prior and baseline medication use, subgroup analyses will be performed. The subgroups include, but are not limited to, the following:

Subgroup	Variant	Definition
Demographics		
Gender		• Male • Female
Race		• White • Asian

Subgroup	Variant	Definition
Demographics		
		• Other
Age at baseline (year)		• < 45 • ≥ 45 and < 65 • ≥ 65
Body weight at baseline (kg)	1 (Categories)	• ≤ 90 • >90
Body weight at baseline (kg)	2 (Quartiles)	• 1st Quartile: ([##] to [##]) • 2nd Quartile: ([##] to [##]) • 3rd Quartile: ([##] to [##]) • 4th Quartile: ([##] to [##])
Body mass index at baseline (kg/m^2)		• Normal [< 25] • Overweight [≥ 25 to < 30] • Obese [≥ 30]
Participating countries/territories		• Eastern Asia • Southeast Asia and Australia • Western Asia • Russia • Ukraine • Poland • Eastern Europe (except Russia, Ukraine, Poland) • Northern and Western Europe • Southern Europe • North America
Baseline disease characteristics		
PsA duration at baseline (year)		• < 1 • ≥ 1 to < 3 • ≥ 3
PsA subtype		• distal interphalangeal joint involvement • polyarticular arthritis with absence of rheumatoid nodules • asymmetric peripheral arthritis • spondylitis with peripheral arthritis
Number of swollen joints at baseline		• < 10 • 10 to 15 • > 15
Number of tender joints at baseline		• < 10 • 10 to 15 • > 15
HAQ-DI score at baseline		• < 1 • 1 to 2 • > 2
CRP at baseline (mg/dL)	1 (Categories)	• < 1 • 1-2 • ≥ 2
CRP at baseline (mg/dL)	2 (Quartiles)	• 1st Quartile: ([##] to [##]) • 2nd Quartile: ([##] to [##]) • 3rd Quartile: ([##] to [##]) • 4th Quartile: ([##] to [##])
Dactylitis at baseline		• Yes • No
Enthesitis at baseline		• Yes • No
PASI score at baseline		• <12

Subgroup	Variant	Definition
Demographics		
		• ≥ 12 to < 20 • ≥ 20
BSA of psoriasis at baseline		• < 3% • ≥ 3% to < 10% • ≥ 10% to < 20% • ≥ 20%
IGA score at baseline		• <2 • ≥2
mvdH-S score at baseline		• ≤median ([##]) • >median ([##])
Erosion score at baseline		• ≤median ([##]) • >median ([##])
#Joints with erosion at baseline		• ≤median ([##]) • >median ([##])
JSN score at baseline		• ≤median ([##]) • >median ([##])
#Joints with JSN at baseline		• ≤median ([##]) • >median ([##])
Prior and baseline medication use		
Use of non-biologic DMARDs (MTX, SSZ, HCQ, LEF) at baseline (based on eCRF)		• Yes • No
Oral corticosteroids at baseline (based on eCRF)		• Yes • No
NSAIDs at baseline		• Yes • No
Number of prior non-biologic treatments including DMARDs, systemic immunosuppressive drugs, and apremilast		• 0 • 1 • 2 • ≥3
Reason for discontinuation of prior DMARDs		• Efficacy - inadequate response (IR) • Safety - contraindication or intolerance (but not IR) • Other
Note that some of the above subgroup cut-off points may be changed if there are no or few participants within a subgroup category		

5.6. Interim Analyses

No interim analysis is planned for this study.

5.6.1. Data Monitoring Committee (DMC)

No DMC is planned for this study.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1: List of Abbreviations

ACR	American College of Rheumatology
AE	Adverse event
ALT	Alanine aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate aminotransferase
BMI	Body mass index
BSA	Body surface area
BUN	Blood urea nitrogen
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
CRP	C-reactive protein
DAPSA	Disease Activity Index for Psoriatic Arthritis
DAS28	Disease Activity Index Score 28
DBL	Database lock(s)
DICOM	Digital Imaging and Communications in Medicine
DLQI	Dermatology Life Quality Index
eCRF	electronic case report form
ED	Early discontinuation
HRV	High radiographic variability
eC-SSRS	electronic Columbia-Suicide Severity Rating Scale
EE	Early Escape
FAS	Full Analysis Set
FCS	Full Conditional Specifications
GDEV	Physician's Global Assessment of Disease Activity
GDPT	Patient's Global Assessment of Disease Activity
GLMM	Generalized Linear Mixed Model
GO	Gross osteolysis
HAQ-DI	HAQ disability index
HRQOL	Health related quality of life
ICE	Intercurrent event
ICF	Informed consent form
IGA	Investigator's Global Assessment
INR	International normalized ratio
IQ	Inter quartile
IR	Inadequate response
IWRS	interactive web response system
JSN	Joint space narrowing
LEI	Leeds Enthesitis Index
LLOQ	Lower limit of quantification
LMP	Low to moderate progression
LTE	Long-term extension
MACE	Major adverse cardiovascular events
MAR	Missing at random
MCAR	Missing completely at random
mCPDAI	modified Composite Psoriatic Disease Activity Index
MDA	Minimal Disease Activity
MedDRA	Medical Dictionary for Regulatory Activities
mFAS	Modified Full Analysis Set
MI	Multiple imputation
MMRM	Mixed-Effect Model Repeated Measures
mNAPSI	modified Nail Psoriasis Severity Index
MNAR	Missing not at random

MTX	Methotrexate
NCI-CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
NP	No progression
NR	Non-responder
NRS	Numeric rating scale
PAIN	Patient's assessment of pain
PASI	Psoriasis Area and Severity Index
PD	Pharmacodynamic(s)
PIC	Pencil in cup
PK	Pharmacokinetic(s)
PGA-F	Physician's Global Assessment of Fingernail Psoriasis
PPAS	Per-Protocol Analysis Set
PsA	Psoriatic arthritis
PsAID-12	PsA Impact of Disease-12
PsARC	Psoriatic Arthritis Responder Criteria
RBC	Red blood cell
RP	Rapid progression
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SDC	Smallest detectable change
SDV	Source Data Verification
SE	Standard error
SJC	Swollen Joint Count
TEAE	Treatment emergent adverse event
TF	Treatment failure(s)
TJC	Tender Joint Count
ULN	Upper limit of normal
VAS	Visual analog scale
vdH-S	Van der Heijde Sharp
VLDA	Very Low Disease Activity
VTE	Venous thromboembolism
WBC	White blood cell
WPAI-SHP	Work Productivity and Activity Impairment Questionnaire - Specific Health Problem

6.2. Appendix 2: Demographics and Baseline Characteristics

Table 16 presents a list of the demographic and baseline variables that will be summarized by intervention group, combined active intervention group, and overall for the following analysis sets: mFAS and FAS (should it differ from mFAS).

Table 16: 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)	
Categorical Variables	
Age (45 years, 45-<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)	
Weight (≤90kg, >90kg)	
BMI (underweight <18.5 kg/m2, normal 18.5-<25 kg/m2, overweight 25-<30 kg/m2, obese >=30 kg/m2)	

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

The baseline characteristics will be summarized for the same analysis sets as the demographic variables. They include, but are not limited to, baseline disease characteristics of PsA (e.g., duration of disease, PsA subtypes, baseline efficacy assessments), medical history, prior exposure to non-biologic medications, prior joint procedures/injections, and baseline medication usage for PsA.

6.3. Appendix 3: Protocol Deviations

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

The study selection criteria will be grouped into the following 5 categories: PsA disease criteria, medication criteria, laboratory criteria, medical history criteria, and other.

Protocol deviation in study intervention administrations includes missing doses, incorrect doses, and treatments administered out of the dosing windows defined in Section 5.1.1.1. Additionally, missed doses due to Major Disruption or Natural Disaster will be summarized.

6.4. Appendix 4: Prior and Concomitant Medications

Prior and Concomitant medications will be coded using an appropriate drug dictionary, which will be defined in the study metadata. 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 concomitant medications will be presented by ATC class and ATC term, 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 1 concomitant medication. In addition, concomitant medications of special interest will be presented. These include non-biologic DMARDs, systemic corticosteorids, and NSAIDs. See Section 6.8 for list of medications in each category.

Prior medications taken for PsA and/or psoriasis (e.g., non-biologic DMARDs, apremilast, immunosuppressives, and NSAIDS) will be summarized by randomized intervention group.

6.5. Appendix 5: Medical History

Number and percentage of participants who had medical histories of interest for PsA will be collected and summarized by intervention group, including:

- Inflammatory Bowel Disease
- Uveitis
- Lower back pain
- Fibromyalgia
- Coronary Artery Disease
- Myocardial Infarction
- Peripheral Vascular Disease
- Transient Ischemic Attack
- Stroke
- Diabetes Mellitus
- Hyperlipidemia
- Hypertension
- Asthma
- Depression
- Chronic Liver Disease (e.g., fatty liver disease, alcohol-induced, cirrhosis)
- Skin Squamous Cell Carcinoma
- Skin Basal Cell Carcinoma
- Gout
- Enthesitis
- Dactylitis

Other medical histories not specified above will be coded using MedDRA and presented by System Organ Class, Preferred Term, and intervention group, separately for histories related to PsA and those not related to PsA.

6.6. Appendix 6: Intervention Compliance

Compliance will be summarized descriptively for the overall study intervention, as well as for guselkumab and placebo separately. Compliance to randomized intervention versus actual intervention will be presented in a summary table, and will be calculated as (the number of injections completed / the number of injections planned * 100).

Note that due to the planned unblinding of sites to study intervention after the Week-48 DBL (core study completion), participants in the guselkumab 100 mg q8w arm will no longer be required to take placebo injections to maintain the blind. Placebo injections would no longer be among planned injections subsequent to the calendar date of the sites' unblinding.

6.7. Appendix 7: Adverse Events of Interest

Adverse events of special interest, as well as other adverse events of interest, will be identified based on criteria specified in the following table.

Type of Adverse Event	MedDRA Terms Search Methodology	Requires Medical Review
Adverse Events of Special Interest		
Malignancy	Malignant tumors (SMQ-narrow scope).	Yes
Active Tuberculosis	HLT of Tuberculosis infections excluding PT of Latent Tuberculosis	Yes
Other Adverse Events of Interest		
Infections	SOC Infections and infestations Note that serious infections, and infections requiring oral or parenteral anti-microbial treatment, are based on this MedDRA determination as well as eCRF checkboxes for serious AE and for requiring oral or parenteral anti-microbial treatment respectively	No
Opportunistic Infections	Opportunistic infections (SMQ-narrow scope)	Yes
ISR	No MedDRA search used. Based completely on eCRF checkbox	No
Anaphylaxis and Hypersensitivity	PTs of Anaphylactic reaction, Anaphylactic shock, Anaphylactoid reaction, Anaphylactoid shock, and Type I Hypersensitivity	No
Serum Sickness Reactions	PTs of serum sickness and Serum sickness-like reaction	No
MACE	**SMQs:** • Myocardial infarction (narrow) • Ischaemic central nervous system vascular conditions (narrow scope) • Haemorrhagic central nervous system vascular conditions (narrow scope) **PTs:** Sudden death	Yes

	AESOC: • Cardiac Disorders (fatal events only) • Vascular Disorders (fatal events only)	
VTEs	Customized MedDRA PTs related to venous thrombosis and embolism involving the deep venous vasculature: Axillary vein thrombosis, Brachiocephalic vein thrombosis, Budd-Chiari syndrome, Deep vein thrombosis, Deep vein thrombosis postoperative, Embolism venous, Hepatic vein thrombosis, Homans' sign positive, Inferior vena cava syndrome, Jugular vein thrombosis, Mahler sign, May-Thurner syndrome, Mesenteric vein thrombosis, Obstetrical pulmonary embolism, Ovarian vein thrombosis, Paget-Schroetter syndrome, Pelvic venous thrombosis, Penile vein thrombosis, Peripheral vein thrombus extension, Peripheral vein thrombosis, Post procedural pulmonary embolism, Postpartum venous thrombosis, Pulmonary embolism, Pulmonary infarction, Pulmonary microemboli, Pulmonary thrombosis, Pulmonary venous thrombosis, Renal vein thrombosis, Spermatic vein thrombosis, Splenic vein thrombosis, Subclavian vein embolism, Subclavian vein thrombosis, Thrombosis corpora cavernosa, Vena cava embolism, Vena cava thrombosis, Venous thrombosis, Venous thrombosis in pregnancy, Venous thrombosis limb, Visceral venous thrombosis	No
Clinically Important Hepatic Disorders	Drug related hepatic disorders - comprehensive search (SMQ – narrow scope) and either SAE or AE leading to discontinuation of study intervention	No

6.8. Appendix 8: Medications of Special Interest

Concomitant medications of special interest are defined as follows:

Concomitant Medication Special Interest Category	**Categories**
Non-biologic DMARD	Methotrexate (MTX), sulfasalazine, leflunomide, hydroxychloroquine, chloroquine, gold preparations, penicillamine, other non-biologic DMARDs
Oral corticosteroids	Oral corticosteroids
NSAIDs	NSAIDs

Prior Medication Special Interest Category	**Categories**
Non-biologic DMARD	MTX, sulfasalazine, leflunomide, hydroxychloroquine, chloroquine, gold preparations, penicillamine, other non-biologic DMARDs
Immunosuppressives	Cyclosporine, azathioprine, mycophenolate mofetil, tacrolimus, other immunosuppressives
Systemic corticosteroids	Systemic corticosteroids
NSAIDs	NSAIDs
Apremilast	Apremilast
Previous medications and therapies for PsA	Non-biologic DMARDs, immunosuppressives, apremilast, systemic corticosteroids, NSAIDs
Previous medications and therapies for PsO	Cyclosporine, Topical, Acitretin , UVB, Apremilast, PUVA

6.9. Appendix 9: Laboratory Toxicity Grading

The grading scale use for lab assessments is based on 'Common Terminology Criteria for Adverse Events (CTCAE) v5.0.

If a laboratory value falls within the grading as specified below but also within the local laboratory normal limits, the value is considered to be normal and will be reset to grade 0.

Pre-baseline measurements will use the same grading ranges as applied to baseline measurements. In case a test has two sets of ranges – one for baseline normal and one for baseline abnormal, the one for baseline normal will be applied for all measurements taken pre-baseline and on baseline.

Text in gray italic in the table is present in the grading scale, but is not applied by Janssen when grading lab data.

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
Blood and lymphatic system disorders					
Anemia	Hemoglobin (Hgb) <LLN - 10.0 g/dL; <LLN - 6.2 mmol/L; <LLN - 100 g/L	Hemoglobin (Hgb) <10.0 - 8.0 g/dL; <6.2 - 4.9 mmol/L; <100 - 80g/L	Hemoglobin (Hgb) <8.0 g/dL; <4.9 mmol/L; <80 g/L; *transfusion indicated*	*Life-threatening consequences; urgent intervention indicated*	Clinical signs and symptoms are not taken into consideration for grading.
Leukocytosis	-	-	>100,000/mm3; >100 x 10e9 /L	*Clinical manifestations of leucostasis; urgent intervention indicated*	Clinical signs and symptoms are not taken into consideration for grading; Added ranges in SI unit (x 10e9 /L)
Investigations					
Activated partial thromboplastin time prolonged	>ULN - 1.5 x ULN	>1.5 - 2.5 x ULN	>2.5 x ULN; *bleeding*	-	Clinical signs and symptoms are not taken into consideration for grading.
Alanine aminotransferase increased	>ULN - 3.0 x ULN if baseline was normal; 1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
Alkaline phosphatase increased	>ULN - 2.5 x ULN if baseline was normal; 2.0 - 2.5 x baseline if baseline was abnormal	>2.5 - 5.0 x ULN if baseline was normal; >2.5 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
Aspartate aminotransferase increased	>ULN - 3.0 x ULN if baseline was normal; 1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
Blood bilirubin increased	>ULN - 1.5 x ULN if baseline was normal; > 1.0 - 1.5 x baseline if baseline was abnormal	>1.5 - 3.0 x ULN if baseline was normal; >1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 10.0 x ULN if baseline was normal; >3.0 - 10.0 x baseline if baseline was abnormal	>10.0 x ULN if baseline was normal; >10.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
CD4 lymphocytes decreased	<LLN - 500/mm3; <LLN - 0.5 x 10e9 /L	<500 - 200/mm3; <0.5 - 0.2 x 10e9 /L	<200 - 50/mm3; <0.2 x 0.05 - 10e9 /L	<50/mm3; <0.05 x 10e9 /L	
Cholesterol high	>ULN - 300 mg/dL; >ULN - 7.75 mmol/L	>300 - 400 mg/dL; >7.75 - 10.34 mmol/L	>400 - 500 mg/dL; >10.34 - 12.92 mmol/L	>500 mg/dL; >12.92 mmol/L	
CPK increased	>ULN - 2.5 x ULN	>2.5 x ULN - 5 x ULN	>5 x ULN - 10 x ULN	>10 x ULN	
Creatinine increased	Creatine Kinase >ULN - 1.5 x ULN	Creatine Kinase >1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	Creatine Kinase >3.0 x baseline; >3.0 - 6.0 x ULN	Creatine Kinase >6.0 x ULN	
Fibrinogen decreased	<1.0 - 0.75 x LLN; if abnormal, <25% decrease from baseline	<0.75 - 0.5 x LLN; if abnormal, 25 - <50% decrease from baseline	<0.5 - 0.25 x LLN; if abnormal, 50 - <75% decrease from baseline	<0.25 x LLN; if abnormal, 75% decrease from baseline; absolute value <50 mg/dL	Ranges defined for "abnormal" are applied only on values < LLN. Grade 0 will be assigned to values > ULN.
GGT increased	>ULN - 2.5 x ULN if baseline was normal; 2.0 - 2.5 x baseline if baseline was abnormal	>2.5 - 5.0 x ULN if baseline was normal; >2.5 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
Haptoglobin decreased	<LLN	-	-	-	
Hemoglobin increased	Increase in >0 - 2 g/dL; Increase in >0 - 20 g/L	Increase in >2 - 4 g/dL; Increase in >20 - 40 g/L	Increase in >4 g/dL; Increase in >40 g/L	-	The increase indicates the level of increase above normal (above ULN). Applied as, e.g. grade 1 (g/dL): >ULN – ULN+2 g/dL; Added ranges in SI unit (g/L).
INR increased	>1.2 - 1.5; *>1 - 1.5 x baseline if on anticoagulation;*	>1.5 - 2.5; *>1.5 - 2.5 x baseline if on anticoagulation;*	>2.5; *>2.5 x baseline if on anticoagulation; bleeding*	-	Concomitant therapy or clinical signs and symptoms are not taken

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
	monitoring only indicated	*dose adjustment indicated*			into consideration for grading.
Lipase increased	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN; *>2.0 - 5.0 x ULN and asymptomatic*	>2.0 - 5.0 x ULN with signs or symptoms; *>5.0 x ULN and asymptomatic*	>5.0 x ULN and with signs or symptoms	"Asymptomatic" ranges are not taken into consideration for grading, i.e. worst case grading is applied.
Lymphocyte count decreased	<LLN - 800/mm3; <LLN - 0.8 x 10e9/L	<800 - 500/mm3; <0.8 - 0.5 x 10e9 /L	<500 - 200/mm3; <0.5 - 0.2 x 10e9 /L	<200/mm3; <0.2 x 10e9 /L	
Lymphocyte count increased	-	>4000/mm3 - 20,000/mm3; >4 - 20 x 10e9 /L	>20,000/mm3; >20 x 10e9 /L	-	Added ranges in SI unit (x 10e9 /L).
Neutrophil count decreased	<LLN - 1500/mm3; <LLN - 1.5 x 10e9 /L	<1500 - 1000/mm3; <1.5 - 1.0 x 10e9 /L	<1000 - 500/mm3; <1.0 - 0.5 x 10e9 /L	<500/mm3; <0.5 x 10e9 /L	Both Neutrophils and segmented neutrophils are graded using these criteria.
Platelet count decreased	<LLN - 75,000/mm3; <LLN - 75.0 x 10e9 /L	<75,000 - 50,000/mm3; <75.0 - 50.0 x 10e9 /L	<50,000 - 25,000/mm3; <50.0 - 25.0 x 10e9 /L	<25,000/mm3; <25.0 x 10e9 /L	
Serum amylase increased	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN; *>2.0 - 5.0 x ULN and asymptomatic*	>2.0 - 5.0 x ULN *with signs or symptoms; >5.0 x ULN and asymptomatic*	>5.0 x ULN *and with signs or symptoms*	"Asymptomatic" ranges are not taken into consideration for grading, i.e. worst case grading is applied.
White blood cell decreased	<LLN - 3000/mm3; <LLN - 3.0 x 10e9 /L	<3000 - 2000/mm3; <3.0 - 2.0 x 10e9 /L	<2000 - 1000/mm3; <2.0 - 1.0 x 10e9 /L	<1000/mm3; <1.0 x 10e9 /L	
Metabolism and nutrition disorders					
Acidosis	pH <normal, but >=7.3	-	pH <7.3	*Life-threatening consequences*	pH <normal is implemented as pH <LLN. Clinical signs and symptoms are not taken into consideration for grading.
Alkalosis	pH >normal, but <=7.5	-	pH >7.5	*Life-threatening consequences*	pH >normal is implemented as pH >ULN. Clinical signs and symptoms are not taken

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
					into consideration for grading.
Hypercalcemia	Corrected serum calcium of >ULN - 11.5 mg/dL; >ULN - 2.9 mmol/L; Ionized calcium >ULN - 1.5 mmol/L	Corrected serum calcium of >11.5 - 12.5 mg/dL; >2.9 - 3.1 mmol/L; Ionized calcium >1.5 - 1.6 mmol/L; *symptomatic*	Corrected serum calcium of >12.5 - 13.5 mg/dL; >3.1 - 3.4 mmol/L; Ionized calcium >1.6 - 1.8 mmol/L; *hospitalization indicated*	Corrected serum calcium of >13.5 mg/dL; >3.4 mmol/L; Ionized calcium >1.8 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hyperkalemia	Potassium >ULN - 5.5 mmol/L	Potassium >5.5 - 6.0 mmol/L; *intervention initiated*	Potassium >6.0 - 7.0 mmol/L; *hospitalization indicated*	Potassium >7.0 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypermagnesemia	Magnesium >ULN - 3.0 mg/dL; >ULN - 1.23 mmol/L	-	Magnesium >3.0 - 8.0 mg/dL; >1.23 - 3.30 mmol/L	Magnesium >8.0 mg/dL; >3.30 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypernatremia	Sodium >ULN - 150 mmol/L	Sodium >150 - 155 mmol/L; *intervention initiated*	Sodium >155 - 160 mmol/L; *hospitalization indicated*	Sodium >160 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypertriglyceridemia	Triglycerides 150 mg/dL - 300 mg/dL; 1.71 mmol/L - 3.42 mmol/L	Triglycerides >300 mg/dL - 500 mg/dL; >3.42 mmol/L - 5.7 mmol/L	Triglycerides >500 mg/dL - 1000 mg/dL; >5.7 mmol/L - 11.4 mmol/L	Triglycerides >1000 mg/dL; >11.4 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypoalbuminemia	Albumin <LLN - 3 g/dL; <LLN - 30 g/L	Albumin <3 - 2 g/dL; <30 - 20 g/L	Albumin <2 g/dL; <20 g/L	*Life-threatening consequences; urgent intervention indicated*	Clinical signs and symptoms are not taken into consideration for grading.
Hypocalcemia	Corrected serum calcium of <LLN - 8.0 mg/dL; <LLN - 2.0 mmol/L;	Corrected serum calcium of <8.0 - 7.0 mg/dL; <2.0 - 1.75 mmol/L;	Corrected serum calcium of <7.0 - 6.0 mg/dL; <1.75 - 1.5 mmol/L;	Corrected serum calcium of <6.0 mg/dL; <1.5 mmol/L;	Clinical signs and symptoms are not taken into consideration for grading.

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
	Ionized calcium <LLN - 1.0 mmol/L	Ionized calcium <1.0 - 0.9 mmol/L; *symptomatic*	Ionized calcium <0.9 - 0.8 mmol/L; *hospitalization indicated*	Ionized calcium <0.8 mmol/L; *life-threatening consequences*	
Hypoglycemia	Glucose <LLN - 55 mg/dL; <LLN - 3.0 mmol/L	Glucose <55 - 40 mg/dL; <3.0 - 2.2 mmol/L	Glucose <40 - 30 mg/dL; <2.2 - 1.7 mmol/L	Glucose <30 mg/dL; <1.7 mmol/L; *life-threatening consequences; seizures*	Clinical signs and symptoms are not taken into consideration for grading. Urine glucose is not graded.
Hypokalemia	*Potassium <LLN - 3.0 mmol/L*	*Symptomatic with* Potassium <LLN - 3.0 mmol/L; intervention indicated	Potassium <3.0 - 2.5 mmol/L; *hospitalization indicated*	Potassium <2.5 mmol/L; *life-threatening consequences*	"Symptomatic" ranges are applied for grade 2, grade 1 not assigned, i.e. worst case applied. Clinical signs and symptoms are not taken into consideration for grading of grade 3 and 4.
Hypomagnesemia	Magnesium <LLN - 1.2 mg/dL; <LLN - 0.5 mmol/L	Magnesium <1.2 - 0.9 mg/dL; <0.5 - 0.4 mmol/L	Magnesium <0.9 - 0.7 mg/dL; <0.4 - 0.3 mmol/L	Magnesium <0.7 mg/dL; <0.3 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hyponatremia	Sodium <LLN - 130 mmol/L	*Sodium 125-129 mmol/L and asymptomatic*	*Sodium 125-129 mmol/L symptomatic;* *120-124 mmol/L regardless of symptoms* Sodium <130-120 mmol/L	Sodium <120 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading. Worst case ("<130-120 mmol/L" for grade 3 added by Janssen) is applied across grade 2/3 ranges: 120-129 mol/L assigned to grade 3, grade 2 not used.
Renal and urinary disorders					
Proteinuria	1+ proteinuria;	**Adult:** 2+ and 3+ proteinuria;	**Adult:** 4+ proteinuria;	-	In case both 24-h urine collection and dipstick

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes

	urinary protein ≥ULN - <1.0 g/24 hrs;
	urinary protein ≥ULN - <1000 mg/day
		urinary protein 1.0 - <3.5 g/24 hrs;
		urinary protein 1000 - <3500 mg/day
		
		Pediatric:
		Urine P/C (Protein/Creatinine) ratio 0.5 - 1.9;
		Urine P/C (Protein/Creatinine) 56.5 – 214.7 g/mol
			urinary protein >=3.5 g/24 hrs;
			urinary protein >=3500 mg/day;
			
			Pediatric:
			Urine P/C (Protein/Creatinine) ratio >1.9;
			Urine P/C (Protein/Creatinine) >214.7 g/mol
					are collected, then worst case is taken, as opposed to having 24-h urine collection take precedence over dipstick. Added ranges in SI unit for urinary protein (mg/day) and for urine P/C (g/mol). Pediatric grading is applied to age range [0-18]. Adult grading is applied for ages [>18].

* Grade 0 is assigned to a lab assessment when the lab test is described in the table, but the lab value is not assigned a grade 1 or higher.

6.10. Appendix 10: Rules Applied in Definitions of Endpoints

1. Joint Evaluability Rules for Sign and Symptom Data

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

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

2. Joint Count Adjustment Rule

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

3. LLOQ rule

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

4. Joint Evaluability Rules for Radiographic Data

A joint may be not evaluable due to surgery/joint replacement or radiographically insufficient data for reading. Joints with surgery/joint replacement or with radiographically insufficient data for reading will be considered as not evaluable for joint erosion/JSN.

For a joint that is considered as not evaluable at a given time point, both the joint-level erosion score and the joint-level JSN score will be set to missing at the said time point.

5. Erosion and JSN Score Adjustment Rules

The regional erosion and JSN scores for hands and feet will be determined based on the evaluable joints. For participants who have an incomplete set of evaluable joints at a given time point, each reader's regional erosion and JSN scores at the said time point will be adjusted using the following rules:

Rules for Adjustment of Erosion Scores by Region	
Region	**Adjustment for the incomplete set of evaluable joints**
Hands/Wrists (40 joints)	• If total number of joints evaluable at the given time point is ≥ 20 (ie, 50% of 40), then the erosion score for hands/wrists will be obtained by calculating the average erosion score for hands and wrists and multiplying with 40. • If the total number of joints evaluable at the given time point is < 20, then the erosion score for hands and wrists will be set to missing.
Feet (12 joints)	• If total number of joints evaluable at the given time point is ≥ 6 (i.e., 50% of 12), then the erosion score for feet will be obtained by calculating the average erosion score for feet and multiplying with 12. • If the total number of joints evaluable at the given time point is < 6, then the erosion score for feet will be set to missing.

Rules for Adjustment of JSN Scores by Region	
Region	**Adjustment for the incomplete set of evaluable joints**
Hands/Wrists (40 joints)	• If total number of joints evaluable at the given time point is ≥ 20 (i.e., 50% of 40), then the JSN score for hands/wrists will be obtained by calculating the average JSN score for hands and wrists and multiplying with 40. • If the total number of joints evaluable at the given time point is < 20, then the JSN score for hands and wrists will be set to missing.
Feet (12 joints)	• If total number of joints evaluable at the given time point is ≥ 6 (i.e., 50% of 12), then the JSN score for feet will be obtained by calculating the average JSN score for feet and multiplying with 12. • If the total number of joints evaluable at the given time point is < 6, then the JSN score for feet will be set to missing.

A reader's modified vdH-S score is the sum of the reader's erosion and JNS scores of both hands and feet. If a reader's regional score is missing for any region of erosion or JSN, the reader's modified vdH-S score will be set to missing.

6. Reader Confirmation Rules

For each participant, let Δ1 and Δ2 stand for the change from baseline in modified vdH-S score, respectively, of primary readers 1 and 2 at any post-baseline visit (i.e., Week 24 in Read Campaign 1, Weeks 24 or 48 in Read Campaign 2, and Week 24, 48, or 96 in Read Campaign 3, and Week 24, 48, 96, or 156 in Read Campaign 4). If the absolute difference between Δ1 and Δ2 is greater than or equal to 3 (i.e., $|\Delta 2 - \Delta 1| \geq 3$), or either Δ1 or Δ2 is missing (but not both Δ1 and Δ2 are missing), a confirmation reader (a third reader) will then read all the radiographic images (including baseline and post-baseline images) in that given read campaign from that participant.

7. Reader Selection Rules

For participants who require confirmation (i.e., there are readings from 3 readers), the scores from 2 selected readers will be used in the analysis.

Let Δ1, Δ2, and Δ3 stand for the change from baseline in modified vdH-S score at Week 24, respectively, of primary reader 1 (Reader 1), primary reader 2 (Reader 2), and the confirmation reader (Reader 3). The 2 readers whose scores will be used in the analysis will be selected from the 3 readers based on the criteria specified in the table below.

Rules for Selection of Readers Following Confirmation	
Scenarios based on change from baseline in modified vdH-S score at Week 24	Readers whose scores will be used for the analysis at each visit
$\|\Delta 3 - \Delta 1\| < \|\Delta 3 - \Delta 2\|$	Reader 1 and confirmation reader (Reader 3)
$\|\Delta 3 - \Delta 2\| < \|\Delta 3 - \Delta 1\|$	Reader 2 and confirmation reader (Reader 3)
$\|\Delta 3 - \Delta 1\| = \|\Delta 3 - \Delta 2\|$	Reader 1 and Reader 2
Δ1 is missing but Δ2 and Δ3 are non-missing	Reader 2 and confirmation reader (Reader 3)
Δ2 is missing but Δ1 and Δ3 are non-missing	Reader 1 and confirmation reader (Reader 3)
Δ3 is missing but Δ1 and Δ2 are non-missing	Reader 1 and Reader 2
Both Δ1 and Δ3 are missing, but Δ2 is non-missing	Reader 1 and confirmation reader (Reader 3)
Both Δ2 and Δ3 are missing, but Δ1 is non-missing	Reader 2 and confirmation reader (Reader 3)

8. Smallest Detectable Change

Smallest Detectable Change (SDC) is the smallest change in a score that is considered to be assessed correctly based on the limits of agreement (ie, above the measurement error) (Bruynesteyn 2005).

The SDC in score of interest is determined as follows:

$$SDC = 1.96 * SD / (\sqrt{2} * \sqrt{k}), \text{ where}$$

- SD is the standard deviation of the difference between the 2 selected readers in change from baseline in the score of interest
- k = 2, is the number of readers

6.11. Appendix 11: Summary of Analyses Based on Multiple Imputation

Table 17: Summary of Multiple Imputation Method

Endpoints (Population[a]) Estimand[b]	MI specification	Analysis method/Summary statistics
ACR20 at Week 24 (mFAS) Adjusted Composite Estimand	Multiple imputation with FCS regression of component scores	MIdat_ACR1 (N=200, Seed[d]=18496) • Imputation variables: 7 ACR components from Week 0 - Final scheduled visit of the DBL. • Ancillary variables: Intervention group, randomization strata levels
ACR20 at Week 24 (FAS) Adjusted Composite Estimand	Multiple imputation with FCS regression of component scores	MIdat_ACR2 (N=200, Seed[d]=877681) • Imputation variables: 7 ACR components from Week 0 - Final scheduled visit of the DBL. Ancillary variables: Intervention group, randomization strata levels
ACR20 at Week 24 (mFAS-UKR) Treatment Policy Estimand	Multiple imputation with FCS regression of component scores	MIdat_ACR3 (N=200, Seed[d]=39982) • Imputation variables: 7 ACR components from Week 0 - Final scheduled visit of the DBL. Ancillary variables: Intervention group, randomization strata levels
Modified vdH-S, erosion, JSN (mFAS) Adjusted Treatment Policy Estimand	Multiple imputation with FCS regression of component scores	MIXdat_VDH1 (N=200, Seed[d] =25940) • Imputation variables: hand erosion, foot erosion, hand JSN, and foot JSN from baseline – Final scheduled visit of the DBL. • Ancillary variables[c]: intervention group, randomization stratification factors, and 7 ACR component overtime through Week 24
Modified vdH-S, erosion, JSN (FAS) Adjusted Treatment Policy Estimand	Multiple imputation with FCS regression of component scores	MIXdat_VDH2 (N=200, Seed[d] =12783) • Imputation variables: hand erosion, foot erosion, hand JSN, and foot JSN from baseline – Final scheduled visit of the DBL. • Ancillary variables[c]: intervention group, randomization stratification factors, and 7 ACR component overtime through Week 24
Modified vdH-S, erosion, JSN (mFAS-UKR) Treatment Policy Estimand	Multiple imputation with FCS regression of component scores	MIXdat_VDH3 (N=200, Seed[d] =928374) • Imputation variables: hand erosion, foot erosion, hand JSN, and foot JSN from baseline – Final scheduled visit of the DBL. Ancillary variables[c]: intervention group, randomization stratification factors, and 7 ACR component overtime through Week 24

[a] The population defines which subset of participants the imputation will be performed for.

[b] The handling of ICEs associated with the estimand listed will be applied to the imputation variables and ancillary variables (if post-baseline) prior to imputation.

[c] For the modified vdh-S score which is assessed infrequently compared to other assessments, the 7 ACR components are included in the list of the ancillary variables since they may be related to the mechanism leading to missing data.

[d] The starting seed for FCS regression MI is used to generate a series of imputation seeds using the algorithm: INT((2**31-2)*RANUNI(starting seed)), where each imputation seed will be used for a single imputation. To account for the possibility that some imputations may fail to complete due to out-of-range issues, 200+ initial imputation seeds will be prepared, and the first 200 successful imputations will be used for analysis.

6.12. Appendix 12: Radiographic Image Reading and Scoring

In this study, radiographic images will be read in 4 read campaigns **for efficacy**:

- Read Campaign 1: Baseline and Week 24/ED1
- Read Campaign 2: Baseline, Week 24, and Week 48/ED2
- Read Campaign 3: Baseline, Week 48, and Week 96/ED3
- Read Campaign 4: Baseline, Week 48, Week 96, and Week 156/ED4

Here, the baseline image will typically be taken at the Screening visit. Additionally, the Early Discontinuation (ED) is defined as follows:

ED1=If discontinuation occurs prior to Week 24

ED2= If discontinuation occurs after Week 24, but prior to Week 48

ED3= If discontinuation occurs after Week 48 but prior to Week 96

ED4= If discontinuation occurs after Week 96 but prior to Week 156

If imaging is not performed at the last scheduled time point of each campaign but an ED visit occurs within the analysis window after the projected date from baseline of the missed time point, that ED visit will be included in the reading campaign. ED visits acquired later than that will be read in the subsequent campaign. For example in the case of campaign1 (Week 24), if imaging is not performed at Week 24 but an ED visit with imaging occurs before Week 26 (i.e. within 2 weeks of Week 24), the ED visit will be included in the Week 24 campaign. ED visits occurring after Week 26 (i.e. > 2 weeks from Week 24) will be included in the next campaign.

For Read Campaigns 2 to 4, only those participants who have at least 1 new image taken since the previous read campaign will be read.

Confirmation will be conducted in all read campaigns. During each read campaign, the designated radiographic images will be evaluated independently by the 2 primary readers and, in the case of confirmation, by the confirmation reader. In each read campaign, the readers and their roles will remain the same.

Refer to Section 5.3.5.1.1 and Section 6.10 (Appendix 11) for scoring method and confirmation criteria.

In an effort to avoid the introduction of bias to the reading sessions, readers will be blinded to the following: local site assessments, site and participant ID, medical history and clinical status, treatment assignment, and the results of image evaluations by the other central readers. Additionally, time points will be displayed within the read system in a random order and assigned a blinded time point descriptor (e.g., Time Point A) based on the order in which they are displayed to the readers. This order and randomization scheme will be maintained for each anatomical region and consistent for all readers for each read campaign. Readers are restricted from communicating with study sites that are involved in the protocol for which they are reading.

All study image data received will be processed and saved in Digital Imaging and Communications in Medicine (DICOM) format. During this process, relevant electronic header information (eg, participant identifiers) will be blinded within the digital data set. When presented with an image, the reader will evaluate and score for erosions and joint space narrowing (JSN) of individual joints in a region. Once the reader completes the assessments for erosion/JSN, the scores will be locked and changes to completed assessments will not be permitted.

In order to assess intra-reader variability, images of 30 participants in each read campaign will be randomly selected and re-read by each of the 2 primary readers.

The intra-reader and inter-reader variability will be monitored by the imaging vendor. The targeted intra-class correlation coefficient for status scores is >0.8 for modified vdH-S as well as total erosion and JSN **scores**, and for modified vdH-S **change** is ≥0.5 or smallest detectable change (SDC)≤ 5. If reader agreement thresholds are not met an additional standardization session will be conducted to align the reader performance. If agreement thresholds are still not met further training or replacement of readers will be considered.

Aside from the 4 read campaigns, there is also a separate read on the number of joints with erosions at screening, for **eligibility and stratification** purpose. The results from these reads will not be directly analyzed, only tangentially involved in other analysis as they are used to help determine the randomization strata level. In general, a similar reading process will be used for the **eligibility and stratification** reads as for the **efficacy** reads, although the **eligibility and stratification** reads cannot be blinded to visit, as there is only a single timepoint; additionally, **eligibility and stratification** reads will only assess erosions, not JSN.

For more details regarding imaging acquisition, standardization, reading, and data transfer, refer to the Imaging Charter.

6.13. Appendix 13: Description of Statistical Models

MMRM Model

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 guselkumab group and the placebo group. The model will include treatment group, the interaction terms of visit with treatment group, randomization strata levels, and baseline score 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) Toeplitz 2) first order Autoregressive Moving Average. For analyses through Week 24 the model will include data from all 3 treatment groups through Week 24. The treatment difference between a guselkumab 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.

After Week 24, the MMRM model may still be used to generate LSmeans for each treatment group. However, LSmeans difference between treatments and associated p-values will no longer be calculated.

ANCOVA Model

The ANCOVA model will be used on the change from baseline, under the assumption of MCAR, to test the difference between a guselkumab group and the placebo group. The model will include treatment group, randomization strata levels, and baseline score as explanatory variables. For analyses through Week 24 the model will include data from all 3 treatment groups at Week 24. The treatment difference between a guselkumab 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.

After Week 24, the ANCOVA model may still be used to generate LSmeans for each treatment group. However, LSmeans difference between treatments and associated p-values will no longer be calculated.

GLMM

To account for the missing data for binary endpoints, a GLMM will be used on the response status, under the assumption of MAR, to test the difference between a guselkumab group and the placebo group. The model will include treatment group, the interaction terms of visit with treatment group, and randomization strata levels as explanatory variables. An unstructured covariance matrix for repeated measure within a subject will be used. The logit link will serve as the link function. In case of lack of convergence, empirical structured covariances will be used in the following order until convergence is reached: 1) Toeplitz 2) first order Autoregressive Moving Average. For analyses through Week 24 the model will include data from all 3 treatment groups through Week 24. The proportion difference, its 95% confidence interval, and pvalue will be calculated based on the GLMM.

After Week 24, the GLMM may still be used to generate proportion of response for each treatment group. However, proportion difference between treatments and associated p-values will no longer be calculated.

MLGC Model

To use all available post-treatment data that are collected out-side of the analytical window for Week 24, a mixed effect linear growth model will be fitted to the change from baseline in modified vdH-S score where, effect of treatment on the change from baseline in modified vdH-S score through Week 24 will be estimated by the slope. The fixed effects in the model include the interaction of time with treatment group, and the interaction of time with randomization stratification factors. The model will include a random coefficient for time. Additionally, an intercept term will not be included in the model. Time will be included as a continuous variable. All observed data post treatment in the placebo controlled period will be used.

6.14. Appendix 14: Summary of ICEs and ICE Strategy by Estimand

	ICE Category	ICE Strategy by Estimand		
		Adjusted Composite	**Adjusted Treatment Policy**	**Treatment Policy**
1	Discontinued study intervention injections due to any reason except Natural Disaster or Major Disruption.	• Composite strategy • Consider all subsequent data TF • Priority = high	• Treatment Policy strategy • No action taken, use all observed data • Priority = low	• Treatment Policy strategy • No action taken, use all observed data
2	Initiated or increased the dose of non-biologic DMARDs (MTX, SSZ, HCQ, LEF) or oral corticosteroids over baseline for PsA.	• Composite strategy • Consider all subsequent data TF • Priority = high	• Treatment Policy strategy • No action taken, use all observed data • Priority = low	
3	Initiated protocol prohibited medications/theraies for PsA.	• Composite strategy • Consider all subsequent data TF • Priority = high	• Treatment Policy strategy • No action taken, use all observed data • Priority = low	
4	Discontinued study intervention injections due to Natural Disaster or Major Disruption.	• Hypothetical strategy • Do not use all subsequent data (assuming MAR) • Priority = low	• Hypothetical strategy • Do not use all subsequent data (assuming MAR) • Priority = high	
5	Severe treatment non-compliance due to Natural Disaster or Major Disruption. Defined as: for a given visit, when the total number of doses of study intervention missed due to Natural Disaster or Major Disruption exceeds 30% of the total protocol defined doses from Week 0 up to and including that visit. For Weeks 20, this amounts to ≥2 dose missed	• Hypothetical strategy • Do not use data at subsequent visit (assuming MAR) • Priority = low	• Hypothetical strategy • Do not use data at subsequent visit (assuming MAR) • Priority = high	

	ICE Category	ICE Strategy by Estimand		
		Adjusted Composite	**Adjusted Treatment Policy**	**Treatment Policy**
6	Decided to NOT enter the LTE due to lack of efficacy OR adverse event of worsening of PsA *(only relevant after Week 48)*	• Composite strategy • Consider all data subsequent to Wk48 TF • Priority = high	• Treatment Policy strategy • No action taken, use all observed data • Priority = low	

Note: For continuous endpoints, participants considered TF should be set as change=0; for binary endpoints, participants considered TF should be set as NR.

Note: High priority ICE strategies will supersede low priority ICE strategies should both be applicable to the same visit.

Definitions:

Natural Disaster: site closure, site access restrictions, or lockdowns caused by COVID-19.

Major Disruption: the disruption involving Ukraine and neighboring countries/territories beginning February 24, 2022

7. REFERENCES

Bruynesteyn K, Boers M, Kostense P, et al. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005 64: 179-182.

Cassell SE, Bieber JD, Rich P et al. The modified Nail Psoriasis Severity Index: validation of an instrument to assess psoriatic nail involvement in patients with psoriatic arthritis. The Journal of Rheumatology. 2007;34(1):123-129.

Cella D, Yount S, Sorensen M, et al. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. J Rheum.2005; 32(5):811-819.

Chandran V, Bhella S, Schentag C, Gladman D. Functional Assessment of Chronic Illness Therapy-Fatigue Scale is valid in patients with psoriatic arthritis. Ann Rheum Dis. 2007;66:936-939.

Coates LC, Helliwell PS. Validation of minimal disease activity criteria for psoriatic arthritis using interventional ta. Arthritis Care Res (Hoboken). 2010; 62(7):965-969.

Di Carlo, M, Becciolini A, Lato V et al. The 12-item Psoriatic Arthritis Impact of Disease Questionnaire: Construct Validity, reliability, and interpretability in a Clinical setting. J Rheumatol. 2017;44(3):279-285.

Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology Preliminary Core Set of Disease Activity Measures for Rheumatoid Arthritis Clinical Trials. Arthritis Rheum. 1993; 36(6):729-740.

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

Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. Clin Exp Dermatol. 1994;19(3):210-216.

Fredriksson T, Pettersson U. Severe psoriasis-oral therapy with a new retinoid. Dermatologica. 1978;157(4):238-244.

Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcomes in arthritis. *Arthritis Rheum.* 1980; 23(2):137-145.

Gladman DD, Inman RD, Cook RJ, et al. International spondyloarthritis interobserver reliability exercise--the INSPIRE study: II. Assessment of peripheral joints, enthesitis, and dactylitis. J Rheumatol. 2007;34(8):1740-1745.

Gladman DD, Ziouzina O, Thavaneswaran A, Chandran V. Dactylitis in psoriatic arthritis: prevalence and response to therapy in the biologic era. J Rheumatol. 2013;40(8):1357-1359.

Healy PJ, Helliwell PS. Measuring clinical enthesitis in psoriatic arthritis: assessment of existing measures and development of an instrument specific to psoriatic arthritis. Arthritis Rheum. 2008;59(5):686-691

Helliwell PS, Kavanaugh A. Comparison of composite measures of disease activity in psoriatic arthritis using data from an interventional study with golimumab. Arthritis Care Res. 2014;66(5):749-756.

Khilji FA, Gonzalez M, Finlay AY. Clinical meaning of change in dermatology life quality index scores. Br J Dermatol. 2002;147(suppl 62): 25-54.

Kosinski M, Zhao SZ, Dedhiya S, et al. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum. 2000;43(7):1478–1487.

Lassere M, Boers M, van der Heijde D, et al. smallest detectable difference in radiographical progression. J Rheumatol. 1999; 26(3): 731-739

Mundt JC, Greist JH, Jefferson JW, et al. Prediction of suicidal behavior in clinical research by lifetime suicidal ideation and behavior ascertained by the electronic Columbia-Suicide Severity Rating Scale. J Clin Psych. 2013;74(9):887-893.

Pham T, Guillemin F, Claudepierre P, et al. TNFα antagonist therapy in ankylosing spondylitis and psoriatic arthritis: recommendations of the French Society for Rheumatology. Joint Bone Spine. 2006;73:547–553.

Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011;168(12):1266-1277.

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

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

Smolen JS, 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 der Heijde DM, van Leeuwen MA, van Riel PL, et al. Biannual radiographic assessments of hands and feet in a three-year prospective followup of patients with early rheumatoid arthritis. Arthritis Rheum. 1992;35(1):26-34.

van Riel PL, van Gestel AM, Scott DL. EULAR Handbook of Clinical Assessments in Rheumatoid Arthritis. Alphen Aan Den Rijn, The Netherlands: Van Zuiden Communications B.V.; 2004

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

Yount S, Sorensen MV, Cella D, et all. 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):838-46

Signature

User	Date	Reason
PPD	21-May-2025 21:26:58 (GMT)	Document Approval
PPD	22-May-2025 16:22:50 (GMT)	Document Approval