Janssen Pharmaceutical K.K.

Statistical Analysis Plan

A Multicenter, Randomized, Placebo-controlled, Double-blind, Proof-of-concept Study of Guselkumab in Participants with Systemic Sclerosis

Protocol CNTO1959SSC2001; Phase 2a

CNTO1959 (Guselkumab)

*This study is being conducted by Janssen Pharmaceutical K.K. in Japan. The term "sponsor" is used throughout the protocol to represent Janssen Pharmaceutical K.K.

Status:	Approved
Date:	7 June 2023
Prepared by:	Janssen Pharmaceutical K.K.
Document No.:	EDMS-RIM-330901

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

TABLE OF CONTENTS

1. INTRODUCTION

This protocol CNTO1959SSC2001 is a Phase 2a Proof-of-Concept (PoC) study of Guselkumab in participates with Systemic Sclerosis (SSc).

This Statistical Analysis Plan (SAP) contains definitions of analysis sets, derived variables and statistical methods for all planned analyses for CNTO1959SSC2001 including both main study period and long-term extension (LTE) period.

1.1. Objectives and Endpoints

Objectives and endpoints of this study are listed in Table 1.

Table 1: Objectives and Endpoints

Objectives	Endpoints
Primary	
• To evaluate the efficacy of guselkumab in participants with SSc	• Change from baseline in Modified Rodnan Skin Score (mRSS) at Week 24
Secondary	
• To evaluate the additional efficacy of guselkumab in participants with SSc	• Change from baseline in mRSS at Week 52 • The proportion of participants who experience worsening of mRSS at Week 24 and Week 52 • The proportion of participants achieving a score of 0.6 in American College of Rheumatology Combined Response Index in dcSSc (ACR CRISS) at Week 24 and Week 52 • Change from baseline in forced vital capacity (FVC) and percent predicted FVC at Week 24 and Week 52 • Change from baseline in the measured absolute diffusing capacity of the lung for carbon monoxide (DLCO) and the derived percent predicted DLCO at Week 24 and Week 52 • Change from baseline in digital ulcer counts at Week 24 and Week 52 • Change from baseline in Health Assessment Questionnaire-Disability Index (HAQ-DI) score at Week 24 and Week 52
• To evaluate the safety and tolerability of guselkumab in participants with SSc	• Number and proportion of participants with treatment-emergent adverse events (AEs) and serious adverse events (SAEs) or adverse events of special interest (AESI)

Objectives	Endpoints
	through Week 24, Week 52, Week 76 and Week 104
• To evaluate the pharmacokinetic(s) and immunogenicity of guselkumab	• Serum guselkumab concentrations • Incidence of anti-guselkumab antibody
Exploratory	
• To evaluate the efficacy of guselkumab in SSc	• Time to disease worsening: increase from baseline ≥5 and ≥20% in mRSS or decrease >10% from baseline in FVC % • The proportion of participants who experience FVC decline compared with baseline at Week 24 and Week 52 • The proportion of participants who experience worsening of FVC at Week 24 and Week 52 • Change from baseline in fibrotic change assessed with high-resolution computed tomography (HRCT) at Week 24 and Week 52 in participants with baseline fibrosis/fibrotic changes • The proportion of participants who experience worsening of fibrotic change assessed with HRCT at Week 24 and Week 52 • Change from baseline in Frequency Scale for the Symptoms of Gastroesophageal reflux disease (FSSG) score at Week 24 and Week 52 • Assessment of nailfold capillaries at Week 24 and Week 52 • Change from baseline in Patients global assessment (PGA) at Week 24 and Week 52 • Change from baseline in Physician global assessment (PhGA) at Week 24 and Week 52
• To evaluate the long-term efficacy of guselkumab in participant with SSc	• Change from baseline and Week 52 in mRSS at Week 76 and Week 104 • The proportion of participants who experience worsening of mRSS at Week 76 and Week 104

	• The proportion of participants achieving a score of 0.6 in American College of Rheumatology Combined Response Index in dcSSc (ACR CRISS) at Week 76 and Week 104 • Change from baseline in forced vital capacity (FVC) and percent predicted FVC at Week 76 and Week 104 • The proportion of participants who experience worsening of FVC at Week 76 and Week 104 • Change from baseline in the measured absolute diffusing capacity of the lung for carbon monoxide (DLCO) and the derived percent predicted DLCO at Week 76 and Week 104 • Change from baseline in Health Assessment Questionnaire-Disability Index (HAQ-DI) score at Week 76 and Week 104 • Change from baseline in digital ulcer counts at Week 76 and Week 104 • Change from baseline in fibrotic change assessed with high-resolution computed tomography (HRCT) at Week 76 and Week 104 in participants with baseline fibrosis/fibrotic changes • The proportion of participants who experience worsening of fibrotic change assessed with HRCT at Week 76 and Week 104 • Change from baseline in Frequency Scale for the Symptoms of Gastroesophageal reflux disease (FSSG) score at Week 76 and Week 104
• To evaluate the effect of guselkumab vs placebo on the pharmacodynamic (PD) and SSc disease biomarkers	• Change from baseline levels of skin (optional) and circulating (blood based) biomarkers over time to enable: Evaluation of target engagement in tissue/efficacy pharmacodynamic (PD) readouts • Single nucleotide polymorphisms (SNPs) frequency associated with IL-23 pathway

1.2. Study Design

This is a randomized, double-blind, placebo-controlled, parallel, multicenter, interventional study to evaluate the efficacy of guselkumab in men and women between the ages of 18 to 75 years, inclusive, with a diagnosis of SSc according to American College of Rheumatology (ACR) and European League Against Rheumatism (EULAR) 2013 criteria with a disease duration of ≤36 months (defined as time from first non-Raynaud phenomenon manifestation), and Modified Rodnan Skin Score (mRSS) of ≥10 to ≤22 units.

Participants will be randomly assigned in a 1:1 ratio, based on randomization strata of presence of interstitial lung disease (ILD [yes, no]), baseline mRSS (low [≥10 to ≤15] or high [≥16 to ≤22]), and baseline anti-topoisomerase I antibody status (positive, negative), to 1 of the following treatment groups:

- Guselkumab (Group/Arm A): Guselkumab 400 mg IV administration at Weeks 0, 4, and 8 (induction) followed by guselkumab 200 mg subcutaneous (SC) administration every 4 weeks (Q4W) from Week 12 until Week 48 (maintenance)
- Placebo (Group/Arm B): Matching placebo IV administration at Weeks 0, 4, and 8 (induction) followed by matching placebo SC administration Q4W from Week 12 until Week 48 (maintenance)

The main study will be conducted in 3 phases: a maximum 6-week screening phase, a 52-week double-blind intervention phase, and a safety follow-up phase with a postintervention follow-up visit at 12 weeks after the participant's last dose of study intervention to collect any adverse events (AEs) since the last study visit. The duration of individual participation will be approximately 66 weeks without LTE.

Participants who complete the main study (Week 0 through 52: ie, after the Week 48 evaluation, prior to Week 52 evaluation) and who, in the opinion of the investigator, may benefit from continued treatment, will participate in the LTE by signing the ICF before or at Week 52. The duration of individual participation will be approximately 66 weeks without LTE and 118 weeks with LTE.

- Group/Arm A (guselkumab arm from the main study): Guselkumab 200 mg SC and placebo IV at LTE Weeks 52, 56, and 60 followed by guselkumab 200 mg SC Q4W from LTE Week 64 until Week 100.
- Group/Arm B (placebo arm from the main study): Placebo SC and guselkumab 400 mg IV at LTE Weeks 52, 56, and 60 followed by guselkumab 200 mg SC Q4W from LTE Week 64 until Week 100.

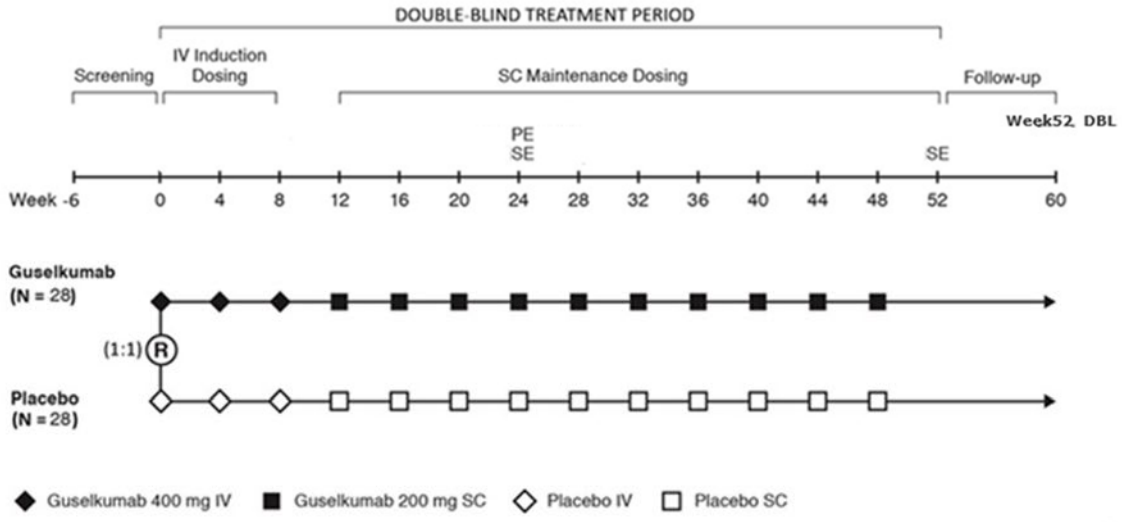
The objective of LTE is to evaluate the long-term safety and efficacy of guselkumab in SSc patients. The LTE begins with the completion of the assessments at the Week 52, and LTE continues through LTE Week 104. The LTE includes a safety follow-up phase with a postintervention follow-up visit at LTE Week 112 (12 weeks after the participant's last dose of study intervention).

Two planned database locks (DBLs) will occur: EOS of the main study (Week 52 visit for participants who entered LTE or safety follow up visit for participants who did not enter LTE,

hereinafter referred to as "Week 52 DBL"), and EOS of LTE (Week 112 visit, hereinafter referred to as "Final DBL").

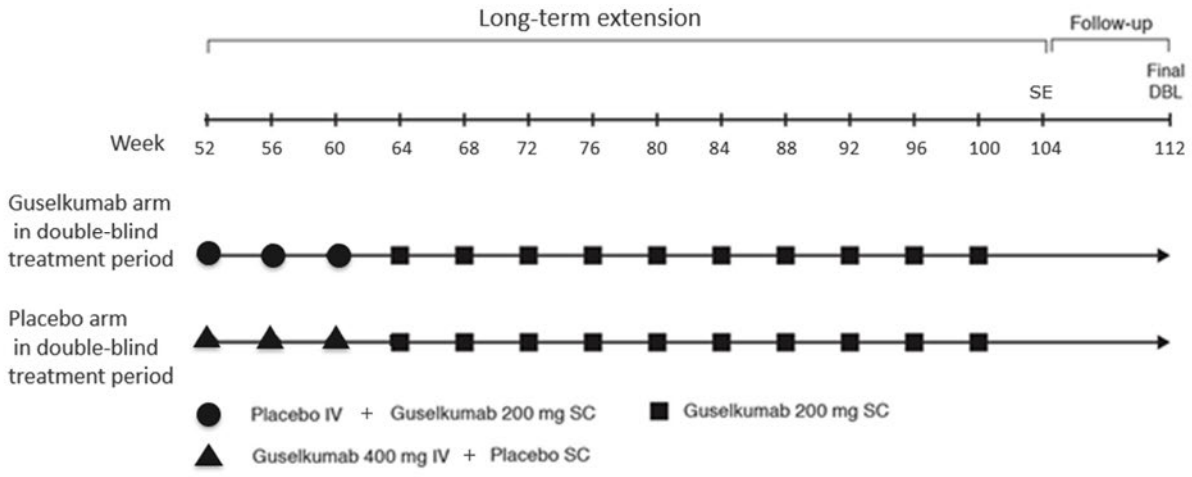
A target of 56 participants will be randomly assigned in this study with 28 participants planned per intervention group.

A diagram of the main study design is provided in the following schema.



Abbreviations: DBL=database lock; IV=intravenous; N=number of participants; PE=primary endpoint; R=randomization; SC=subcutaneous; SE=secondary endpoints

A diagram of the LTE period is provided in the following schema.



Abbreviations: DBL=database lock; IV=intravenous; SC=subcutaneous; SE=secondary endpoints

2. STATISTICAL HYPOTHESES

This PoC study aims at detecting an early efficacy signal of guselkumab in participants with SSc, compared to placebo, to support further clinical development.

The primary efficacy endpoint is defined as the change from baseline in mRSS at Week 24.

- The null hypothesis is that the treatment difference between guselkumab and placebo for the primary endpoint is =0.
- The alternative hypothesis is that the treatment difference between guselkumab and placebo for the primary endpoint is ≠0.

Due to the PoC nature of this study, a 2-sided alpha of 0.2 is allocated for testing this hypothesis. The objective will be considered met if the calculated p-value is less than 0.2.

3. SAMPLE SIZE DETERMINATION

There is lack of positive placebo-controlled trials in dcSSc.

Taking into account the following 4 Phase 2/3 placebo-controlled trials (see Table 2) in SSc with mRSS results reported in the ClinicalTrials.gov registry, a reasonable estimate of the standard deviation (SD) to be used in the sample size calculation of this study has been taken as 8.0.

Table 2: Standard deviations of mRSS in Referenced Phase 2/3 Placebo-controlled Trials in SSc

Trial	Treatment group		Placebo group	
	N	Estimated SD	N	Estimated SD
A Double-Blind, Randomised, Placebo-controlled Trial Evaluating Efficacy and Safety of Oral Nintedanib Treatment for at Least 52 Weeks in Patients With Systemic Sclerosis Associated Interstitial Lung Disease (SSc-ILD) (NCT02597933)	288	4.582	288	4.412
A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Assess the Efficacy and Safety of Tocilizumab Versus Placebo in Patients With Systemic Sclerosis (NCT02453256)	104	8.169	106	8.142
A Phase 2/3, Multicenter, Randomized, Double-blind, Placebo-controlled Study To Assess The Efficacy And Safety Of Tocilizumab Versus Placebo In Patients With Systemic Sclerosis (NCT01532869)	43	8.266	41	8.933

Trial	Treatment group		Placebo group	
	N	Estimated SD	N	Estimated SD
A Phase 2 Study to Evaluate Subcutaneous Abatacept vs. Placebo in Diffuse Cutaneous Systemic Sclerosis- a Double-blind, Placebo-controlled, Randomized Controlled Trial (NCT02161406)	44	7.562	44	7.562

Khanna et al (2006) revealed that a 95% confidence interval (CI) of [-4.7, -1.7] in mRSS change from baseline to 6 months was considered a clinically meaningful change. The lower bound of this 95% CI, -4.7, has been taken as the expected treatment difference in the sample size calculation of this study, assuming guselkumab can achieve a clinically meaningful change in most of guselkumab-treated participants.

Assuming an SD of 8 points, an 80% power at a 2-sided significance level of 0.20, with a 1:1 randomization ratio, to detect a treatment effect of -4.7 points, a total of at least 54 participants will be required.

Allowing for up to 2 participants randomized but without postbaseline efficacy assessments, approximately 56 (28 each) randomized participants are planned. In order to evaluate the interobserver and intraobserver variability of the mRSS to support Phase 3 planning, it is desirable for each center to randomize 3 participants at least.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

For purposes of analysis, the following analysis sets are defined:

Table 3: Definition of Analysis Sets

Analysis Sets	Description
Enrolled	All participants who sign the ICF.
Randomized	All participants who were randomized in the study.
Full Analysis Set (FAS)	All randomized participants who received at least 1 dose (complete or partial) of study intervention. Note: Efficacy and participant information analyses will be based on the randomized intervention groups, regardless of the intervention they actually received.
Full Analysis Set for LTE Period	All randomized participants who received at least 1 dose (complete or partial) of study intervention during LTE period.
Safety	All randomized participants who received at least 1 dose (complete or partial) of study intervention. Note: Safety will be analyzed based on the intervention they received, regardless of the intervention groups to which they are assigned.
Safety Analysis Set for LTE Period	All participants who received at least 1 dose (complete or partial) of study intervention during LTE period.
PK	All participants who received at least 1 complete administration of guselkumab and have at least 1 observed post dose PK data.
Immunogenicity	All participants who received at least 1 administration of guselkumab and have at least 1 observed post dose immune response data.
PK/PD Analysis Set for mRSS	All participants who were randomized to guselkumab and received at least 1 complete administration of guselkumab during main study period, have evaluable data of both PK and PD (mRSS) at the same visit which includes Week 0, and at least one postbaseline timepoint at Week 12, or Week 24 or Week 52 visit.
PK/PD Analysis Set for ACR CRISS	All participants who were randomized to guselkumab and received at least 1 complete administration of guselkumab during main study period, have evaluable data of both PK and PD (ACR CRISS) at the same visit of at least one postbaseline timepoint at Week 12, or Week 24 or Week 52 visit.

5. STATISTICAL ANALYSES

5.1. General Considerations

Descriptive summary statistics, such as n, mean, standard deviation (SD), median, inter quantile range, minimum, and maximum for continuous variables, and counts and percentages for discrete variables will be used to summarize data as appropriate unless otherwise specified.

5.1.1. Level of Significance

As a PoC study, a 2-sided alpha of 0.2 is allocated for all hypothesis testing throughout the study. As the reference of the internal GO/NO-GO decision making, 95% CIs as well as 80% CIs will be provided for primary endpoint analysis and secondary endpoint analysis.

5.1.2. Missing Data

Unless otherwise specified, a missing-at-random (MAR) missing data mechanism is assumed for all efficacy endpoints and there will be no imputation of missing data prior to the analysis.

5.1.3. Intervention Group

The population will be the FAS defined in Section 4 for all the efficacy analysis, and the assigned treatment group for each participate will be used regardless of the actual treatment received.

All the safety analyses will be performed on safety analysis set, and by the actual treatment participants received regardless of the assigned treatment.

5.1.4. Visit Windows

Unless otherwise specified, nominal visits will be used for all by visit analyses. The study visits scheduled after randomization should occur at the time delineated in the Schedule of Activities as described in the protocol.

5.2. Participant Dispositions

Screened participants and the number of screen failures will be summarized overall in all enrolled participants.

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

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

Listings of participants will be provided for the following categories:

- Participants who discontinued study intervention
- Participants who terminated study prematurely

5.3. Primary Endpoint Analysis

The primary endpoint is the change from baseline in mRSS at Week 24. This section outlines the definitions and analysis of this primary endpoint.

5.3.1. Definition of Endpoint

The mRSS is a validated physical examination method for estimating skin induration. It correlates with biopsy measures of skin thickness and reflects prognosis and visceral involvement, especially in early disease. It is scored on 0 (normal) to 3 (severe induration) ordinal scales over 17 body areas, with a maximum score of 51 and is used to categorize severity of SSc (Khanna 2007).

This assessment should be performed by a physician who is experienced and trained in skin scoring. To prevent interobserver variability, the same physician must perform skin scoring for the same participant throughout the entire study.

5.3.2. Estimands

1. Primary Estimand (Composite Estimand)

The primary estimand is defined as the following components:

Primary Trial Objective: to evaluate the efficacy of guselkumab in participants with SSc

Estimand Scientific Question of Interest: what is the efficacy, as is measured by change from baseline in mRSS at Week 24 for participants who are assigned to guselkumab versus placebo

Study Intervention:

- Experimental: guselkumab
- Control: placebo

Population: participants between 18 to 75 years (inclusive) with diagnosis of SSc according to ACR and EULAR 2013 criteria with a disease duration of ≤36 months (defined as time from first non-Raynaud phenomenon manifestation), and mRSS ≥10 to ≤22 units

Variable: change from baseline in mRSS at Week 24. Participants with an intercurrent event before Week 24 will have a zero change from baseline in the mRSS since the intercurrent event occurs, regardless of the observed data.

Summary measure (Population-level summary): difference in least-square means (LSMs) of change from baseline in mRSS at Week 24 between guselkumab vs placebo

Intercurrent events and their corresponding strategies:

Intercurrent Events	Strategy for Addressing Intercurrent Events and Its Description
A prohibited change in SSc medication (for details, see Appendix 11)	**Composite Strategy:** A participant with this intercurrent event is considered to have a zero change at and after the event occurred.
Discontinuation of study intervention due to lack of efficacy or an AE of Systemic Sclerosis	**Composite Strategy:** Same as above

2. Supplementary Estimand (Hypothetical Estimand)

This supplementary estimand will be used to complement the primary estimand. The components of the supplementary estimand are the same as those for the primary estimand with the exception of definitions in Variable and Intercurrent event strategies (**hypothetical strategy**), which are described as follows:

Variable: change from baseline in mRSS at Week 24. Participants with an intercurrent event before Week 24 will have their mRSS set to missing since the intercurrent event occurs.

Intercurrent events and their corresponding strategies:

Intercurrent Events	Strategy for Addressing Intercurrent Events and Its Description
A prohibited change in SSc medication (for details, see Appendix 11)	**Hypothetical Strategy:** A participant with this intercurrent event is considered to have missing data at and after the event occurred.
Discontinuation of study intervention due to lack of efficacy or an AE of Systemic Sclerosis	**Hypothetical Strategy:** Same as above

3. Supplementary Estimand (Treatment Policy Estimand)

In this estimand, the components are almost the same as those for the primary estimand with the exception of definitions in Variable and Intercurrent event strategies (treatment policy strategy), which are described as follows:

Variable: change from baseline in mRSS at Week 24

Intercurrent events and their corresponding strategies:

Intercurrent Events	Strategy for Addressing Intercurrent Events and Its Description
A prohibited change in SSc medication (for details, see Appendix 11)	**Treatment Policy Strategy:** The intercurrent event does not affect the outcome, and the data collected at and after the event will be used for analysis.
Discontinuation of study intervention due to lack of efficacy or an AE of Systemic Sclerosis	**Treatment Policy Strategy:** Same as above

5.3.3. Analysis Methods

The primary analysis will use a Mixed-Effect Model Repeated Measure (MMRM) model. Under the assumption of MAR, the missing data will be accounted for through correlation of repeated measures in the model.

The MMRM model includes treatment group, baseline mRSS, stratification factors (presence of ILD [yes, no] and baseline anti-topoisomerase I antibody status [positive, negative]), visit, treatment group by visit interaction, and baseline mRSS by visit interaction as fixed effects. Stratification factor of baseline mRSS is not included as baseline mRSS is already addressed in the model as a continuous covariate. Treatment effects will be estimated based on least-square means (LSMs) of the differences. The p-values for the LSM differences along with the 2-sided 80% and 95% CIs will be presented. The degree of freedom will be calculated as the error degrees of freedom for the model minus the number of extra random effects estimated.

An unstructured covariance matrix for repeated measures within a participant will be used. 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. If the normality assumption is in question, an appropriate transformation will be used.

5.3.4. Supplementary Analyses

5.3.4.1. Supplementary Analysis 1 (Hypothetical Estimand)

The analysis will be performed using the Hypothetical Estimand. The treatment comparison will use the similar MMRM model as in the primary analysis. LSMs, together with the 2-sided 80% CIs for the LSMs and p-value will be calculated based on the MMRM.

5.3.4.2. Supplementary Analysis 2 (Treatment Policy Estimand)

The analysis will be performed using the Treatment Policy Estimand. The treatment comparison will use the similar MMRM model as in the primary analysis. LSMs, together with the 2-sided 80% CIs for the LSMs and p-value will be calculated based on the MMRM.

5.3.5. Sensitivity Analyses

If at least one data through Week 24 is missing and p-value is <0.2 with the primary analysis, a sensitivity analysis for primary estimand (Composite Estimand) will be performed, using multiple imputation.

Step #1: Utilize a Markov Chain Monte Carlo (MCMC) approach to impute intermediate (nonmonotone) missing data and convert the dataset into a monotone missing data pattern. One hundred monotone datasets will be imputed using the seed 12345, by treatment group.

Step #2: After imputing nonmonotone data, the remaining missing value will be imputed using the PROC MI monotone regression procedure. This approach imputes missing data in a sequential manner, for each time point, a regression model based on all available data at this timepoint (either observed or imputed) is fitted and used to impute values still missing at this timepoint. The ancillary variable includes in the model will be treatment group, the stratification factors, and the mRSS scores by visit. The seed is 54321. In this way, one thousand complete datasets are generated.

Step #3: Analyze each of the 100 imputed datasets using the same model specified for the primary analysis and the results will be combined using the PROC MIANALYZE procedure.

The tipping point analysis will be implemented to assess the robustness of conclusions. A constant delta is added or subtracted to the imputed data in the direction of lack of efficacy for guselkumab and in the direction of improvement for placebo. A range of evenly spaced deltas will be used to adjust imputed values in the guselkumab and placebo group independently in order to produce a grid of delta adjustments to missing values in both groups. The size of the deltas will be chosen pragmatically, so that adjustments that lead to p-value >= 0.2 encompassed in the grid. In particular, assuming MAR on placebo, ie, no delta adjustment, an adjustment on the guselkumab arm that just tips the analysis into p-value >= 0.2 will be included.

For each pair of delta adjustments, the point estimate of treatment effect, and corresponding p-value combined from results from 1000 adjusted datasets through the PROC MIANALYZE procedures will be presented.

Ukyo and Noma proposed the permutation-based inference methods for the analyses using MMRM that is considered to be more appropriate under small sample settings (Ukyo 2019). Their method will be applied to examine the robustness of the primary analysis result.

Step #1: Permutate the treatment group indicator of the mRSS dataset. The seed is defined as 127.

Step #2: Calculate the *t*-statistics ($T^{(b)}\left(\beta_1^{null}\right)$) for the coefficient ($\beta_1$) that indicates treatment effect by the MMRM used for the primary analysis using the mRSS dataset prepared in Step #1.

Step #3: Repeat Step #1 and #2 for B times (B=1000 in this analysis) to obtain the *t*-statistics. Note that the *t*-statistics obtained in this procedure follow the empirical null distribution estimated by the actual dataset. Therefore, statistical test and inference are considered valid even if the sample size is small and longitudinal dataset is incomplete.

The p-value will be calculated as the following equation, where $T(\beta_1)$ is the *t*-statistics obtained by the original mRSS dataset:

$$p = \frac{1}{B+1}\left\{1 + \sum_{b=1}^{B} I\left\{\left|T^{(b)}(\beta_1^{null})\right| > |T(\beta_1)|\right\}\right\}$$

After calculating the 10% ($T_{0.1}(\beta_1^{null})$) and 90% ($T_{0.9}(\beta_1^{null})$) quantiles of the *t*-statistics obtained in Step #3, the 2-sided 80% CIs for β_1 will be β where the following equation holds:

$$T_{0.1}(\beta_1^{null}) \leq \frac{\hat{\beta}_1 - \beta}{\sqrt{Var(\hat{\beta}_1)}} \leq T_{0.9}(\beta_1^{null})$$

5.3.6. Subgroup Analysis

Descriptive statistics of the primary efficacy endpoint will be presented by the subgroups as follows:

- Gender: Male, Female
- Age: <65 years, ≥65 years
- Weight: <=median, >median
- Corticosteroid use at baseline: yes, no
- Presence of ILD: Yes, No
- Baseline mRSS: Low [≥10 to ≤15], High [≥16 to ≤22]
- Baseline anti-topoisomerase I antibody status: Positive, Negative
- Disease duration at baseline: 0-18 months, >18-36 months

5.4. Secondary Endpoints Analysis

5.4.1. Multiplicity Adjustment

No adjustments for multiple comparisons will be made across all the primary and secondary efficacy endpoints. Nominal p-values will be presented for efficacy analyses.

5.4.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

- Change from baseline in mRSS at Week 52
- The proportion of participants who experience worsening of mRSS at Week 24 and Week 52
- The proportion of participants achieving a score of 0.6 in American College of Rheumatology Combined Response Index in dcSSc (ACR CRISS) at Week 24 and Week 52

- Change from baseline in forced vital capacity (FVC) and percent predicted FVC at Week 24 and Week 52
- Change from baseline in the measured absolute diffusing capacity of the lung for carbon monoxide (DLCO) and the derived percent predicted DLCO at Week 24 and Week 52
- Change from baseline in digital ulcer counts at Week 24 and Week 52
- Change from baseline in Health Assessment Questionnaire-Disability Index (HAQ-DI) score at Week 24 and Week 52

5.4.2.1. Definition of Endpoints

Worsening of mRSS

An increase from baseline ≥5 points and ≥20% in mRSS.

American College of Rheumatology Combined Response Index in dcSSC (ACR CRISS)

ACR CRISS is a composite response index for clinical trials in early dcSSc. Application of ACR CRISS algorithm in a randomized clinical trial is a 2-step process.

Step 1: if a patient develops any of the following conditions, they will be assigned a probability of improving equal to 0.0. Investigators will evaluate if the participants have the criterion.

- new scleroderma renal crisis
- decline in percent predicted FVC ≥15% relative to baseline confirmed by a second FVC within one month, HRCT to confirm ILD (if previous HRCT did not show ILD) and FVC <80% of predicted attributable to systemic sclerosis
- new onset of left ventricular failure (defined as left ventricular ejection fraction ≤45%) requiring treatment, attributable to systemic sclerosis
- new onset of pulmonary arterial hypertension on right-sided heart catheterization requiring treatment, attributable to systemic sclerosis

Step 2: for the remaining patients, compute the predicted probability of improving for each participant using the following equation (equation to derive predicted probabilities from a logistic regression model):

$$\frac{exp\ [-5.54 - 0.81 * \Delta_{MRSS} + 0.21 * \Delta_{FVC\%} - 0.40 * \Delta_{Pt-glob} - 0.44 * \Delta_{MD-glob} - 3.41 * \Delta_{HAQ-DI}]}{1 + exp\ [-5.54 - 0.81 * \Delta_{MRSS} + 0.21 * \Delta_{FVC\%} - 0.40 * \Delta_{Pt-glob} - 0.44 * \Delta_{MD-glob} - 3.41 * \Delta_{HAQ-DI}]}$$

where, Δ_{MRSS} indicates the change in mRSS from baseline to follow-up, $\Delta_{FVC\%}$ denotes the change in percent predicted FVC from baseline to follow-up, $\Delta_{Pt-glob}$ indicates the change in patient global assessment, $\Delta_{MD-glob}$ denotes the change in physician global assessment, and Δ_{HAQ-DI} is the change in HAQ-DI. All changes are absolute change.

Participants for which the predicted probability is greater or equal to 0.60 are considered improved while participants for which the predicted probability is below 0.60 are considered non-improved.

Forced Vital Capacity (FVC)

FVC and % predicted FVC are measurements of lung function and will be performed locally.

Diffusing Capacity of the Lung for Carbon Monoxide (DLCO)

DLCO is the measurement of lung function and will be reported locally.

% predicted DLCO (hemoglobin-corrected) is only collected at screening, while for other visits, it will be derived based on following steps:

1) Compute Body Surface Area (BSA) per formula as below:

$$BSA\ (m2)\ =\ Weight\ (kg)^{\ 0.425}\ \times\ Height\ (cm)^{\ 0.725}\ \times\ 0.007184$$

The Weight of the same visit should be used, if Weight of the visit is missing then the value closest to date of DLCO measurement will be used.

2) Compute the predicted DLCO (ml/min/mm Hg) per formula as below:

Male:

$$\text{Predicted DLCO (ml/min/mm Hg)}\ =\ 15.5\ \times\ \text{BSA}\ -\ 0.238\ \times\ \text{age}\ +\ 6.8$$

Female:

$$\text{Predicted DLCO (ml/min/mm Hg)}\ =\ 15.5\ \times\ \text{BSA}\ -\ 0.117\ \times\ \text{age}\ +\ 0.5$$

3) Compute the predicted DLCO (Hgb-corrected) (ml/min/mm Hg) per formula as below:

Male:

$$\text{DLCO (ml/min/mm Hg)}\ \times\ ((10.22 + \text{Hgb}\ )/(1.7\ \times\ \text{Hgb}))$$

Female:

$$\text{DLCO (ml/min/mm Hg)}\ \times\ ((9.38 + \text{Hgb}\ )/(1.7\ \times\ \text{Hgb}))$$

The Hgb (g/dL) of the same visit from central lab will be used.

4) Compute the % predicted DLCO (Hgb-corrected) per formula as below:

$$\%\ \text{predicted DLCO (Hgb corrected)} = \frac{\text{Predicted DLCO (Hgb corrected) (ml/min/mm Hg)}}{\text{Predicted DLCO (ml/min/mm Hg)}} \times 100$$

Health Assessment Questionnaire -Disability Index (HAQ-DI)

The Stanford HAQ is a brief self-report questionnaire assessing physical function pertaining to activities of daily living across 8 domains: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities. The Stanford HAQ translated into Japanese with culturally

appropriate modifications of the arising, eating, and reach category questions will be used in the study. (Matsuda 2003).

5.4.2.2. Estimands

The following describes the attributes of the estimands for the secondary efficacy endpoints.

Study Intervention:

- Experimental: guselkumab
- Control: placebo

Population: participants between 18 to 75 years (inclusive) with diagnosis of SSc according to ACR and EULAR 2013 criteria with a disease duration of ≤36 months (defined as time from first non-Raynaud phenomenon manifestation), and mRSS ≥10 to ≤22 units

Variables and Population-level Summary:

Variable (Endpoint)	Population-level Summary
Change from baseline in mRSS at Week 52	Difference in LSMs of change from baseline in mRSS at Week 52 between guselkumab vs placebo
The proportion of participants who experience worsening of mRSS at Week 24	Odds ratio of proportion of participants who experience worsening of mRSS at Week 24 between guselkumab vs placebo
The proportion of participants who experience worsening of mRSS at Week 52	Odds ratio of proportion of participants who experience worsening of mRSS at Week 52 between guselkumab vs placebo
The proportion of participants achieving a score of 0.6 in ACR CRISS at Week 24	Odds ratio of proportion of participants who achieve a score of 0.6 in ACR CRISS at Week 24 between guselkumab vs placebo
The proportion of participants achieving a score of 0.6 in ACR CRISS at Week 52	Odds ratio of proportion of participants who achieve a score of 0.6 in ACR CRISS at Week 52 between guselkumab vs placebo
Change from baseline in FVC and percent predicted FVC at Week 24	Difference in LSMs of change from baseline in FVC and % predicted FVC at Week 24 between guselkumab vs placebo
Change from baseline in FVC and percent predicted FVC at Week 52	Difference in LSMs of change from baseline in FVC and % predicted FVC at Week 52 between guselkumab vs placebo
Change from baseline in the measured absolute DLCO and the derived percent predicted DLCO at Week 24	Difference in LSMs of change from baseline in the measured absolute DLCO and the derived % predicted DLCO at Week 24 between guselkumab vs placebo

Variable (Endpoint)	Population-level Summary
Change from baseline in the measured absolute DLCO and the derived percent predicted DLCO at Week 52	Difference in LSMs of change from baseline in the measured absolute DLCO and the derived % predicted DLCO at Week 52 between guselkumab vs placebo
Change from baseline in digital ulcer counts at Week 24	Difference in LSMs of change from baseline in digital ulcer counts at Week 24 between guselkumab vs placebo
Change from baseline in digital ulcer counts at Week 52	Difference in LSMs of change from baseline in digital ulcer counts at Week 52 between guselkumab vs placebo
Change from baseline in HAQ-DI score at Week 24	Difference in LSMs of change from baseline in HAQ-DI score at Week 24 between guselkumab vs placebo
Change from baseline in HAQ-DI score at Week 52	Difference in LSMs of change from baseline in HAQ-DI score at Week 52 between guselkumab vs placebo

Intercurrent events and their corresponding strategies:

The intercurrent events for all key secondary endpoints at Week 24 are the same as those in the primary estimand, as follows:

1. A prohibited change in SSc medication (for details, see Appendix 11)
2. Discontinuation of study intervention due to lack of efficacy or an AE of Systemic Sclerosis

As rescue medication is available for participants after Week 24, an additional intercurrent event is defined for secondary efficacy endpoints at Week 52.

3. Initiation of rescue medication after Week 24

Intercurrent events 1, 2 and 3 are all handled by the **composite strategy** to assume that the occur of intercurrent events indicates an unfavorable outcome. For continuous endpoints, the change from baseline will be zero; and for binary endpoints, the participants are considered as not achieving an improvement (ACR CRISS score >= 0.6) or experiencing worsening (an increase from baseline ≥5 points and ≥20% in mRSS).

5.4.2.3. Analysis Methods

For continuous endpoint as change from baseline in mRSS at Week 52, an MMRM model with treatment group, baseline mRSS, stratification factors (presence of ILD [yes, no] and baseline anti-topoisomerase I antibody status [positive, negative]), visit, treatment group by visit interaction, and baseline mRSS by visit interaction as fixed effects will be applied. The estimates of the treatment difference between guselkumab and placebo will be provided by the difference in the least squares means (LSMs), as well as the 2-sided 80% and 95% CIs for the differences and p-values.

For continuous endpoints other than change from baseline in mRSS at Week 52, an MMRM model with treatment group, baseline value of the endpoint, stratification factors (presence of ILD [yes, no], baseline mRSS (low [≥10 to ≤15], high [≥16 to ≤22]) and baseline anti-topoisomerase I antibody status [positive, negative]), visit, treatment group by visit interaction, and baseline value of the endpoint by visit interaction as fixed effects will be applied. The estimates of the treatment difference between guselkumab and placebo will be provided by the difference in the least squares means (LSMs), as well as the 2-sided 80% and 95% CIs for the differences and p-values. The degree of freedom will be calculated as the error degrees of freedom for the model minus the number of extra random effects estimated.

An unstructured covariance matrix for repeated measures within a participant will be used. 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. If the normality assumption is in question, an appropriate transformation will be used.

For binary endpoints such as proportion of participants achieving an ACR CRISS score of 0.6, a logistic regression model that includes factors for treatment group and stratification factors (presence of ILD [yes, no], baseline mRSS (low [≥10 to ≤15] or high [≥16 to ≤22]) and baseline anti-topoisomerase I antibody status [positive, negative]) will be applied provide the odds ratios, p-values, as well as the 2-sided 80% and 95% CIs for comparisons between guselkumab and placebo.

5.4.2.4. Supplementary Analysis for Change from Baseline in MRSS at Week 52

The analysis is performed on the supplementary estimand of change from baseline in mRSS at Week 52 as in Section 5.4.2.2. The similar estimand as described in Section 5.4.2.2 is built except that intercurrent event “initiation of rescue therapy medication after Week 24” will be handled by treatment policy. An MMRM model similar as described in Section 5.4.2.3 is built and 80% CIs will be calculated.

5.5. Tertiary/Exploratory Efficacy Endpoints Analysis

The exploratory efficacy endpoints are listed as follows:

- Time to disease worsening: increase from baseline ≥5 and ≥20% in mRSS or decrease >10% from baseline in FVC %
- The proportion of participants who experience FVC decline compared with baseline at Week 24 and Week 52
- The proportion of participants who experience worsening of FVC at Week 24 and Week 52
- Change from baseline in fibrotic change assessed with high-resolution computed tomography (HRCT) at Week 24 and Week 52 in participants with baseline fibrosis/fibrotic changes, including:

1) Change from baseline in quantitative lung fibrosis (QLF) score by whole lung and each lobe at Week 24 and Week 52 in participants with baseline QLF score of whole lung >0.
2) Change from baseline in quantitative interstitial lung disease (QILD) score by whole lung and each lobe at Week 24 and Week 52 in participants with baseline QILD score of whole lung >0.
3) Change from baseline in the absolute volume of consolidation, emphysema-like, ground glass, honeycombing, normal lung, reticulation texture pattern by whole lung and each lobe at Week 52.

- The proportion of participants who experience worsening of fibrotic change assessed with HRCT at Week 24 and Week 52, including:
  1) The proportion of participants who experience worsening of fibrotic change by whole lung and each lobe assessed with QLF of HRCT at Week 24 and Week 52, in participants with baseline QLF score of whole lung >0.
  2) The proportion of participants who experience worsening of fibrotic change by whole lung and each lobe assessed with QILD of HRCT at Week 24 and Week 52, in participants with baseline QILD score of whole lung >0.
- Change from baseline in frequency scale for the symptoms of gastroesophageal reflux disease (FSSG) score at Week 24 and Week 52
- Assessment of nailfold capillaries at Week 24 and Week 52, including:
  1) Change from baseline in average total number of capillaries/ length of nailfold; average total number of capillaries/ length of nailfold = sum (total number of capillaries of each finger) / sum (length of nailfold of each finger)
  2) Change from baseline in average number of enlarged capillaries (20-50µm)/ length of nailfold; average number of enlarged capillaries (20-50µm)/ length of nailfold = sum (Number of enlarged capillaries (20-50µm) of each finger) / sum (length of nailfold of each finger)
  3) Change from baseline in average Number of giant capillaries (>50µm)/ length of nailfold; average Number of giant capillaries (>50µm)/ length of nailfold = sum (Number of giant capillaries (>50µm) of each finger) / sum (length of nailfold of each finger)
  4) Change from baseline in average total number of abnormal formed capillaries/ length of nailfold; average Total number of abnormal formed capillaries/ length of nailfold = sum (Total number of abnormal formed capillaries of each finger) / sum (length of nailfold of each finger)
  5) Change from baseline in average number of nailfold bleeding/ length of nailfold; average Number of nailfold bleeding/ length of nailfold = sum (Number of nailfold bleeding of each finger) / sum (length of nailfold of each finger)
  6) Change from baseline in capillary pattern (shift table, based on worst capillary pattern)
  7) Change from baseline in average Blood velocity (mm/s)

- Change from baseline in patient global assessment (PGA) at Week 24 and Week 52
- Change from baseline in physician global assessment (PhGA) at Week 24 and Week 52

And LTE efficacy endpoints:

- Change from baseline and Week 52 in mRSS at Week 76 and Week 104
- The proportion of participants who experience worsening of mRSS at Week 76 and Week 104
- The proportion of participants achieving a score of 0.6 in ACR CRISS at Week 76 and Week 104
- Change from baseline in FVC and percent predicted FVC at Week 76 and Week 104
- The proportion of participants who experience worsening of FVC at Week 76 and Week 104
- Change from baseline in the measured absolute DLCO and the derived percent predicted DLCO at Week 76 and Week 104
- Change from baseline in digital ulcer counts at Week 76 and Week 104
- Change from baseline in fibrotic change assessed with HRCT at Week 76 and Week 104 in participants with baseline fibrosis/fibrotic changes, including:
    1) Change from baseline in QLF score by whole lung and each lobe at Week 76 and Week 104 in participants with baseline QLF score of whole lung >0.
    2) Change from baseline in QILD score by whole lung and each lobe at Week 76 and Week 104 in participants with baseline QILD score of whole lung >0.
- The proportion of participants who experience worsening of fibrotic change assessed with HRCT at Week 76 and Week 104, including:
    1) The proportion of participants who experience worsening of fibrotic change by whole lung and each lobe assessed with QLF of HRCT at Week 76 and Week 104, in participants with baseline QLF score of whole lung >0.
    2) The proportion of participants who experience worsening of fibrotic change by whole lung and each lobe assessed with QILD of HRCT at Week 76 and Week 104, in participants with baseline QILD score of whole lung >0.
- Change from baseline in Frequency Scale for the FSSG score at Week 76 and Week 104

5.5.1. Definition of Endpoints

Time to Disease Worsening

Time to disease worsening is calculated as the time from the date of randomization to the date of first occurrence of disease worsening, defined as: an increase from baseline ≥5 and ≥20% in mRSS or decrease >10% from baseline in FVC %.

FVC Decline Compared with Baseline

FVC decline compared with baseline is defined as a decrease in postbaseline visits compared with baseline in FVC.

Worsening of FVC

Worsening of FVC is defined as an absolute change of % predicted FVC <-10.

High-resolution Computed Tomography (HRCT) Assessment

Lung involvement is one of the most important cause of morbidity and mortality in patients with SSc. Because of recent advancement of computer-based scoring, analysis of CT images is drawing attention as a method to assess fibrotic changes in ILD. Texture-based analysis and computer vision-based approaches can be applied to imaging data, and used to evaluate the extent of lung fibrosis in chest CT images. The main assessed parameters QLF, which included fibrotic reticulation patterns only; quantitative honeycomb (QHC); and quantitative ground glass (QGG). QILD scores represented the total ILD pattern and consisted of the sum of all three scores (ie, QLF+QGG+QHC). The score will be reported by central image reading.

Worsening of Fibrotic Change

Worsening of fibrotic change will be assessed by whole lung and each lobe with different parameters QLF and QILD separately, which is defined as increases of 2% or more compared to baseline score.

Symptoms of Gastroesophageal Reflux Disease (FSSG)

The FSSG is a succinct questionnaire to evaluate esophageal symptoms of gastroesophageal reflux disease (GERD) which is validated based on endoscopic esophagitis. The original version of FSSG consists of the most prevalent 7 acid-reflux related and 5 dysmotility-related symptoms of GERD with higher scores being more indicative of underlying GERD. Each score was determined as follows: 0 = never, 1 = occasionally, 2 = sometimes, 3 = often, and 4 = always. A negative change from baseline indicates improvement.

Nailfold Capillaroscopy

Nailfold capillaroscopy is a non-invasive method to visualize the nailfold capillaries and assess microvascular morphology. Nailfold capillary abnormalities are evaluated by using nailfold capillaroscopy. The typical changes of the nailfold capillaries during SSc include the total number of capillaries, capillary dimension, capillary morphology, hemorrhages, blood flow velocity in capillaries, and length of nailfold.

Patient Global Assessment (PGA)

Participant's overall health will be assessed using the patient-reported outcome instrument of PGA. The PGA would rate participant's overall health related to SSc in the past 7 days, using a 10 cm visual analogue scale where 0 = excellent and 10 = extremely poor. A negative change from baseline indicates improvement.

Physician Global Assessment (PhGA)

PhGA rated participant's overall health related to SSc in the past 7 days, using a 10 cm visual analogue scale where 0 = excellent and 10 = extremely poor. A negative change from baseline indicates improvement.

5.5.2. Analysis Methods

The analyses of exploratory efficacy endpoints will be based on the FAS.

Descriptive statistics, such as n, mean, SD, median, IQ range, minimum and maximum for continuous variables and counts and percentages for discrete variables will be used to summarize data when applicable. The analysis will be based on the observed data; therefore, no imputation of missing data will be performed.

All statistical testing will be performed between treatment groups at the 2-sided 0.2 significance level.

Continuous efficacy endpoints collected at multiple post baseline time points (ie, the endpoints of change from baseline) will be compared using a MMRM model. The model will include all available data through Week 24 (for endpoints at Week 24) or through Week 52 (for endpoints at Week 52).

The explanatory variables of the MMRM model will include treatment group, baseline value of the endpoint, stratification factors (presence of ILD, baseline mRSS and baseline anti-topoisomerase I antibody status), visit, treatment group by visit interaction, and baseline value of the endpoint by visit interaction as fixed effects. The estimates of the treatment difference between treatment groups will be provided by the difference in LSMs. The 80% CI for the differences in LSMs will be calculated based on the MMRM.

The analysis of change from baseline in fibrotic change assessed with HRCT at Week 24 and nailfold capillaries change at Week 24 will be performed using ANCOVA (analysis of covariance) model with baseline value of the endpoint as covariate, and stratification factors (presence of ILD, baseline mRSS and baseline anti-topoisomerase I antibody status) and treatment group as factors.

A logistic regression model will be used to compare the proportion of participants achieving selected endpoints (eg, worsening of % predicted FVC) between guselkumab and placebo. The analyses will include treatment group, stratification factors (presence of ILD, baseline mRSS and baseline anti-topoisomerase I antibody status) as factors. The same intercurrent events and composite strategy defined in Section 5.4.2.2 will be applied to following endpoints as well:

- The proportion of participants who experience FVC decline compared with baseline through Week 52
- The proportion of participants who experience worsening of FVC through Week 52

A summary of baseline values and change from baseline at Week 24 and Week 52 will be presented for nailfold capillaroscopy.

As to time to event endpoint, ie, time to disease worsening, it will be summarized by Kaplan-Meier method, with the median, 25th, and 75th percentiles and 95% CI for the quartiles. The comparison between guselkumab and placebo will use a Cox proportional hazards model adjusting for the stratification factors at randomization. An 80% CI for the hazard ratio (HR) will be produced. The

same intercurrent events and handling strategy defined in Section 5.4.2.2 will be applied to this endpoint as well, the detailed definition of censoring and event is described as below:

No.	Situation	Date of Event or Censoring	Outcome
1	Disease worsening: an increase from baseline ≥5 and ≥20% in mRSS or decrease >10% from baseline in FVC %	First occurrence of disease worsening	Event
2	A prohibited change in SSc medication on or before Week 52	First occurrence of prohibited change	Event
3	Discontinuation of study intervention due to lack of efficacy or an AE of SSc before Week 52	Date of discontinuation of study intervention	Event
4	Initiation of rescue medication after Week 24 and through Week 52	Date of initiation of rescue medication	Event
5	Other	Date of last mRSS or FVC assessment through Week 52	Censored

Note: if participants had more than one items of 1,2,3,4 occurred, the date of first occurrence will be considered as event date.

Unless otherwise specified, LTE efficacy endpoints will only be summarized using descriptive statistics.

5.6. Other Efficacy Analyses

Descriptive statistics of primary, secondary and exploratory efficacy endpoints will be provided by study intervention at each scheduled visit through week 24/52/76/104 if applicable. These endpoints will be analyzed by the method described in Section 5.3.3, 5.4.2.3 and 5.5.2 respectively.

LS mean and standard error of LS mean, together with the LS mean difference between treatment groups will be calculated for the continuous endpoints, while for the binary endpoints, the odds ratio will be calculated.

Listings of primary, secondary and exploratory efficacy endpoints will be provided.

5.7. Safety Analyses

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

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

5.7.1. Extent of Exposure

The number and percentage of participants who receive study interventions will be summarized by visit.

Descriptive statistics for duration of study interventions will be summarized. Study intervention duration (weeks) is defined as ((date of last dose of study intervention – date of first dose of study intervention) +1)/7.

For IV doses, the number (%) of participants with a study intervention modification during the infusion will be summarized by intervention group and visit.

Descriptive statistics will be presented for the following parameters:

- Number of administrations
- Cumulative total dose

Intervention compliance will be summarized descriptively. See Appendix 7 for further details.

5.7.2. Adverse Events

The verbatim terms used in the CRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any AE occurring at or after the initial administration of study intervention is considered to be treatment emergent. If the event occurs on the day of the initial administration of study intervention, and either event time or time of administration are missing, then the event will be assumed to be treatment emergent. If the event date is recorded as partial or completely missing, then the event will be considered to be treatment emergent unless it is known to be prior to the first administration of study intervention based on partial onset date or resolution date. All reported treatment-emergent adverse events will be included in the analysis. For each adverse event, the number and percentage of participants who experience at least 1 occurrence of the given event will be summarized by intervention group.

Summary tables will be provided for treatment-emergent adverse events in the main study period and LTE period:

- AEs
- Serious AEs (SAEs)
- AEs leading to discontinuation of study intervention
- AEs by severity
- AEs by relationship to study intervention
- AEs leading to drug interruption

- Infusion reaction
- Injection site reaction
- Covid-19 related AEs and SAEs

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

- Had SAEs
- Had AEs leading to discontinuation of study intervention

Incidence of other treatment-emergent adverse events of special interest will be summarized. See Appendix 8 for list of adverse events in each category.

Deaths will be displayed by actual intervention received. Frequencies for the following parameters will be included in the summary table:

- Number of participants who died
- Cause of death
- Relationship to study intervention (yes/no)

A listing of participants who died will be provided.

5.7.3. Additional Safety Assessments

5.7.3.1. Clinical Laboratory Tests

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

Descriptive statistics will be presented for selected chemistry, hematology, and urinalysis (pH and specific gravity) laboratory tests at scheduled time points.

Change from baseline over time through Week 52 will be summarized for selected chemistry, hematology, and urinalysis (pH, and specific gravity) tests and displayed by intervention group. A box plot of change from baseline over time through Week 52 will be provided for the following laboratory tests: hematology: hemoglobin, platelets, total WBC, absolute lymphocytes, and absolute neutrophils; chemistry: ALT, AST, Alkaline phosphatase and total bilirubin.

Applicable lab results are to be graded based on National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 5.0. The laboratory tests not included in the NCI-CTCAE criteria will not be presented in the corresponding tables or listings.

Summary of maximum postbaseline NCI-CTCAE toxicity grade through Week 52 will be presented. Additionally, shift table from baseline laboratory value to the worst toxicity grade through Week 52 in chemistry and hematology tests with NCI-CTCAE will be summarized for the following lab parameters: hematology: hemoglobin, platelets, total WBC, absolute lymphocytes, and absolute neutrophils; chemistry: ALT, AST, Alkaline phosphatase and total bilirubin.

Proportion of participants with maximum postbaseline ALT and AST through Week 52 will be provided for the categories:

- >1 to ≤3 x ULN
- >3 to ≤5 x ULN
- >5 to ≤10 x ULN
- >10 to ≤20 x ULN
- >20 x ULN

Besides, proportion of participants with total bilirubin >2 x ULN and either AST or ALT ≥3 x ULN at the same timepoint will also be provided.

In addition, proportion of participants with maximum postbaseline alkaline phosphatase through Week 52 will be provided for the categories:

- ≤1 x ULN
- >1 to ≤1.5 x ULN
- >1.5 to ≤2.5 x ULN
- >2.5 to ≤5 x ULN
- >5 to ≤20 x ULN
- >20 x ULN

Above analysis will also be summarized through LTE period.

Listings of participants with any abnormal postbaseline laboratory values of NCI-CTCAE grade ≥2 or any abnormal postbaseline laboratory values of reference range will also be provided.

5.7.3.2. Vital Signs and Physical Examination Findings

Continuous vital sign parameters including temperature, respiratory rate, weight, pulse and blood pressure (systolic and diastolic) will be summarized at each assessment time point. Change from baseline will be summarized through Week 52 and through end of LTE period. Descriptive statistics will be presented.

Incidence of markedly abnormal vital signs during intervention, as defined in Table 4, will be summarized. A listing of participants with markedly abnormal vital signs will be presented, along with a listing of all vital sign measurements.

Table 4: Markedly Abnormal Vital Signs

Vital Sign	Criteria
Pulse	>120 bpm and with ≥15 bpm increase from baseline
	<50 bpm and with ≥15 bpm decrease from baseline
Systolic blood pressure	>180 mm Hg and with ≥20 mm Hg increase from baseline
	<90 mm Hg and with ≥20 mm Hg decrease from baseline
Diastolic blood pressure	>105 mm Hg and with ≥15 mm Hg increase from baseline
	<50 mm Hg and with ≥15 mm Hg decrease from baseline
Temperature	>38 ºC
	< 36 ºC
Respiratory rate	>24 breaths per minute
	<12 breaths per minute

5.7.3.3. Electrocardiogram

The interpretation of the ECGs as determined by a qualified physician (investigator or qualified designee) will be displayed by the number and percentage of participants meeting the normality criteria. The interpretation will be summarized over time.

Listings will be produced for all ECG data including unscheduled visit data. A listing of clinically relevant ECG abnormalities will also be provided.

5.8. Other Analyses

5.8.1. Pharmacokinetics

Blood samples for measuring serum guselkumab concentrations will be collected from all participants at the specified visits as indicated in the Schedule of Activities in protocol. PK analyses will be performed on the PK analysis set (Section 4).

Serum guselkumab concentrations summary and analysis will be based on the observed data; therefore, no imputation of missing data will be performed.

All PK data including actual sampling time will be listed. All serum guselkumab concentrations below the lowest quantifiable sample concentration of the assay (BQL) or missing data will be labeled as such in the concentration data listing or statistical analysis system dataset. All participants and samples excluded from the analysis will be clearly documented (eg, unknown or unreliable drug intake information).

Serum guselkumab concentrations over time from Week 0 through Week 52 and from Week 52 to Week 104 will be summarized in tables for guselkumab treatment group for each scheduled visit with arithmetic mean, SD, coefficient of variation (%CV), median, minimum, maximum, 25% quartile and 75% quartile. Serum guselkumab concentrations over time from Week 52 through Week 104 will be summarized by intervention group.

Serum guselkumab concentrations over time from Week 0 to Week 52 and from Week 52 to 104 will be also summarized in tables by the following subgroups:

- Baseline body weight category (≤ median / >median)
- Antibodies to guselkumab status by Subject ADA status who have at least one positive ADA sample or no positive ADA sample through Week 52/104.

PK concentration data from Week 0 to Week 52 and from Week 52 to 104 will be also displayed graphically by following figures.

- Median (IQR) of serum guselkumab concentration time profiles through Week 52 and from Week 52 to 104 (in linear and semi-log scales) with nominal sampling time (for analysis from Week 52 through Week 104, shown by intervention group).
- Median (IQR) of serum guselkumab concentration time profiles through Week 52 and from Week 52 to 104 (in linear and semi-log scales) with nominal sampling time by Subject ADA status who have at least one positive ADA sample or no positive ADA sample (for analysis from Week 52 to Week 104, shown by intervention group).

For each population including participants who performed self-study intervention administration at Week 84, 92 or 100, serum guselkumab concentrations at Week 88, 96 or 104, respectively, will be summarized in table by following group: guselkumab arm in DB period and placebo arm in DB period. Also, serum guselkumab concentration at Week 76 will be summarized for each population (ie, participants who performed self-study intervention administration at Week 84, 92 or 100 and whose serum concentrations were summarized at 88, 96 or 104).

For descriptive statistics of serum concentrations of guselkumab, the following data handling rules will be applied:

- Participants will be excluded from the PK analysis if their data do not allow for accurate assessment of the PK. In particular, all serum concentration summaries will exclude, from the time of occurrence, data collected for participants who 1) discontinue guselkumab, 2) skip an infusion or injection, 3) receive an incomplete infusion or injection, 4) receive an incorrect infusion or injection, 5) receive an additional infusion or injection, and/or 6) receive commercial guselkumab. Exclusion data from PK analysis due to such inadequate administration will be specified by PK analyst. In addition, PK samples taken outside the scheduled visit window (±7 days for all scheduled visits) will be excluded from the summaries. Post-infusion samples that occurred more than 24 hours after infusion will be also excluded.
- Serum concentrations of guselkumab data will be calculated based on the number of participants with observed data, including BQL, at each sampling time.
- When more than half (>50%) of the serum concentrations of guselkumab are BQL at each scheduled time point, mean, median, minimum and 25% quartile will be shown as ‘BQL’, and SD and %CV and 75% quartile will be shown as ‘NC’ (not calculated). Maximum observed value will be presented as maximum.
- When all serum concentration data are BQL at each scheduled time point, mean and median will be reported as ‘BQL’; SD and %CV are reported as ‘NC’ (not calculated); IQ range, maximum and minimum will be reported as ‘BQL’.

- When the number of serum concentrations data of guselkumab at each scheduled time point is less than or equal to 2, N, mean and median will be calculated, and SD, %CV, 25% quartile and 75% quartile will be shown as 'NC'. Minimum and maximum will be shown as 'NC' (quantifiable data=1) or will be reported as observed including BQL (quantifiable data =2).
- At the time point where no observation is obtained, 'NA (not applicable)' is reported.
- Data of samples with no information about the sampling date and time and/or the drug administration (time and dosage) will be excluded from descriptive statistics.
- Sampling time at Week 0 predose will be substituted with "0" in the figure. The data point at which mean of serum guselkumab concentration is BQL will be treated as lower limit of quantification (LLOQ) value (eg, 0.01 µg/mL) in graph.
- Serum guselkumab concentrations below the LLOQ will be imputed as zero in the summary statistics.

The elapsed time from last dosing of next PK sampling point will be described in listing.

5.8.2. Immunogenicity

Immunogenicity analyses will be performed on the Immunogenicity analysis set (Section 4). The anti-guselkumab antibodies summary and analysis will be based on the observed data; therefore, no imputation of missing data will be performed.

The incidence of positive antibodies, neutralizing antibody (Nabs, as appropriate) and the maximum titers of antibodies to guselkumab from Week 0 to Week 24/52/104 will be summarized by the intervention group.

The summary of following endpoints at the following visit will be presented by the intervention group and Subject ADA status who have at least one positive ADA sample or no positive ADA sample evaluated at the same visit:

- Change from baseline in mRSS at Week 24/52
- ACR CRISS responder (participants who achieve a score ACR CRISS>=0.6) at Week 24/52
- Injection-Site Reactions status through Week 52/104

If appropriate, these summary by category of titer level for positive participants will also be considered.

A listing of individual immune response including antibody status (negative/positive) for both sample level and participant level, Nab status, serum maximum titers, serum guselkumab concentration, Injection-Site Reactions status and change from baseline in mRSS at all sampling time points for all participants will be provided. In addition, self-study intervention administration status will also be included in the listing.

5.8.3. Pharmacodynamics

Serum will be collected from all participants to assess PD markers associated with guselkumab as well as markers related to SSc. Measurements may include but are not limited to serum levels of

IL 17A, IL-17F, IL-22, beta defensin-2 (BD-2), and SAA. The samples will be collected at Weeks 0, 4, 12, 16, 24, 28, and 48 and analyzed by immunoassays. Broader proteomic profiling (eg, via Olink method) for biomarker discovery may be performed. Additionally, serum samples should also be collected at the final visit from participants who discontinued study intervention or were withdrawn from the study.

Blood samples for the isolation of PBMCs will also be collected for subsequent immunophenotyping analyses by multiparameter flow cytometry or mass cytometry (CyTOF) analysis to measure immune cell populations before and during treatment. The samples will be collected at Weeks 0, 4, 24, and 48. Gene expression analysis may also be performed; this may include single cell RNA-seq profiling. Additionally, serum samples for isolation of PBMCs should also be collected at the final visit from participants who discontinued study intervention or were withdrawn from the study.

Results of pharmacodynamic/biomarker analyses will be presented in a separate report.

5.8.4. Pharmacokinetic/Pharmacodynamic Relationships

Pharmacokinetic/Pharmacodynamic analyses will be based on the PK/PD analysis set for mRSS or for ACR CRISS (Section 4). PK and PD data from those whose PK samples are excluded from the summaries of PK analysis at Week 12, 24 and 52 will be excluded from the PK/PD analysis at Week 12, 24 and 52, respectively.

The change from baseline score in mRSS at Week 12, 24 and 52 versus serum guselkumab concentrations (same visit point) will be tabulated by quartile of guselkumab concentration, for each schedule visit.

This relationship will be also explored graphically by drawing a bar-chart whose height represents the change from baseline score in mRSS total score for each quartile of the median of serum guselkumab concentration (<=1st quartile, >1st quartile and <= 2nd quartile, >2nd quartile and <= 3rd quartile, >3rd quartile).

The serum guselkumab concentrations at Week 12, 24, and 52 will be summarized by quartile of the change from baseline score in mRSS (same visit point) for participants randomized to guselkumab at Week 0, for each schedule visit.

This relationship will be also explored graphically by drawing a bar-chart whose height represents the median of serum guselkumab concentration for each quartile of the change from baseline score in mRSS total score (<=1st quartile, >1st quartile and <= 2nd quartile, >2nd quartile and <= 3rd quartile, >3rd quartile).

If data permits, relationship between the proportions of ACR CRISS responders who achieve a score ACR CRISS ≧ 0.6 at Week 12, 24 and 52 and serum guselkumab concentrations (same visit point) for participants randomized to guselkumab will be tabulated by quartile of guselkumab concentration, for each schedule visit.

This relationship will be also explored graphically by drawing a bar-chart whose height represents the proportion of responder for each guselkumab concentration quartile (<=1st quartile, >1st quartile and <= 2nd quartile, >2nd quartile and <= 3rd quartile, >3rd quartile).

The difference in guselkumab concentrations over time between responders and non-responders of following endpoints will be also assessed for PK/PD analysis set by summarizing serum guselkumab concentrations over time through Week 24/52 by scheduled visit and following response status of endpoints. The difference in guselkumab concentrations over time between response status will be also explored graphically by drawing a median (IQ range) of serum guselkumab concentration time profiles through Week 24/52 with nominal sampling time, by the intervention group and following response status of endpoints.

- An increase from baseline ≥5 points and ≥20% in mRSS achieve/not achieve at Week 24/52
- ACR CRISS responder/ non-responder at Week 24/52

Other efficacy measures and/or relevant PD endpoints, including biomarkers in blood or in skin biopsies, versus serum guselkumab concentrations may be explored graphically when appropriate. If any visual trend is observed, additional analysis may be conducted if deemed necessary.

5.9. Interim Analyses

An interim analysis will be performed when all participants have completed EOS of the main study. All sponsor personnel will be unblinded at the analyses for internal decision-making and/or interactions with health authorities.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 List of Abbreviations

ACR	American College of Rheumatology
ACR CRISS	American College of Rheumatology Combined Response Index in diffuse cutaneous systemic sclerosis
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse event of special interest
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
BQL	Below quantification limit
CI	Confidence interval
COVID-19	Coronavirus disease 2019
DBL	Database lock
dcSSc	Diffuse cutaneous systemic sclerosis
DILI	Drug-induced Liver Injury
DLCO	Diffusing capacity of the lung for carbon monoxide
ECG	Electrocardiogram
eCRF	Electronic case report form
EOS	End of study
EULAR	European League Against Rheumatism
FAS	Full analysis set
FOIA	Freedom of Information Act
FSSG	Frequency Scale for the Symptoms of Gastroesophageal reflux disease
FVC	Forced vital capacity
GCP	Good Clinical Practice
GERD	Gastroesophageal reflux disease
HAQ-DI	Health Assessment Questionnaire-Disability Index
Hgb	Hemoglobin
HR	Hazard ratio
HRCT	High resolution computed tomography
IL	Interleukin
ILD	Interstitial lung disease
IQR	Interquartile range
LLOQ	Lower limit of quantification
LSM	Lease square mean
LTE	Long term extension
MAR	Missing at random
MCMC	Markov Chain Monte Carlo
MI	Multiple imputation
MMRM	Mixed-Effect Model Repeated Measure
mRSS	Modified Rodnan Skin Score
NAb	Neutralizing antibodies
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
PGA	Patient global assessment
PhGA	Physician global assessment
PoC	Proof of concept
PK	Pharmacokinetic(s)
PD	Pharmacodynamic
Q4W	Every 4 weeks
SAE	Serious adverse event
SAP	Statistical analysis plan
SC	Subcutaneous
SD	Standard deviation
SNP	Single nucleotide polymorphisms
SSc	Systemic sclerosis

ULN	Upper Limit of Normal
WBC	White blood cell
WHO-DD	World Health Organization Drug Dictionary

6.2. Appendix 2 Changes to Protocol-Planned Analyses

Not applicable.

6.3. Appendix 3 Demographics and Baseline Characteristics

The number of participants in each analysis set will be summarized by intervention group and overall.

Table 5 presents a list of the demographic variables that will be summarized by intervention group and overall, for the FAS.

Table 5: Demographic Variables

Continuous Variables:	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median, range [minimum and maximum] and IQ range).
Weight (kg)	
Height (cm)	
Body Mass Index (BMI) (kg/m^2)	
Categorical Variables	
Age (<65 years, and >=65 years)	Frequency distribution with the number and percentage of participants in each category.
Sex (male, female, undifferentiated)	
Race[a] (Asian, Multiple, Other)	

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

Table 6 presents a list of the baseline disease characteristic variables that will be summarized by intervention group and overall, for the FAS.

Table 6: Baseline Disease Characteristic Variables

Continuous Variables:	**Summary Type**
Duration of SSc (months)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum] and IQ range).
mRSS	
HAQ-DI	
FSSG	
Physician's Global VAS	
Patients' Global VAS	
% predicted FVC	
% DLCO	
Categorical Variables	
Duration of SSc (≤18 months, > 18 months)	Frequency distribution with the number and percentage of participants in each category.
Use of corticosteroids at baseline (Yes, No)	
Presence of Digital Ulcer (Yes, No)	
mRSS (10-15, 16-22)	
Presence of ILD (Yes, No)	
Baseline Anti-topoisomerase I Antibody (Yes, No)	

The use of alcohol and tobacco at baseline in participants will also be collected and summarized.

6.4. Appendix 4 Protocol Deviations

In general, the following list of major protocol deviations may have the potential to impact participants' rights, safety or well-being, or the integrity and/or result of the clinical study. Participants with major protocol deviations will be identified prior to database lock and the participants with major protocol deviations will be summarized by category for different study period (main study period and LTE period).

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

In particular, major and minor protocol deviations related with COVID-19 will be collected and summarized separately. Separate listings will also be presented for both COVID-19 induced major and minor protocol deviations.

6.5. Appendix 5 Prior and Concomitant Medications

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

Summary of prior medications includes prior biologic treatment history and prior medication history used to treat SSc. Prior biologic treatment history will be summarized by intervention group and type of medication (ie, IL-23, IL-12/23, IL-17, IL-6, CD20 and other). Prior medication history for SSc will also be summarized by intervention group and medication term.

Medication taken for SSc at baseline will also be captured in eCRF, and summary will be provided by type of medication and intervention group.

Summaries of concomitant medications will be presented by ATC term and intervention group for different study period (main study period and LTE period). The proportion of participants who receive each concomitant medication will be summarized as well as the proportion of participants who receive at least 1 concomitant medication.

Listings will be provided for participants with prior medications, baseline medications for SSc, concomitant medications and rescue medications separately.

A summary of rescue medications given to participants after week 24 will be provided by visit.

6.6. Appendix 6 Medical History

Medical history (including general medical history, medical history of interest and systemic sclerosis conditions of interest) will be listed.

6.7. Appendix 7 Intervention Compliance

Compliance to randomized intervention versus actual intervention will be presented in a summary table.

6.8. Appendix 8 Adverse Events of Special Interest

Adverse events of special interest are defined as below:

AE Special Interest Category	Criteria
Drug-induced Liver Injury (DILI)	HEPATIC DISORDERS (SMQ) and SUB_SMQ1='DRUG RELATED HEPATIC DISORDERS - COMPREHENSIVE SEARCH (SMQ)' and SCOPE='Narrow'

AE Special Interest Category	Criteria
Malignancy	AE type='Malignancy' as assessed by investigator in CRF
Active TB	AE type='Active TB' as assessed by investigator in CRF
ILD	AE type='ILD' as assessed by investigator in CRF
Other	AE type='Other' as assessed by investigator in CRF

AE Special Interest Category	Criteria
Hypersensitivity Reactions	HYPERSENSITIVITY (SMQ) and SCOPE='Narrow'
Opportunistic Infections	OPPORTUNISTIC INFECTIONS (SMQ) and SCOPE='Narrow'
Adverse Cardiovascular Events	AESDTH (AE leading to death) = 'Y' or UPCASE(AEBODSYS) in ('CARDIAC DISORDERS','NERVOUS SYSTEM DISORDERS','VASCULAR DISORDERS')

After applying above handling rule for Hypersensitivity Reactions, Opportunistic Infections and Adverse Cardiovascular Events, the selected AEs will be sent to clinician for review to flag "Yes" if AE is considered as AESI.

6.9. Appendix 9 Medications of Special Interest

Not applicable.

6.10. Appendix 10 Laboratory Toxicity Grading

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

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

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

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

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

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
					"normal baseline" are applied.
Blood bilirubin increased	>ULN - 1.5 x ULN if baseline was normal; > 1.0 - 1.5 x baseline if baseline was abnormal	>1.5 - 3.0 x ULN if baseline was normal; >1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 10.0 x ULN if baseline was normal; >3.0 - 10.0 x baseline if baseline was abnormal	>10.0 x ULN if baseline was normal; >10.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
CD4 lymphocytes decreased	<LLN - 500/mm3; <LLN - 0.5 x 10e9 /L	<500 - 200/mm3; <0.5 - 0.2 x 10e9 /L	<200 - 50/mm3; <0.2 x 0.05 - 10e9 /L	<50/mm3; <0.05 x 10e9 /L	
Cholesterol high	>ULN - 300 mg/dL; >ULN - 7.75 mmol/L	>300 - 400 mg/dL; >7.75 - 10.34 mmol/L	>400 - 500 mg/dL; >10.34 - 12.92 mmol/L	>500 mg/dL; >12.92 mmol/L	
CPK increased	>ULN - 2.5 x ULN	>2.5 x ULN - 5 x ULN	>5 x ULN - 10 x ULN	>10 x ULN	
Creatinine increased	Creatine Kinase >ULN - 1.5 x ULN	Creatine Kinase >1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	Creatine Kinase >3.0 x baseline; >3.0 - 6.0 x ULN	Creatine Kinase >6.0 x ULN	
Fibrinogen decreased	<1.0 - 0.75 x LLN; if abnormal, <25% decrease from baseline	<0.75 - 0.5 x LLN; if abnormal, 25 - <50% decrease from baseline	<0.5 - 0.25 x LLN; if abnormal, 50 - <75% decrease from baseline	<0.25 x LLN; if abnormal, 75% decrease from baseline; absolute value <50 mg/dL	Ranges defined for "abnormal" are applied only on values < LLN. Grade 0 will be assigned to values > ULN.
GGT increased	>ULN - 2.5 x ULN if baseline was normal; 2.0 - 2.5 x baseline if baseline was abnormal	>2.5 - 5.0 x ULN if baseline was normal; >2.5 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
Haptoglobin decreased	<LLN	-	-	-	
Hemoglobin increased	Increase in >0 - 2 g/dL; Increase in >0 - 20 g/L	Increase in >2 - 4 g/dL; Increase in >20 - 40 g/L	Increase in >4 g/dL; Increase in >40 g/L	-	The increase indicates the level of increase above normal (above ULN). Applied as, eg, grade 1 (g/dL): >ULN – ULN+2 g/dL; Added ranges in SI unit (g/L).

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
INR increased	>1.2 - 1.5; *>1 - 1.5 x baseline if on anticoagulation; monitoring only indicated*	>1.5 - 2.5; *>1.5 - 2.5 x baseline if on anticoagulation; dose adjustment indicated*	>2.5; *>2.5 x baseline if on anticoagulation; bleeding*	-	Concomitant therapy or clinical signs and symptoms are not taken into consideration for grading.
Lipase increased	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN; *>2.0 - 5.0 x ULN and asymptomatic*	>2.0 - 5.0 x ULN with signs or symptoms; *>5.0 x ULN and asymptomatic*	>5.0 x ULN and with signs or symptoms	"Asymptomatic" ranges are not taken into consideration for grading, ie, worst case grading is applied.
Lymphocyte count decreased	<LLN - 800/mm3; <LLN - 0.8 x 10e9/L	<800 - 500/mm3; <0.8 - 0.5 x 10e9 /L	<500 - 200/mm3; <0.5 - 0.2 x 10e9 /L	<200/mm3; <0.2 x 10e9 /L	
Lymphocyte count increased	-	>4000/mm3 - 20,000/mm3; >4 - 20 x 10e9 /L	>20,000/mm3; >20 x 10e9 /L	-	Added ranges in SI unit (x 10e9 /L).
Neutrophil count decreased	<LLN - 1500/mm3; <LLN - 1.5 x 10e9 /L	<1500 - 1000/mm3; <1.5 - 1.0 x 10e9 /L	<1000 - 500/mm3; <1.0 - 0.5 x 10e9 /L	<500/mm3; <0.5 x 10e9 /L	Both Neutrophils and segmented neutrophils are graded using these criteria.
Platelet count decreased	<LLN - 75,000/mm3; <LLN - 75.0 x 10e9 /L	<75,000 - 50,000/mm3; <75.0 - 50.0 x 10e9 /L	<50,000 - 25,000/mm3; <50.0 - 25.0 x 10e9 /L	<25,000/mm3; <25.0 x 10e9 /L	
Serum amylase increased	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN; *>2.0 - 5.0 x ULN and asymptomatic*	>2.0 - 5.0 x ULN *with signs or symptoms;* *>5.0 x ULN and asymptomatic*	>5.0 x ULN *and with signs or symptoms*	"Asymptomatic" ranges are not taken into consideration for grading, ie, worst case grading is applied.
White blood cell decreased	<LLN - 3000/mm3; <LLN - 3.0 x 10e9 /L	<3000 - 2000/mm3; <3.0 - 2.0 x 10e9 /L	<2000 - 1000/mm3; <2.0 - 1.0 x 10e9 /L	<1000/mm3; <1.0 x 10e9 /L	
Metabolism and nutrition disorders					
Acidosis	pH <normal, but >=7.3	-	pH <7.3	*Life-threatening consequences*	pH <normal is implemented as pH <LLN. Clinical signs and symptoms are not taken into consideration for grading.

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
Alkalosis	pH >normal, but <=7.5	-	pH >7.5	*Life-threatening consequences*	pH >normal is implemented as pH >ULN. Clinical signs and symptoms are not taken into consideration for grading.
Hypercalcemia	Corrected serum calcium of >ULN - 11.5 mg/dL; >ULN - 2.9 mmol/L; Ionized calcium >ULN - 1.5 mmol/L	Corrected serum calcium of >11.5 - 12.5 mg/dL; >2.9 - 3.1 mmol/L; Ionized calcium >1.5 - 1.6 mmol/L; *symptomatic*	Corrected serum calcium of >12.5 - 13.5 mg/dL; >3.1 - 3.4 mmol/L; Ionized calcium >1.6 - 1.8 mmol/L; *hospitalization indicated*	Corrected serum calcium of >13.5 mg/dL; >3.4 mmol/L; Ionized calcium >1.8 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hyperkalemia	Potassium >ULN - 5.5 mmol/L	Potassium >5.5 - 6.0 mmol/L; *intervention initiated*	Potassium >6.0 - 7.0 mmol/L; *hospitalization indicated*	Potassium >7.0 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypermagnesemia	Magnesium >ULN - 3.0 mg/dL; >ULN - 1.23 mmol/L	-	Magnesium >3.0 - 8.0 mg/dL; >1.23 - 3.30 mmol/L	Magnesium >8.0 mg/dL; >3.30 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypernatremia	Sodium >ULN – 150 mmol/L	Sodium >150 – 155 mmol/L; *intervention initiated*	Sodium >155 – 160 mmol/L; *hospitalization indicated*	Sodium >160 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypertriglyceridemia	Triglycerides 150 mg/dL - 300 mg/dL; 1.71 mmol/L - 3.42 mmol/L	Triglycerides >300 mg/dL – 500 mg/dL; >3.42 mmol/L - 5.7 mmol/L	Triglycerides >500 mg/dL – 1000 mg/dL; >5.7 mmol/L - 11.4 mmol/L	Triglycerides >1000 mg/dL; >11.4 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypoalbuminemia	Albumin <LLN - 3 g/dL; <LLN - 30 g/L	Albumin <3 - 2 g/dL; <30 - 20 g/L	Albumin <2 g/dL; <20 g/L	*Life-threatening consequences; urgent intervention indicated*	Clinical signs and symptoms are not taken into consideration for grading.

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
Hypocalcemia	Corrected serum calcium of <LLN - 8.0 mg/dL; <LLN - 2.0 mmol/L; Ionized calcium <LLN - 1.0 mmol/L	Corrected serum calcium of <8.0 - 7.0 mg/dL; <2.0 - 1.75 mmol/L; Ionized calcium <1.0 - 0.9 mmol/L; *symptomatic*	Corrected serum calcium of <7.0 - 6.0 mg/dL; <1.75 - 1.5 mmol/L; Ionized calcium <0.9 - 0.8 mmol/L; *hospitalization indicated*	Corrected serum calcium of <6.0 mg/dL; <1.5 mmol/L; Ionized calcium <0.8 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypoglycemia	Glucose <LLN - 55 mg/dL; <LLN - 3.0 mmol/L	Glucose <55 - 40 mg/dL; <3.0 - 2.2 mmol/L	Glucose <40 - 30 mg/dL; <2.2 - 1.7 mmol/L	Glucose <30 mg/dL; <1.7 mmol/L; *life-threatening consequences; seizures*	Clinical signs and symptoms are not taken into consideration for grading. Urine glucose is not graded.
Hypokalemia	*Potassium <LLN - 3.0 mmol/L*	*Symptomatic with* Potassium <LLN - 3.0 mmol/L; intervention indicated	Potassium <3.0 - 2.5 mmol/L; *hospitalization indicated*	Potassium <2.5 mmol/L; *life-threatening consequences*	"Symptomatic" ranges are applied for grade 2, grade 1 not assigned, ie, worst case applied. Clinical signs and symptoms are not taken into consideration for grading of grade 3 and 4.
Hypomagnesemia	Magnesium <LLN - 1.2 mg/dL; <LLN - 0.5 mmol/L	Magnesium <1.2 - 0.9 mg/dL; <0.5 - 0.4 mmol/L	Magnesium <0.9 - 0.7 mg/dL; <0.4 - 0.3 mmol/L	Magnesium <0.7 mg/dL; <0.3 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hyponatremia	Sodium <LLN – 130 mmol/L	*Sodium 125-129 mmol/L and asymptomatic*	*Sodium 125-129 mmol/L symptomatic; 120-124 mmol/L regardless of symptoms* Sodium <130-120 mmol/L	Sodium <120 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading. Worst case ("<130-120 mmol/L" for grade 3 added by Janssen) is applied across grade 2/3 ranges: 120-129 mol/L

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
					assigned to grade 3, grade 2 not used.
Renal and urinary disorders					
Proteinuria	1+ proteinuria; urinary protein ≥ULN - <1.0 g/24 hrs; urinary protein ≥ULN - <1000 mg/day	**Adult:** 2+ and 3+ proteinuria; urinary protein 1.0 - <3.5 g/24 hrs; urinary protein 1000 - <3500 mg/day **Pediatric:** Urine P/C (Protein/Creatinine) ratio 0.5 - 1.9; Urine P/C (Protein/Creatinine) 56.5 – 214.7 g/mol	**Adult:** 4+ proteinuria; urinary protein >=3.5 g/24 hrs; urinary protein >=3500 mg/day; **Pediatric:** Urine P/C (Protein/Creatinine) ratio >1.9; Urine P/C (Protein/Creatinine) >214.7 g/mol	-	In case both 24-h urine collection and dipstick are collected, then worst case is taken, as opposed to having 24-h urine collection take precedence over dipstick. Added ranges in SI unit for urinary protein (mg/day) and for urine P/C (g/mol). Pediatric grading is applied to age range [0-18]. Adult grading is applied for ages [>18].

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

6.11. Appendix 11 A Prohibited Change in SSC Medication

- New start of prohibited and restricted therapies during the study which are defined in the protocol Section 6.8:
  - IL-23 inhibitor therapy (including but not limited to guselkumab, risankizumab, tildrakizumab, brazikumab, mirikizumab)
  - tocilizumab
  - sirukumab, sarilumab, mavrilimumab, abatacept, belimumab
  - IL-12/23 inhibitors (ustekinumab)
  - IL-17 inhibitors (secukinumab, ixekizumab, brodalumab)
  - tyrosine kinase inhibitor (nintedanib)
  - Any live viral or live bacterial vaccination
  - Rituximab
  - Any cytotoxic agents (cyclophosphamide, chlorambucil, nitrogen mustard, or other alkylating agents)
  - Intravenous immunoglobulin (IVIG), apheresis therapy (plasmapheresis or leukopheresis)
  - Systemic immunosuppressive agents (including but not limited to cyclosporine A, azathioprine, tacrolimus, sirolimus, methotrexate, sulfasalazine, leflunomide with cholestyramine washout or mycophenolate mofetil/mycophenolic acid)
  - Intramuscular, intra-articular, intrabursal, epidural, intralesional or IV GCs
  - Pirfenidone
  - Systemic and topical Janus kinase inhibitors (eg, tofacitinib, upadacitinib)
  - Other biologics
  - Any investigational agents/treatments
- Treat with rescue therapy before Week 24, including:
  1) Oral Corticosteroid ≤20 mg/day of prednisolone or equivalent
  2) Immunomodulator (Azathioprine), within the approved dose for SSc
- New start of oral Corticosteroid (prednisolone or equivalent) before Week 24.
- Significant increase of oral Corticosteroid (prednisolone or equivalent) before Week 24.

7. REFERENCES

Khanna D, Furst DE, Hays RD, et al. Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study. Ann Rheum Dis. 2006;65(10):1325-1329.

Khanna D, Merkel PA. Outcome measures in systemic sclerosis: an update on instruments and current research. Curr Rheumatol Rep. 2007;9(2):151-157.

Matsuda Y, Singh G, Yamanaka H, et al. Validation of a Japanese Version of the Stanford Health Assessment Questionnaire in 3,763 Patients With Rheumatoid Arthritis. Arthritis Rheum. 2003;49(6):784–788.

Ukyo Y and Noma H. Permutation Inference Methods for the MMRM (Mixed-Effects Model for Repeated Measures) in Incomplete Longitudinal Data Analysis. Jpn J Biomet 2019;40(1):15-34.