Janssen Research & Development *

Statistical Analysis Plan

A Phase 3, Multicenter, Randomized, Double-blind, Placebo-controlled Study Evaluating the Efficacy and Safety of Guselkumab Administered Subcutaneously in Subjects with Active Psoriatic Arthritis including those Previously Treated with Biologic Anti-TNFα Agent(s)

Protocol CNTO1959PSA3001; Phase 3
AMENDMENT 3

CNTO1959 (guselkumab)

Status: Approved
Date: 21 March 2019
Prepared by: Janssen Research & Development, LLC
Document No.: EDMS-ERI-146207736

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

AMENDMENT HISTORY

AMENDMENT 1 (APPROVAL DATE: MAY 25, 2018)

Summary of Major Changes and Modifications

This statistical analysis plan (SAP) was amended to implement the following major changes and modifications to the original SAP:

1. Modified treatment failure (TF) rules

2. Modified multiplicity control method to separate US and rest of the world

    - Rest of the world: modified overall control for all endpoints and both dose levels to control 2 doses for the primary endpoints, and control within dose level for primary and all major secondary endpoints

    - US: added overall control for all endpoints and both dose levels to control selective endpoints and both dose levels by removing highly correlated major secondary endpoints from controlled list

3. Modified the Multiple Imputation (MI) methods

    - Impute the continuous components for binary endpoints instead of the composite binary endpoint itself

    - Use the Full Conditional Specifications (FCS) to replace the imputation first imputed non-monotone missing values and then imputed monotone missing values.

    - Added Appendix 4 to detail the MI methods for endpoints with possibility to use MI method

4. Modified or added definitions for estimands and list of efficacy endpoints using each of the estimands

    - Modified the composite estimand for binary endpoints with 4 components for clarity and expanded endpoints using composite estimand from primary and major secondary to all binary endpoints

    - Modified the treatment policy estimand with 4 components for clarity and expanded endpoints using this estimand from primary endpoint only to also including major secondary endpoints of change from baseline in HAQ score, DAS28(CRP), SF36-PCS and SF36-MCS at Week 24

    - Replaced the treatment efficacy estimand with the composite estimand for continuous endpoints and used this estimand for all continuous endpoints

- Added an expanded composite estimand to incorporate expanded treatment failure rules and used this estimand as a supplementary analysis for the primary endpoint

- Added a Per-Protocol estimand and used this estimand as a supplementary analysis for primary endpoint

5. Added and modified analysis methods for efficacy endpoints

- Modified MMRM model by adding the interactions of the fixed effects to use a model with full parameters, and added cLDA method if more than 2% subjects have missing baseline measurements.

- For continuous major secondary endpoints, added analysis methods when the data have extreme departure from normality

- Changed analysis method of other binary endpoints from a GLMM model or logistic regression to a CMH test

- Modified or added sensitivity/supplementary analyses for primary endpoint of ACR 20 response at Week 24

    a) For the composite estimand, when the main analysis is based on non-responder imputation, replace the sensitivity analysis based on MAR tipping point analysis to Exhaustive Scenario Imputation tipping analysis to cover all possible scenarios

    b) Added analyses based on the expanded composite estimand

    c) Added analyses based on the Per-Protocol estimand

- Added a summary table for analysis methods at the end of each section by the category of efficacy endpoints

- Removed the endpoints related to mCPDAI and related analyses since DLQI, a component of mCPDAI, is not collected

6. Added a section for analysis of suicidal ideation and behavior data

Additionally, non-content related wording or formatting changes were made to enhance clarity and to generate 2 consistent SAPs.

AMENDMENT 2

Summary of Major Changes and Modifications

This statistical analysis plan (SAP) was amended to implement the following modifications to the original SAP:

1. Section 2.4.
    a. Revised the categories for the race subgroup - #1b
    b. Instead of subgrouping by geographical region defining a subgroup called participating countries - #1g.
    c. Added subgroups #2l, #3g, #3h, and #3i.
2. Added clarification to treatment failure (section 2.5) criterion #1 to include adverse events caused by worsening of PsA.
3. Modified multiplicity control method (section 5.2.2).
4. Added clarification to section 5.2.3.3 regarding the imputation of each variable within its possible range of values.
5. Added a supplementary analysis for IGA response based on the treatment policy estimand to section 5.4.3.3 and Table 6.
6. Modified assessment of normality of residuals from MMRM/cLDA models by using Q-Q plots in addition to the Shapiro-Wilks test (sections 5.1 and 5.4.1.2, and Table 4).
7. Modified the enthesitis scoring for subjects with an incomplete set of 6 evaluated sites at a particular visit (section 5.4.4.1). Removed the use of non-missing site score based imputation to adjust for sites with missing scores.
8. Modified analysis methods for enthesitis and dactylitis related endpoints (sections 5.4.4.2, 5.4.5.2, and 5.5.1.2) in accordance with the multiplicity control method (section 5.2.2).
9. Added an assessment for the presence of tender dactylitis (section 5.5.1.1.6)
10. Added the definition for low or very lose disease activity based on the PASDAS score (section 5.5.1.1.7) and a corresponding analysis to Table 5.
11. Added the definition of low disease activity based on the GRACE index score (section 5.5.1.1.8) and a corresponding analysis to Table 5.
12. Added the definition of remission or low disease activity based on the DAPSA score (section 5.5.1.1.9) and a corresponding analysis to Table 5.

13. Added the definition of very low disease activity (section 5.5.1.1.10) and a corresponding analysis to Table 5.

14. Added the requirement of BASDAI>0 at baseline for the analysis of the change from baseline in BASDAI score (section 5.5.1.1.11 and Table 5).

15. Deleted DLQI related endpoints from section 5.5.2 as these had been erroneously included in this Statistical Analysis Plan.

16. Added clarification to section 6.4, regarding reporting of abnormalities or changes during physical examination as adverse events.

17. Separated joint evaluability rules for surgical joint procedures and injections administered due to PsA or non-PsA reasons in Appendix 1 (Joint Evaluability Rules for Sign and Symptom Data).

18. Added clarifications to the summary of multiple imputation methods in Appendix 4.

    a. Imputation variables to be used for MIdataset2.

    b. Use of enthesitis-4, which is the tender entheses count based on 4 sites (left and right achilles tendon insertion, and left and right humeral epicondyle lateral) instead of 6 sites at baseline and Week 2, in the imputation of enthesitis scores.

    c. Change from baseline at Week 24 in enthesitis score will be calculated among subjects with at least one tender enthesis at baseline.

    d. Added footnotes “c” and “d”.

AMENDMENT 3

Summary of Changes in Amendment 3 to the SAP

1. Revised the definition of subgroups (section 3.4):

    a. Age at baseline (year): < 45, ≥ 45 and < 65, ≥ 65 (revised definition)
    Age at baseline (year): < 65, ≥ 65 (previous definition)
    b. Clarified that the following disease characteristics are assessed at baseline: PsA duration, number of swollen joints, number of tender joints, HAQ-DI, dactylitis, enthesitis, PASI, BSA, and IGA.
    c. Clarified that the subgroup of use of non-biologic DMARDs at baseline (Yes, No) and prior anti-TNFα use (Yes, No) are based on the eCRF data and not the IWRS randomization factors.
    d. Removed the category of "any DMARDs" from the subgroup of non-biologic DMARDs at baseline. The revised definition includes only three categories: None, MTX, non-MTX DMARDs.
    e. Edited the descriptors for the reason of discontinuation of prior DMARDs and prior anti-TNFα use.

2. Revised the definition of the Treatment Failure Criteria (Section 2.5). The TF criteria are expanded by considering discontinuation of study agent due to other reasons (in addition to lack of efficacy) as a criterion for TF. This is reflected in Criterion 1 which now states - Discontinued study agent injections due to any reason.

3. Added the following paragraph to Section 4:

    Subject information will be summarized by randomized treatment group based on the FAS1. Treatment exposure (e.g., number of treatment administrations, and cumulative dose received) and duration of study follow-up time will be summarized by treatment actually received based on the safety analysis set (Section 4.4).

4. Added the following information to Section 5.1 for binary response efficacy endpoints:

    The Mantel Fleiss criterion will be used to determine the appropriateness of using the CMH test at each visit for each treatment pair under comparison. If the Mantel Fleiss criterion is not satisfied the Fisher's exact test will be used instead of the CMH test to compare the two treatment groups.

    For analyses based on Multiple Imputation, 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[27] transformation is used to pool the p-values from each imputed data set.

5. For major secondary efficacy endpoints the analysis will be performed using an ANCOVA model based on MI data and will no longer use a Mixed Effect Repeated Measures or Constrained Longitudinal Data Analysis Model. The previously planned sensitivity analyses using an ANOVA model based on the van der Waerden normal score will not be performed. These changes to the plan are reflected in the edits in Section 5.1 for major secondary continuous efficacy endpoints. Added details regarding the calculation of the confidence intervals and p-values.

6. Revisions to section 5.2.2.1 (US-Specific Multiplicity Adjustment for Testing Procedures).

7. Added section 5.2.2.2 (Global Multiplicity Adjustment for Testing Procedure)

8. Added the following to Section 5.2.3.3.1:

   For categorical endpoints (IGA response, enthesitis resolution, dactylitis resolution), the above imputation will be performed for the respective scores on a continuous scale, then rounded to the nearest integer prior to deriving the response or resolution.

9. Revised the definition of the Treatment Failure Criteria (Section 2.5). The TF criteria are expanded by considering discontinuation of study agent due to other reasons (in addition to lack of efficacy) as a criterion for TF. This is reflected in criterion 1 which now states - Discontinued study agent injections due to any reason.

10. Added a definition for the alternative composite estimand (Section 5.2.4.2).

11. Added the following to Section 5.2.4.3:

    The Treatment Policy Estimand will be analyzed for all major secondary endpoints (Table 4a), and selected endpoints (Table 7a) analyzed at Week 16.

12. In section 5.3.2.1, for the primary endpoint, deleted the supplementary analyses based on the previously defined expanded composite estimand, and added a supplementary analysis based on the alternative composite estimand. This is also reflected in Table 3.

13. Added clarification to Section 5.4 about the analysis sets to be used for psoriasis response of IGA, change from baseline in enthesitis score, resolution of enthesitis, change from baseline in dactylitis score. and resolution of dactylitis.

14. Section 5.4.1.2 reflects changes to the analysis methods for HAQ-DI score as those described in item 5 (above) and in Section 5.1.

15. Sections 5.4.6.2 and 5.4.7.2 reflect changes as those described in item 5 above. These changes are also reflected in Table 4, which summarizes analyses related to major secondary endpoints.

16. Added section 5.4.9 (Additional Tipping Point Analyses) that describes the tipping point analyses to be performed on the major secondary endpoints based on the Treatment Policy Estimand.

17. In section 5.5.1.2, added that change from baseline in HAQ-DI, DAS28 (CRP), dactylitis, and enthesitis scores will be analysed using an ANCOVA model on MI data.

18. Added 4 endpoints - Change from baseline in FACIT-fatigue score at Week 24 by ACR 20 response (primary endpoint) at Week 24, Proportion of subjects who achieve ≥ 4-point improvement from baseline in FACIT-fatigue score at Week 24 by ACR 20 response (primary endpoint) at Week 24, Proportion of subjects who achieve an improvement of ≥ 3 points in PROMIS-29 domain scores by visit through Week 24, and Proportion of subjects who achieve an improvement of ≥ 5 points in PROMIS-29 domain scores by visit through Week 24 – to Section 5.5.3 and Table 7.

19. In Section 5.5.3.2:

    - added that the change from baseline in SF-36 PCS score and SF-36 MCS score will be also be analyzed using an ANCOVA model on MI data.

    - Added a description of additional analyses for FACIT-Fatigue.

20. Added section 5.5.4, that explains tipping point analyses to be performed for endpoints other than the primary and major secondary endpoints.

21. Added the following 3 references:

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

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

L. Valeri and T. J. VanderWeele (2013), " Meidation Analysis Allowing for Exposure-Mediator Interactions and Causal Interpretation: Theoretical Assumptions and Implementation with SAS and SPSS Macros", Psychological Methods, Vol. 18, No.2: 137-150.

22. Added description of Mediation Analysis to Appendix 3.

23. Revisions to Appendix 4:

a. Changed "at" to "through" for endpoints HAQ-DI score, DAS28 (CRP) score, Enthesitis score, Dactylitis score, PCS score, and MCS score.

b. Added MI for IGA change from baseline through Week 24.

c. Edit to footnote "a":

i. MIdataset2 is different from MIdataset1 as the joint scores used in DAS 28 (CRP) are different from those used in ACR. (current footnote)

ii. MIdataset2 is different from MIdataset1 as it is based on a different Estimand. The same, MIdataset2, data set can be used for both HAQ-DI and DAS28(CRP) analyses. (previous footnote).

d. Edits to footnote "d":

i. Enthesitis-4 is the tender entheses count based on 4 sites (left and right achilles tendon insertion, and left and right humeral epicondyle lateral) instead of 6 sites. Only baseline, Week 2, Week 4 are included as ancillary variables. (current footnote)

ii. Enthesitis-4 is the tender entheses count based on 4 sites (left and right achilles tendon insertion, and left and right humeral epicondyle lateral) instead of 6 sites at baseline and Week 2. (previous footnote)

24. Revised Appendix 2.

Note that changes to the analytic approaches in these amended SAPs also apply to the analysis plan for the Integrated Summary of Efficacy.

ABBREVIATIONS

ACR	American College of Rheumatology
AE	adverse event
ALT	Alanine aminotransferase
AMDF	Arithmetic Mean of the Desirability Function
ANA	Antinuclear antibodies
ANCOVA	analysis of covariance
ANOVA	Analysis of variance
ARMA	Autoregressive moving average
AST	Aspartate aminotransferase
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BMI	Body mass index
BSA	Body surface area
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
CRP	C-reactive protein
CSR	Clinical Study Report
DAPSA	Disease Activity Index for Psoriatic Arthritis
DAS28	Disease Activity Index Score 28
DBL	database lock
DIP	Distal interphalangel
DLQI	Dermatology Life Quality Index
DMARDS	Disease-Modifying Antirheumatic Drugs
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	electronic case report form
EE	Early escape
EQ-5D	EuroQol 5 Dimensions Health-Related Quality of Life
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy - Fatigue
FAS	Full Analyses Set
FCS	Full Conditional Specification
GDEV	Physician's Global Assessment of Disease Activity
GDPT	Patient's Global Assessment of Disease Activity
GMS	Global Medical Safety
GRAPPA	Group for Research and Assessment of Psoriasis and Psoriatic Arthritis
HAQ-DI	Health Assessment Questionnaire Disability Index
HCQ	Hydroxychloroquine
ITT	Intent-to-Treat
IWRS	interactive web response system
LEF	Leflunomide
LEI	Leeds Enthesitis Index
LLOQ	Lower limit of quantification
LS means	Least square means
MCP	Metacarpophalangeal Joint
MCS	Mental Component Score
MDA	Minimal Disease Activity
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed-effect repeated measures
MTX	Methotrexate
NCI-CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
NRI	Non-responder imputation
NSAID	non-steroidal anti-inflammatory drug
PAIN	Patient's assessment of pain
PASDAS	Psoriatic ArthritiS Disease Activity Score
PASI	Psoriasis Area and Severity Index

PCS	Physical Component Summary
PsARC	Modified Psoriatic Arthritis Responder Criteria
PD	Pharmacodynamic
PIP	proximal interphalangeal
PK	pharmacokinetic(s)
PRO	patient reported outcome
PROMIS	Patient-Reported Outcomes Measurement Information System
PsA	Psoriatic Arthritis
q4w	every 4 weeks
q8w	every 8 weeks
RA	Rheumatoid arthritis
RBC	Red blood cell
SAE	serious adverse event
SAP	Statistical Analysis Plan
SC	subcutaneous
SD	standard deviation
SF-36	36-Item Short Form Survey Instrument
SJC	Swollen joint count
SPARCC	Spondyloarthritis Research Consortium of Canada
SSG	Statistical support group
SSZ	Sulfasalazine
SUA	Serious unexpected adverse event
TB	Tuberculosis
TF	Treatment failure
TJC	Tender joint count
TNF	tumor necrosis factor
VAS	Visual analog scale
WBC	White blood cell

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for the analysis of efficacy, safety, pharmacokinetics (PK), pharmacodynamics (PD), Immunogenicity and health related quality of life in the CNTO1959PSA3001 study.

1.1. Trial Objectives

Primary Objective

The primary objective of this study is to evaluate the efficacy of SC administration of guselkumab 100 mg in subjects with active psoriatic arthritis (PsA) by assessing the reduction in signs and symptoms of PsA.

Secondary Objectives

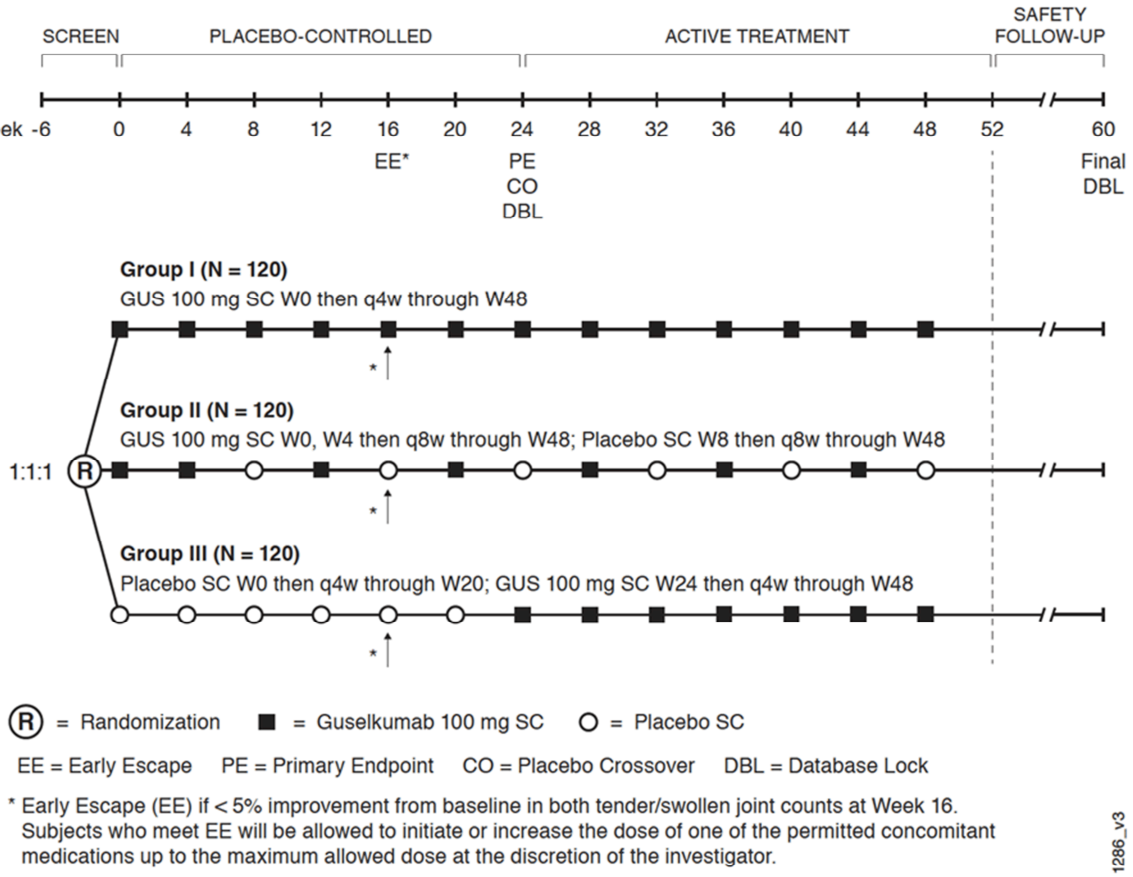
The secondary objectives are to assess the following for guselkumab treatment in subjects with active PsA:

- Efficacy in improving psoriatic skin lesions, physical function, health-related quality of life, and patient-reported health outcomes
- Safety
- PK, PD, and immunogenicity

1.2. Trial Design

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, 3-arm study of guselkumab in subjects with active PsA who had inadequate response to standard therapies (eg, non-biologic DMARDs, apremilast, or NSAIDs). In addition, subjects (approximately 30%) may have been previously treated with up to 2 anti-TNFα agents. An overview of the study design is provided in Figure 1.

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



At Week 0, approximately 360 subjects who have met the study inclusion and exclusion criteria are to be randomized in a blinded fashion in a 1:1:1 ratio to 1 of the following 3 treatment groups using permuted block randomization stratified by baseline use of non-biologic DMARD (MTX, SSZ, HCQ, LEF) use (yes, no) and prior exposure to anti-TNFα agents (yes, no):

- **Group I** (n=120): Subjects will receive SC guselkumab 100 mg every 4 weeks (q4w) from Week 0 through Week 48.
- **Group II** (n=120): Subjects will receive SC guselkumab 100 mg at Weeks 0 and 4, then q8w (Weeks 12, 20, 28, 36, and 44) and placebo injections at other visits (Weeks 8, 16, 24, 32, 40, 48) to maintain the blind.
- **Group III** (n=120): Subjects will receive SC placebo q4w from Week 0 to Week 20, and will crossover at Week 24 to receive guselkumab 100 mg q4w through Week 48.

Through the study, stable doses of concomitant NSAIDs, oral corticosteroids, and selected non-biologic DMARDs (limited to MTX, SSZ, HCQ, LEF, see Table 1) will be allowed but are not required. Subjects should not initiate any new treatment for PsA through Week 60, except at Week 16. At Week 16, subjects in all treatment groups who have < 5% improvement from baseline in both swollen and tender joint counts will be considered as meeting early escape (EE) criteria and will, be allowed to initiate or increase the dose of one of the permitted concomitant medications up to the maximum dose allowed as specified in Table 1, as selected by the investigator.

Table 1: Permitted Concomitant Medications for PsA and Maximum Doses Allowed During the Study

Permitted Concomitant Medications for PsA[a,b]	**Maximum Dose Allowed**
NSAIDs and other analgesics	Marketed dose approved in the country where the study is being conducted
Oral corticosteroids	Equivalent to 10 mg/day of prednisone
Non-biologic DMARDS:	
Methotrexate (MTX)[c]	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] Subjects may not be receiving more than one non-biologic DMARD from baseline through Week 60.

[c] It is recommended that all subjects 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.

At Week 24, all subjects who are still on treatment in the placebo group (Group III) will cross over to receive guselkumab 100 mg SC q4w from Week 24 through Week 48. Subjects in the guselkumab groups (Groups I and II) will remain on the same dose regimen at Week 24 and after, that is, there will be no change to the dose regimen.

Subjects will be followed for adverse events (AE) and serious adverse events (SAE) up to 12 weeks following the last study treatment administration.

The end of the study is defined as the last visit for the last subject. The last visit is the last safety follow-up visit for subjects who complete the Week 48 study treatment and for subjects who terminate study treatment prior to Week 48.

There are 2 planned database locks (DBLs) in this study. Each of these 2 DBLs will result in a study report. The first DBL will occur when all subjects randomized in this study have either completed the Week 24 assessments or terminated study participation prior to the Week 24 visit (referred to as **Week 24 DBL** hereafter). The second DBL will occur when all subjects

randomized in this study have either completed the Week 60 assessments or terminated study participation prior to the Week 60 visit [referred to as **Final (Week 60) DBL** hereafter].

The primary endpoint of this study is the proportion of subjects who achieve a 20% improvement from baseline in the American College of Rheumatology criteria (ACR 20) at Week 24 (refer to Section 5.3 for endpoint definition and analyses). This endpoint was chosen because it is well-accepted by regulatory authorities and the clinical PsA community. Additionally, there are 13 major secondary endpoints in the study (Section 5.4 – definitions and analyses methods). The primary and major secondary endpoints will be analyzed at Week 24 DBL.

The study will remain blinded for the duration of the trial, until after the Final (Week 60) DBL. Selected Sponsor personnel will be unblinded at Week 24 DBL for the purposes of performing data analysis and data review. The details regarding the Sponsor personnel that will be unblinded at Week 24 DBL will be provided in the Data Release Plans.

An independent, external Data Monitoring Committee (DMC) will monitor the safety of the study in an unblinded fashion on a regular basis and whenever deemed necessary. Additional details related to DMC are provided in Section 3.2.

1.3. Statistical Hypotheses for Trial Objectives

The primary endpoint of this study is proportion of subjects who achieved an ACR 20 response at Week 24 (refer to Section 5.3 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. 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 subjects who achieved an ACR 20 response at Week 24 (***primary hypothesis***); and
2. 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 subjects who achieved an ACR 20 response at Week 24 (***major 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 are 13 major secondary endpoints in this study (refer to Section 5.4 for the list of these endpoints and their definitions). The hypotheses related to the major secondary endpoints (***all are major secondary hypotheses***) are provided in Appendix 2.

For hypothesis testing order and multiplicity adjustment, refer to Section 5.2.2.

1.4. Sample Size Justification

The sample size was chosen based on the data from a Sponsor's recent PsA study, CNTO1275PSA3002 that included the subjects previously treated with biologic anti-TNFα agents. The ACR 20 response rates at Week 24 in CNTO1275PSA3002 study were 20.2%, 43.7% and 43.8%, respectively, for the placebo, ustekinumab 45 mg, and ustekinumab 90 mg treatment groups. In order to ensure a statistical power of > 90% at a 2-sided α-level of 0.05, assuming a 40% ACR 20 response rate in the guselkumab group and a 20% ACR 20 response rate in the placebo group at Week 24, a total of 360 subjects are planned to be randomized in a 1:1:1 ratio to each of treatment groups. Table 2 provides the power evaluation of various assumptions.

Table 2: Statistical Power for the ACR 20 Response Rate Comparing to Placebo at Week 24

	Treatment group	Sample size	ACR 20 response	Δ (difference)	Power
1	Placebo	120	20%	20%	93%
	Guselkumab 100 mg	120	40%		
2	Placebo	120	25%	20%	91%
	Guselkumab 100 mg	120	45%		
3	Placebo	120	20%	25%	99%
	Guselkumab 100 mg	120	45%		
4	Placebo	120	25%	25%	98%
	Guselkumab 100 mg	120	50%		

1.5. Randomization and Blinding

1.5.1. Randomization

A central randomization is to be implemented in this study using an interactive web response system (IWRS). When a subject is eligible for randomization at a study site, the randomization requestor at that study site will contact the IWRS using the requester's own user identification and personal identification number and provide the relevant subject details to uniquely identify that subject. Based on a computer-generated randomization schedule prepared before the study under the supervision of the Sponsor, the IWRS will then assign a unique treatment code, which will dictate the treatment assignment and matching study agent kit for that subject.

To assure relatively even treatment balance within each stratum defined by baseline use of non-biologic DMARD (yes, no) and prior exposure to anti-TNFα agents (yes, no) randomization at Week 0 is determined using permuted block randomization stratified by baseline use of non-biologic DMARD (yes, no) and prior exposure to anti-TNFα agents (yes, no). Specific details are provided in the IWRS Specification Unbinding Randomization for CNTO1959PAS3001 (JAN333).

At Week 24, all subjects who were randomized to placebo at Week 0 and who are still on study treatment at Week 24 will be switched to receive guselkumab 100 mg q4w by the IWRS to receive guselkumab 100 mg q4w starting from Week 24.

1.5.2. Maintenance of the Blind

The study blind will be maintained for the duration of the study, until after the Final (Week 60) DBL at end of study.

To maintain the study blind, the study agent container will have a multipart label with directions for use and other information, but not the identity of the study agent, on each part. One part of the label is designed to be torn off, separated from the study agent container, and attached to the subject's source documents. The rest of the label will remain affixed to the study agent container. Thus, the study agent assigned to a subject will be linked between the container and the subject without breaking the study blind. The study agent kit number will be entered in the electronic case report form (eCRF) or other equivalent data capture method when the drug is administered. The investigator will not be provided with treatment codes, but the codes will be maintained within the IWRS, which has the functionality to allow the investigator to break the blind for an individual subject when there is a need based on medical judgment.

Data that may potentially unblind the treatment assignment (i.e., study agent serum concentrations, antibodies to study agent) will be handled with special care to ensure that the integrity of the blind is maintained and the potential for bias is minimized. This may include making special provisions, such as segregating the data in question from view by the investigators, clinical team, or others as appropriate.

An investigator may be unblinded to a given subject's treatment allocation when specific emergency treatment would be dictated by knowing the treatment status of the subject. In such cases, the investigator may determine the identity of the treatment by contacting the IWRS provider. It is strongly recommended that the investigator contact the Sponsor or its designee if possible to discuss the particular situation prior to unblinding via IWRS. Telephone contact with the Sponsor or its designee will be available 24 hours per day, 7 days per week. In the event that the investigator is unable to contact the Sponsor, or emergency unblinding is considered medically necessary, the investigator may determine the identity of the treatment via /IWRS. However, the Sponsor must be informed as soon as possible. The date, time, and reason for the unblinding must be documented in the appropriate section of the eCRF, and in the source document. The documentation received from the /IWRS indicating the code break must be retained with the subject's source documents in a secure manner (e.g., sealed envelope) so as to not unblind the treatment assignment to the subject, the study site, or Sponsor personnel. The investigator is also advised not to reveal the study treatment assignment to the subject, the study site, or Sponsor personnel. Subjects who have had their treatment assignment unblinded are expected to continue to return for scheduled evaluations. Further study agent administrations should be discussed with the study responsible physician.

A given subject's treatment assignment may be unblinded to the Sponsor, Institutional Review Board/Independent Ethics Committee (IRB/IEC), and site personnel to fulfill regulatory reporting requirements for suspected unexpected serious adverse reactions (SUSARs). A separate code break procedure will be available for use by Janssen Global Medical Safety (GMS) to allow for unblinding of individual subjects to comply with specific requests from regulatory or health authorities.

A number of prospectively identified Sponsor individuals will be unblinded at the Week 24 DBL for the purposes of performing data analysis and review. Identification of sponsor personnel who will have access to the unblinded data at subject-level and who will have access to the unblinded data at group-level will be documented in the Data Release Plan prior to unblinding at the Week 24 DBL. Investigative sites and subjects will remain blinded to treatment assignment for the duration of the study, till after the Final (Week 60) DBL.

An independent, external DMC will monitor the safety of the study in unblinded fashion on a regular basis and whenever deemed necessary. In addition, the Sponsor Medical Monitor will review safety data in a blinded manner as the study is ongoing. The DMC's roles and responsibilities, the safety data for DMC review, and other related information (such as, the general procedures, communications, etc.) are defined and documented in the DMC charter.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Visit Windows

2.1.1. Visit Windows for Dosing and PK Analysis

All post-baseline visits through Week 24 will have a visit window of ± 4 days. All the visits after Week 24 through Week 52 will have a visit window of ± 7 days (1 Week). The final safety visit will have a visit window of 14 days (± 2 weeks).

Study agent administrations must always be at least 14 days apart.

For PK analyses, if a subject 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.

2.2. Pooling Algorithm for Analysis Centers

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

2.3. Analysis Sets

2.3.1. Efficacy Analysis Set(s)

2.3.1.1. Full Analysis Set 1 (Week 0 – Week 24)

The full analysis set 1 (FAS1) includes all randomized subjects who received at least 1 dose (complete or partial) of study agent. This analysis set will be used for the efficacy analyses of endpoints through Week 24.

2.3.1.2. Full Analysis Set 2 (Week 24 – Week 52)

The full analysis set 2 (FAS2) includes all randomized subjects who were still on study treatment at Week 24. This analysis set will be used for the efficacy analysis of endpoints from Week 24 through Week 52.

2.3.1.3. Per-Protocol Analysis Set (Week 0 – Week 24)

The per-protocol analysis set (PPAS) includes all subjects in FAS1 who met all inclusion and exclusion criteria and had no major protocol deviation 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. Subjects to be excluded from this analysis will be identified prior to Week-24 DBL and unblinding.

In the efficacy analyses, subjects will be analyzed according to the randomized treatment groups they were assigned to, regardless of the treatments they actually received.

2.3.2. Safety Analysis Set

The safety analysis set includes all subjects who received at least 1 (partial or complete) dose of study agent, i.e., the treated population.

In the safety analyses, subjects will be analyzed according to the treatment they actually received, regardless of the treatments they are randomized to.

2.3.3. PK Analysis Set

The PK analysis set includes all subjects who received at least 1 complete dose of guselkumab and had at least 1 valid blood sample drawn for PK analysis.

In the PK analyses, subjects will be analyzed according to the treatment they actually received, regardless of the treatments they are randomized to.

2.3.4. Immunogenicity Analysis Set

The immunogenicity analysis set includes all subjects who received at least 1 (partial or complete) dose of guselkumab and who have at least 1 sample obtained after their first dose of guselkumab for the detection of antibodies to guselkumab.

In the immunogenicity analyses, subjects will be analyzed according to the treatment they actually received, regardless of the treatments they are randomized to.

2.3.5. PD Analysis Set

The PD analysis set includes all subjects who received at least 1 (partial or complete) dose of study agent, i.e., the treated population.

In the PD analyses, subjects will be analyzed according to the treatment they actually received, regardless of the treatments they are randomized to.

2.4. Definition of Subgroups

To evaluate the consistency in the primary efficacy endpoint (proportion of subjects who achieve ACR 20 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:

1. **Demographic subgroups**
   a. Gender: Male, Female
   b. Race: White, Other
   c. Age at baseline (year): < 45, ≥ 45 and < 65, ≥ 65
   d. Body weight at baseline (kg): ≤ 90, >90
   e. Body weight at baseline (kg): (quartiles)
   f. Body mass index (BMI) at baseline: Normal [< 25], Overweight [≥ 25 to < 30], Obese [≥ 30]
   g. Participating countries: Poland, Russia, Ukraine, Western countries (USA, Canada, Spain, and Germany), Other countries (combining countries in Asia and Eastern European countries other than Poland, Russia and Ukraine)
2. **Baseline disease characteristics subgroups**
   a. PsA duration at baseline (year): < 1, ≥ 1 to < 3, ≥ 3
   b. PsA subtype: distal interphalangeal joint involvement, polyarticular arthritis with absence of rheumatoid nodules, arthritis mutilans, asymmetric peripheral arthritis, spondylitis with peripheral arthritis
   c. Number of swollen joints at baseline: < 10, 10 -15, > 15
   d. Number of tender joints at baseline: < 10, 10 -15, > 15
   e. HAQ-DI at baseline: < 1, 1 - 2, > 2
   f. CRP at baseline (mg/dL): < 1, 1-2, ≥ 2
   g. CRP at baseline (mg/dL): (quartiles)
   h. Dactylitis at baseline: Yes, No
   i. Enthesitis at baseline: Yes, No
   j. PASI at baseline: <12, 12-20, >20
   k. BSA affected by psoriasis at baseline: <3, 3-10%, 10-20%, >20%
   l. IGA at baseline: <2, ≥ 2
3. **Prior and baseline medication use subgroups:**
   a. Prior anti-TNFα use (eCRF): Yes, No
   b. Number of prior anti-TNFα agents: 1, 2.
   c. Reason for discontinuation for prior anti-TNFα: Efficacy - inadequate response (IR), Safety - contraindication or intolerance (but not IR), Other
   d. Use of non-biologic DMARDs (MTX, SSZ, HCQ, LEF) at baseline (eCRF): Yes, No
   e. Oral corticosteroids at baseline: Yes, No
   f. NSAIDs at baseline: Yes, No
   g. Number of prior non-biologic treatments including DMARDs, systemic immunosuppressive drugs, and apremilast: 0, 1, 2, ≥ 3.
   h. Non-biologic DMARDS (MTX, SSZ, HCQ, LEF) at baseline: None, MTX, non-MTX DMARDs

i. 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 subjects within a subgroup category.

2.5. Treatment Failure

A subject will be considered a treatment failure (TF) from the earliest date that the subject meets any of the following TF criteria onward through Week 24:

1. Discontinued study agent injections due to any reason.
2. Terminated study participation due to any reason.
3. Initiated or increased the dose of non-biologic DMARDs (MTX, SSZ, HCQ, LEF) or oral corticosteroids over baseline for PsA.
4. Initiated protocol prohibited medications/therapies for PsA.

Note that all subjects who meet criterion 2 will always have met criterion 1.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

3.1. Interim Analysis

No interim analysis is planned for this study.

3.2. Data Monitoring Committee

An independent, external DMC has been established to monitor data on an ongoing basis to ensure the continuing safety of the subjects while participating in this study. All subjects are to be followed by the DMC from the first dose administered through at least the Week 24 DBL. The DMC consists of 3 members (including 2 medical experts in the relevant therapeutic areas and 1 statistician) who are independent of the Sponsor. None of the members is participating in the current study. An independent statistical support group (SSG), not affiliated to the sponsor, supports the DMC and serves the liaison between the DMC and the sponsor.

The major function of the DMC is to monitor the safety of the study agent by reviewing the serious adverse events (SAEs) each month and by periodically reviewing the interim study safety data every 4-months. After each review, the DMC is to make recommendations regarding the continuation of the study or, in the event that any unanticipated serious events occur, placing the study on hold or stopping the study.

The DMC will have access to unblinded data and review tabulated safety summaries (if appropriate) and any additional data as deemed necessary during the conduct of the study. No formal statistical hypothesis testing is planned.

The content of the safety summaries, the DMC role and responsibilities and the general procedures (including communications) and their recommendations on the study conduct are

defined and documented in the DMC charter, which will be finalized prior to the first DMC review.

In addition, during the study, the Sponsor's study responsible physician (or designee) will regularly review blinded safety data from the sites and notify the DMC and appropriate Sponsor personnel of any issues.

4. SUBJECT INFORMATION

4.1. Subject information will be summarized by randomized treatment group based on the FAS1. Treatment exposure (e.g., number of treatment administrations, and cumulative dose received) and duration of study follow-up time will be summarized by treatment actually received based on the safety analysis set (Section 4.4).Demographics and Baseline Characteristics

Demographic and baseline characteristic variables will be descriptively summarized by randomized treatment group. No formal statistical comparison is planned. P-values will not be provided.

Demographic variables to be summarized will include sex, race, age, height, baseline weight, and baseline body mass index (BMI). Baseline characteristic variables to be summarized will include, but not be limited to, baseline disease characteristics of PsA (e.g., duration of disease, PsA subtypes, baseline efficacy assessments), medical history, prior medication exposure, prior joint procedures/injections, and baseline medication usage for PsA.

In addition, balance between randomized treatment groups will also be explored for the subpopulations used for analyses below:

- subjects with baseline ≥ 3% body surface area (BSA) psoriatic involvement and an IGA score of ≥ 2 (mild) at baseline (for skin disease related endpoint analyses).
- subjects with enthesitis at baseline
- subjects with dactylitis at baseline

4.2. Disposition Information

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

Disposition will also include tabulations, by randomized treatment group, of the number of subjects who discontinued study agent administration early and the primary reasons for discontinuation of study agent administration early, and the number of subjects who discontinued study participation early and the primary reason for discontinuation of study participation early. Tabulations by randomized treatment group will also be provided for subjects who met EE criteria at Week 16 and for subjects who met 1 or more TF criteria as defined in Section 2.5.

In addition, subjects who were randomized but never treated and subjects who met any TF criteria will be presented in data listings.

4.3. Treatment Compliance

Treatment compliance will be assessed by the number of study agent administrations completed versus the number of study agent administrations planned.

Tabulations of the number of subjects by the study agent lot(s) and by treatment assigned versus treatment received will also be provided.

4.4. Treatment Exposure and Study Follow-up

The overall treatment exposure (e.g., duration of study treatment, number of treatment administrations, and cumulative dose received) and duration of study follow-up time will be summarized by treatment actually received **based on the safety analysis set,** ie, all treated subjects.

4.5. Protocol Deviations

Subjects with major protocol deviations, defined as having the potential to impact subjects' rights, safety or well-being, or the integrity and/or result of the clinical trial, will be listed or descriptively summarized by randomized treatment group. The major protocol deviations will be grouped into the following 5 categories:

- 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.

Subjects with protocol deviations in study agent administration will be summarized by type of deviations and by treatment group.

4.6. Prior and Concomitant Medications

Prior medications taken for PsA and psoriasis (e.g., DMARDs, oral corticosteroids and NSAIDS) will be summarized by randomized treatment group. Concomitant medications taken for PsA will also be summarized by randomized treatment group.

5. EFFICACY

5.1. 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, treatment comparisons will generally be performed using a Cochran-Mantel-Haenszel (CMH) test stratified by baseline use of non-biologic DMARD (yes, no) and prior exposure to anti-TNFα agents (yes, no). 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.[26] In these analyses, subjects with missing data will be imputed as not achieving the response as described in Section 5.2.3.1.

The Mantel Fleiss criterion will be used to determine the appropriateness of using the CMH test at each visit for each treatment pair under comparison. If the Mantel Fleiss criterion is not satisfied the Fisher's exact test will be used instead of the CMH test to compare the two treatment groups.

For analyses based on Multiple Imputation, 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[27] transformation is used to pool the p-values from each imputed data set.

Major Secondary Continuous Efficacy Endpoints

For the major secondary continuous endpoints (Section 5.4) and related continuous efficacy endpoints in Section 5.5, treatment comparisons will be performed using an analysis of covariance (ANCOVA) model based on MI data. The MI method will be applied to impute the missing value under the assumption of missing at random (MAR). 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 treatment groups for performing ANCOVA within each imputation dataset. The confidence interval is based on the critical values from the t-distribution.

The ANCOVA model will be based on the original scale and will include treatment group, baseline score, baseline use of non-biologic DMARD (yes, no) and prior exposure to anti-TNFα agents (yes, no) as the explanatory factors. The model will include data from all the 3 treatment groups.

Other Continuous Efficacy Endpoints

For all other continuous efficacy endpoints, treatment comparisons will be performed using a MMRM model (Appendix 3) or a cLDA Model (Appendix 3). The model will include all available data from the 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 LSmeans. The 95% CIs for the differences in LSmeans and p-values will be calculated.

In addition, graphical data displays (eg, line plots) and subject listings may also be used to summarize/present the data.

5.2. Analysis Specifications

5.2.1. Level of Significance

5.2.2. The overall type I error will be controlled among the primary and major secondary endpoints at 5% as specified in Section 5.2.2. Multiplicity Adjustment for Testing Procedures

This study has 1 primary endpoint (proportion of subjects who achieved an ACR 20 response at Week 24) and 13 major secondary endpoints. With 14 endpoints and 2 treatment comparisons for each of these endpoints, there are a total of 28 hypotheses to be tested. Refer to Appendix 2 for the list of hypotheses to be tested.

The primary hypothesis in this study is that 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 subjects who achieved an ACR 20 response at Week 24. Due to regional differences in regulatory requirements on multiplicity control from health authorities, 2 multiplicity control procedures are pre-specified. Details for the US-specific multiplicity adjustment are provided in Section **Error! Reference source not found.** and for the Global(ex-US) multiplicity adjustment procedure are provided in Section 5.2.2.2.

With respect to endpoints of enthesitis and dactylitis, the important soft tissue manifestations of PsA, the resolution of enthesitis and resolution of dactylitis endpoints will be tested by combining data from studies CNTO1959PSA3001 and CNTO1959PSA3002 to provide a more robust comparison with greater power than using data from only 1 study of CNTO1959PSA30021 or CNTO1959PSA3002..

5.2.2.1. US-Specific Multiplicity Adjustment for Testing Procedures

The multiple comparison procedure pre-specified to address the requirement of the Food and Drug Administration (FDA) of the United States (US) for family-wise control of the primary and major secondary endpoints is described in this section. Per FDA review comment dated 03OCT2017 [***IND 124177 – Guselkumab (CNTO 1959) – Guidance regarding statistical analysis plans for CNTO1959PSA3001 and CNTO1959PSA3002 submitted on August 10, 2017***], only the primary and major secondary endpoints that are important and access different attributes of disease will be controlled.

The overall Type I error of the treatment comparisons of both doses versus placebo for the primary and the 7 selected major secondary endpoints will be controlled at a significance level of ≤ 0.05. Refer to Appendix 2 for the 16 hypotheses to be tested and controlled with the graphical procedure as specified in Figure 3. For all endpoints specified in the graphical procedure, both adjusted and nominal p-values will be provided. In the instance that an adjusted p-value is not significant, the nominal p-value must only be interpreted as supportive.

The method of controlling the type I error across the testing of the hypotheses is described below and in Figure 3.

- In addition to the primary endpoint, ACR20 at Week 24, 3 major secondary endpoints are identified as important and assess different attributes of the disease:Proportion of subjects who achieved a psoriasis IGA response at Week 24 among the subjects with $\geq$ 3% BSA psoriatic involvement and an IGA score of ≥ 2 at baseline

- Change from baseline in HAQ-DI score at Week 24

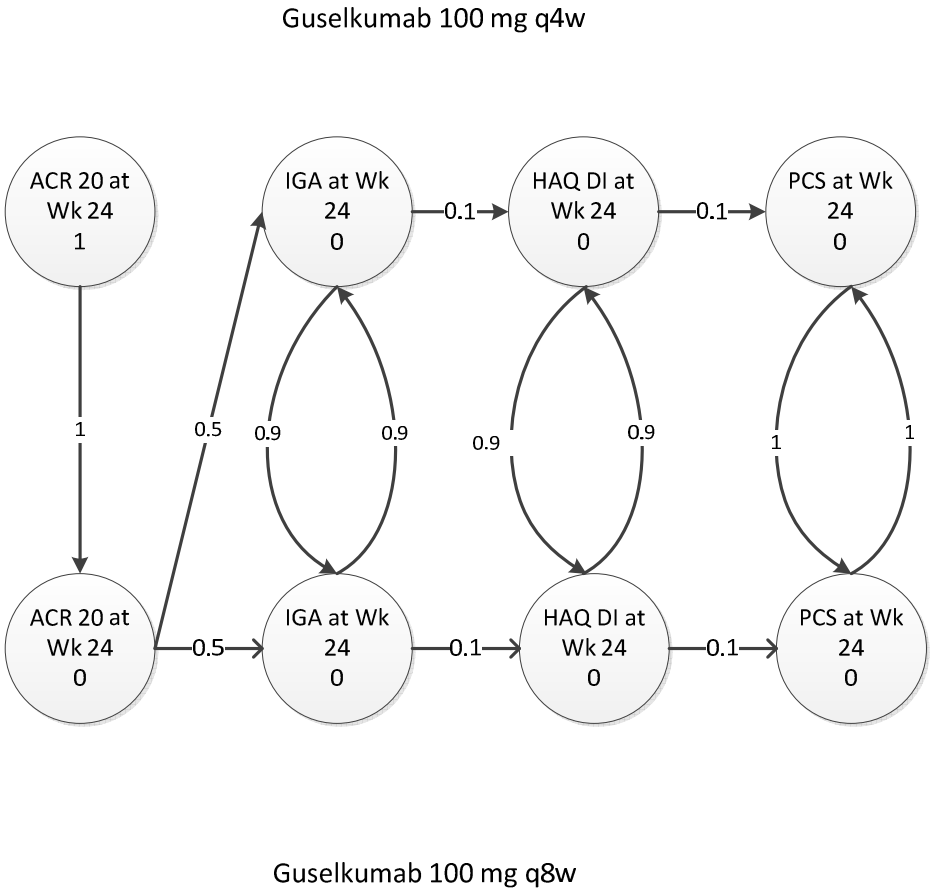
- Change from baseline in SF-36 PCS at Week 24

The overall Type I error of the treatment comparisons of both doses versus placebo for the primary and the 3 selected major secondary endpoints will be controlled at a significance level of ≤ 0.05. Refer to Appendix 2 for the 8 hypotheses to be tested and controlled with the graphical procedure as specified in Figure 3. For all endpoints specified in the graphical procedure, both adjusted and nominal p-values will be provided. In the instance that an adjusted p-value is not significant, the nominal p-value must only be interpreted as supportive.

The method of controlling the type I error across the testing of the hypotheses is described below and in Figure 3.

Nominal p-values will be reported for the major secondary endpoints of the resolution of enthesitis, resolution of dactylitis, the change from baseline in the dactylitis, enthesitis, and SF-36 MCS score at Week 24. Refer to Appendix 2 for the list of these 10 hypotheses. The treatment comparisons are underpowered for these hypotheses.

Figure 2: US-Specific Multiple Comparison Procedure



The major secondary endpoints of ACR 20 at Week 16, ACR 50 at Week 16, ACR 50 at Week 24, ACR 70 at Week 24, and change from baseline in DAS28 at Week 24 are closely related to the primary endpoint (ACR 20 at Week 24) in assessing PsA signs and symptoms. When ACR 20 at Week 24 is significant for a guselkumab group versus the placebo group, these 5 major secondary endpoints will be formally tested for that guselkumab group versus the placebo group at a 2-sided α-level of 0.05. Otherwise, the p-values for the treatment group comparisons will be

considered nominal. A total of 10 hypotheses tests fall in this category of weakly controlled. Refer to Appendix 2 for the list of these 10 hypotheses.

Nominal p-values will be provided for treatment comparisons on the following major secondary endpoints:

- proportion of subjects with resolution of dactylitis at Week 24
- Proportion of subjects with resolution of enthesitis at Week 24
- Change from baseline in the dactylitis at Week 24
- Change from baseline in the enthesitis at Week 24
- Change from baseline in the SF-36 MCS score at Week 24

Refer to Appendix 2 for the list of these 10 hypotheses. The treatment comparisons are underpowered for these hypotheses.

•

5.2.2.2. Global Multiplicity Adjustment for Testing Procedures

There are 2 hypotheses for the primary endpoint with 2 treatment comparisons:

- The Guselkumab 100 mg q4w group is superior to placebo group as measured by ACR20 response at Week 24 (**primary hypothesis**).
- The Guselkumab 100 mg at Weeks 0, 4, and then q8w group is superior to placebo group as measured by ACR20 response at Week 24.

These 2 hypotheses will be tested in a fixed sequence in the order specified above. The overall type 1 error of treatment comparison for the primary endpoint be controlled at a significance level of $\leqslant$ 0.05. If the primary hypothesis achieves statistical significance at a 2-sided α-level of 0.05, the study will be considered positive. If the primary hypothesis does not achieve statistical significance (ie, the Guselkumab 100mg q4w dose group comparison against placebo is not significant), the treatment group comparison for the second hypothesis will not be formally tested and the p-value for the treatment group comparison of the Guselkumab 100mg at Weeks 0 ,4, and then q8w group against placebo will be considered nominal..

For each Guselkumab dose group, the overall Type I error of treatment comparison for primary and the selected major secondary endpoints will be controlled at a significance level of $\leq$ 0.05 as specified below:

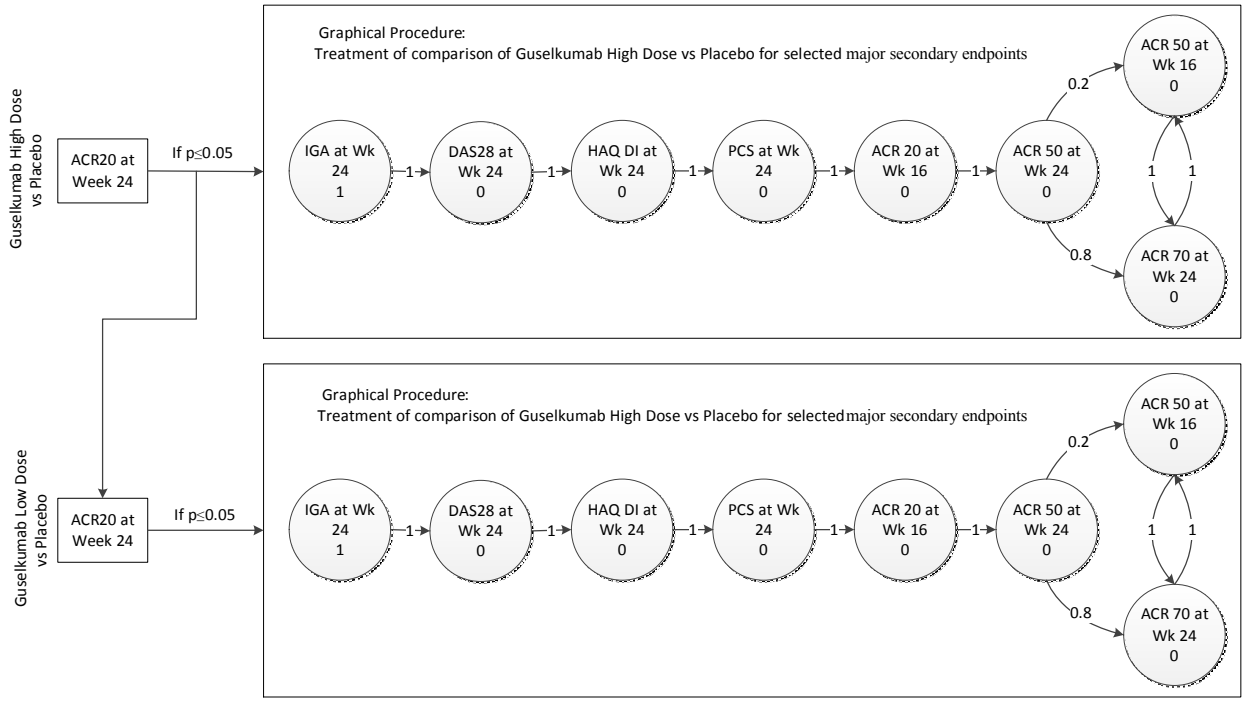
- The hypotheses for the selected major secondary endpoints for a Guselkumab dose group will be formally tested only if the primary endpoint is significant for that Guselkumab dose group. For each Guselkumab dose group, the overall Type I error of treatment comparison for the selected major secondary endpoints will be controlled at a significance level of $\leq$ 0.05 according to the graphical procedure shown in Figure 4. Refer to Appendix 2 for the 18 hypotheses for the major secondary endpoints to be tested

and controlled as specified in Figure 4. For all endpoints specified in the graphical procedure, both adjusted and nominal p-values will be provided. In the instance that an adjusted p-value is not significant, the nominal p-value must only be interpreted as supportive.

- If the primary endpoint is not significant for that Guselkumab dose group, all treatment group comparisons for the major secondary endpoints for that Guselkumab dose group will not be formally tested and the p-values for the treatment group comparisons will be considered nominal.

Nominal p-values will be reported for the major secondary endpoint of the resolution of enthesitis, and dactylitis, the change from baseline in the dactylitis, enthesitis, and SF-36 MCS score. Refer to Appendix 2 for the list of these 10 hypotheses. The treatment comparisons are underpowered for these hypotheses.

Figure 4: Global Multiple Comparison Procedure



5.2.3. Data Handling Rules

5.2.3.1. Missing Data Non-Responder Imputation (NRI)

For response efficacy endpoints, subjects with missing response status will be considered non-responders.

5.2.3.2. Missing Data Exhaustive Scenario Imputation

For selected response efficacy endpoints, the exhaustive scenario tipping point analyses will be performed to evaluate the deviation of the imputation of all missing data as non-responder, by varying the amount of non-responder and responder imputation for missing data.

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'.

5.2.3.3. Multiple Imputation (MI)

Under the assumption of missing at random (MAR), multiple imputation (MI) will be used to impute the missing data for the continuous/ordinal measurements. The missing data will be imputed using the predicted value from an imputation model using the Full Conditional Specification (FCS) regression method for any missing pattern. Each variable will be restricted to only impute within its possible range of values (eg, HAQ Score may only be imputed to within 0-3. The explanatory variables in the imputation model include imputation variables and ancillary variables as specified in Appendix 4.

The number of imputations (N) and the starting seeds are also specified in Appendix 4.

For the composite continuous endpoints (such as DAS28), the above imputation will be performed on components with missing data and then composite score will be derived based on such imputed components.

For categorical endpoints (IGA response, enthesitis resolution, dactylitis resolution), the above imputation will be performed for the respective scores on a continuous scale, then rounded to the nearest integer prior to deriving the response or resolution.

For the composite binary endpoints (such as ACR 20), the above imputation will be performed on each component with missing data and then response status will be determined based on such imputed components.

The treatment comparisons between a guselkumab group versus the placebo group will be performed using the analysis method specified for each of the N imputation datasets. The analysis results from all the N imputation datasets will be combined according to Rubin[20] and the p-value for testing the treatment difference will be obtained.

5.2.3.4. MI for Binary Endpoints for Tipping Point Sensitivity Analysis

For selected binary endpoints, the tipping point analyses based on imputed data by MI will be performed to evaluate the impact of missing data when deviating from 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 subjects on guselkumab have worse outcomes than subjects on placebo.

5.2.3.5. Missing Data as Missing for Continuous Endpoints

For continuous endpoints, when treatment comparisons are performed using a Mixed-Effect Model Repeated Measures (MMRM) or Constrained Longitudinal Data Analysis (cLDA) Model (Appendix 3), all available data from the 3 treatment groups through Week 24 will be included. Missing data will not be imputed. Under the assumption of MAR, the missing data will be accounted for through correlation of repeated measures in the model.

5.2.3.6. Missing Data Handling for Tipping Point Sensitivity Analysis for Continuous Endpoints

For the change from baseline in selected endpoints, tipping point analyses based on imputed data by MI will be performed to evaluate the impact of missing data when deviating from the MAR assumption

- A delta (e.g., Dg =0.2, Dp =0.1) will be added to the imputed value for each subject with missing value from the MI depending on whether the subject 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 subjects on guselkumab have worse outcomes than subjects on placebo.

5.2.4. Estimands

The same population, ie, all subjects with active psoriatic arthritis including those previously treated with biologic anti-TNFα who meet inclusion/exclusion criteria will be used for all estimands defined below. The FAS1 (Section 2.3.1.1) or a subset of FAS1 so that the endpoint measurement is meaningful, will be used to analyze data.

5.2.4.1. Composite Strategy

The Composite Strategy assesses the treatment effects not only based on the variable measurements, but also based on intercurrent events defined in TF criteria. When subjects maintain their background PsA medications at their baseline levels, the values based on the variable measurements will be used. If a subject met any of the TF criteria the subject will be a non-responder for response variables and will have a score of no improvement for continuous variables for signs and symptoms. This estimand acknowledges that meeting the TF criteria is an unfavorable outcome.

Variables:

- Binary: The endpoint (e.g. ACR 20) is defined as responders who had not met any TF criteria prior to the specific visit at which the endpoint was assessed.

- Continuous: The endpoint is defined as change from baseline score prior to meeting TF criteria and 0 (no improvement) after meeting TF criteria. This is based on the placebo response observed in prior PsA studies: the over-time mean improvements in the placebo group is generally greater than 0 by a meaningful amount. Therefore, no change is considered an extremely unfavorable outcome.

Intercurrent Events: The intercurrent event is captured through variable definitions.

Population level summary:

- Binary: difference in proportion of responders between guselkumab group and placebo group.
- Continuous: difference in mean changes between guselkumab group and placebo group.

The Composite Estimand will be analyzed for all efficacy endpoints through Week 24.

5.2.4.2. Alternative Composite Strategy

The Alternative Composite Strategy is similar to the Composite Strategy described in Section 5.2.4.1, however, discontinuation of study agent due to reasons other than lack of efficacy (including adverse events caused by worsening of PsA) are not treatment failure. In other words, the intercurrent events include TF criterion 2, 3, 4, and a subset of criterion 1 from those listed in section 2.5.

This estimand is only used at Week 24, the last visit of the placebo-controlled period.

The Alternative Composite Estimand (Binary) will be analyzed for ACR 20 at Week 24.

The Alternative Composite Estimand (Binary) will be analyzed for ACR 20 at Week 24.

5.2.4.3. Treatment Policy Strategy

The treatment policy strategy is to use all observed data collected for the endpoint. 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. This is the supplementary strategy aiming to achieve a robust treatment effect for regulatory decision making for primary and some major secondary endpoints.

Variable:

- Binary: response endpoint (e.g. ACR 20)
- Continuous: **c**hange from baseline

Intercurrent Events: Regardless of the intercurrent event of meeting TF criteria

Population level summary:

- Binary: difference in proportion of responders between guselkumab group and placebo group.
- Continuous: difference in mean changes between guselkumab group and placebo group.

The Treatment Policy Estimand will be analyzed for all major secondary endpoints (Table 4a), and selected endpoints (Table 7a) analyzed at Week 16.

5.2.4.4. Per-Protocol Strategy

The per-protocol strategy is to use the last assessment while on randomized treatment up to a specific visit (e.g., Week 24). For subjects who met TF criteria 3 or 4 prior to the said visit, the last assessment while on randomized treatment is the last assessment prior to meeting TF criteria 3 or 4. This is the supplementary strategy aiming to achieve a robust treatment effect for regulatory decision making for the primary endpoint of ACR 20 response at Week 24.

Variable: last measurement while on randomized treatment prior to meeting TF criteria 3 and 4 (Section 2.5) or up to a specific visit (e.g., Week 24), whichever is earlier, ***where***,

> while on randomized treatment is defined as the time from first administration of randomized treatment through onset of TF criteria 3 or 4 being met or through 4 weeks after last administration of randomized treatment up to the said specific visit (e.g., Week 24), whichever is earlier.

Intercurrent Events: the intercurrent event is captured through variable definitions.

Population level summary:

- Binary: difference in proportion of responders between guselkumab group and placebo group.

The Per-Protocol Strategy Estimand will be analyzed for ACR20 at Week 24.

5.3. Primary Efficacy Endpoint(s)

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

5.3.1. Definition

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.[3,4] An ACR20 response is defined as:

1. ≥ 20% improvement from baseline in both tender joint count (68 joints) [TJC68] and swollen joint count (66 joints) [SJC66]

AND

2. ≥ 20% improvement from baseline in at least 3 of the following 5 assessments:

    a. Patient's Assessment of Pain (VAS) [PAIN]

    b. Patient's Global Assessment of Disease Activity (arthritis, VAS) [GDPT]

    c. Physician's Global Assessment of Disease Activity (VAS) [GDEV]

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

    e. C-reactive protein (CRP)

Following are the definitions of each of the forgoing disease assessment criteria (components) that are used in the determination of ACR20 response:

1. Tender Joint Count 68 (TJC68): a total number of tender joints among the 68 joints evaluated for tenderness. Each of the 68 joints will be evaluated for tenderness, categorized as tender or not tender. Joint evaluability rules specified in Appendix 1 for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 68 joint set, joint count adjustment rules described in Appendix 1 will be applied in determining the ultimate count of tender joints.
2. Swollen Joint Count 66 (SJC66): a total number of swollen joints among the 66 joints evaluated for swelling. (Note: The 2 hip joints are excluded from swelling assessment.) Each of the 66 joints will be evaluated for swelling, categorized as swollen or not swollen. Joint evaluability rules specified in Appendix 1 for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 66 joint set, joint count adjustment rules described in Appendix 1 will be applied in determining the ultimate count of swollen joints.
3. Patient's Assessment of Pain (PAIN): a measure from 0 (no pain) to 10 (the worst possible pain) on a 10-unit VAS.
4. Patient's Global Assessment of Disease Activity (arthritis, GDPT): a measure from 0 (very well) to 10 (very poor) on a 10-unit VAS.
5. Physician's Global Assessment of Disease Activity (GDEV): a measure from 0 (no arthritis activity) to 10 (extremely active arthritis) on a 10-unit VAS.
6. HAQ-DI: a measure of difficulty a subject may have in accomplishing tasks in 8 functional areas. For additional details, please refer to the definition of HAQ-DI in Section 5.4.1.1.
7. C-reactive protein (CRP): a lab parameter measured in mg/dL. LLOQ rule specified in Appendix 1 will be applied to values < LLOQ.

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

5.3.2. Analysis Methods

5.3.2.1. Primary Analysis

The primary endpoint will be analyzed at Week 24 DBL based on the Composite Estimand (Section 5.2.4.1). In the primary efficacy analysis, data from all subjects in FAS1 (Section 2.3.1) will be analyzed according to randomized treatment group regardless of the treatment actually received.

The primary endpoint to be analyzed is proportion of subjects who achieved an ACR 20 response at Week 24 and who did not meet any TF criteria (Section 2.5) prior to Week 24. Subjects who met any TF criteria prior to Week 24 will be considered non-responders at Week 24 regardless of the observed ACR 20 response status.

The treatment difference between each guselkumab group versus the placebo group will be tested using a CMH test stratified by baseline use of non-biologic DMARD (yes, no) and prior exposure to anti-TNFα agents (yes, no). 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.

Data Handling Rules

- **Missing Data NRI** rules (Section 5.2.3.1) will be applied, i.e., subjects with missing data for ACR 20 response at Week 24 will be considered ACR 20 non-responders at Week 24.

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

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.

5.3.2.2. Sensitivity and Supplementary Analyses

1. To evaluate the robustness of the Composite Estimand regarding the assumption of all missing data as non-responder, **a sensitivity analysis** with the exhaustive 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 5.2.3.2). A 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 non-responders 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. To support regulatory decision making, the Treatment Policy Estimand (Section 5.2.4.3) will also be evaluated as **a supplementary analysis**. In this analysis, however, the observed ACR 20 response for all subjects will be used regardless of whether or not TF criteria are met prior to Week 24, and the missing ACR 20 response for all subjects will be imputed by MI method (Section 5.2.3.3) under the assumption of MAR. Treatment comparisons for each imputation data set will be based on a CMH test stratified by baseline use of non-biologic DMARD (yes, no) and prior exposure to anti-TNFα agents (yes, no). And then the analysis results from the N imputation datasets will be combined, according to Rubin[20] and the p-value for testing the treatment difference will be obtained.
3. Two-dimensional tipping point analyses based on MI imputed data will be performed as **a supplemental analysis** to assess the robustness for Treatment Policy Estimand regarding the assumption of MAR. 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 5.2.3.4). The same analysis method as in sensitivity and supplementary analysis 2 will be applied for the pairs of deltas. The analysis will be done for pairs of delta values include the scenarios where subjects on guselkumab have worse outcomes than subjects on placebo.
4. In addition, the Alternative Composite Estimand will also be evaluated as a supplemental analysis (section 5.2.4.2). Subjects with missing data will be considered non-responders. The same analysis method as that used for the primary analysis will be applied.
5. **A supplemental analysis** will be performed similar to the primary analysis (Section 5.3.2.1), however, based on the Per-Protocol Strategy Estimand (Section 5.2.4.4). In this analysis, ACR response status will be determined based on the last assessment up to Week 24 while subjects on randomized treatment or the last non-missing assessment prior to meeting TF criteria 3 or 4, whichever is earlier. ***Note*** that this analysis will use the Per-Protocol Analysis Set (PPAS).

5.3.2.3. Subgroup Analyses

Subgroup analyses will be performed using a logistic regression model to evaluate treatment consistency in proportion of subjects 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 2.4. Odds ratios and the corresponding 95% CIs will also be provided for each of subgroups. In addition, the p-values for interaction of the treatment groups and the subgroups will also be provided when a subgroup has at least 2 categories.

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

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

Table 3 below provides an overview on all the analyses related to the primary endpoint of ACR 20 response at Week 24, the estimands, the analysis sets, the data handling rules to be used,

and the analysis methods and summary statistics. ***Note*** that, for subgroup analyses, the analysis sets are the individual subgroups (Section 2.4) of FAS1.

Table 3: Summary of Analyses Related to the Primary Endpoint of ACR 20 Response at Week 24

Analysis (Analysis Set)	Missing data	Analysis method/Summary statistics
Analyses based on **Composite Estimand**, in which, subjects meeting any TF criteria (defined in Section 2.5) prior to Week 24 will be considered as ACR 20 non-responders at Week 24.		
Primary Analysis (FAS1)	Subjects with missing data are considered to be non-responders	• Response rates • Treatment difference in response rates and 95% CI • P-value from the CMH test (stratified by randomization stratification factors) for treatment comparison
Sensitivity Analysis (FAS1)	Subjects with missing data are considered to be non-responders	• Exhaustive tipping point analysis • The chi-squared test to compare treatment groups. • Analysis results to be presented graphically
Subgroup Analyses (Individual subgroup levels defined in Section 2.4)	Subjects with missing data are considered to be non-responders	• Response rates • Odds ratio and 95% CI for treatment comparison • P-value from logistic regression for the interaction of treatment group and subgroup variable • Analysis results to be presented in forest plots
Analyses based on **Treatment Policy Estimand**, in which, all observed data will be used, regardless of whether or not the subjects meet any TF criteria.		
Supplementary Analysis 1 (FAS1)	Multiple imputation with FCS regression of component scores	• Response rates • Treatment difference in response rates and 95% CI • P-value from the CMH[a] test (stratified by randomization stratification factors) for treatment comparisons.
Supplementary Analysis 2 (FAS1)	Multiple imputation with FCS regression of component scores	• Tipping point analysis. • The CMH[a] test (stratified by randomization stratification factors) to compare treatment groups. • Analysis results to be presented graphically
Analyses based on **Alternative Composite Estimand**, in which, subjects meeting any TF criteria (defined in Section 2.5) except discontinuation of study agent due to reasons other lack of efficacy, prior to Week 24 will be considered as ACR 20 non-responders at Week 24..		
Supplementary Analyses 3 (FAS1)	Subjects with missing data are considered to be non-responders	• Response rates • Treatment difference in response rates and 95% CI • P-value from the CMH test (stratified by randomization stratification factors) for treatment comparison
Analyses based on **Per-Protocol Strategy Estimand**, in which, ACR 20 response is determined based on the last measurement while subjects on randomized treatment up to Week 24 or prior to meeting TF criteria 3 or 4, whichever is earlier.		
Supplementary Analysis 4 (PPAS)	Not Applied	• Response rates • Treatment difference in response rates and 95% CI • P-value from the CMH test (stratified by randomization stratification factors) for treatment comparison
[a] When combining analysis results for CMH test, Wilson-Hilferty transformation will be applied to the test statistics to achieve an approximate normal distribution.		

5.4. Major Secondary Endpoints

The major secondary endpoints in this study are:

1. Change from baseline in HAQ-DI score at Week 24.

2. Proportion of subjects who achieve an ACR 50 response at Week 24.
3. Proportion of subjects with a psoriasis response of an Investigator Global Assessment (IGA) (ie, an IGA psoriasis score of 0 [cleared] or 1 [minimal]) AND ≥2-grade reduction from baseline) at Week 24 among the subjects with ≥3% body surface area (BSA) psoriatic involvement and an IGA score of ≥2 (mild) at baseline.
4. Proportion of subjects who achieve an ACR 20 response at Week 16.
5. Proportion of subjects with resolution of enthesitis at Week 24 among the subjects with enthesitis at baseline.
6. Proportion of subjects with resolution of dactylitis at Week 24 among the subjects with dactylitis at baseline.
7. Change from baseline in enthesitis score (based on Leeds Enthesitis Index [LEI]) at Week 24 among the subjects with enthesitis at baseline.
8. Change from baseline in dactylitis scores at Week 24 among the subjects with dactylitis at baseline.
9. Change from baseline in SF-36 Physical Component Score (PCS) at Week 24.
10. Change from baseline in DAS28 (CRP) at Week 24.
11. Change from baseline in SF-36 Mental Component Score (MCS) at Week 24.
12. Proportion of subjects who achieve an ACR 50 response at Week 16.
13. Proportion of subjects who achieve an ACR 70 response at Week 24.

This section outlines the definition and analyses of these major secondary endpoints. All the secondary endpoints will be analyzed at Week 24 DBL according to the randomized treatment groups. Data from all subjects in FAS1 (Section 2.3.1) will be included with the following exceptions:

the analysis of the psoriasis response of IGA will be based on FAS1 among the subjects with a ≥ 3% BSA psoriatic involvement and an IGA score of ≥ 2 (mild) at baseline

the analysis of change from baseline in enthesitis score change and resolution of enthesitis will be based on FAS1 among the subjects with enthesitis at baseline

the analysis of change from baseline in dactylitis score and resolution of dactylitis will be based on FAS1 among the subjects with dactylitis at baseline

5.4.1. Change from Baseline in HAQ-DI Score at Week 24

5.4.1.1. Definition

HAQ disability index (HAQ-DI) score is an evaluation of the functional status for a subject.[6] The 20-question instrument assesses the degree of difficulty a person has in accomplishing tasks in 8 functional areas (dressing, arising, eating, walking, hygiene, reaching, gripping, and activities of daily living). Responses in each functional area are scored from 0, indicating no difficulty, to 3,

indicating inability to perform a task in that area (i.e., lower scores are indicative of better functioning).

The HAQ-DI score is the sum of computed category scores divided by the number of categories answered. The HAQ-DI score will not be computed if the subject does not have scores for at least 6 of the 8 categories.

The scoring algorithm (including adjusting the use of aids or devices) is provided by http://patienteducation.stanford.edu/research/haq20.html and is detailed in a separate document.

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

5.4.1.2. Analysis Methods

The change from baseline in HAQ–DI score at Week 24 will be analyzed at Week 24 DBL based on the Composite Estimand (Section 5.2.4.1). In this analysis, data from all subjects in FAS1 (Section 2.3.1) will be analyzed according to randomized treatment group regardless of the treatment received.

Analysis of the change from baseline in HAQ–DI score at Week 24 will be performed using an ANCOVA model based on MI data. The MI method will be applied to impute, under the assumption of MAR, the missing change score of HAQ-DI from baseline at Week 24 defined in the Composite Estimand (i.e., change from baseline with no change assumed at and after meeting TF criteria). 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.

To support regulatory decision making, a supplementary analysis will be provided in FAS1 based on the Treatment-Policy-Estimand. Analysis of the change from baseline in HAQ–DI score at Week 24 will be performed using an ANCOVA model based on MI data similar to the main analysis.

5.4.2. Proportion of Subjects with ACR 50 Response at Week and 24, Proportion of Subjects with ACR 20 Response at Week 16, Proportion of Subjects with ACR 50 Response at Week 16, and Proportion of Subjects with ACR 70 Response at Week 24

5.4.2.1. Definition

ACR 50 and ACR 70 responses are defined similarly to ACR 20 response (Section 5.3.1), except that the improvement threshold of 20% from baseline in ACR 20 response is replaced by 50% and 70%, respectively.

5.4.2.2. Analysis Methods

Data from all subjects in FAS1 (Section 2.3.1) will be analyzed according to randomized treatment group regardless of the treatment actually received.

The same analysis method as described in Section 5.3.2.1 will be applied to the treatment comparisons on proportion of subjects with ACR 50 response at Week 24, proportion of subjects with ACR 20 response at Week 16, proportion of subjects with ACR 50 response at Week 16, and proportion of subjects with ACR 70 response at Week 24. The analyses are based on the Composite Estimand (Section 5.2.4.1).

5.4.3. Proportion of Subjects Who Achieve a Psoriasis IGA Response at Week 24 Among the Subjects with ≥3% BSA Psoriatic Involvement and an IGA Score of ≥2 (mild) at Baseline

5.4.3.1. Definition

The Investigator's Global Assessment (IGA) documents the investigator's assessment of the subject's psoriasis at a given time point. Overall lesions are graded for induration, erythema, and scaling. The subject's psoriasis is assessed as cleared (0), minimal (1), mild (2), moderate (3), or severe (4).

A psoriasis IGA response is defined as an IGA psoriasis score of 0 (cleared) or 1 (minimal) AND ≥ 2-grade reduction from baseline in the IGA psoriasis score.

A psoriasis IGA response is only applicable to subjects who had an IGA psoriasis score of ≥ 2 (mild) at baseline.

5.4.3.2. Analysis Methods

Data from all subjects in FAS1 who had ≥ 3% BSA psoriatic involvement and an IGA score of ≥ 2 (mild) at baseline will be included and analyzed according to the randomized treatment groups.

The same analysis method as described in Section 5.3.2.1 will be applied to the treatment comparisons on proportion of subjects with a psoriasis IGA response at Week 24. The analyses are based on the Composite Estimand (Section 5.2.4.1).

5.4.4. Change from Baseline in Enthesitis Score at Week 24 and Proportion of subjects with Resolution of Enthesitis at Week 24 in Subjects with Enthesitis at Baseline

5.4.4.1. Definition

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).[10] The LEI was developed to assess enthesitis in subjects with PsA and evaluates the presence (score of 1) or absence (score of 0) of pain by applying local pressure to the following entheses:

- Lateral epicondyle humerus, left and right

- Medial femoral condyle, left and right
- Achilles tendon insertion, left and right

The enthesitis index score is a total score of the 6 evaluated sites as listed above, with range from 0 to 6. For subjects who have an incomplete set of 6 evaluated sites, the enthesitis index score based on the observed data cannot be calculated.

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

Subject with enthesitis at baseline are those subjects with at least at one tender enthesis among the 6 sites include in the LEI.

Subjects with resolution of enthesitis are those subjects who had at least one tender enthesis at baseline and none at the analysis visit among the 6 sites included in the LEI.

5.4.4.2. Analysis Methods

For this endpoint, using data from the CNTO1959PSA3001 study, treatment groups will be compared in the CNTO1959PSA3002 study and will be reported with the analysis results from that study.

5.4.5. Change from Baseline in Dactylitis Scores at Week 24 and Proportion of subjects with Resolution of Dactylitis at Week 24 in Subjects with Dactylitis at Baseline

5.4.5.1. Definition

Presence and severity of dactylitis will be assessed in both hands and feet using a scoring system from 0 to 3 (0 – no dactylitis, 1 – mild dactylitis, 2 – moderate dactylitis, and 3 – severe dactylitis).[8,9]

The results for each digit are summed to produce a final dactylitis score with a range from 0 to 60.

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

Subject with dactylitis at baseline are those subjects with a dactylitis score > 0 at baseline.

Subjects with resolution of dactylitis are those subjects who had a dactylitis score > 0 at baseline and a score of 0 at the analysis visit.

5.4.5.2. Analysis Methods

For this endpoint, using data from the CNTO1959PSA3001 study, treatment groups will be compared in the CNTO1959PSA3002 study and will be reported with the analysis results from that study.

5.4.6. Change from Baseline in SF-36 PCS Score at Week 24 and Change from Baseline in SF-36 MCS score at Week 24

5.4.6.1. Definition

The questionnaire of the 36-item short form health survey (SF-36) is a health-related quality of life instrument with 36 questions, developed as part of the Rand Health Insurance Experiment. Version 2 will be used. This instrument consists of 8 multi-item scales (domains).[24]

- Limitations in physical functioning due to health problems
- Limitations in usual role activities due to physical health problems
- Bodily pain
- General mental health (psychological distress and well-being)
- Limitations in usual role activities due to personal or emotional problems
- Limitations in social functioning due to physical or mental health problems
- Vitality (energy and fatigue)
- General health perception

Each of these 8 scales (domains) is scored from 0 to 100 with higher scores indicating better health. Based on the scale scores, the 2 summary scores, physical component summary (PCS) score and mental component summary (MCS) score, will be derived. These summary scores are also scaled with higher scores indicating better health.

Note that, the SF-36 scale scores will be derived based on the algorithm and software provided by the developer.[21] The norm-based score will be used for data analysis and derived by using the 2009 general US population with missing value estimated based on Advanced Data Recovery with Missing Score Estimator (MSE) provided in the user manual.[21]

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

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

5.4.6.2. Analysis Methods

Data from all subjects in FAS1 will be included and analyzed according to the randomized treatment groups.

The same analysis methods as described in Section 5.4.1.2 (ANCOVA on MI data) will be applied to treatment comparisons on the change from baseline in SF-36 PCS at Week 24 and change from baseline in SF-36 MCS at Week 24, based on the Composite Estimand (Section 5.2.4.1).

5.4.7. Change from Baseline in DAS28 (CRP) at Week 24

5.4.7.1. Definition

The Disease Activity Index Score 28 using CRP [DAS28 (CRP)][22] is a derived score combining tender joints (28 joints), swollen joints (28 joints), CRP, and Patient's Global Assessment of Disease Activity. The 28 joints evaluated for swelling and tenderness are shoulder, elbow, wrist, MCP1, MCP2, MCP3, MCP4, MCP5, PIP1, PIP2, PIP3, PIP4, PIP5 joints of the upper right and upper left extremities as well as the knee joints of the lower right and lower left extremities.

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

DAS28 (CRP) = 0.56* SQRT(TJC28) + 0.28*SQRT(SJC28) + 0.36* Ln ($CRP_{mg/L}$ +1) + 0.014*$GDPT_{mm}$ + 0.96, where

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

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

3. $CRP_{mg/L}$: CRP in mg/L. In the calculation of DAS28 (CRP) value, the natural logarithm of ($CRP_{mg/L}$ + 1) is used. LLOQ rule specified in Appendix 1 will be applied to values < LLOQ.

4. $GDPT_{mm}$: Patient's Global Assessment of Disease Activity scaled from 0 (very well) to 100 (very poor) on a 100-unit VAS for the calculation of DAS28 (CRP) value.

If any of the components required for computing the DAS28 (CRP) value is missing, the DAS 28 (CRP) score will be set to missing.

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

5.4.7.2. Analysis Methods

Data from all subjects in FAS1 will be included and analyzed according to the randomized treatment groups.

The same analysis methods as described in Section 5.4.1.2 (ANCOVA on MI data) will be applied for treatment comparisons on change from baseline in DAS28 (CRP) at Week 24, based on the Composite Estimand (Section 5.2.4.1).

5.4.8. Summary of Analyses Related to Major Secondary Endpoints

Table 4 below provides an overview on all the analyses related to the major secondary endpoints, the estimands, the analysis sets, the data handling rules to be used, and the analysis methods and summary statistics.

Table 4: Summary of Analyses Related to Major Secondary Endpoints

Endpoints (Analysis Set)	Missing data	Analysis method/Summary statistics
Analyses of **Response Endpoints based on Composite Estimand**, in which, subjects meeting any TF criteria (defined in Section 2.5) prior to a visit (e.g., Week 16) will be considered as non-responders at the said visit (e.g., Week 16).		
• ACR 50 Response at Week 24 (FAS1) • ACR 70 Response at Week 24 (FAS1) • ACR 20 Response at Week 16 (FAS1) • ACR 50 at Week 16 (FAS1) • IGA Response at Week 24 (FAS1 with ≥3% BSA psoriatic involvement and an IGA score of ≥2 at baseline)	Subjects with missing data are considered to be non-responders	• Response rates • Treatment difference in response rates and 95% CI • P-value from the CMH test (stratified by randomization stratification factors) for treatment comparison
Analyses of **Continuous Endpoints based on Composite Estimand**, in which, subjects meeting any TF criteria (defined in Section 2.5) prior to a visit (e.g., Week 24) will be considered as no change (no improvement) from baseline at the said visit (e.g., Week 24).		
Change from baseline at Week 24 in: • HAQ-DI score (FAS1) • DAS28(CRP) (FAS1) • SF-36 PCS score (FAS1) • SF-36 MCS score (FAS1)	MI (Section 5.2.3.3)	**Main Analysis:** • Descriptive summary statistics • ANCOVA for each MI dataset and then analysis results across all MI datasets combined • LS mean (95% CI) for each treatment group and for differences between groups. • P-values for treatment comparisons

5.4.9. Additional Tipping Point Analyses

To support regulatory decision making per the FDA review comment dated 14MAR2019 [***IND 124177 - Guselkumab (CNTO 1959) - Comments for IND 124177***], additional tipping point analyses will be performed for each major secondary endpoint to evaluate the impact when the missing data deviate from the MAR assumption. Analyses will be performed for the Treatment Policy Estimand (Section **Error! Reference source not found.**), where observed data is included regardless of whether or not TF criteria are met prior to the time at which the endpoint is assessed.

Missing data in **binary endpoints** will first be imputed using multiple imputation, then a pair of deltas will be added to the predicted response or resolution rates of each missing subject from the MI method depending on guselkumab or placebo group to new MI datasets (Section 5.2.3.4). The same CMH test as that used for the main analysis of the endpoint will be used for each imputation set, and the analysis results from the N imputation datasets will be combined for each pair of deltas, according to Rubin[20], and the p-values for testing the treatment difference will be obtained. When combining analysis results for the CMH test, the Wilson-Hilferty transformation[27] will be applied to the test statistics to achieve an approximate normal distribution.

Missing data for **continuous endpoints** will first be imputed using multiple imputation, then a pair of deltas will be added to the imputed values from the MI method depending on guselkumab or placebo group (Section 5.2.3.6). The same ANCOVA model as that used for the main analysis of the endpoint will be used for each imputation set, and the analysis results from the N imputation datasets will be combined for each pair of deltas, according to Rubin[20], and the p-values for testing the treatment difference will be obtained.

Table 4a below summarizes the tipping point analyses for major secondary endpoints. For the summary of additional tipping point analyses other than for primary and major secondary endpoints see Section 5.5.5.

Table 4a: Summary of Tipping Point Analyses for Major Secondary Endpoints

Endpoints (Analysis Set)	Missing data	Analysis method
Supplementary Analyses of **Continuous Endpoints based on Treatment Policy Estimand**, in which all observed data will be used, regardless of whether or not the subjects meet any TF criteria.		
Change from baseline at Week 24 in: • HAQ-DI score (FAS1) • SF-36 PCS score (FAS1) • SF-36 MCS score (FAS1) • DAS28(CRP) (FAS1) • Enthesitis (LEI) among subjects with enthesitis (LEI) at baseline (FAS1) • Dactylitis among subjects with dactylitis at baseline (FAS1)	MI with FCS regression (Section 5.2.3.6, **Error! Reference source not found.**)	• Two-dimensional tipping point analysis • Treatment comparison using ANCOVA model • Analysis results to be presented graphically

Table 4a: Summary of Tipping Point Analyses for Major Secondary Endpoints		
Endpoints (Analysis Set)	**Missing data**	**Analysis method**
Supplementary Analyses of **Binary Endpoints based on Treatment Policy Estimand** in which, all observed data will be used, regardless of whether or not the subjects meet any TF criteria.		
• ACR 50 Response at Week 24 (FAS1) • ACR 70 Response at Week 24 (FAS1) • ACR 20 Response at Week 16 (FAS1) • ACR 50 at Week 16 (FAS1) • IGA Response at Week 24 (FAS1 with ⩾3% BSA psoriatic involvement and an IGA score of ⩾2 at baseline) • Resolution of enthesitis (LEI) at Week 24 among subjects with enthesitis (LEI) at baseline • Resolution of dactylitis at Week 24 among subjects with dactylitis at baseline	MI with FCS regression (Section **Error! Reference source not found.**, **Error! Reference source not found.**)	• Two-dimensional tipping point analysis. • The CMH[a] test (stratified by randomization stratification factors) to compare treatment groups. • Analysis results to be presented graphically
[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.		

5.5. Other Efficacy Variable(s)

In addition to the primary and major secondary endpoints, other efficacy analyses related to reduction of signs and symptoms and physical function, skin disease and health related quality of life will be analyzed.

This section outlines the definition and analyses of the other efficacy endpoints.

5.5.1. Efficacy Endpoints Related to Reduction of Signs and Symptoms and Physical Function

5.5.1.1. Definition

5.5.1.1.1. ACR Related Endpoints

Other efficacy endpoints related to ACR include:

- Proportion of subjects who achieve an ACR 20 response by visit through Week 52
- Proportion of subjects who achieve an ACR 50 response by visit through Week 52
- Proportion of subjects who achieve an ACR 70 response by visit through Week 52
- ACR components by visit through Week 52
- Percent change (improvement) from baseline in ACR components by visit through Week 52
- Proportion of subjects who maintain an ACR 20 response at Week 52 among the subjects who achieved an ACR 20 response at Week 24.
- Proportion of subjects who maintain an ACR 50 response at Week 52 among the subjects who achieved an ACR 50 response at Week 24.

- Proportion of subjects who maintain an ACR 70 response at Week 52 among the subjects who achieved an ACR 70 response at Week 24.

For definition of **ACR 20 response**, refer to Section 5.3.1. For definition of **ACR 50 response** and **ACR 70 response**, refer to Section 5.4.2.1.

5.5.1.1.2. HAQ-DI Related Endpoints

The other efficacy endpoints related to HAQ-DI include:

- Change from baseline in HAQ-DI score by visit overtime through Week 52.
- Proportion of subjects who achieve a HAQ-DI response (i.e., a ≥ 0.35 improvement from baseline in HAQ-DI score) by visit overtime through Week 52 among those subjects with HAQ-DI score ≥ 0.35 at baseline.
- Proportion of subjects who maintain a HAQ-DI response at Week 52 among the subjects who achieved a HAQ-DI response at Week 24.

For definitions of **HAQ-DI** and **change from baseline in HAQ-DI score**, refer to Section 5.4.1.1.

HAQ-DI responders are subjects who achieved ≥ 0.35 improvement in HAQ-DI score. ***Note*** that a ≥ 0.35 improvement in HAQ-DI score is considered a clinically meaningful improvement in PsA.[14] ***The responders are only applicable to subjects whose baseline HAQ-DI score ≥ 0.35.***

5.5.1.1.3. DAS28 Related Endpoints

The other efficacy endpoints related to DAS28 (CRP) include:

- Proportion of subjects with DAS28 (CRP) response by visit overtime through Week 52
- Proportion of subjects with DAS28 (CRP) remission by visit overtime through Week 52
- Change from baseline in DAS28 (CRP) by visit overtime through Week 52

For definitions of **DAS28 (CRP),** refer to Section 5.4.7.1.

DAS28 Response for DAS28 (CRP) is as defined in Table below.[23]

DAS28 (CRP) Response Criteria

DAS28 (CRP) at the visit	**Improvement from Baseline**		
	> 1.2	**> 0.6 – 1.2**	**≤ 0.6**
≤ 3.2	Good response	Moderate response	No response
> 3.2 – 5.1	Moderate response	Moderate response	No response
> 5.1	Moderate response	No response	No response

DAS28 (CRP) remission is defined as a DAS28 (CRP) value of < 2.6 at a visit.

5.5.1.1.4. Responders Based on Modified Psoriatic Arthritis Responder Criteria (PsARC)

A subject will be considered a responder based on Modified Psoriatic Arthritis Responder Criteria (PsARC) if the subject meets at least 2 of the following criteria at a visit, including at least 1 of the 2 joint criteria and with no deterioration in the other criteria.[18] No deterioration means ≥ 0% improvement.

- ≥ 30% decrease in swollen joint count (SJC66)
- ≥ 30% decrease in tender joint count (TJC68)
- ≥ 20% improvement in Patient's Global Assessment of Disease Activity (arthritis) on a VAS
- ≥ 20% improvement in Physician's Global Assessment of Disease Activity on a VAS

5.5.1.1.5. Enthesitis Related Endpoints

Refer to Section 5.4.4.1 for definitions of **enthesitis score**, **change from baseline in enthesitis score**, and **subjects with resolution of enthesitis** based on LEI.

Enthesitis will also be assessed by the Spondyloarthritis Research Consortium of Canada (SPARCC).[15] The SPARCC developed a measure for enthesitis in general spondyloarthritis which evaluates the presence or absence of pain by applying local pressure to the following entheses:

- Supraspinatus, insertion, left and right
- Medial epicondyle humerus, left and right
- Lateral epicondyle, humerus left and right
- Greater trochanter, left and right
- Quadriceps –to-Patella, left and right
- Patellar-tibia, left and right
- Achilles tendon insertion, left and right
- Plantar fascia, left and right

Tenderness on examination is recorded as either present (1) or absent (0) for each of the 16 sites for an overall score range of 0–16.

- **Change from baseline in enthesitis score (**based on SPARCC**)** measures the change in enthesitis, where a negative change indicates an improvement and a positive change indicates a worsening.

5.5.1.1.6. Dactylitis Related Endpoints

Refer to Section 5.4.5.1 for definitions of **dactylitis score**, **change from baseline in dactylitis score**, and **subjects with resolution of dactylitis**.

- In addition, the presence of tender dactylitis will be assessed by the number of fingers and toes with tender dactylitis, with a range of 0 to 20. The change from baseline in this tender dactylitis score will be assessed over time through Week 52.

5.5.1.1.7. PASDAS Related Endpoints

The Psoriatic ArthritiS Disease Activity Score (PASDAS) is a derived score combining Patient's Global Assessment of Disease Activity (arthritis and psoriasis, on a 100-unit VAS), Physician's Global Assessment of Disease Activity (on a 100-unit VAS), swollen joint count (66 joints), tender joint count (68 joints), CRP (mg/L), enthesitis based on LEI (scaled to a 0–6 range), tender dactylitis count (scoring each digit from 0–3 and recoding to 0–1, where any score > 0 equaled 1), and the PCS score of the SF-36 health survey.[11,13]

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

$PASDAS = 1.5*\{[0.18* SQRT(GDEV_{mm}) + 0.159* SQRT(GDPTS_{mm}) - 0.253 \times SQRT(SF\_PCS) + 0.101 * Ln (SJC66 +1) + 0.048 * Ln (TJC68 + 1) + 0.23 * Ln (ENTHE + 1) + 0.377 * Ln (DACTY + 1) + 0.102* Ln (CRP_{mg/L} +1)] + 2\}$, where

1. $GDEV_{mm}$: Physician's Global Assessment of Disease Activity on a 100-unit VAS.

2. $GDPTS_{mm}$: Patient's Global Assessment of Disease Activity (arthritis and psoriasis) on a 100-unit VAS.

3. SF-PCS: PCS score of the SF-36 health survey.

4. SJC66: a total number of swollen joints among the 66 joints evaluated for swelling. Each of the 66 joints will be evaluated for swelling, categorized as swollen or not swollen. Joint evaluability rules specified in Appendix 1 for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 66 joint set, joint count adjustment rules described in Appendix 1 will be applied in determining the ultimate count of swollen joints.

5. TJC68: a total number of tender joints among the 68 joints evaluated for tenderness. Each of the 68 joints will be evaluated for tenderness, categorized as tender or not tender. Joint evaluability rules specified in Appendix 1 for overwriting joint evaluation will be applied to those joints with joint injection(s)/surgical joint procedure(s). For subjects with any joint not evaluable in the 68 joint set, joint count adjustment rules described in Appendix 1 will be applied in determining the ultimate count of tender joints.

6. ENTHE: enthesitis score based on LEI with a range from 0 to 6.

7. DACTY: dactylitis count (scoring each digit from 0–3 and recoding to 0–1, where any score > 0 equaled 1).

8. $CRP_{mg/L}$: CRP in mg/L. LLOQ rule specified in Appendix 1 will be applied to values < LLOQ.

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

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

Subjects with low or very low disease activity based on the PASDAS score are those subjects who have a PASDAS score less than or equal to 1.9, or greater than 1.9 and less than or equal to 3.2, respectively.

5.5.1.1.8. GRACE Index

GRAppa Composite scorE (GRACE) Index score is a composite score of Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA), derived as GRACE Index = (1 - AMDF) x 10, where, AMDF is the Arithmetic Mean of the Desirability Function (AMDF).[13,12]

AMDF is calculated by transforming the following variables using predefined algorithms and expressing the total score as a mean with a score range of 0 – 1, where 1 indicates a better state than 0:[13,12]

- Tender joint count (68 joints)
- Swollen joint count (66 joints)
- HAQ-DI score
- Patient's Global Assessment of Disease Activity (arthritis and psoriasis) on a 100-unit VAS
- Patient's Assessment of Skin Disease Activity on a 100-unit VAS
- Patient's Global Assessment of Disease Activity (arthritis) on a 100-unit VAS
- Psoriasis Area and Severity Index PASI score (Section 5.5.2.1.1)
- Psoriatic Arthritis Quality of Life Index (PsAQOL) score, which is derived as follows:

  PsAQOL = 25.355 + (2.367 × HAQ-DI) – (0.234 × SF-PCS) – (0.244 × SF-MCS) *where*,

HAQ-DI: HAQ-DI score (0-3)

SF-PCS: PCS score of SF-36 Health Survey

SF-MCS: MCS score of SF-36 Health Survey

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

Subjects with low disease activity based on the GRACE index score are those subjects who have a GRACE index score less than or equal to 2.3.

5.5.1.1.9. DAPSA Related Endpoints

The Disease Activity Index for Psoriatic Arthritis (DAPSA) score is a derived score combining swollen joint count (66 joints), tender joint count (68 joints), CRP (mg/dL), Patient's Assessment

of Pain (on a 10-unit VAS), and Patient's Global Assessment of Disease Activity (arthritis, on a 10-unit VAS).[13]

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

DAPSA = STC66 + TJC68 + CRP (mg/dL) + PAIN + GDPT, *where*

- SJC66 and TJC68 are defined similarly as in DAS28 (CRP) (Section 5.4.7.1).
- $CRP_{mg/dL}$: CRP in mg/dL. LLOQ rule specified in Appendix 1 will be applied to values < LLOQ.
- PAIN: Patient's Assessment of Pain on a 10-unit VAS.
- GDPT: Patient's Global Assessment of Disease Activity (arthritis) on a 10-unit VAS.

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

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

Subjects in remission, or with low disease activity based on the DAPSA score are those subjects who have a DAPSA score less than or equal to 4, or greater than 4 and less than or equal to 14, respectively.

5.5.1.1.10. Minimal Disease Activity (MDA) and Very Low Disease Activity (VLDA)

The PsA minimal disease activity (MDA) criteria are a composite of 7 outcome measures used in PsA. A subject is considered as having achieved the PsA MDA at a visit if the subject has fulfilled at least 5 of the following 7 criteria at that visit:[2]

- Tender joint count (68 joints) ≤ 1
- Swollen joint count (66 joints) ≤ 1
- Psoriasis activity and severity index ≤1
- Patient's Assessment of Pain ≤ 15 on a 100-unit VAS
- Patient's Global Assessment of Disease Activity (arthritis and psoriasis) ≤ 20 on a 100-unit VAS
- HAQ-DI score ≤ 0.5
- Tender entheseal points ≤1 (LEI index score ≤ 1)

A subject is considered as having achieved VLDA at a visit if the subject has fulfilled all 7 criteria described above at that visit.

5.5.1.1.11. BASDAI Related Endpoints

The Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) was developed as a subject self-assessment for and consists of 6 questions relating to the 5 major symptoms of ankylosing spondylitis.[7] Only subjects with spondylitis with peripheral arthritis as their primary arthritic presentation of PsA will complete the BASDAI using a 10-unit VAS to indicate the degree of their symptoms over the past week on the following criteria:

A. Fatigue on a 10-unit VAS
B. Spinal pain on a 10-unit VAS
C. Joint pain on a 10-unit VAS
D. Enthesitis on a 10-unit VAS
E. Qualitative morning stiffness on a 10-unit VAS
F. Quantitative morning stiffness on a 10-unit VAS

The BASDAI score = 0.2 * (A + B + C + D + 0.5*[E + F]). If any of the components required for computing the BASDAI score is missing, the BASDAI score will be set to missing.

Higher BASDAI scores indicate greater disease severity and a score decrease of 50% or 2 points is considered **clinically meaningful**.[25]

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

Note that change from baseline in BASDAI score, and **≥ 20%, ≥ 50%, ≥70 %, and ≥ 90% improvement from baseline in BASDAI score** are only applicable to subjects with spondylitis and peripheral joint involvement as their primary arthritic presentation of PsA and BASDAI >0 at baseline.

5.5.1.2. Analysis Method

Endpoints through Week 24 will be analyzed at Week 24 DBL based on FAS1 (Section 2.3.1.1) with the following exceptions:

HAQ-DI response, among FAS1 subjects with HAQ-DI score ≥ 0.35 at baseline.

BASDAI: among FAS1 subjects with spondylitis and peripheral joint involvement as their primary arthritic presentation of PsA and BASDAI >0 at baseline.

Endpoints from Week 24 through Week 52 will be analyzed at Final (Week 60) DBL based on FAS2 (Section 2.3.1.2).

All endpoints will be descriptively summarized by treatment groups. Treatment comparisons will be performed by visit through Week 24. Nominal p-values and 95% CIs for the difference between each guselkumab group and placebo group will be provided. No treatment comparison will be performed after Week 24.

Note that for endpoints of Enthesitis and Dactylitis:

- Analyses using data of CNTO1959PSA3001 study only will be performed and all the p-values will be considered nominal (including those at Weeks 16 and 24).

- However, analyses using pooled data from studies CNTO1959PSA3001 and CNTO1959PSA3002 will be analyzed in the CNTO1959PSA3002 study (refer to the SAP CNTO136PSA3002 – Amendment 3). The results from the pooled data analyses will be reported along with the analysis results from that study.

For **binary response endpoints**, treatment comparisons will be performed using a CMH test as described in Section 5.3.2.1. For **continuous endpoints**, treatment comparisons will be performed using an ANCOVA model on the MI data for HAQ-DI, DA 28 (CRP), dactylitis, and enthesitis endpoints as described in Section 5.1 and Appendix 3.

The following Table 5 outlines the other efficacy endpoints related to reduction of signs and symptoms and physical function, the estimands, the methods for analyses, and the data handling rules to be used.

Table 5: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Estimand	Analysis Set	Missing Data Rules	Analysis Methods
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL					
1	Proportions of subjects who achieved an ACR 20, ACR 50, and ACR 70 responses	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
2	ACR components	NA	FAS1	-	• Summarized descriptively
3	Percent change from baseline in ACR components	NA	FAS1	-	• Summarized descriptively
4	Change from baseline in HAQ-DI score	Composite (Section 5.2.4.1)	FAS1	MI Section (5.2.3.3)	• Descriptive summary statistics • ANCOVA for each MI dataset and then analysis results across all MI datasets combined • LS mean (95% CI) for each treatment group and for differences between groups. o P-values for treatment comparisons

Table 5: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Estimand	Analysis Set	Missing Data Rules	Analysis Methods
5	Proportion of HAQ-DI responders (≥0.35 Improvement from baseline in HAQ-DI score)	Composite (Section 5.2.4.1)	FAS1 whose baseline HAQ-DI score ≥ 0.35	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
6	Proportions of subjects with DAS28 (CRP) responses	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
7	Proportions of subjects with DAS28 remission	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
8	Change from baseline in DAS28 (CRP)	Composite (Section 5.2.4.1)	FAS1	MI Section (5.2.3.3)	• Descriptive summary statistics • ANCOVA for each MI dataset and then analysis results across all MI datasets combined • LS mean (95% CI) for each treatment group and for differences between groups. ○ P-values for treatment comparisons
9	Proportions of subjects with PsARC response	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
10	Proportions of subjects with resolution of enthesitis	Composite (Section 5.2.4.1)	FAS1 with enthesitis at baseline	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
11	Change from baseline in enthesitis score	Composite (Section 5.2.4.1)	FAS1 with enthesitis at baseline	MI Section (5.2.3.3)	• Descriptive summary statistics • ANCOVA for each MI dataset and then analysis results across all MI datasets combined • LS mean (95% CI) for each treatment group and for

Table 5: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Estimand	Analysis Set	Missing Data Rules	Analysis Methods
					differences between groups. ○ P-values for treatment comparisons
12	Proportions of subjects with resolution of dactylitis	Composite (Section 5.2.4.1)	FAS1 with dactylitis at baseline	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
13	Change from baseline in dactylitis score	Composite (Section 5.2.4.1)	FAS1 with dactylitis at baseline	MI Section (5.2.3.3)	• Descriptive summary statistics • ANCOVA for each MI dataset and then analysis results across all MI datasets combined • LS mean (95% CI) for each treatment group and for differences between groups. ○ P-values for treatment comparisons
14	Change from baseline in PASDAS	Composite (Section 5.2.4.1)	FAS1	-	• Summarized descriptively • MMRM model for ○ LS mean (SD) ○ Treatment difference in LS mean (95% CI) ○ P-value of comparing LS mean
15	Change from baseline in GRACE index score	Composite (Section 5.2.4.1)	FAS1	-	• Summarized descriptively • MMRM model for ○ LS mean (SD) ○ Treatment difference in LS mean (95% CI) ○ P-value of comparing LS mean
16	Change from baseline in DAPSA score	Composite (Section 5.2.4.1)	FAS1	-	• Summarized descriptively • MMRM model for ○ LS mean (SD) ○ Treatment difference in LS mean (95% CI) ○ P-value of comparing LS mean
17	Proportions of subjects with minimal disease activity	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics

Table 5: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Signs & Symptoms and Physical Function

	Endpoint	Estimand	Analysis Set	Missing Data Rules	Analysis Methods
18	Proportions of subjects with ≥20%, ≥50%, ≥70%, and ≥90% improvement from baseline in BASDAI score	Composite (Section 5.2.4.1)	FAS1 with spondylitis and peripheral joint involvement as their primary arthritic presentation of PsA and BASDAI>0 at baseline	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
19	Change from baseline in BASDAI score	Composite (Section 5.2.4.1)	FAS1 with spondylitis and peripheral joint involvement as their primary arthritic presentation of PsA and BASDAI>0 at baseline	-	• Summarized descriptively • MMRM model for o LS mean (SD) o Treatment difference in LS mean (95% CI) o P-value of comparing LS mean
20	Proportion of subjects with low or very low disease activity based on PASDAS	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value
21	Proportion of subjects with low disease activity based on GRACE score	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value
22	Proportion of subjects with low disease activity or remission based on DAPSA	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value
23	Proportions of subjects with very low disease activity	Composite (Section **Error! Reference source not found.**)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics

'NA' = Not Applicable; '–' indicates Missing Data Rules not to be applied.

ENDPOINTS BY VISIT FROM WEEK 24 THROUGH WEEK 52 AT Final (WEEK 60) DBL

- Subjects in FAS2 will be the analysis population.
- All the endpoints will be descriptively summarized by visit.
- No treatment comparison will be performed.
- No data handling rules will be applied and all analyses will be based on observed data.

5.5.2. Efficacy Endpoints Related to Skin Disease

Other efficacy endpoints related to skin disease include:

- Proportions of subjects who achieve a PASI 75, 90 and 100 responses by visit through Week 52 among the subjects with ≥ 3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.Proportion of subjects with an IGA score of 0 (cleared) by visit through Week 52 among the subjects with ≥ 3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.
- Change from baseline in PASI score by visit through Week 52 among the subjects with ≥ 3% BSA psoriatic involvement and an IGA score of ≥ 2 (mild) at baseline.
- Proportion of subjects who achieve both PASI 75 and ACR 20 responses by visit through Week 52 among the subjects with ≥3% BSA psoriatic involvement and an IGA score of ≥ 2 (mild) at baseline.
- Proportion of subjects who achieve both PASI 75 and modified PsARC response by visit through Week 52 among the subjects with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline.

This section outlines the definitions and analyses for the above skin disease endpoints.

5.5.2.1. Definitions

5.5.2.1.1. PASI Related Endpoints

The Psoriasis Area and Severity Index (PASI) is a system used for assessing and grading the severity of psoriatic lesions and their response to therapy.[5] In the PASI system, the body is divided into 4 regions: neck and head, trunk (including axillae and groin), upper extremities, and lower extremities (including buttocks), which account for 10%, 30%, 20%, and 40% of the total BSA, respectively. Each of these 4 regions is assessed separately for erythema, induration and scaling on a scale of 0 - 4. The PASI score is a continuous endpoint calculated as follows:

PASI = 0.1 (Eh + Ih + Sh) Ah + 0.3 (Et + It + St) At + 0.2 (Eu + Iu + Su) Au + 0.4 (El + Il + Sl) Al, where

- E, I, S are, respectively, the state of erythema (E), induration (I) and scaling (S) assessed on a scale of 0 = none, 1 = slight, 2 = moderate, 3 = severe, and 4 = very severe.
- h, t, u, l are the 4 body regions in the PASI system: neck and head (h), trunk (t), upper extremities (u), and lower extremities (l), respectively.
- A is the area of involvement for psoriatic lesions. The scale for estimating the area of involvement is outlined below:

  0 = no involvement
  1 = 1% to 9% involvement
  2 = 10% to 29% involvement
  3 = 30% to 49% involvement
  4 = 50% to 69% involvement
  5 = 70% to 89% involvement
  6 = 90% to 100% involvement

- Coefficients 0.1, 0.3, 0.2, and 0.4 corresponds to that head (h), trunk (t), upper extremities (u) and lower extremities (l) account for 10%, 30%, 20%, and 40% of the total BSA, respectively.

If any of the components required for computing the PASI score is missing, the PASI score will be set to missing. The PASI score ranges from 0 to 72, with a higher score indicating a more severe disease.

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

PASI 75 response is defined as ≥ 75% improvement from baseline in PASI scores.

PASI 90 response and **PASI 100 response** (and 50 response) are defined similarly, with improvement threshold replaced by 90% and 100% respectively.

5.5.2.2. Analysis Methods

Endpoints through Week 24 will be analyzed at Week 24 DBL based on FAS1 (Section 2.3.1.1) who had ≥ 3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline; endpoints from Week 24 through Week 52 will be analyzed at Final (Week 60) DBL based on FAS2 (Section 2.3.1.2) who had ≥ 3% BSA psoriatic involvement and an IGA score of ≥ 2 (mild) at baseline.

All endpoints will be descriptively summarized by treatment groups. Treatment comparisons will be performed by visit through Week 24. Nominal p values and 95% CIs for the difference between each guselkumab group and placebo group will be provided. No treatment comparison will be performed after Week 24.

For **binary response endpoints**, treatment comparisons will be performed using a CMH test as described in Section 5.3.2.1. For **continuous endpoints**, treatment comparisons will be performed using an MMRM model as described in Section 5.1 and Appendix 3.

The following Table 6 outlines the other efficacy endpoints related to skin disease, the estimands, the analysis sets, the data handling rules to be used, and the methods for analyses.

Table 6: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Skin Disease

	Endpoint	Estimand	Analysis Set	Missing Data Rules	Analysis Methods
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL					
1	Proportions of subjects who achieve a PASI 75, 90, 100 responses	Composite (Section 5.2.4.1)	FAS1 who had ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald

Table 6: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for Endpoints of Skin Disease					
	Endpoint	**Estimand**	**Analysis Set**	**Missing Data Rules**	**Analysis Methods**
					statistics
2	Proportions of subjects who achieve who achieve both PASI 75 and ACR 20 responses	Composite (Section 5.2.4.1)	FAS1 who had ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
3	Proportions of subjects who achieve who achieve both PASI 75 and modified PsARC responses	Composite (Section 5.2.4.1)	FAS1 who had ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
4	Proportions of subjects with an IGA score of 0 (cleared)	Composite (Section 5.2.4.1)	FAS1 who had ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics

5	Change from baseline in PASI scores	Composite (Section 5.2.4.1)	FAS1 who had ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline	-	• Summarized descriptively
					• MMRM model
					  ○ LS mean (SD)
					  ○ Treatment difference in LS mean (95%CI)
					  ○ P-value of comparing LS mean

'–' indicates Missing Data Rules not to be applied.					
ENDPOINTS BY VISIT FROM WEEK 24 THROUGH WEEK 52 AT FINAL (WEEK 60) DBL					
• Subjects in FAS2 will be the analysis population. • All the endpoints will be descriptively summarized by visit. • No treatment comparison will be performed. • No data handling rules will be applied and all analyses will be based on observed data.					

5.5.3. Other Efficacy Endpoints Related to Health-Related Quality of Life and Health Economics

In this study, health related quality of life (HRQOL) measures include questionnaires of SF-36 health survey, Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue), and Work Productivity and Activity Impairment (PROMIS-29).

Other efficacy endpoints related to HRQOL include:

- Change from baseline in SF-36 PCS score by visit through Week 52.
- Change from baseline in SF-36 MCS score by visit through Week 52.

- Change from baseline in domain scores of SF-36 scales by visit through Week 52.
- Proportion of subjects who achieve ≥ 5-point improvement from baseline in SF-36 MCS score by visit through Week 52.
- Proportion of subjects who achieve ≥ 5-point improvement from baseline in SF-36 PCS score by visit through Week 52.
- Change from baseline in FACIT-Fatigue score by visit through Week 52.
- Proportion of subjects who achieve ≥ 4-point improvement from baseline in FACIT-Fatigue scores by visit through Week 52.
- Change from baseline in FACIT-fatigue score at Week 24 by ACR 20 response (primary endpoint) at Week 24.
- Proportion of subjects who achieve ⩾ 4-point improvement from baseline in FACIT-fatigue score at Week 24 by ACR 20 response (primary endpoint) at Week 24.
- Change from baseline in PROMIS-29 scores by visit through Week 52.
- Proportion of subjects who achieve an improvement of ≥ 3 points in PROMIS-29 domain scores by visit through Week 24.
- Proportion of subjects who achieve an improvement of ≥ 5 points in PROMIS-29 domain scores by visit through Week 24.

This section outlines the definitions and analyses for the above HRQOL endpoints.

5.5.3.1. Definition

5.5.3.1.1. SF-36 Questionnaire

The SF-36 questionnaire is defined in Section 5.4.6.1.

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

5.5.3.1.2. FACIT-Fatigue Questionnaire

The FACIT-Fatigue is a questionnaire that assesses self-reported tiredness, weakness, and difficulty conducting usual activities due to fatigue.[1] The questionnaire consists of 13 questions that assess a subject's level of fatigue and tiredness over the last 7 days. Each question is graded on a 5-point scale (0 - 4); and accordingly, the total FACIT-Fatigue scores can range from 0 to 52, with lower score reflecting more fatigue and higher scores reflecting less fatigue. Note that, when at least 7 questions are answered, the score can be calculated and should be adjusted by the number of available questions.

Although not developed for PsA, the FACIT-Fatigue has been used to assess fatigue in clinical trials of subjects with RA and has demonstrated sensitivity to change in these subjects. In rheumatology, a change of 4 points is considered clinical meaningful and has been used as response definition in the RA population.[1]

Change from baseline in FACIT-Fatigue scores measures the change in fatigue where a positive change indicates an improvement and a negative change indicates a worsening.

5.5.3.1.3. PROMIS 29

PROMIS-29 profile instrument is intended for adults (ages 18+). It is a collection of short forms containing 4 items for each of seven PROMIS domains (Depression, Anxiety, Physical Function, Pain Interference, Fatigue, Sleep Disturbance, and Ability to Participate in Social Roles and Activities).[16] PROMIS-29 also includes an additional pain intensity 0-10 numeric rating scale (NRS). The PROMIS-29 Profile is universal rather than disease-specific. They assess all domains over the past seven days except for Physical Function which has no timeframe specified. The raw score of each domain is converted into a standardized score with a mean of 50 and a standard deviation (SD) of 10 (T-Score). The standardized T-score is reported as the final score for each participant. Pain Intensity is presented as raw responses (0-10). For PROMIS domains of Depression, Anxiety, Physical Function, Pain Interference, Fatigue, a score of 50 is the average for the United States general population with a standard deviation of 10, because testing was performed on a large sample of the general population. However, the other two domains (Ability to Participate in Social Roles and Activities and Sleep Disturbance) were not centered in a national sample. For these two domains, a score of 50 represents the average of the calibration sample which was generally more enriched for chronic illness, and a score of 50 likely represents somewhat sicker people than the general population. A higher PROMIS T-score represents more of the concept being measured. For negatively-worded concepts like Anxiety, a T-score of 60 is one SD worse than average. By comparison, an Anxiety T-score of 40 is one SD better than average. However, for positively-worded concepts like Physical Function, a T-score of 60 is better than average while a T-score of 40 is better.

5.5.3.2. Analysis Method

Endpoints through Week 24 will be analyzed at Week 24 DBL based on FAS1 (Section 2.3.1.1); endpoints from Week 24 through Week 52 will be analyzed at Final (Week 60) DBL based on FAS2 (Section 2.3.1.2).

All endpoints will be descriptively summarized by treatment groups. Treatment comparisons will be performed by visit through Week 24. Nominal p values and 95% CIs for the difference between each guselkumab group and placebo group will be provided. No treatment comparison will be performed after Week 24.

For **binary response endpoints**, treatment comparisons will be performed using a CMH test as described in Section 5.3.2.1. For **continuous endpoints**, treatment comparisons will be performed using ANCOVA on the MI data for MCS and PCS analyses as described in Section 5.4.1.2, and an MMRM or cLDA model as described in Section 5.1 and Appendix 3.

The following Table 7 outlines the other efficacy endpoints related to skin disease, the estimands, the analysis sets, the data handling rules to be used, and the methods for analyses.

Table 7: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for HRQOL Endpoints

	Endpoint	Estimand	Analysis Set	Missing Data Rules	Analysis Methods
ENDPOINTS BY VISIT THROUGH WEEK 24 AT WEEK-24 DBL					
1	Change from baseline in SF-36 PCS score	Composite (Section 5.2.4.1)	FAS1	MI (Section 5.2.3.3)	• Descriptive summary statistics • ANCOVA for each MI dataset and then analysis results across all MI datasets combined • LS mean (95% CI) for each treatment group and for differences between groups. o P-values for treatment comparisons
2	Change from baseline in SF-36 MCS score	Composite (Section 5.2.4.1)	FAS1	MI (Section 5.2.3.3)	• Descriptive summary statistics • ANCOVA for each MI dataset and then analysis results across all MI datasets combined • LS mean (95% CI) for each treatment group and for differences between groups. o P-values for treatment comparisons
3	Change from baseline in SF-36 domain scores of SF-36 scale	Composite (Section 5.2.4.1)	FAS1	-	• Summarized descriptively • MMRM model o LS mean (SD) for o Treatment difference in LS mean (95%CI) o P-value of comparing LS mean
4	Proportion of subjects who achieve ≥ 5-point improvement from baseline in SF-36 PCS	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
5	Proportion of subjects who achieve ≥ 5-point improvement from baseline in SF-36 MCS	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
6	Change from baseline in FACIT-Fatigue	Composite (Section 5.2.4.1)	FAS1	-	• Summarized descriptively • MMRM model o LS mean (SD) for o Treatment difference in LS mean (95%CI) o P-value of comparing LS mean
7	Change from baseline in FACIT-Fatigue at Week 24 by ACR 20 response at Week 24(primary	Composite (Section **Error! Reference source not found.**)	FAS1	-	• Summarized descriptively • MMRM model o LS mean (SD) for o Treatment difference in LS mean (95%CI)

Table 7: Summary of Estimands, Analysis Sets, Data Handling Rules, and Analysis Methods for HRQOL Endpoints

	Endpoint	Estimand	Analysis Set	Missing Data Rules	Analysis Methods
	endpoint)				• P-value of comparing LS mean
8	Proportions of subjects who achieve ⩾ 4-point improvement from baseline in FACIT-Fatigue score at Week 24 by ACR 20 response at Week 24 (primary endpoint)	Composite (Section **Error! Reference source not found.**)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
9	Proportions of subjects who achieve ≥ 4-point improvement from baseline in FACIT-Fatigue score	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics

10	Change from baseline in PROMIS-29 score	Composite (Section 5.2.4.1)	FAS1	-	• Summarized descriptively
					• MMRM model for
					  ○ LS mean (SD)
					  ○ Treatment difference in LS mean (95% CI)
					  ○ P-value of comparing LS mean

11	Proportions of subjects who achieve ⩾ 3-point improvement from baseline in PROMIS-29 domain scores	Composite (Section 5.2.4.1)	FAS1	NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
12	Proportions of subjects who achieve ⩾ 5-point improvement from baseline in PROMIS-29 domain scores	Composite (Section 5.2.4.1)		NRI (Section 5.2.3.1)	• Summarized descriptively • CMH test for p-value • Treatment difference in response rates and 95% CI calculated based on the Wald statistics
'–' indicates Missing Data Rules not to be applied.					
ENDPOINTS BY VISIT FROM WEEK 24 THROUGH WEEK 52 AT FINAL (WEEK 60) DBL					
• Subjects in FAS2 will be the analysis population. • All the endpoints will be descriptively summarized by visit. • No treatment comparison will be performed. • No data handling rules will be applied and all analyses will be based on observed data.					

Additional Analyses for FACIT-Fatigue

Subgroup analysis comparing change in fatigue score between treatment and control will be conducted in ACR20 non-responder subgroup and ACR20 responder subgroup separately. Propensity score-based methods will be used to account for potential imbalances in baseline covariates. Covariates to be used in the propensity score models will include baseline fatigue score and selected demographics and disease related variables such as age, gender, BMI, PsA duration, physician global assessment, patient global assessment, HAQ-DI score, pain assessment, swollen joints 66 and tender joints 68. Logistic regression or Covariate Balancing Propensity Score (CBPS) models will be used to estimate the propensity score.

A mediation analysis will be performed to examine the mediating role of 24-week ACR20 response on change from baseline in fatigue score at week 24 provided both endpoints demonstrate a statistically significant difference between treatment arms (Appendix 3).

5.5.4. Other Tipping Point Analyses

Additional tipping point analyses aside from those for primary and major secondary endpoints will be performed to evaluate the impact when the missing data deviate from the MAR assumption and are summarized in the below Table 7a. The analysis methods used are the same as those described in Table 4a.

Table 7a: Summary of Tipping Point Analyses Other Than for Primary and Major Secondary Endpoints

Endpoints (Analysis Set)	Missing data	Analysis method
Supplementary Analyses of **Continuous Endpoints based on Treatment Policy Estimand**, in which all observed data will be used, regardless of whether or not the subjects meet any TF criteria.		
Change from baseline at Week 16 in: • HAQ-DI score (FAS1) • SF-36 PCS score (FAS1) • SF-36 MCS score (FAS1) • DAS28(CRP) (FAS1) • Enthesitis (LEI) among subjects with enthesitis (LEI) at baseline (FAS1) • Dactylitis among subjects with dactylitis at baseline (FAS1)	MI with FCS regression (Section 5.2.3.6, **Error! Reference source not found.**)	• Two-dimensional tipping point analysis • Treatment comparison using ANCOVA model • Analysis results to be presented graphically
Supplementary Analyses of **Binary Endpoints based on Treatment Policy Estimand**, in which all observed data will be used, regardless of whether or not the subjects meet any TF criteria.		
• ACR 70 Response at Week 16 (FAS1) • IGA Response at Week 16 (FAS1 with ⩾3% BSA psoriatic involvement and an IGA score of ⩾2 at baseline) • Resolution of enthesitis (LEI) at Week 16 among subjects with enthesitis (LEI) at baseline • Resolution of dactylitis at Week 16 among subjects with dactylitis at baseline	MI with FCS regression (Section **Error! Reference source not found.**, **Error! Reference source not found.**)	• Two-dimensional tipping point analysis. • The CMH[a] test (stratified by randomization stratification factors) to compare treatment groups. • Analysis results to be presented graphically

[a] When combining analysis results for the CMH test, the Wilson-Hilferty transformation will be applied to the test statistics to

Table 7a: Summary of Tipping Point Analyses Other Than for Primary and Major Secondary Endpoints		
Endpoints (Analysis Set)	**Missing data**	**Analysis method**
achieve an approximate normal distribution.		

6. SAFETY

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.

In all the safety analysis, subjects who were randomized and received at least 1 (partial or complete) dose of study agent administration will be included and analyzed according to the treatment they actually received, regardless of the treatments they are randomized to. No formal statistical comparison is planned.

6.1. Safety Table Presentation

There are 2 DBLs in this study, respectively, at Week 24 and End of Study (Week 60). 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 and through End of Study (Week 60). Tabular summaries of safety events for key study periods are in general presented as follows:

6.1.1. Summaries Through Week 24

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

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

The above treatment 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	###	###	###	###

6.1.2. For subjects who started treatment with placebo only but later received any amount of guselkumab prior to Week 24 (due to inadvertency), 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 subjects had been receiving placebo only will be included in the data summaries through Week 24. Summaries Through End of Study (Week 60)

Safety data through End of Study (Week 60) will be analyzed according to the following treatment groups:

1. **Placebo:** Subjects who received placebo only. Follow-up will be based on the period that the subject was on placebo from the first dose up to end of study (Week 60).

    a. For subjects 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:** Subjects 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 End of Study (Week 60).

    - All the safety events/measurements that occurred on and after the first dose of guselkumab will be included in this group.

3. **Guselkumab 100 mg at Weeks 0, 4, and then q8w:** Subjects 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 end of study.

4. **Guselkumab 100 mg q4w:** Subjects who received guselkumab 100 mg q4w prior to Week 24. Follow-up will be from the first dose through end of study.

5. **Guselkumab 100 mg q4w Combined:** Combined 2 and 4.

6. **All Guselkumab Combined:** Combined 2, 3, and 4.

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

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

6.2. Adverse Events

AEs that occurred any time over the study will be reported and coded using MedDRA. Analyses of AEs will be performed on those events that are considered treatment emergent. Treatment emergent AEs (TEAEs) are those AEs that occurred after the start of initial study agent administration and those AEs that were present at baseline but worsened in severity after the start of initial study agent administration.

TEAEs will be summarized by MedDRA system organ class, preferred term, and actual treatment group. The numbers of subjects reporting at least 1 event of the following AE categories will be summarized:

- Any TEAEs
- Treatment emergent serious AEs (SAEs)
- TEAEs with severe intensity
- TEAEs that led to permanent discontinuation of study agent administration
- Treatment emergent infections
- Treatment emergent serious infections
- Treatment emergent infections requiring oral or parenteral anti-microbial treatment
- Injection-site reactions
- Anaphylactic reactions or serum sickness reactions

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

In addition to the summary tables, a by-subject listing will be provided for deaths that occurred during the study and, respectively, for the following TEAEs:

1. SAEs
2. AEs that led to permanent discontinuation of study agent administration
3. Anaphylactic reactions or serum sickness reactions
4. Malignancies
5. Serious infections including TB

6.3. Clinical Laboratory Tests

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

- **Hematology**: bands, basophils, eosinophils, hemoglobin, hematocrit, lymphocytes, monocytes, neutrophils, platelets, red blood cell (RBC) count and white blood cell (WBC) count
- **Clinical chemistry**: albumin, alkaline phosphatase, 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, total protein, uric acid

The following analyses will be performed as appropriate by actual treatment group:

- Plots of the observed values and changes from baseline over time for selected clinical laboratory parameters
- Number of subjects with post-baseline values by maximum toxicity grade according to National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI-CTCAE) for parameters with NCI-CTCAE criteria defined
- Listings of subjects with any post-baseline lab value of NCI-CTCAE toxicity Grade 3 or higher

6.4. Vital Signs and Physical Examination Findings

Vital signs will be measured at visits as per the schedule of events in the protocol. Descriptive statistics of the observed value and change from baseline of the vital signs will be summarized by treatment group. The numbers of subjects with any markedly abnormal post-baseline measurements will also be summarized over time by treatment group. The criteria for markedly abnormal vital signs are defined in the following table.

Abnormalities or changes in severity noted during physical examination will be reported as adverse events and are included in the analysis of adverse events.

Parameter	Markedly Abnormal Low	Markedly Abnormal High
Systolic blood pressure	Absolute value ≤ 90 mmHg and a decrease from baseline by ≥ 20 mmHg	Absolute value ≥ 180 mmHg and an increase from baseline by ≥ 20 mmHg
Diastolic blood pressure	Absolute value ≤ 50 mmHg and a decrease from baseline by ≥ 15 mmHg	Absolute value ≥ 105 mmHg and an increase from baseline by ≥ 15 mmHg
Pulse	Absolute value ≤ 50 bpm and a decrease from baseline ≥ 15 bpm	Absolute value ≥ 120 bpm and an increase from baseline ≥ 15 bpm

6.5. Electrocardiogram

This section does not apply to this study.

6.6. Other Safety Parameters

6.6.1. 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 subject self-report questionnaire, including standardized questions, follow-up prompts, error handling routines, and scoring conventions.[17,19] 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 60.

Subjects 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 subjects 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 subject completes prior to study agent administration.

At the conclusion of each assessment, the site will receive an eC-SSRS Findings Report from the eC-SSRS vendor. Positive reports are generated from the eC-SSRS vendor for ANY of the following findings:

- Ideation Level 4: Some intent to act, no plan
- Ideation Level 5: Specific plan and intent
- Behaviors: Actual Suicide Attempts
- Behaviors: Interrupted Attempts
- Behaviors: Aborted Attempts
- Behaviors: Preparatory actions

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.

Suicidal ideation and behavior will be categorized as follows, with higher scores indicating greater severity:

Suicidal Ideation (1-5)

1 = Wish to be Dead

2 = Non-specific Active Suicidal Thoughts

3 = Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act

4 = Active Suicidal Ideation with Some Intent to Act, without Specific Plan

5 = Active Suicidal Ideation with Specific Plan and Intent

Suicidal Behavior (6-10)

6 = Preparatory Acts or Behavior

7 = Aborted Attempt

8 = Interrupted Attempt

9 = Actual Attempt (non-fatal)

10 = Completed Suicide

The baseline is defined as the most severe/maximum score at screening and Week 0. Suicidal ideation and behavior will be analyzed by the most severe/maximum post baseline outcome. In addition, a shift table from baseline to post-baseline will also be provided. Subjects with positive (i.e., score > 0) ideation and behavior will be presented in a data listing.

7. PHARMACOKINETICS/PHARMACODYNAMICS

7.1. Pharmacokinetics

Pharmacokinetics (PK) samples for measuring serum guselkumab concentrations will be collected from all subjects at the specified visits as shown in the schedule of events of the protocol. Serum samples will also be collected at the final visit from subjects who terminate study participation early. Samples must be collected before study agent administration at visits when a study agent administration is scheduled. A random venous blood sample for population PK analysis will be collected from all subjects on any day between Weeks 4 to 12, except on the days of the scheduled study visit at Weeks 4, 8, and 12. Additionally, this blood sample must be collected at least 24 hours prior to or after the actual time of study agent administration at Weeks 4, 8, or 12 (ie, it cannot be collected within 24 hours before or after study agent administration).

All PK analyses will be based on the PK Analysis Set (Section 2.3.3). Subjects will be analyzed according to the treatment groups that they actually received. No imputation for missing concentration data will be performed.

For analysis on serum guselkumab concentrations, descriptive statistics, including arithmetic mean, SD, median, interquartile range, minimum and maximum will be calculated, where appropriate, by treatment group at each serum sampling time. 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 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 2.1.1), 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.

Population PK analyses will be performed to characterize the disposition characteristics of guselkumab based on the guselkumab concentration data from PsA studies. Data may be combined with other selected studies to support a relevant structural model. The population pharmacokinetic approach will also be used to identify significant covariates such as demographic characteristics (including but not limited to body weight, ethnic origin, sex, and age) and concomitant medications in subjects with PsA. A detailed analysis plan for population PK analysis will be developed separately, and a stand-alone technical report will be written to summarize the results of the population PK analysis.

PK analyses presentation

PK analyses will be summarized through the following time periods:

- Through Week 24
- Through End of Study (Week 60)

For the analyses, a subject is included in one and only one treatment group on the basis of the treatment regimen followed. The description of treatment groups are as follows:

1. **Placebo → Guselkumab 100 mg q4w**: Subjects randomized to placebo and were switched over to Guselkumab 100 mg at Week 24 (for summary through Week 60 only).
2. **Guselkumab 100 mg at Week 0, Week 4, and then q8w**: Subjects randomized to guselkumab 100 mg at Week 0, Week 4, and then q8w and received guselkumab 100 mg q8w throughout with an additional dose at Week 4.

3. **Guselkumab 100 mg q4w**: Subjects randomized to guselkumab 100 mg q4w and received guselkumab 100 mg q4w throughout.

7.2. Immunogenicity (Antibodies to Guselkumab)

Blood samples will be collected to examine the formation of antibodies to guselkumab at the specified visits as shown in the schedule of events of the protocol. Serum samples will also be collected at the final visit from subjects who terminate study participation early.

The antibodies to guselkumab will be summarized based on Immunogenicity Analysis Set (Section 2.3.4). Subjects will be analyzed according to the treatment groups that they actually received. 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 ACR 50, PASI) 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 trough serum guselkumab concentrations over time by antibody to guselkumab status

Immunogenicity analyses presentation

Immunogenicity analyses will be summarized through the following time periods:

- Through Week 24
- Through End of Study (Week 60)

For the immunogenicity analyses, the description of treatment groups is as follows:

1. **Placebo → Guselkumab 100 mg q4w**: Subjects randomized to placebo and were switched over to Guselkumab 100 mg at Week 24 (for summary through Week 60 only).
2. **Guselkumab 100 mg at Week 0, Week 4, and then q8w:** Subjects randomized to guselkumab 100 mg at Week 0, Week 4, and then q8w and received guselkumab 100 mg q8w throughout with an additional dose at Week 4.
3. **Guselkumab 100 mg q4w**: Subjects randomized to guselkumab 100 mg q4w and received guselkumab 100 mg q4w throughout.

4. **Guselkumab Combined:** all subjects who received guselkumab in groups 1, 2 and 3.

7.3. Pharmacodynamics

Samples for serum biomarkers will be collected for all subjects as indicated in the Time and Events Schedule of the protocol. The analyses on PD biomarkers are to better understand the biology of PsA, to provide a biological assessment of the subjects' response to treatment with guselkumab, to analyze differences between responders and non-responders, and to determine if the markers can be used to classify subjects as potential responders prior to treatment.

All PD analyses will be based on the PD Analysis Set (Section 2.3.5). Subjects will be analyzed according to the treatment groups that they actually received. No imputation for missing concentration data will be performed. The PD analyses results will be provided in an independent technical report.

7.4. 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.

REFERENCES

1. 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.
2. 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.
3. 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.
4. 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.
5. Fredriksson T, Pettersson U. Severe psoriasis-oral therapy with a new retinoid. Dermatologica. 1978;157(4):238-244.
6. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcomes in arthritis. Arthritis Rheum. 1980;23(2):137-145.
7. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol. 1994;21(12):2286-2291.
8. 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.
9. 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.
10. 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.
11. Helliwell PS, FitzGerald O, Fransen J, et al. The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project). Ann Rheum Dis. 2013;72(6):986-991.
12. Helliwell PS, FitzGerald O, Fransen J. Composite disease activity and responder indices for psoriatic arthritis: A report from the GRAPPA 2013 Meeting on development of cutoffs for both disease activity states and response. J Rheum. 2014;41(6):1212-1217.
13. 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.
14. 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.
15. Maksymowych WP, Mallon C, Morrow S, et al. Development and validation of the Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index. Ann Rheum Dis 2009;68:948 –53.
16. Maksymowych WP, Mallon C, Morrow S, Shojania K, Olszynski WP, Wong RL, et al. Development and validation of the Spondyloarthritis
17. 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.
18. 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.
19. 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.
20. Rubin DB, Multiple Imputations for Nonresponse in Surveys. New York: John Wiley & Sons. 1987.

21. Saris-Baglama RN, Dewey CJ, Chisholm GB, et al. QualityMetric Health OutcomesTM Scoring Software 4.5 User's Guide. QualityMetric Health Outcomes Solutions. 2011.

22. van der Linden MA. Disease Activity Scores using C-reactive protein. April 5, 2004. Available at: http://www.umcn.nl/scientist/afdelingen/reumatologie/the_das/the_das_crp. Accessed 19 August 2005.

23. 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.; 2000:40.

24. Ware JE, Kosinski M, Bjorner JB, et al. User's manual for the SF-36v2 Health Survey (second edition). Lincoln, RI: QualityMetric Incorporated. 2007.

25. Zochling J, van der Heijde D, Burgos-Vargas R, et al. ASAS/EULAR recommendations for the management of ankylosing spondylitis. Ann Rheum Dis. 2006;65(4):442-452.

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

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

28. L. Valeri and T. J. VanderWeele (2013), Meidation Analysis Allowing for Exposure-Mediator Interactions and Causal Interpretation: Theoretical Assumptions and Implementation with SAS and SPSS Macros, Psychological Methods, Vol. 18, No.2: 137-150.

APPENDICES

APPENDIX 1: RULES APPLIED IN DEFINITIONS OF ENDPOINTS

1. Joint Evaluability Rules for Sign and Symptom Data

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

- For subjects having a joint injection and/or surgical joint procedure prior to the date of randomization, the affected joints will be analyzed according to the impact of the joint injection and/or surgical joint procedure on the evaluability of the involved joints.
    - If a joint is considered un-evaluable at baseline due to certain procedure/injection performed prior to the date of randomization, the joint will be considered un-evaluable throughout the study.
- For subjects undergoing surgical joint procedures for the treatment of PsA during the study, the affected joints will be considered as swollen and tender from the date of procedure onwards.
- For subjects undergoing surgical 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 subjects 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 subjects 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 subjects who have an incomplete set of evaluable joints the joint count/score will be adjusted to the total number joints of interest (e.g., 68 joints for tenderness and 66 joints for swelling) by dividing the number of affected joints by the number of evaluable joints and multiplying by the total number joints of interest.

3. LLOQ rule

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

APPENDIX 2: STATISTICAL HYPOTHESES

Hypothesis	Global	US Specific
Primary Endpoints		
H1. Guselkumab 100 mg q4w SC is superior to placebo as assessed by the proportion of subjects achieving an ACR 20 response at Week 24 ***(primary hypothesis)***	Controlled as in figure 2	Controlled as in figure 3
H2. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by the proportion of subjects achieving an ACR 20 response at Week 24	Controlled as in figure 2	Controlled as in figure 3
Major Secondary Endpoints		
H3. Guselkumab 100 mg q4w SC is superior to placebo as assessed by_proportion of subjects who achieved a psoriasis IGA response at Week 24 among the subjects with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline	Controlled as in figure 2	Controlled as in figure 3
H4. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by_proportion of subjects who achieved a psoriasis IGA response at Week 24 among the subjects with ≥3% BSA psoriatic involvement and an IGA score of ≥2 (mild) at baseline	Controlled as in figure 2	Controlled as in figure 3
H5. Guselkumab 100 mg q4w SC is superior to placebo as assessed by change from baseline in DAS28 (CRP) at Week 24	Controlled as in figure 2	Weakly-controlled
H6. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by change from baseline in DAS28 (CRP) at Week 24	Controlled as in figure 2	Weakly-controlled
H7. Guselkumab 100 mg q4w SC is superior to placebo as assessed by change from baseline in HAQ-DI score at Week 24	Controlled as in figure 2	Controlled as in figure 3
H8. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by change from baseline in HAQ-DI score at Week 24	Controlled as in figure 2	Controlled as in figure 3
H9. Guselkumab 100 mg q4w SC is superior to placebo as assessed by change from baseline in SF-36 PCS at Week 24	Controlled as in figure 2	Controlled as in figure 3
H10. Guselkumab 100 mg at Week 0, Week 4, then q8w	Controlled as	Controlled as

SC is superior to placebo as assessed by change from baseline in SF-36 PCS at Week 24	in figure 2	in figure 3
H11. Guselkumab 100 mg q4w SC is superior to placebo as assessed by proportion of subjects who achieved an ACR 50 response at Week 24	Controlled as in figure 2	Weakly-controlled
H12. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by proportion of subjects who achieved an ACR 50 response at Week 24;	Controlled as in figure 2	Weakly-controlled
H13. Guselkumab 100 mg q4w SC is superior to placebo as assessed by proportion of subjects achieved an ACR 20 response at Week 16	Controlled as in figure 2	Weakly-controlled
H14. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by proportion of subjects achieved an ACR 20 response at Week 16	Controlled as in figure 2	Weakly-controlled
H15. Guselkumab 100 mg q4w SC is superior to placebo as assessed by proportion of subjects who achieve an ACR 70 response at Week 24	Controlled as in figure 2	Weakly-controlled
H16. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by proportion of subjects who achieve an ACR 70 response at Week 24	Controlled as in figure 2	Weakly-controlled
H17. Guselkumab 100 mg q4w SC is superior to placebo as assessed by proportion of subjects who achieve an ACR 50 response at Week 16	Controlled as in figure 2	Weakly-controlled
H18. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by proportion of subjects who achieve an ACR 50 response at Week 16	Controlled as in figure 2	Weakly-controlled
H19. Guselkumab 100 mg q4w SC is superior to placebo as assessed by proportion of subjects with resolution of enthesitis at Week 24 among the subjects with enthesitis at baseline	Nominal	Nominal
H20. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by proportion of subjects with resolution of enthesitis at Week 24 among the subjects with enthesitis at baseline	Nominal	Nominal
H21. Guselkumab 100 mg q4w SC is superior to placebo as assessed by proportion of subjects with resolution of dactylitis at Week 24 among the subjects with dactylitis at baseline	Nominal	Nominal

H22. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by proportion of subjects with resolution of dactylitis at Week 24 among the subjects with dactylitis at baseline	Nominal	Nominal
H23. Guselkumab 100 mg q4w SC is superior to placebo as assessed by change from baseline in SF-36 MCS at Week 24	Nominal	Nominal
H24. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by change from baseline in SF-36 MCS at Week 24	Nominal	Nominal
H25. Guselkumab 100 mg q4w SC is superior to placebo as assessed by change from baseline in enthesitis score (based on Leeds Enthesitis Index [LEI]) at Week 24 among the subjects with enthesitis at baseline	Nominal	Nominal
H26. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by change from baseline in enthesitis score (based on Leeds Enthesitis Index [LEI]) at Week 24 among the subjects with enthesitis at baseline	Nominal	Nominal
H27. Guselkumab 100 mg q4w SC is superior to placebo as assessed by change from baseline in dactylitis scores at Week 24 among the subjects with dactylitis at baseline	Nominal	Nominal
H28. Guselkumab 100 mg at Week 0, Week 4, then q8w SC is superior to placebo as assessed by change from baseline in dactylitis scores at Week 24 among the subjects with dactylitis at baseline	Nominal	Nominal

APPENDIX 3: DESCRIPTION OF STATISTICAL MODELS

1. Mixed-effect Model Repeated Measures (MMRM)

To account for the missing data for continuous endpoints of change from baseline, a Mixed-Effect Model Repeat Measures (MMRM) will be used, under the assumption of MAR, to test the difference between a guselkumab group and the placebo group. The explanatory variables of the model will include treatment group, an interaction term of visit with treatment group, an interaction term of visit with baseline use of non-biologic DMARDs (yes, no), an interaction term of visit with prior exposure to anti-TNFα agents (yes, no), and an interaction term of visit with baseline score. 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. 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.

2. Constrained Longitudinal Data Analysis (cLDA) Model

To account for the missing data for continuous endpoints when 2% or more subjects have baseline score missing, a Constrained Longitudinal Data Analysis (cLDA) model may be performed based on the original measurement score rather than on the change score, under the assumption of MAR, to test the difference between a guselkumab group and the placebo group. The explanatory variables of the model will include treatment group, an interaction term of visit with treatment group, an interaction term of visit with baseline use of non-biologic DMARDs (yes, no), and an interaction term of visit with prior exposure to anti-TNFα agents (yes, no). 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. 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 LS means. The 95% CIs for the differences in LS means and p-values will be calculated based on the cLDA model.

3. Mediation Analysis

Mediation analysis will be performed to examine the mediating role of 24-week ACR20 response on change from baseline in fatigue score at week 24 provided both endpoints demonstrate a statistically significant difference between treatment arms. In the mediation analysis framework, natural direct effect can be conceived of as independent treatment effect on the outcome (i.e. change in fatigue score) that is above and beyond its effect on the mediator (i.e. ACR20 response); controlled direct effect can be conceived of as the independent treatment

effect on the outcome controlling the mediator at a fixed level; and the natural indirect effect can be conceived of as a treatment effect on the outcome that is mediated by its effect on the mediator.

The parametric counterfactual approach implemented by Valeri and VanderWeele (2013)[35] will be used to estimate the controlled direct effect (CDE), natural direct effect (NDE) and natural indirect effect (NIE). Let Y_α and Mα denote the values of the outcome and mediator that would have been observed had the exposure A been set to level α and let Yαm denote the value of the outcome that would have been observed had the exposure A and mediator M been set to levels αand m respectively. In the analysis of binary exposure, we will set α=1 for treated arm and α=0 for controlled arm. The average controlled direct effect (CDE) comparing treated (α =1) to control (α =0) given the mediator fixing to level m is defined by $CDE(m) = E(Y_{1m} - Y_{0m})$. The average natural direct effect (NDE) is defined by $NDE = E(Y_{1M_0} - Y_{0M_0})$. The average natural indirect effect is defined by $NIE = E(Y_{1M_1} - Y_{1M_0})$.

Covariates considered in this mediation analysis will include baseline fatigue score and demographics parameters and disease related parameters such as age, gender, BMI, PsA duration, physician Global Assessment (GDEV), patient global assessment (GDPT), HAQ-DI score, pain assessment (PAIN), Swollen Joints 66 (SJC66) and tender joints 68 (TJC68). We will also allow exposure-mediator interaction in this analysis.

Bootstrap will be used to obtain final effect estimates and their corresponding confidence intervals and P values. Model and variable selection may be conducted as sensitivity analysis. Alternative implementations of mediation analysis may also be explored as part of sensitivity analysis.

APPENDIX 4: SUMMARY OF MULTIPLE IMPUTATION METHODS

Endpoints	MI specification	Analysis method/Summary statistics
ACR 20 response at Week 24 **Supplementary analysis 1** **Supplementary analysis 2**	Multiple imputation with FCS regression of component scores	MIdataset1 (N=200, Seed [c] =4362478) • Imputation variables: 7 ACR components from Week 0 - 24 • Ancillary variables: Treatment group, randomization stratification factors
Change from baseline through Week 24 in: • HAQ-DI Score • DAS28(CRP) Score	Multiple imputation with FCS regression of component scores	MIdataset2[a] (N=200, Seed [c] =4237590) • Imputation variables: 5 ACR components other than joint counts from Week 0 - Week 24, tender joint counts based on 28 joints, swollen joint counts based on the 28 joints. • Ancillary variables: Treatment group, randomization stratification factors
Change from baseline through Week 24 in Enthesitis score (LEI) among subjects with at least one tender enthesis at baseline	Multiple imputation with FCS regression of enthesitis scores	MIdata3 (N=200, Seed [c] =6509723) Note: The MI is done only for the subset of subjects who have enthesitis at baseline and not all subjects in the study population. • Imputation variables: enthesitis score from Weeks 0 – 24. • Ancillary variables [b]: Treatment group, randomization stratification factors, 7 ACR components from Weeks 0 – 24, enthesitis-4[d]
Dactylitis change from baseline through Week 24 among subjects with dactylitis at baseline	Multiple imputation with FCS regression of dactylitis scores	MIdata4 (N=200, Seed [c] =1284097) Note: The MI is done only for the subset of subjects who have dactylitis at baseline and not all subjects in the study population. • Imputation variables: dactylitis score from Weeks 0 – 24. • Ancillary variables [b]: Treatment group, randomization stratification factors, 7 ACR components from Weeks 0 – 24
Change from baseline through Week 24 in: • SF-36 PCS score • SF-36 MCS score	Multiple imputation with FCS regression of PCS, MCS scores	MIdata5 (N=200, Seed [c] =890473) • Imputation variables: PCS and MCS scores from Weeks 0 – 24 • Ancillary variables [b]: Treatment group, randomization stratification factors, 7 ACR components from Weeks 0 - 24
IGA change from baseline through Week 24 among subjects with ⩾3% BSA psoriatic involvement and an IGA score of ⩾2 at baseline	Multiple imputation with FCS regression of IGA scores	MIdata6 (N=200, Seed[e] = 413249) • Imputation variables: IGA scores from Weeks 0 – 24 Ancillary variables [b]: Treatment group, randomization stratification factors, 7 ACR components from Weeks 0 - 24

[a] MIdataset2 is different from MIdataset1 as the joint scores used in DAS 28 (CRP) are different from those used in ACR.

[b]An ancillary variable may be removed if its correlation with an indicator variable that determines missing-ness of the variable to be imputed is low or, there are too many missing values for the ancillary variable within the subgroup of incomplete cases for the variable to be imputed. All the 7 ACR components (including all measurements from baseline through week 24) are in the list of the ancillary variables since they may be related to the mechanism of the

missing data.

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

[d] Enthesitis-4 is the tender entheses count based on 4 sites (left and right achilles tendon insertion, and left and right humeral epicondyle lateral) instead of 6 sites. Only baseline, Week 4 are included as ancillary variables.