SANOFI

NCT03600818

STATISTICAL ANALYSIS PLAN

A randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of sarilumab in patients with polymyalgia rheumatica

SAR153191-EFC15160

STATISTICIAN: [redacted]

DATE OF ISSUE: 20-May-2021

Total number of pages: 88

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

ADA:	antidrug antibody
AEs:	adverse events
AESIs:	adverse events of special interest
ALP:	alkaline phosphatase
ALT:	alanine aminotransferase
ANA:	antinuclear antibodies
AST:	asparate aminotransferase
BMI:	body mass index
CRF:	case report form
CSR:	clinical study report
ds-DNA:	double-stranded DNA
HbA1c:	hemoglobin A1c
HDL-C:	high density lipoprotein cholesterol
HLGT:	high level group term
HLT:	high level term
ID:	identification
IMP:	investigational medicinal product
LDH:	lactate dehydrogenase
LDL-C:	low density lipoprotein cholesterol
LLN:	lower limit of normal
LLOQ:	lower limit of quantification
LLT:	lower level term
LS:	least square, least square
MD-VAS:	physician global assessment of disease activity- visual analog scale
MedDRA:	Medical Dictionary for Regulatory Activities
NCEP ATPIII:	National Cholesterol Education Program Adult Treatment Panel III
PT:	preferred term
RBC:	red blood cell
SAEs:	serious adverse events
SMQ:	standardized MedDRA query
SOC:	system organ class
TEAE:	treatment-emergent adverse event
ULN:	upper limit of normal
WBC:	white blood cell
WHO-DD:	World Health Organization - Drug Dictionary

1 OVERVIEW AND INVESTIGATIONAL PLAN

This statistical analysis plan (SAP) provides a comprehensive and detailed description of statistical strategy and methodology to be used to analyze the data from sarilumab EFC15160 study in patients with active polymyalgia rheumatica (PMR).

Due to the ongoing pandemic of COVID-19 (the disease caused by the novel coronavirus), the operations of sarilumab (Kevzara®) clinical trial EFC15160 have been significantly disrupted in various countries and sites. Consequently, the screening/recruitment activities for this study have been suspended since March 2020. Due to the protracted recruitment timeline and internal portfolio reprioritization, in July 2020, Sanofi decided not to re-open the screening/recruitment activities and to discontinue the study. This decision is not related to any safety issues arising from the administration of sarilumab.

A total of 118 patients (out of 280 planned patients) have been randomized in the study. Despite the decision to discontinue the study, all patients already recruited will be allowed to continue in the study to complete the 52-week treatment period and the 6 weeks post-treatment follow up per protocol.

1.1 STUDY DESIGN AND RANDOMIZATION

Original study design

This is a multicenter, randomized, double-blind, placebo-controlled 52-week Phase 3 study with a 6-week post-treatment follow-up phase, evaluating the efficacy and safety of sarilumab in patients with active polymyalgia rheumatica (PMR).

Patients with active PMR who meet the entry criteria will be randomized in a 1:1 ratio to either

- Group 1: sarilumab 200 mg every two weeks (q2w) with a 14-week prednisone taper, or
- Group 2: sarilumab matching placebo q2w with a 52-week prednisone taper.

All patients will receive sarilumab 200 mg or its matching placebo for 52 weeks, while the initial dose of prednisone for both groups will be 15 mg/day for the first 2 weeks after randomization (baseline). To ensure the double blinding of the corticosteroid (CS) regimen thereafter, once daily prednisone and/or prednisone matching placebo will be provided according to the protocol defined CS tapering regimen (Appendix B) as summarized below.

- Group 1: From Week 2 to Week 13, patients will receive gradually decreasing dose levels of prednisone (prednisone or combinations of prednisone and its matching placebo). From Week 14 onwards, patients without flare will receive prednisone matching placebo.
- Group 2: From Week 2 to Week 51, patients will receive gradually decreasing dose levels of prednisone (prednisone or combinations of prednisone and its matching placebo).

During the initial 12 weeks of prednisone taper, treatment for one flare before Week 12 is permitted if it can be successfully treated with a low dose (≤5 mg/day) prednisone add-on taper regimen (completed prior to Week 12) and provided that all other sustained remission parameters are met. For patients who experience a disease flare and are in need of rescue therapy (such as CS) as per investigator judgment, during the course of the study, they may continue administration of sarilumab or matching placebo in a double-blinded fashion for the full duration of the 52-week treatment period only if CS are used as rescue therapy. Treatment with non-biological immunosuppressive drugs (such as alkylating agents, hydroxychloroquine, CsA, MMF, AZA, etc.) is not permitted during the study, unless used for the purpose of rescue therapy.

For the management of pre-defined laboratory abnormalities, investigators may temporarily hold the study drug (sarilumab or its matching placebo) and/or permanently reduce the dose of sarilumab (or its matching placebo) to 150 mg q2w in a continuous blinded manner.

There will be a post treatment follow-up visit 6 weeks after the end of treatment visit. All patients who prematurely and permanently discontinue study treatment will be asked to return to the site for study assessments as per protocol until end of study (EOS) evaluation. The last patient last visit will occur when the last patient has completed the 6-week follow up period (EOS). The end of the clinical trial is defined as the last patient's last visit.

Approximately 280 patients are planned to be randomized into 2 treatment groups of 140 patients per group.

Study discontinuation plan

As mentioned in Section 1, due to the protracted recruitment timeline and internal portfolio reprioritization, Sanofi decided to terminate the EFC15160 study in July 2020. All randomized patients continued to complete the study per protocol, and the last patient last visit (EOS) was achieved on May 19, 2021.

1.2 OBJECTIVES

1.2.1 Primary objectives

The primary objective of this study is to evaluate the efficacy of KEVZARA® (sarilumab) in patients with PMR as assessed by the proportion of patients with sustained remission at Week 52 for sarilumab with a 14-week CS tapering regimen as compared to placebo with a 52-week CS tapering regimen (see Section 2.1.3.1 for definition of sustained remission).

1.2.2 Secondary objectives

- To demonstrate the efficacy of sarilumab (14 week taper of CS) compared to placebo (52 week taper of CS) in patients with PMR with regards to:
  - Clinical responses (such as components of sustained remission, disease remission rates, time to first disease flare) over time.

- Cumulative CS (including prednisone) exposure.

- To assess the safety (including immunogenicity) and tolerability of sarilumab in patients with PMR.
- To measure sarilumab concentrations from serum of patients with PMR.
- To assess the effect of sarilumab on reducing glucocorticoid toxicity as measured by the composite glucocorticoid toxicity index (GTI) questionnaire.

1.2.3 Other objectives

- To assess the effect of sarilumab on physician assessment of disease activity as measured by a visual analogue scale (MD-VAS).
- To assess the effect of sarilumab on a variety of patient-reported outcome (PRO) concepts, including fatigue (as measured by FACIT-fatigue scale), health status (as measured by EQ-5D-3L and SF-36v2), physical function (as measured by HAQ-DI, pain (as measured via HAQ-DI by a visual analogue scale [VAS]) and patient assessment of disease activity (as measured via HAQ-DI by a VAS).
- To assess the correlation between the ESR/CRP levels on remission status.
- To characterize the disease activity of PMR patients while on steroid taper or sarilumab treatment in a subset of patients using comprehensive approaches to evaluate circulating immune cell types.
- To characterize the disease activity of PMR patients while on steroid taper or sarilumab treatment by evaluating circulating proteins, genetics and gene expression.

1.3 DETERMINATION OF SAMPLE SIZE

The sample size calculations are based on the primary efficacy variable of sustained remission at Week 52 compared between the sarilumab + 14-week prednisone taper and placebo + 52-week prednisone taper groups. There are no prior controlled study data to establish the placebo + 52-week prednisone taper remission rate in PMR. The hypothesis to be tested on the primary endpoint is that sarilumab + 14-week prednisone taper is superior to placebo + 52-week prednisone taper based on the proportion of patients achieving sustained remission at Week 52. A 25% difference from placebo + 52-week prednisone taper response rate is considered clinically relevant, and the sample size of 140 per group provides at least 90% power to detect such a difference regardless of the placebo + 52-week prednisone taper response rate (i.e., assuming 5% to 50% placebo + 52-week prednisone taper response rates), using a two-sided $\chi 2$ test at a significance level of 0.01. Data from the GiACTA trial in a related population with GCA suggest that such a treatment difference is achievable (1).

The above sample size calculation was performed for the full enrollment (140 per group). The total number of randomized patients is 118 at the time of the premature termination of the study. Randomized patients could continue in the study to complete the 52-week treatment period per protocol.

Based on protocol amendment 2, the alpha level has been changed from 0.01 to 0.05. The power calculations provided in Table 1 below are based on a sample size of 118 and alpha level of 0.05. The sample size of 59 per group provides at least 85% and 95% power to detect a 25% and 30% difference in response rates between sarilumab + 14-week prednisone taper and placebo + 52-week prednisone taper groups respectively, assuming response rates between 5% to 15% in the placebo + 52-week prednisone taper group based on two-sided $\chi 2$ test at a significance level of 0.05.

Table 1: Power calculations based on total sample size of 118 randomized patients

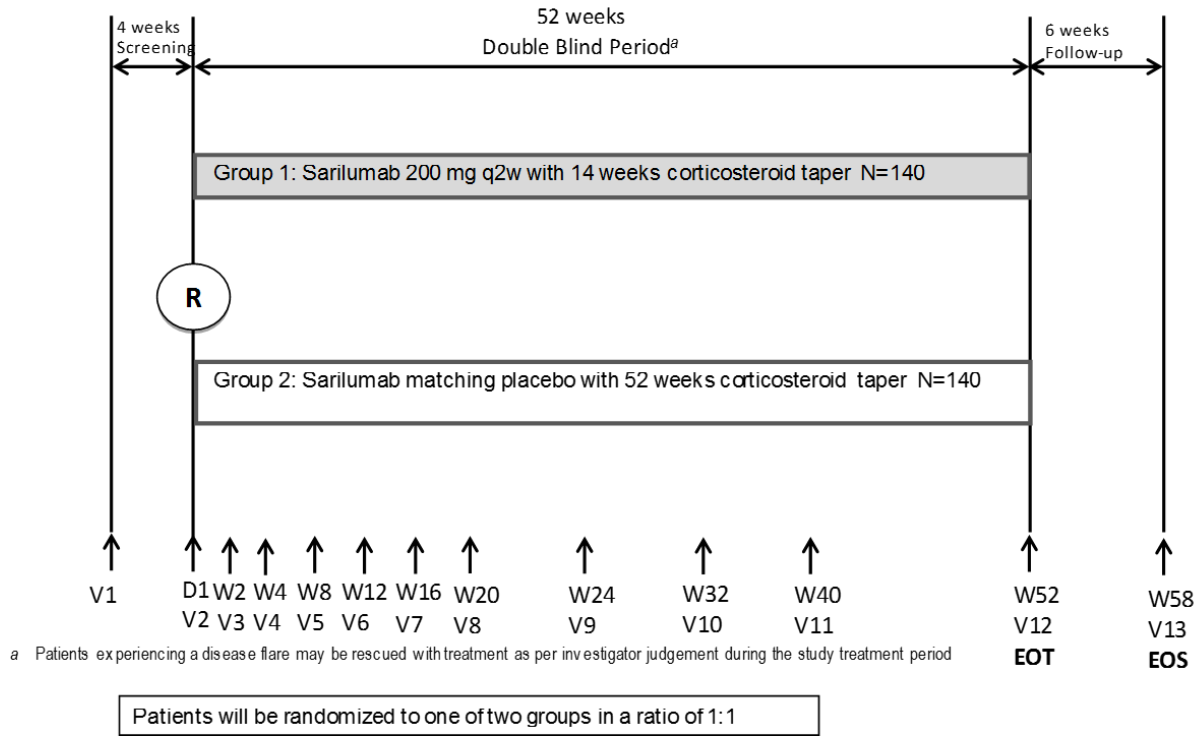
Assumed sustained remission rate for placebo+ 52-week prednisone taper	Sample Size per group	Power assuming an absolute 25% between group difference	Power assuming an absolute 30% between group difference
5%	59	95.6%	98.8%
10%	59	91.2%	97.2%
15%	59	86.9%	95.4%

Calculations were made using nQuery Advisor 7.0 based on two-sided $\chi 2$ test at a significance level of 0.05.

1.4 STUDY PLAN

The schedule of the safety, efficacy, pharmacokinetic (PK), and immunogenicity assessments from Section 1.2 of the study protocol is provided in Appendix A. The graphical study design is provided in Figure 1.

Figure 1 – Graphic study design



1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

In general, the statistical methods remain unchanged by the protocol amendment 02 except for revisions made to sample size and power calculations as listed below:

Amended protocol 02 (19 April 2021):

As a result of the inability to recruit due to the COVID19 pandemic, the study was terminated prematurely resulting in a change in the total expected number of patients from 280 to 118 (Section 1.3).

- Statistical significance level changed from 0.01 to 0.05 due to the change in total expected number of patients.
- Power calculations revised due to the change in total expected sample size to 118, and the change in statistical significance level to 0.05.
- 99% confidence intervals revised to 95% confidence intervals associated with the change in statistical significance level to 0.05.

1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN

Due to the premature termination of the study, a total of 118 patients (42%) of the 280 patients planned were randomized. The randomized patients could continue in the study to complete the treatment period and the 6 weeks post-treatment follow up per protocol. Hence, there are no changes to the efficacy analysis plan per protocol.

The COVID-19 pandemic has also impacted data collection as patients may not be able to visit clinical sites for endpoint assessments. Additional patient level information at visits that were impacted by the pandemic situations were captured for assessments performed in visits that were delayed, not done, partially done on site, or partially done by phone. Major or critical COVID-19 impacted deviations per protocol were also recorded.

To assess the impact of the pandemic on the study, COVID-19 related summaries are added as below.

a) Patients who permanently discontinue study treatment or study due to COVID-19 reasons will be summarized in patient disposition table

b) Patient disposition by visit according to trial impact due to COVID-19

c) Critical or major deviations summary according to COVID-19 impact

d) Number (%) of patients experiencing at least one treatment-emergent COVID-19 related adverse event by primary SOC and PT

In addition, if > 10% patients are impacted by the COVID-19 pandemic, below summaries/analyses will be generated by COVID-19 impacted population (Yes, No):

a) Extent of exposure to SC IMP

b) Treatment compliance for SC IMP

c) Number (%) of patients with TEAEs by primary SOC and PT, and

d) Analysis of the primary endpoint of sustained remission at Week 52.

The population with and without trial impact (disruption) due to COVID-19 is defined in Section 2.3.4.

In addition, upon meeting with Japan Health Authorities (PMDA) in April 2019, data for the components to calculate the PMR activity score (PMR-AS) were subsequently collected and the analysis plan is described in Section 2.4.4.3.

2 STATISTICAL AND ANALYTICAL PROCEDURES

Data analyses will be performed in the double-blind (DB) 52-week treatment period and presented by two treatment groups (sarilumab 200 mg q2w with 14-week CS taper, and placebo with 52-week CS taper). Safety data will be summarized for the treatment and 6-week follow-up period.

2.1 ANALYSIS ENDPOINTS

2.1.1 Demographic and baseline characteristics

The baseline value is defined as the last available value prior to the first dose of study medication, SC IMP. For patients randomized but not treated, the baseline value is the last available value up to randomization. For PRO assessments, baseline value is defined as the last available value up to and including the day of administration of the first dose of SC IMP.

All baseline safety and efficacy parameters (apart from those listed below) are presented along with the on-treatment summary statistics in the efficacy and safety sections (Section 2.4.4 and Section 2.4.5).

Demographic characteristics

The following demographic characteristics will be summarized separately by treatment and overall:

- Sex (Male, Female)
- Race (White, Black or African American, Asian, Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native, Unknown, Not reported)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Unknown, Not reported)
- Region (Region 1 [Western countries], Region 2 [South America], and Region 3 [Rest of the World]); definitions of the regions are provided in Section 2.5.7
- Age (years)
- Age group (≥50 and <65, ≥65 and <75, ≥75 and <85, ≥85 years)
- Weight (kg)
- Weight group (<60, ≥60 and <100, ≥100 kg)
- Body mass index (BMI, kg/m^2)
- BMI group (<25, ≥25 and < 30, ≥30 kg/m^2)
- Smoking status (never, former, current)
- Alcohol habits (never, occasional, monthly, weekly, daily)

Medical or surgical history

Medical or surgical history includes all the relevant findings within the lifetime of the patient.

This information will be coded to a "Lower Level Term (LLT)", "Preferred Term (PT)", "High Level Term (HLT)", "High Level Group Term (HLGT)", and associated primary "System Organ Class (SOC)" using the version of Medical Dictionary for Regulatory Activities (MedDRA) currently in effect at Sanofi at the time of database lock.

Disease characteristics at baseline

The following baseline disease characteristics as well as the other baseline characteristics will be summarized separately by treatment group:

- PMR signs and symptoms:
  - Morning stiffness (≤ 45 minutes, > 45 minutes)
  - Pain in the shoulders, neck and/or hip girdles
  - Limited range of motion of the shoulders and /or hip girdles (no upper limb elevation, elevation below the shoulder girdle, elevation up to the shoulder girdle, elevation above the shoulder girdle)
  - Constitutional symptoms (low grade fever, unintentional weight loss, fatigue, malaise, loss of appetite)
  - Other features judged by the investigator to be consistent with PMR disease activity
- Number of patients with prior PMR flares (≤6 months, 6 to 12 months or >12 months prior to screening) and number of prior flares per patient
- Prednisone dose at baseline
- Highest dose and duration of corticosteroid taper taken for PMR from 24 weeks prior to screening up to baseline
- Physician global assessment of disease activity (MD-VAS)
- C-reactive protein (CRP) at baseline
- ESR at baseline
- Glucocorticoid toxicity at baseline*

*The baseline glucocorticoid toxicity is established by a composite index calculated from weighted scores assigned by toxicities. The scores are described in Appendix J.

Any technical details related to computation, dates, and imputations for missing dates are described in Section 2.5.

2.1.2 Prior or concomitant medications

All medications taken within a certain period of time before randomization and until the end of the study, including vaccines taken before screening, biological/non-biological disease modifying anti-rheumatic drugs [DMARDs] and immunosuppressive agents taken for patients with symptomatic PMR disease are to be reported in the case report form (CRF) pages. The highest dose and duration of all corticosteroids taper taken for PMR from 24 weeks prior to screening up to baseline and all corticosteroids taken during the study are also reported.

All medications will be coded using the World Health Organization – Drug Dictionary (WHO-DD) version currently in effect at Sanofi at the time of database lock.

- Prior medications are those the patient used prior to first investigational medicinal product (IMP) intake. Prior medications can be discontinued before first dosing or can be ongoing during treatment phase.
- Concomitant medications are any treatments received by the patient concomitantly to the IMP, during the treatment-emergent adverse event period (Section 2.1.4).

A given medication can be classified as a prior medication, as a concomitant medication, and/or as a post-treatment medication.

For patients who experience a disease flare and are in need of rescue therapy, study patients prescribed with rescue corticosteroids (non-investigational medicinal product) per investigator judgement during the course of the study will continue administration of subcutaneous (SC) IMP (sarilumab or its matching placebo) for the full duration of the 52-week treatment period.

Any technical details related to computation, dates, imputation for missing dates are described in Section 2.5.

2.1.3 Efficacy endpoints

2.1.3.1 Primary efficacy endpoint(s)

The primary endpoint is the proportion of patients achieving sustained remission at Week 52, where the sustained remission at Week 52 is defined as having met all of the following parameters:

1. Achievement of disease remission not later than Week 12,
2. Absence of disease flare from Week 12 through Week 52,
3. Sustained reduction of CRP (to <10 mg/L, with an absence of successive elevations to ≥10 mg/L) from Week 12 through Week 52, and
4. Successful adherence to the prednisone taper from Week 12 through Week 52.

Components of primary endpoint are described as below:

1. **Achievement of disease remission not later than Week 12**

Disease remission is defined as resolution of signs and symptoms of PMR, and normalization of CRP (<10 mg/L). Note that a single CRP elevation (≥10 mg/L) is not considered absence of remission unless CRP remained elevated (≥10 mg/L) at the next study visit.

The status of normalization of CRP (<10 mg/L) will be determined based on the last two non-missing post-baseline CRP values measured up to Week 12. If at least one of them is <10 mg/L, then it is considered as normalization of CRP.

Signs and symptoms of disease

Evaluation of clinical signs and symptoms by the Efficacy Assessor at every study visit according to the schedule of assessment will include the following:

- Morning stiffness and/or pain, in the neck, shoulder and/or hip girdles
- Limited range of motion of the shoulders and/or hip girdles
- Constitutional symptoms, such as fatigue, weight loss and low-grade fever
- Other features judged by the clinician-investigator to be consistent with a PMR flare

The resolution of signs and symptoms of PMR is identified in eCRF "PMR Assessment" page with the question "Based upon the patient's relevant medical history and the clinical assessment and responses provided above, does the patient have active PMR?" ticked as "No". The resolution of signs and symptoms of PMR by Week 12 will be determined based on the last non-missing post-baseline PMR assessment measured up to Week 12.

Patients who took rescue CS due to active PMR prior to Week 12 or who permanently withdraw from the study treatment prior to Week 12 will be considered as not achieved disease remission by Week 12.

Note that during the initial 12 weeks of prednisone taper, treatment for one flare before Week 12 is permitted if it can be successfully treated with a low dose (≤5 mg/day) prednisone add-on taper regimen (completed prior to Week 12) and provided that all other sustained remission parameters are met.

2. **Absence of disease flare from Week 12 through Week 52.**

Flare is defined as either 1) recurrence of signs and symptoms attributable to active PMR plus an increase in CS dose due to PMR, or 2) elevation of ESR attributable to active PMR plus an increase in CS dose due to PMR. Increase in CS dose is defined as:

- Any dose increase during the protocol defined steroid taper, or
- Re-initiation of prednisone therapy after the protocol defined taper has been completed.

Patients with disease flare are identified as those with the reason for treatment ticked "rescue therapy" in eCRF "Steroid Medication" page and to the question "If Rescue Therapy, please specify" ticked either "Active PMR" or "Elevated ESR Attributable to Active PMR".

3. **Sustained reduction of CRP (to <10 mg/L, with an absence of successive elevations to ≥10 mg/L) from Week 12 through Week 52**

The status of normalization of CRP from Week 12 through Week 52 will be determined based on the CRP values measured at Week 16, Week 20, Week 24, Week 32, Week 40 and Week 52. If there are two or more consecutive visits with CRP≥10 mg/L, then it is categorized as no normalization of CRP. Intermittent missing CRP at Week 16, Week 20, Week 24, Week 32, Week 40 or Week 52 will be imputed using last value carried forward (LOCF) approach up to the last non-missing measurement.

4. **Successful adherence to the prednisone taper from Week 12 through Week 52**

Successful adherence to the prednisone taper from Week 12 through Week 52 is defined as patients who did not take rescue therapy from Week 12 through Week 52 and any excess prednisone (beyond the per protocol CS tapering regimen) with a cumulative dose of ≤100 mg (or equivalent), such as those employed to manage AE not related to PMR. The cumulative dose of excess prednisone use will be counted from baseline to Week 52 (does not include add-on prednisone taken prior to Week 12).

In addition, patients who permanently withdraw from the study treatment prior to Week 52 will be considered as not having achieved component 2, 3, or 4 for the primary endpoint.

Blinding of CRP and ESR assessment results

Investigators, including Efficacy Assessors, and patients will remain blinded to CRP and ESR results (except screening and baseline). Safety Assessors will also be blinded on post baseline CRP but will have access to ESR results.

ESR kits will be provided by the central laboratory while the test will be performed locally at site, results (the distance in millimeters (mm) that red blood cells has descended in 1 hour) will be blinded to Investigators including Efficacy Assessors and staff directly involved in efficacy assessment of study patients.

2.1.3.2 Secondary efficacy endpoint(s)

The following efficacy variables will be assessed:

- Components of sustained remission composite measure at Week 52
- Total cumulative corticosteroid (including prednisone) dose over during the treatment period
- Time to first PMR flare during the treatment period
- GTI Cumulative worsening score (CWS) and Aggregate improvement score (AIS) at Week 52.

Below are descriptions of these variables. Details related to imputation of GTI scores are described in Section 2.5.2.

Components of sustained remission composite measure at Week 52

1. Proportion of patients who achieved disease remission by Week 12.
2. Proportion of patients who have no disease flare from Week 12 through Week 52.
3. Proportion of patients who have normalization of CRP (to <10 mg/L, with an absence of successive elevations to ≥10 mg/L) from Week 12 through Week 52.
4. Proportion of patients who successfully adhere to the prednisone taper from Week 12 through Week 52.

Total cumulative corticosteroid (including prednisone) dose during the treatment period

The total cumulative corticosteroid includes prednisone taper regimen per protocol, add-on prednisone prior to Week 12 or use of commercial prednisone during the study treatment period such as those employed to manage AE not related to PMR, and rescue corticosteroids dose prescribed for rescue therapy over the treatment period for PMR disease only.

Time to first PMR flare during the treatment period

The duration (in days) from randomization to first PMR flare after clinical remission and up to 52 weeks. Patients who never achieve remission will be censored at randomization day; and those who achieved clinical remission and never flared will be censored at the end of treatment assessment date up to Week 52.

The definition of PMR flare is provide in Section 2.1.3.1 primary efficacy endpoint.

Clinical remission is defined as resolution of signs and symptoms and normalization of CRP (<10 mg/L). Intermittent missing CRP at visit will be imputed using last value carried forward (LOCF) approach. If missing signs and symptoms, then clinical remission is not achieved.

Glucocorticoid toxicity index (GTI) scores and components

The GTI is intended to capture glucocorticoid (GC)-related morbidity and GC-sparing ability of other therapies (2).

The GTI is composed of two components, the Composite GTI and the Specific List (Table 2).

The Composite GTI consists of nine domains (body mass index (BMI), glucose tolerance, blood pressure, Hyperlipidemia, Bone Health (BMD), steroid myopathy, Skin steroid-related Toxicity, Neuropsychiatric – steroid related symptoms and infection) on common steroid-related adverse events anticipated to occur in ≥5% of patients treated in a clinical trial involving glucocorticoids. The Specific List consists of 23 items (11 domains) that are mainly less common but serious adverse effects, serving as a complementary tool to the Composite GTI to ensure that all important elements of glucocorticoid toxicity were captured.

The domains of the Composite GTI and the specific list of the GTI will be assessed at baseline, Week 12, Week 24, Week 40 and Week 52, except for bone density (BMD) which will be assessed at Baseline and Week 52 only.

For patients who withdraw from the study prior to Week 24, BMD will not be measured at the EOT visit as no substantial changes are expected within 6 months of treatment. As a result, missing BMD score will be imputed as 'no change'. The BMD measured at the end of treatment (EOT) visit will be included in the definition of cumulative GTI scores (i.e., CWS and AIS as defined below) at Week 52 regardless when the EOT visit occurred.

Scoring of the Glucocorticoid Toxicity Index (GTI)

Domain-specific score for each patient at a visit is defined as the weight assigned based on the change in severity level as described for each domain in Table 2.

Composite GTI score for each patient at a visit is the sum of 9 domain-specific scores at each visit.

Cumulative GTI score is the sum of composite GTI score across visits. Two cumulative GTI scores, CWS and AIS, will be defined.

- **Cumulative worsening score (CWS)**

  CWS is designed to assess cumulative GC toxicity, regardless of whether the toxicity has lasting effects or is transient. New toxicities that occur are added, but toxicities that appear to resolve on follow-up are not removed. Thus, the CWS serves as a lasting record of GC toxicity observed and can only increase or remain the same over time.

  CWS will be lower in the treatment arm compared to the placebo arm at the end of trial, if the treatment is effective at decreasing GC toxicity over time.

- **Aggregate improvement score (AIS)**

  Patients are anticipated to have some GC toxicity at baseline, the AIS would be important in establishing that the new therapy is effective at diminishing any baseline GC toxicity over time. With the AIS, toxicities can be removed if improvement occurs, and added if worsening occurs.

  AIS decrease over time means that the treatment is effective in decreasing GC toxicity level, compared to placebo which may show fluctuation of CIS over time. Negative score reflects that toxicities present at baseline (or occurring during the trial) resolved over the treatment period.

The steps to calculate the CWS and AIS at Week 52 are described below.

Step 1: Calculate domain-specific worsening score (for computing CWS in Step 2) and improvement score (for computing AIS in Step 2) at each visit.

For Domains without sub-items (Body Weight, Glucose Metabolism, Blood Pressure, Lipid Metabolism, Bone Health and Steroid Myopathy),

a) For worsening score, only consider positive domain-specific score at each visit. For domain with a negative domain-specific score, the worsening score will be 0.

b) For improvement score, all domain-specific scores (could be positive or negative) at each visit will be considered.

The Skin and Neuropsychiatric domains both have sub-items. For the Skin domain, these are Acneiform Rash, Easy Bruising, Hirsutism, Atrophy/Striae, and Erosions/tears/ulcerations. For the Neuropsychiatric domain, these are Insomnia, Depression, Mania, and Cognitive Impairment.

a) For worsening score, only the item with the highest weight is scored for any GTI interval (typically three months) with the CWS. As an example, if neither insomnia nor depression were present at the baseline visit but there is now mild Insomnia and moderate depression present at follow-up, then only the moderate depression is scored (+63 points).

b) For improvement score, improvement as well as worsening can be recorded. Because it is conceivable that one item might improve while another worsens, the item of greatest improvement (highest absolute weight) and the item of greatest worsening (highest weight) are recorded for given GTI interval.

For Infection Domain, the worsening score and improvement score handle the scoring of infections differently, because the CWS and AIS reflect reciprocal measures of GC toxicity.

a) For worsening score, the most severe infection in every GTI interval is scored.

b) For improvement score, only the most severe infection occurring over the course of the entire trial is scored.

Step 2: Calculate composite worsening score and improvement score at each visit by summing the 9 domain-specific score.

a) For composite worsening score, sum the 9 domain-specific worsening scores.

b) For composite improvement score, sum the 9 domain-specific improvement scores.

Step 3: Calculate cumulative GTI scores by summing composite scores across visits.

a) For CWS at Week 52, summing composite GTI worsening score at Week 12, Week 24, Week 40, and Week 52.

b) For AIS at Week 52, summing composite GTI improvement score at Week 12, Week 24, Week 40, and Week 52.

Table 2 – The Glucocorticoid Toxicity Index

Composite GTI	Item weight	Specific List[1]
1.BMI		
Decrease by >/= 5 BMI units	-36	Major increase in BMI
Decrease by >2 but <5 BMI units	-21	
No significant change (+/- 2 BMI units)	0	
Increase of >2 to <5 BMI units	21	
Increase of 5 or more BMI units	36	

Composite GTI	Item weight	Specific List[1]
2.Glucose tolerance		
Improvement in HbA1c AND decrease in medication	-44	Diabetic retinopathy
Improvement in HbA1c OR decrease in medication	-32	Diabetic nephropathy
No significant change	0	Diabetic neuropathy
Increase in HbA1c OR increase in medication	32	
Increase in HbA1c AND increase in medication	44	
3.Blood pressure		
Improvement in BP AND decrease in medication	-44	Hypertensive emergency
Improvement in BP OR decrease in medication	-19	Posterior reversible encephalopathy syndrome
No significant change	0	
Increase in BP OR increase in medication	19	
Increase in BP AND increase in medication	44	
4.Hyperlipidemia		
Decrease in LDL AND decrease in medication	-30	
Decrease in LDL OR decrease in medication	-10	
No significant change	0	
Increase in LDL OR increase in medication	10	
Increase in LDL AND increase in medication	30	
5.Bone Health (BMD)		
Increase in BMD (gain of more than 3%)	-29	Major decrease in bone density
No significant change in BMD (+/- 3%)	0	Insufficiency fracture
Decrease in BMD (loss of more than 3%)	29	
6.Steroid myopathy		
Moderate weakness to none	-63	Severe steroid myopathy
Moderate to Mild weakness	-54	
Mild weakness to none	-9	
No significant change	0	
None to Mild weakness (without functional limitation)	9	
Mild to Moderate weakness	54	
None to Moderate weakness (with functional limitation)	63	
7.Skin steroid-related Toxicity		
Decrease in Skin Toxicity – Moderate to None	-26	Severe skin toxicity
Decrease in Skin Toxicity – Moderate to Mild	-18	
Decrease in Skin Toxicity – Mild to None	-8	
No significant change	0	
Increase in Skin Toxicity – None to Mild	8	
Increase in Skin Toxicity – Mild to Moderate	18	
Increase in Skin Toxicity – None to Moderate	26	
8.Neuropsychiatric – steroid related symptoms		
Decrease in NP Toxicity – Moderate to None	-74	Psychosis
Decrease in NP Toxicity – Moderate to Mild	-63	GC-induced violence
Decrease in NP Toxicity – Mild to None	-11	

Composite GTI	Item weight	Specific List[1]
No significant change	0	Other severe neuropsychiatric symptoms
Increase in NP Toxicity – None to Mild	11	
Increase in NP Toxicity – Mild to Moderate	63	
Increase in NP Toxicity – None to Moderate	74	
9.Infection		
No infection	0	Grade IV infection
Oral or vaginal candidiasis or non-complicated zoster (<Grade3)	19	Grade V infection
Grade 3, 4, or 5 infection	93	

[1] Maximum weight is given for specific list events in Blood pressure, Steroid myopathy, Skin toxicity, Neuropsychiatric and Infection domains at each visit. For the rest of domains, specific list events will not affect the corresponding domain specific score.

2.1.3.3 Exploratory endpoint(s)

Following recommendation from the meeting with Japan Health Authority (PMDA) in April 2019, exploratory analysis will be performed for disease activity in PMR, quantified by the PMR activity score (PMR-AS).

PMR-AS is calculated as: CRP (mg/dl) + VAS p (0-10) + VAS ph (0-10) + (MST (min) x 0.1) + EUL (3-0).

Visual analogue scale for pain (VAS p) is the patient's assessment of pain from HAQ-DI assessment, and the visual analogue scale for physician's assessment (VAS ph) is the physician's global assessment of disease activity, MD-VAS. The exact duration of morning stiffness (MST) and the ability to elevate the upper limbs (EUL: 3 = none, 2 = below shoulder girdle, 1 = up to shoulder girdle, 0 = above shoulder girdle) are captured in the "PMR assessment" eCRF. CRP value is from laboratory measurements.

PMR-AS will be assessed at baseline, Week 12, Week 24, and Week 52.

Note that only binary MST duration was collected (≤ 45mins OR > 45mins) prior to the collection of exact MST duration implemented in June 2019. Missing exact MST duration will be imputed using the median of MST duration at visit for each MST category (≤45mins and >45mins).

2.1.4 Safety endpoints

The safety analysis will be based on the reported adverse events (AEs), serious AEs (SAEs) and adverse events of special interest (AESIs) including but not limited to neutropenia, thrombocytopenia, elevations in hepatic enzymes leading to permanent discontinuation and tuberculosis, and other opportunistic infections. Other safety information, such as clinical laboratory data and vital signs will also be assessed.

Observation period

The observation period will be divided into 4 epochs:

- The **screening** epoch is defined as the time from the signed informed consent date up to the time prior to first dose of SC IMP.
- The **treatment** epoch is defined as the time from the first dose of the IMP to the last dose of the SC IMP in the DB period +13 days.
- The **residual treatment** epoch is defined as the time from the end of TREATMENT epoch to the last dose of the SC IMP +60 days.
- The **post-treatment** epoch is defined as the time after the last dose of SC IMP +60 days.

The rationale of having "last dose of SC IMP +60 days" as the end of follow-up in the residual treatment epoch is to capture the events observed within an approximately 5 half-lives of the last sarilumab SC IMP.

The treatment-emergent adverse event (TEAE) period includes both **treatment** and **residual treatment** epochs.

2.1.4.1 Adverse events variables

Adverse event observation period

- Pre-treatment AEs are AEs that developed or worsened or became serious from the signed informed consent date up to first administration of SC IMP
- TEAEs are AEs that developed or worsened or became serious during the TEAE period
- Post-treatment AEs are AEs that developed or worsened or became serious during the post-treatment period

All AEs (including SAEs and AESIs) will be coded to a "LLT", "PT", "HLT", "HLGT", and associated primary "SOC" using the version of MedDRA currently in effect at Sanofi at the time of database lock.

The occurrence of AEs (including SAEs and AESIs) will be recorded from the time of signed informed consent until the end of the study.

Any SAE brought to the attention of the Investigator at any time after the end of the study for the patient and considered by him/her to be caused by the IMP with a reasonable possibility, should be reported to the monitoring team.

Adverse event of special interest (AESI)

Adverse events of special interest will be flagged in the database. The AESIs and its search criteria are listed in Table 3 below:

Table 3 – Protocol defined AESIs and search criteria

AESI flag	Search criteria
Clinically significant infections including: • Opportunistic infections • Tuberculosis • Infections requiring prolonged medication • Infections requiring parenteral treatment	Reported by the investigators on the AE CRF page with AESI box ticked.
Lab abnormalities including: • ALT increase leading to permanent discontinuation • ANC decrease leading to permanent discontinuation • Thrombocytopenia leading to permanent discontinuation	
ALT >= 3xULN	
Pregnancy	CRF checkbox: Pregnancy
Symptomatic Overdose	CRF checkbox: Symptomatic Overdose

2.1.4.2 Deaths

The deaths per the observation period will be summarized.

- Death on-treatment: deaths occurring during the treatment period (i.e., treatment Epoch)
- Death during follow-up: deaths occurring during the residual treatment period
- Death post-treatment: deaths occurring during the post-treatment period

2.1.4.3 Laboratory safety variables

Clinical laboratory data will include hematology, clinical chemistry, and urinalysis. Clinical laboratory values will be converted to standard international units; international units will be used in all listings and tables. In addition, the lipids parameters will also be summarized in US units.

Hematology and chemistry data will be collected at screening, baseline, Weeks 4, 12, 24, 40, and 52, and/or early termination unless otherwise specified. The laboratory parameters will be classified as follows:

- Hematology
  - **Red blood cells and platelets and coagulation:** hemoglobin, hematocrit, platelet count, red blood cell (RBC) count, and RBC morphology (if RBC count is abnormal).
  - **White blood cells:** white blood cell (WBC) count and WBC differential (neutrophils, lymphocytes, monocytes, basophils, eosinophils).
- Clinical chemistry
  - **Metabolism:** fasting lipids (total cholesterol, high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), ratio of HDL-C versus LDL-C (HDL-C:LDL-C), triglycerides), total proteins, CRP, fasting glucose, and hemoglobin A1c (HbA1c).
  - **Electrolytes:** sodium, potassium, chloride, calcium, phosphate, and bicarbonate.

- **Renal function:** creatinine, creatinine clearance, uric acid, and blood urea nitrogen (BUN).
- **Liver function:** alanine aminotransferase (ALT)/SGPT, aspartate aminotransferase (AST)/SGOT, alkaline phosphatase (ALP), albumin, total bilirubin, conjugated and unconjugated bilirubin, and lactate dehydrogenase (LDH).
- **Pregnancy test:** Serum β-hCG for all women of childbearing potential (screening only).
- **TB screen:** QuantiFERON®-TB Gold evaluation.
- **Hepatitis screen:** Hepatitis B and C serology, human immunodeficiency virus serology.

Urine samples will be collected as follows:

- **Urinalysis – dipstick:** pH, specific gravity, blood, glucose, protein, nitrates, leukocyte esterase, and bilirubin. If any parameter on the dipstick is abnormal, a urine sample was to be sent to the central laboratory for testing.

A urine pregnancy test was done for all women of childbearing potential at baseline, all visits up to Week 52 or early termination.

2.1.4.4 Vital signs variables

Vital signs variables include: heart rate, systolic blood pressure, and diastolic blood pressure, height, weight, and body temperature.

2.1.4.5 Antinuclear antibodies

Antinuclear antibodies (ANA) will be measured at baseline and EOT visits.

2.1.5 Immunogenicity endpoints

The immunogenicity blood samples will be collected before injection during the same visit according to the Study Flow Chart (Appendix A).

Samples analyzed in the antidrug antibody (ADA) assay will be categorized as either positive or negative, and in the case of a positive result will be further characterized as either neutralizing or non-neutralizing.

ADA positive patient is defined as patient with at least one treatment-emergent or treatment-boosted ADA positive sample during the TEAE period, where

- Treatment-emergent ADA positive patient is defined as a patient with non-positive assay (meaning negative or missing) response at baseline but with a positive assay response during the TEAE period.

- Treatment-boosted ADA positive patient is defined as a patient with a positive ADA assay response at baseline and with at least a 4-fold increase in titer compared to baseline during the TEAE period.

ADA negative patient is defined as patient without a treatment-emergent or treatment-boosted ADA positive sample during the TEAE period.

A treatment-emergent positive response is further classified as persistent or transient.

Persistent ADA response: treatment-emergent ADA detected at two or more consecutive sampling time points during the TEAE period, where the first and last ADA positive samples are separated by a period of at least 16 weeks. Also persistent in case last sample analyzed is positive.

Transient ADA response: A treatment-emergent response that is not determined to be persistent.

2.1.6 Pharmacokinetic variables

The blood samples for the determination of functional sarilumab concentrations in serum will be collected before injection during the same visit according to the Study Flow Chart (refer to Section 1.2 in the study protocol) and at or near the onset and completion of the occurrence of a SAE. In addition, post-dose sample will be taken 4-7 days after the Week 24 (Visit 9).

2.1.7 Pharmacodynamic/genomics endpoints

Blood samples (DNA and genetic RNA) will be collected from each consenting patient to participate in the pharmacogenomics substudy. Details of any analysis will be described in a separate plan.

2.1.8 Other efficacy endpoints

Quality-of-life endpoints

Other efficacy endpoints in the clinical outcome assessments will be summarized. Patients are asked to complete the following patient-reported outcome (PRO) questionnaires described below at Visit 2 (baseline visit), Visit 6 (Week 12), Visit 9 (Week 24), and Visit 12 (Week 52).

Patient-Reported Outcomes

Short Form 36 Version 2 (SF-36 V2)

The SF-36 V2 is a multi-purpose, short-form health survey with 36 questions. It yields scores for eight domains (Physical Functioning, Role-Physical, Bodily pain, General health, Vitality, Social Functioning, Role-Emotional, and Mental Health, each scale is scored from 0 to 100 where higher scores indicate better health and well-being), as well as two summary measures of physical and mental health: the Physical Component Summary (PCS) and Mental Component Summary (MCS).

The scoring process is summarized below:

- Enter item response data.
- Recode item response values.
- Determine health domain scale raw scores.
- Transform health domain scale raw scores to 0 to 100 scores.
- Transform health domain scale 0 to 100 to normal-based scores.
- Score physical and mental component summary measures.

The following table shows the construction and summary measures of the SF-36 scales:

Table 4 – SF-36 V2 measurement model

Summary measures	Scales	Items
Physical Component Summary (PCS)*	Physical Functioning (PF)	3a, 3b, 3c, 3d, 3e, 3f, 3g, 3h, 3i, 3j
	Role-Physical (RP)	4a, 4b, 4c, 4d
	Bodily Pain (BP)	7, 8
	General Health (GH)	1, 11a, 11b, 11c, 11d
Mental Component Summary (MCS)*	Vitality (VT)	9a, 9e, 9g, 9i
	Social Functioning (SF)	6, 10
	Role-Emotional (RE)	5a, 5b, 5c
	Mental Health (MH)	9b, 9c, 9d, 9f, 9h

*MCS and PCS derived from the eight scales

The score of each of the 36 items is collected in CRF. Appendix H provides the SAS code to calculate the eight scales, the two summary measure scores and the standardized summary scores.

The PCS and MCS summary measure scores will be computed if at least 50% of the component scales are available. The scale scores will be computed if at least 50% of items are available within the corresponding scale. The missing items will be imputed by the mean of available items.

Change from baseline in SF-36 scores (physical component summary score and mental component summary score as well as the eight domains) at each visit will be analyzed.

Functional Assessment of Chronic Illness Therapy Fatigue (FACIT-Fatigue)

The FACIT-Fatigue will be used to assess fatigue. The FACIT-Fatigue is a 13-item questionnaire rated 0 to 4 originally developed to measure fatigue in patients with cancer and is widely used in rheumatoid arthritis patients to demonstrate good consistency and sensitivity to change. The patient will be asked to answer 13 questions rated 0 to 4 (0 = not at all, 1 = a little bit, 2 = somewhat, 3 = quite a bit, 4 = very much). The fatigue scale has 13 items, with 52 as the highest possible score. A higher score in the fatigue scale corresponds to a lower level of fatigue and indicates better Quality of Life.

To calculate the FACIT-fatigue score, the response scores on negatively phrased questions are reversed and then the 13 item responses are added. Eleven items with responses have their scores reversed (item score = 4 – response, if the response is not missing), and two items (items 7 and 8) have their responses unchanged (Appendix F). All items are added so that higher scores correspond to less fatigue. In cases where individual questions are skipped, scores are prorated using the average of other answers in the scale.

FACIT-Fatigue = 13 * [sum (reversed items) + sum (items 7 and 8)] / number of answered items

EuroQol 5 Dimension 3 Level (EQ-5D-3L)

The EQ-5D-3L is a standardized, generic measure of health outcome (3). EQ-5D is designed for self-completion by patients. Instructions to respondents are included in the questionnaire. The EQ5D is specifically included to address concerns regarding the health economic impact of RA, which have been considered in cost effectiveness arguments. The EQ-5D-3L comprises five questions on mobility, self-care, pain, usual activities, and psychological status with three possible answers for each item (1 = no problem, 2 = moderate problems, 3 = severe problems) and a vertical visual analog scale that allows the patients to indicate their health state today that can range from 0 (worst imaginable) to 100 (best imaginable).

The 5 dimensional 3-level systems are converted into a single index utility score. Values for the 243 theoretically possible health states defined by the EuroQol classification are calculated using a regression model and weighted according to the social preferences of the GB population. The minimum value for the single index utility score is -0.594, which corresponds to a level 3 (severe problems) for mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The maximum value for this index is 1.0, which corresponds to a full health (level 1 [no problem] for mobility, self-care, usual activities, pain/discomfort and anxiety/depression).

The VAS records the respondent's self-rated health on a vertical visual analogue scale. The VAS "thermometer" has endpoints of 100 (Best imaginable health state) at the top and 0 (Worst imaginable health state) at the bottom. This information can be used as a quantitative measure of health outcomes as judged by the individual respondents. EQ 5D self-reported VAS data generates information on the self-perceived overall health-related quality of life.

Appendix G provides the SAS code to derive the index utility score using GB based population.

Health Assessment Questionnaire Disease Index (HAQ-DI)

- ***Patient's Assessment of Physical function***

The HAQ-DI is a standardized questionnaire developed to assess physical functional status in adults with arthritis but is now commonly used among many rheumatologic conditions. The HAQ-DI, with the past week as the time frame, focuses on whether the respondent "is able to…" do the activity and covers eight categories in 20 items: dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities, for which there are at least two questions by category. The four responses for the HAQ-DI questions are graded as follows: without any difficulty=0;

with some difficulty=1; with much difficulty=2; and unable to do=3. To calculate the Standard HAQ-DI Score (With Aids/Devices), there are 3 steps:

1. Sum the eight category scores by using the highest sub-category score from each category.

- For example, in the category EATING there are three sub-category items. A patient responds with a 1, 2, and 0, respectively; the category score is 2.

2. Adjust for use of aids/devices and/or help from another person when indicated.

- When there are NO aids or devices or help indicated for a category, the category's score is not modified.
- When aids or devices or help ARE indicated by the patient:
  - Adjust the score for a category by increasing a 0 or a 1 to a 2.
  - If a patient's highest score for that sub-category is a 2 it remains a 2, and if a 3, it remains a 3.
  - The data entered at field "Other specify" will not be used for score adjustment.

3. Divide the summed category scores by the number of categories answered (must be a minimum of 6) to obtain a HAQ-DI score of 0 to 3 (3=worst functioning).

A HAQ-DI score cannot be calculated validly when there are scores for less than six of the eight categories. If there are less than 6 categories completed, a HAQ-DI cannot be computed, whether the missing categories are due to missing values or they do not apply to the respondent. HAQ-DI disability index scoring ranges between 0 and 3 with higher score indicating greater disability. A high HAQ-DI score has been found to be a strong predictor of morbidity and mortality in RA. A 0.22 unit difference is considered clinically meaningful.

In addition to the above, the HAQ-DI has two additional questions to measure pain and global assessment described below.

- **Patient's Assessment of Pain**

Patients will be requested to indicate their pain intensity due to PMR using a 100 mm horizontal VAS to the question *"How much pain have you had IN THE PAST WEEK?"*, where 0 is considered "No pain" and 100 "the worst pain you can imagine".

To obtain the patient score, using a metric ruler, measure the distance in centimeters from the left anchor (at base zero) to their mark and multiply by 0.2. This converts the number of centimeters into the appropriate score and yields a score from 0 to 3. For example, the mark is at 8 centimeters – 8 x 0.2 = a pain score of 1.6.

In case a 0-100 mm scale is used, pain severity coding translations follow below:

PAIN SEVERITY CODING TRANSLATIONS

Measurement (m) = Score	Measurement (m) = Score [continue]
0 = 0	7.8 – 8.2 = 1.6
0.1 – 0 .7 = 0.1	8.3 – 8.7 = 1.7
0.8 – 1.2 = 0.2	8.8 – 9.2 = 1.8
1.3 – 1.7 = 0.3	9.3 – 9.7 = 1.9
1.8 – 2.2 = 0.4	9.8 – 10.2 = 2.0
2.3 – 2.7 = 0.5	10.3 – 10.7 = 2.1
2.8 – 3.2 = 0.6	10.8 – 11.2 = 2.2
3.3 – 3.7 = 0.7	11.3 – 11.7 = 2.3
3.8 – 4.2 = 0 .8	11.8 – 12.2 = 2.4
4.3 – 4.7 = 0 .9	12.3 – 12.7 = 2.5
4.8 – 5.2 = 1.0	12.8 – 13.2 = 2.6
5.3 – 5.7 = 1.1	13.3 – 13.7 = 2.7
5.8 – 6.2 = 1.2	13.8 – 14.2 = 2.8
6.3 – 6.7 = 1.3	14.3 – 14.7 = 2.9
6.8 – 7.2 = 1.4	14.8 – 15.0 = 3.0

- **Patient's Global Assessments of Disease Activity**

Patient's global assessments of their current disease activity will be rated on an anchored 100 mm horizontal VAS, with the given instruction to *"Please rate how well you are doing on a scale of 0 to 100 (0 represents "very well" and 100 represents "very poor" health)"*.

It is scored similarly to the VAS Pain Scale.

Clinician-Reported Outcomes

Physician's Global Assessments of Disease Activity – Visual Analog Scale [MD-VAS]

Physician's global assessments of the patient's current disease activity will be assessed on an anchored 100 mm horizontal VAS where 0 is considered the best disease activity (no disease activity) and 100 the worst (most disease activity).

2.2 DISPOSITION OF PATIENTS

This section describes patient disposition for both patient status and the patient populations.

Screened patients are defined as all patients who signed the informed consent.

Randomized patients consist of all patients who signed informed consent with a treatment kit number allocated and recorded in the Interactive Voice Response System/Interactive Web Response System database, regardless of whether the treatment kit was used or not.

For patient status in the DB treatment period, the total number of patients for each one of the following categories will be presented in the clinical study report (CSR) using a summary table:

- Screened patients
- Screen failure patients and reasons for screen failure
- Non-randomized but treated patients
- Randomized patients
- Randomized but not treated patients
- Randomized and treated patients
- Patients who completed 52-week SC IMP study treatment period
- Patients who discontinued from 52-week study treatment period and the reasons for permanent SC IMP treatment discontinuation
- Patients who completed the study
- Patients who discontinued the study and the reasons for study discontinuation
- Patients who took CS rescue therapy
- Patients who took add-on prednisone
- Subject status at end of study visit

For all categories of patients (except for the screened and non-randomized categories) percentages will be calculated using the number of randomized patients as the denominator. Reasons for SC IMP treatment or study discontinuation, such as if it is related to COVID-19 will be presented in tables giving numbers and percentages by treatment group. Patient disposition by visit according to trial impact due to COVID-19 (visits that were delayed, not done, partially done on site, or partially done by phone, or not impacted) will also be provided.

A patient is considered as lost to follow-up at the end of the study if he/she is not assessed at the end of study visit.

All critical or major deviations potentially impacting efficacy analyses, randomization and drug dispensing irregularities and other major/critical deviations will be summarized in tables giving numbers and percentages of deviations by treatment groups. Critical or major deviations according to COVID19 impact will also be summarized.

Additionally, the analysis populations for safety, efficacy, PK and immunogenicity will be summarized in a table by patient counts based on the randomized populations (See definitions in Section 2.3).

- Efficacy population
- Safety population
- PK population

- ADA population

2.2.1 Randomization and drug dispensing irregularities

Randomization and drug-dispensing irregularities occur whenever:

1. A randomization is not in accordance with the protocol-defined randomization method, such as a) an ineligible patient is randomized, b) a patient is randomized based on an incorrect stratum, c) a patient is randomized twice, or d) in a dynamic randomization scheme the treatment assignment is, in fact, not random, due to a computer program error.

   OR

2. A patient is dispensed an IMP kit not allocated by the protocol-defined randomization, such as a) a patient at any time in the study is dispensed a different treatment kit than as randomized (which may or may not contain the correct-as-randomized IMP), or b) a nonrandomized patient is treated with IMP reserved for randomized patients.

Randomization and drug-dispensing irregularities will be monitored throughout the study and reviewed on an ongoing basis.

All randomization and drug-dispensing irregularities will be considered as protocol deviations and documented in the CSR. If the number of irregularities is large enough to make a tabular summary useful, the irregularities will be categorized and summarized among randomized patients (number and percentages). Nonrandomized, treated patients will be described separately.

Randomization and drug-dispensing irregularities to be prospectively identified include but are not limited to:

Table 5 – Randomization and Drug Allocation Irregularities

Randomization and Drug Allocation Irregularities
Kit dispensation without IVRS transaction
Erroneous kit dispensation
Kit not available
Randomization by error
Subject randomized twice
Forced randomization
Subject switched to another site

IVRS=Interactive Voice Response System

2.3 ANALYSIS POPULATIONS

The randomized population includes any patient who has been allocated to a randomized treatment regardless of whether the treatment kit was used.

Patients treated without being randomized will not be considered randomized and will not be included in any efficacy population.

For any patient randomized more than once, only the data associated with the first randomization will be used in any analysis population. The safety experience associated with any later randomization will be assessed separately.

The safety experience of patients treated and not randomized will be reported separately, and these patients will not be in the safety population.

2.3.1 Efficacy populations

The primary analysis population will be the ITT population as defined in Section 2.3.1.1.

2.3.1.1 Intent–to-treat population

The intent-to-treat (ITT) population is the randomized population analyzed according to the treatment group allocated by randomization.

2.3.2 Safety population

The safety population is defined as randomized population who received at least 1 dose or part of a dose of the SC IMP, analyzed according to the treatment actually received.In addition:

- Nonrandomized but treated patients will not be part of the safety population; however, their safety data will be presented separately
- Randomized patients for whom it is unclear whether they took the IMP will be included in the safety population as randomized
- For patients receiving more than one SC IMP during the trial, the treatment group allocation for as-treated analysis will be the treatment group for which the patient received the majority of doses. In such cases, individual patient listings will be provided for each treatment group with treatment start and end date for each treatment exposed.

2.3.3 Pharmacokinetics and ADA populations

The PK population will consist of all patients in the safety population with at least one post-dose, non-missing serum sarilumab concentration value. The ADA population will consist of all patients in the safety population with at least one post-dose, evaluable ADA sample. Patients will be analyzed according to the treatment actually received.

2.3.4 Population with and without trial impact (disruption) due to COVID-19

The population without trial impact (disruption) due to COVID-19 is defined as patients without any critical or major deviation related to COVID-19 or did not permanently discontinue study treatment due to COVID-19. Population with trial impact due to COVID-19 are patients with any critical or major deviation related to COVID-19 or permanently discontinue study treatment due to COVID-19. These populations will be used to summarize treatment exposure, treatment compliance, subgroup analysis of primary endpoint and safety only if > 10% patients are impacted by the COVID-19 pandemic.

2.4 STATISTICAL METHODS

2.4.1 Demographics and baseline characteristics

Continuous data will be summarized using the number of available data, mean, standard deviation (SD), median, minimum, and maximum for each treatment group. Categorical and ordinal data will be summarized using the number and percentage of patients in each treatment group.

Parameters will be summarized based on the randomized populations.

Parameters described in Section 2.1.1 will be summarized by treatment group and overall using descriptive statistics.

Medical and surgical history will be summarized by SOC and PT sorted by internationally agreed order of SOC and by the decreasing frequency of PT within SOC in total group (including all the randomized patients). P-values on demographic and baseline characteristic data will not be calculated.

No specific description of the safety parameters will be provided at baseline. If relevant, the baseline values will be described along with each safety analysis.

2.4.2 Prior or concomitant medications

The prior and concomitant medications will be summarized based on the randomized populations as follows:

- Prior medications discontinued before the first dose of IMP in the DB period
- Prior medications that continued at the time of the first dose of IMP in the DB period
- Concomitant medications

Medications will be summarized by treatment group according to the WHO-DD dictionary, considering the first digit of the anatomic category (ATC) class (anatomic category) and the generic names. All ATC codes corresponding to a medication will be summarized. Patients will be counted once in each ATC category (anatomic or therapeutic) linked to the medication; therefore, patients may be counted several times for the same medication.

In addition, the following summaries will also be provided:

- Vaccines (prior)
- Other prior DMARDs/immunosuppressive agents taken since diagnosis of PMR and discontinued before the first dose of IMP

The table for prior medications will be sorted by decreasing frequency of anatomic category followed by all other generic names based on the overall incidence across treatment groups. In

case of equal frequency regarding anatomic categories (respectively, generic names), alphabetical order will be used.

The tables for concomitant medications will be sorted by decreasing frequency of anatomic category followed by all other generic names based on the incidence of the total group (including all the randomized patients). In case of equal frequency regarding anatomic categories (respectively, generic names), alphabetical order will be used.

2.4.3 Extent of investigational medicinal product exposure and compliance

The extent of SC IMP exposure and compliance will be assessed and summarized by actual treatment group within the safety population and in populations with and without trial impact due to COVID-19 (Section 2.3.2 and Section 2.3.4).

2.4.3.1 Extent of investigational medicinal product exposure

The extent of SC IMP exposure will be assessed by the duration of SC IMP exposure and actual dose information.

The duration of SC IMP exposure is defined as last DB dose date – first DB dose date +14 days, regardless of unplanned intermittent discontinuations (see Section 2.5.3 for calculation in case of missing or incomplete data).Duration of SC IMP exposure will be summarized descriptively as a quantitative variable (number, mean, SD, median, minimum, and maximum). In addition, duration of treatment exposure will also be summarized categorically by numbers and percentages for each of the following categories and cumulatively according to these categories: ≤4 weeks, >4 and ≤8 weeks, >8 and ≤12 weeks, >12 and ≤16 weeks, >16 and ≤20 weeks, >20 and ≤24 weeks, >24 and ≤32 weeks, >32 and ≤40 weeks, >40 and ≤52 weeks and >52 weeks. The cumulative duration of treatment exposure, defined as the sum of patients' duration of treatment exposure and expressed in patient years, will also be provided.

2.4.3.2 Compliance

A given administration will be considered noncompliant if the patient did not take the planned dose of treatment as required by the protocol. No imputation will be made for patients with missing or incomplete data.

Percentage of compliance for SC IMP will be defined as the number of administrations the patient was compliant divided by the total number of administrations the patient was planned to take on or before the last dose date during the DB treatment period.

Percentage of compliance for oral IMP compliance rate will be defined as the total number of days patient with partial or complete oral IMP intake divided by total number of days between the first oral IMP intake date and day prior to rescue therapy start date or EOT date which is earlier.

Treatment compliance percentages will be summarized descriptively as quantitative variables (number, mean, SD, median, minimum, and maximum). The percentage of patients whose compliance is <80% will be summarized.

Cases of overdose (administering two or more doses of sarilumab or its matching placebo in less than 11 calendar days) will constitute AESI and will be listed as such. More generally, dosing irregularities will be listed in Section 2.2.1.

2.4.4 Analyses of efficacy endpoints

The primary and secondary efficacy endpoints will be analyzed using the ITT population. For patients who reduced dose to sarilumab 150mg during the course of the study, a listing of key efficacy endpoint variables will be provided.

2.4.4.1 Analysis of primary efficacy endpoint(s)

Fisher's exact test at 5% significance level will be performed to test the null hypothesis that the proportion of patients who achieved sustained remission at Week 52 in each treatment group (Group 1 = sarilumab + 52-week prednisone taper; Group 2 = placebo + 14-week prednisone taper) is the same, against the alternative hypothesis that the proportion is not the same between the treatment groups:

$$H_0: \rho_{Group1} - \rho_{Group2} = 0$$

$$H_1: \rho_{Group1} - \rho_{Group2} \neq 0$$

The number and proportion of patients achieving sustained remission at Week 52 in each treatment group will be presented, with the 95% exact confidence interval for the proportion difference and p-value from Fisher's exact test (2-sided).

Sensitivity analysis excluding acute phase reactants:

Sensitivity analysis will be performed for the primary endpoint by removing the acute phase reactants (CRP and ESR) from the definition to mitigate against the possibility of biasing the results due to the known PD effect of sarilumab on acute phase reactants.

Patients achieving sustained remission at Week 52 excluding acute phase reactants is defined by having met all of the following parameters:

- Achievement of disease remission not later than Week 12. **Disease remission** is defined as resolution of signs and symptoms of PMR.
- Absence of disease flare from Week 12 through Week 52. **Flare** is defined as recurrence of signs and symptoms attributable to active PMR plus an increase in CS dose due to PMR.
- Successful adherence to the prednisone taper from Week 12 through Week 52.

The sensitivity analysis will be performed in the same fashion as for the primary endpoint. Descriptive statistics (i.e., counts and percentages) will be provided for the three individual components.

Sensitivity analysis excluding patients who discontinued treatment as they were incorrectly diagnosed as PMR or newly diagnosed as RA while on study:

The diagnosis of PMR can be challenging as symptom characteristics often overlap, making it difficult to decipher differential diagnosis of PMR. Sensitivity analysis of primary endpoint will be performed to remove patients who discontinued study treatment due to a change in PMR disease diagnosis.

Sensitivity analyses for missing data in the primary endpoint:

Tipping point analysis will be performed. The number of responders to impute in each treatment group will vary from zero to up to the total number of patients with missing primary endpoint outcomes. For each imputation combination, a p-value based on Fisher's exact test will be calculated. The goal is to identify scenarios leading to non-significant results (p-value $\geq$0.05).

Patients who receive rescue therapy will be considered as non-responders. Withdrawn patients who had a disease flare prior to withdrawal will also be considered as non-responders as this represents the true outcome had they remained in the study. Withdrawn patients who did not experience a disease flare prior to withdrawal will be considered as missing and will be sequentially imputed for the tipping point analysis.

Subgroup analyses: To assess the consistency in treatment effects across subgroup levels, subgroup analyses will be conducted for the primary endpoint with respect to:

- Gender (Male versus Female)
- Race (Caucasian/White versus all other races)
- Region (Region 1 [Western countries], Region 2 [South America], Region 3 [Rest of the world])
- Age (<70, ≥70)
- Baseline weight (<60, ≥60 and <100, ≥100 kg)
- BMI (<25, ≥25 and <30, ≥30 kg/m2)
- Baseline CRP (≤median, >median)
- Baseline ESR (≤median, >median)
- Baseline CS dose (≤median, >median)
- Number of relapses prior to screening (1, > 1 relapse)
- COVID-19 impacted population (Yes, No) – will only be performed if >10% of patients were impacted by COVID-19. The population without trial impact (disruption) due to COVID-19 is defined as patients without any critical or major deviation related to COVID-19 or did not permanently discontinue study treatment due to COVID-19.

Within each subgroup, the number and proportion of patients achieving sustained remission at Week 52 by treatment groups will be presented together with the proportion difference between sarilumab and placebo groups and its 95% confidence interval.

Subgroup analyses will also be performed with logistic regression models for binary status of patients achieving sustained remission at Week 52, with treatment group, subgroup variable and the subgroup-by treatment interaction as covariates. A p-value for the test of interaction will be provided.

Forest plots of treatment differences (proportion difference) between sarilumab and placebo at Week 52 and corresponding CIs for each subgroup will be provided.

2.4.4.2 Analyses of secondary efficacy endpoint(s)

Components of sustained remission composite measure at Week 52

Descriptive statistics (i.e., counts and percentages) will be provided for the following binary variables by treatment groups:

- Patients who achieve disease remission by Week 12
- Patients who have absence of disease flare from Week 12 through Week 52
- Patients who have normalization of CRP (decrease to <10 mg/L) with sustained normalization from Week 12 through Week 52
- Patients who successfully adhere to the prednisone taper from Week 12 through Week 52

Total cumulative corticosteroid (including prednisone) dose

Descriptive statistics including the number of subjects, mean, and median of the total cumulative prednisone (or equivalent) dose will be presented for each treatment group. The difference between the medians and the corresponding 95% non-parametric confidence interval along with p-value from the non-parametric Wilcoxon rank-sum test between the treatment groups will also be provided.

Both expected and actual cumulative CS dose will be presented. The expected cumulative dose is based on CS tapering regimen up to end of treatment, assuming that the taper was continued without error while actual cumulative CS dose will be calculated as the total cumulative CS used for PMR disease up to the end of treatment, including expected prednisone in tapering regimen per protocol, add-on prednisone, CS used in rescue therapy and the use of commercial prednisone. Patients who received increased prednisone or equivalent in rescue therapy will be included in their original treatment groups.

In the scenario that a patient has already taken external CS for the disease before randomization on the day of randomization, the CS dose taken will be considered as the first dose of oral IMP. The first dose of prednisone per tapering regimen will be skipped, and the patient will take the second dose of prednisone on the next day.

Time to first PMR flare during the treatment period

Time (from randomization) to the first PMR flare after clinical remission up to Week 52 will be analyzed by Kaplan-Meier method providing the median, 25th and 75th percentiles (where possible) along with the 95% CI for the median. The comparison between the treatment groups will be performed using Cox proportional hazards model. Hazard ratio and the corresponding 95% CI and p-value will be provided.

Composite GTI and Specific List

GTI analyses

Descriptive statistics including number, mean, SD, SE, median, minimum, and maximum will be provided by treatment group and by visit for domain-specific GTI score, composite GTI total score, composite worsening score and composite improvement score. For CWS and AIS at Week 52, analysis of covariance (ANCOVA) model with treatment groups and baseline glucocorticoid toxicity score as fixed effects, will be used to test the difference of the least-square means (LS means) between the treatment groups. Difference in LS means and the corresponding 95% CI along with the p-value will be provided.

In addition, descriptive statistics (i.e., counts and percentages) will be provided for not worsening (CWS = 0), worsening (CWS > 0), improved (AIS < 0) and not improved (AIS ≥ 0) at Week 52. Odds ratios and exact 95% CIs of not worsening versus worsening and improved versus not improved will be provided for the comparison between the sarilumab and placebo groups.

Analyses of items in Specific list

Number and percentage of patients with each item in specific list will be summarized by treatment groups and by visit.

2.4.4.3 Analyses of exploratory endpoint(s)

Descriptive statistics will be provided for PMR-AS and its components at baseline, Week 12, Week 24, and Week 52.

The change from baseline at Week 52 for PMR-AS will be analyzed with mixed model repeated measures (MMRM) approach. The repeated-measures analysis will be based on the restricted maximum likelihood method assuming an unstructured covariance structure to model the within-subject errors. The model, including treatment, visit, and treatment -by -visit interaction, baseline and baseline -by -visit interaction as covariates, will be used to test the difference of least square means (LS means) between treatment groups in the change from baseline for PMR-AS at Week 52. The difference in LS means between treatment groups, the corresponding 95% CI and the p-value will be provided. No imputations will be performed for missing values.

Change from baseline at each visit will be plotted by treatment group for PMR-AS.

2.4.4.4 Multiplicity issues

If the primary endpoint reaches statistical significance, the secondary endpoint for total cumulative CS dose will be tested next.

2.4.5 Analyses of safety data

The summary of safety results will be presented by treatment groups.

General common rules

All safety analyses will be performed on the safety population as defined in Section 2.3.2, unless otherwise specified, using the following common rules:

- Safety data in patients who do not belong to the safety population (eg, exposed but not randomized) will be listed separately
- The baseline value is defined as the last available value prior to the first dose of study medication.
- The potentially clinically significant abnormality (PCSA) values are defined as abnormal values considered medically important by the Sponsor according to predefined criteria/thresholds based on literature review and defined by the Sponsor for clinical laboratory tests, vital signs, and ECG (PCSA in BTD-009536 version 3, see Appendix C.
- PCSA criteria will determine which patients had at least 1 PCSA during the TEAE period, taking into account all evaluations performed during the TEAE period, including nonscheduled or repeated evaluations. The number of all such patients will be the numerator for the on-treatment PCSA percentage
- The treatment-emergent PCSA denominator by group for a given parameter will be based on the number of patients assessed for that given parameter in the treatment-emergent adverse event period by treatment group on the safety population.
- For quantitative safety parameters based on central laboratory/reading measurements, descriptive statistics will be used to summarize results and change from baseline values by visit and treatment group. Summaries will include the endpoint value. The endpoint value is commonly defined as the value collected at the same day/time of the last dose of investigational product. If this value is missing, this endpoint value will be the closest one prior to the last dose intake.
- Analyses of the safety variables will be essentially descriptive, and no systematic testing is planned.

2.4.5.1 Analyses of adverse events

Generalities

The primary focus of AE reporting will be on TEAEs. Pre-treatment and post-treatment AEs will be described separately.

If an AE date/time of onset (occurrence, worsening, or becoming serious) is incomplete, an imputation algorithm will be used to classify the AE as pre-treatment, or TEAE, or post-treatment AE. The algorithm for imputing date/time of onset will be conservative and will classify an AE as treatment-emergent unless there is definitive information to determine it is pre-treatment or post-treatment. Details on classification of AEs with missing or partial onset dates are provided in Section 2.5.3.

Adverse event incidence tables will present by SOC, HLGT, HLT, and PT, sorted in SOC internationally agreed order and then alphabetical order, and present the number (n) and percentage (%) of patients experiencing an AE by treatment group. Multiple occurrences of the same event in the same patient will be counted only once in the tables within a treatment phase. The denominator for computation of percentages is the safety population within each treatment group.

The tables by primary SOC and PT will be sorted by internationally agreed order of SOC and decreasing frequency of PT within SOC. Sorting will be based on results for the sarilumab 200mg arm.

Analysis of all treatment-emergent adverse events

The following TEAE summaries will be generated for the safety population.

- Overview of TEAEs, summarizing number (%) of patients with any
  - TEAE
  - Serious TEAE
  - TEAE leading to death
  - TEAE leading to permanent treatment discontinuation
- All TEAEs by primary SOC, HLGT, HLT, and PT, showing number (%) of patients with at least one TEAE sorted by SOC internationally agreed order. The other level (HLGT, HLT, PT) will be presented in an alphabetic order.
- All TEAEs by primary SOC and PT, showing number (%) of patients with at least one TEAE, sorted by SOC internationally agreed order and decreasing incidence of PTs within SOC. This sorting order will be applied to all other tables, unless otherwise specified.
- If > 10% patients are impacted by the COVID-19 pandemic: All TEAEs by primary SOC and PT by COVID-19 impact subgroup, showing number (%) of patients with at least one TEAE by COVID-19 impact subgroup.
- All COVID-19 TEAEs by primary SOC and PT, showing number (%) of patients with at least one COVID-19 related TEAE.
- All TEAEs by primary SOC, showing number (%) of patients with at least one TEAE, sorted by internationally agreed order of primary SOC.

- All TEAEs by relationship, presented by primary SOC, HLGT, HLT, and PT, showing number (%) of patients with at least one TEAE, sorted by SOC internationally agreed order. The other level (HLGT, HLT, PT) will be presented in an alphabetic order.

- All TEAEs by maximal severity, presented by primary SOC and PT, showing number (%) of patients with at least one TEAE by severity (ie, mild, moderate, or severe), sorted by sorting order defined above.

Analysis of all treatment emergent serious adverse event(s)

- All treatment-emergent SAEs by primary SOC, HLGT, HLT, and PT, showing number (%) of patients with at least one serious TEAE, sorted by SOC internationally agreed order. The other level (HLGT, HLT, PT) will be presented in an alphabetic order.
- All treatment-emergent SAEs by primary SOC and PT, showing number (%) of patients with at least one TEAE
- All treatment-emergent SAEs by primary SOC, showing number (%) of patients with at least one TEAE, sorted by internationally agreed order of primary SOC

Analysis of all treatment-emergent adverse event(s) leading to treatment discontinuation

- All TEAEs leading to treatment discontinuation, by primary SOC, HLGT, HLT, and PT, showing number (%) of patients sorted by SOC internationally agreed order. The other level (HLGT, HLT, PT) will be presented in an alphabetic order.
- All TEAEs leading to treatment discontinuation, by primary SOC and PT, showing number (%) of patients with at least one TEAE
- All TEAEs leading to treatment discontinuation, by primary SOC, showing number (%) of patients with at least one TEAE, sorted by internationally agreed order of primary SOC

Analysis of all treatment-emergent adverse event(s) leading to dose reduction

- For the management of pre-defined laboratory abnormalities, sarilumab dose can be reduced to 150 mg q2w. All treatment-emergent adverse events leading to dose reduction, by primary SOC, HLGT, HLT and PT, showing the number (%) of patients, sorted by the sorting order defined above

Analysis AESI

Summaries of AESI defined by the search criteria:

- All treatment-emergent AESIs, by AESI category and PT, showing number (%) of patients, sorted by decreasing incidence of PT within each AESI category.
- Within each treatment-emergent AESI category, at a minimum the data display will include:
  - Overview summary
  - Treatment duration summary
  - Incidence by patient and by event

- Serious TEAEs
- TEAEs leading to treatment discontinuation
- TEAEs leading to death
- Possibly related TEAEs

2.4.5.2 Deaths

The incidence of death will be generated on the safety populations.

- Number (%) of patients who died by study period (on-treatment, treatment residual, and post-treatment).
- Death in non-randomized patients or randomized but not treated patients.
- TEAE leading to death (death as an outcome on the AE CRF page as reported by the Investigator) by primary SOC and PT, showing number (%) of patients sorted by internationally agreed order of SOC and decreasing frequency of PT within a SOC in the Sarilumab 200mg q2w treatment group.

Listings will be provided for all deaths with flags indicating during treatment or post-treatment status.

2.4.5.3 Analyses of laboratory variables

The summary statistics (including number, mean, median, SD, minimum and maximum) of all laboratory variables (central laboratory values and changes from baseline) will be calculated for each visit or study assessment (baseline, each post baseline time point, endpoint) by treatment group. This section will be organized by biological function as specified in Section 2.1.4.3.

In addition, for all laboratory parameters, the mean value and mean change from baseline at each scheduled visit during the treatment period will be plotted by treatment group.

The incidence of PCSAs (list provided in Appendix C) at any time during TEAE period will be summarized by biological function and treatment group whatever the baseline level and/or according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

For lipids, the National Cholesterol Education Program Adult Treatment Panel III (NCEP ATPIII) classification will be used for the baseline.

The incidence of abnormal laboratory values at any time during the TEAE period will be summarized in shift tables by biological function and treatment group whatever the baseline level and/or according to the following baseline status categories:

- Normal/Missing
- Abnormal high according to the normal range
- Abnormal low according to the normal range

Time to onset of the initial LDL-C elevation (≥4.1 mmol/L, ≥4.9 mmol/L), time to onset of the initial total cholesterol elevation (≥6.2 mmol/L), time to onset of the initial Grade 3 or Grade 4 neutropenia (neutrophil count <1.0 Giga/L), time to onset of the initial platelet count <100 Giga/L, time to onset of the initial lymphocyte count <1.0 Giga/L, time to onset of the initial ALT elevation (>3 x ULN, >5 x ULN), time to onset of the initial AST elevation (>3 x ULN, >5 x ULN), and time to onset of initial total bilirubin >2 x ULN will be analyzed using Kaplan-Meier estimates, using the midpoint of the time interval between the first assessment showing the elevation and the previous assessment, presented by treatment group.

Listings will be provided with flags indicating the out of range values as well as the PCSA values.

Neutrophils

The incidence of neutropenia by maximal grade (lowest absolute neutrophil count reported during the TEAE period) will be summarized. The 4 grades are defined as below:

- Grade 1: ≥1.5 Giga/L-LLN.
- Grade 2: ≥1.0 Giga/L- <1.5 Giga/L.
- Grade 3: ≥0.5 Giga/L- <1.0 Giga/L.
- Grade 4: <0.5 Giga/L.

Thrombocytopenia

The incidence of thrombocytopenia by maximal grade (lowest platelet count reported during the TEAE period) as defined below will be summarized.

- ≥75 Giga/L-100 Giga/L.
- ≥50 Giga/L-<75 Giga/L.
- ≥25 Giga/L-<50 Giga/L.
- <25 Giga/L.

Hepatic disorders

The liver function tests, namely AST, ALT, ALP, and total bilirubin are used to assess possible drug-induced liver toxicity. The proportion of patients with PCSA values at any post-baseline visit by baseline status will be displayed by treatment group for each parameter. The proportion of patients with PCSA values at any post baseline visit may be displayed by duration of exposure for each treatment group.

A graph of distribution of peak values of ALT versus peak values of total bilirubin will also be presented. Note that the ALT and total bilirubin values are presented on a logarithmic scale. The graph will be divided into 4 quadrants with a vertical line corresponding to 3x ULN for ALT and a horizontal line corresponding to 2x ULN for total bilirubin.

2.4.5.4 Analyses of vital sign variables

The summary statistics (including number, mean, median, SD, minimum, and maximum) of all vital signs variables (vital signs values and changes from baseline) will be calculated for each visit or study assessment (baseline, each post baseline time point, endpoint) by treatment group.

The incidence of PCSAs at any time during the TEAE period will be summarized by treatment group whatever the baseline level and/or according to the following baseline status categories:

- Normal/Missing
- Abnormal according to PCSA criterion or criteria

Listings will be provided with flags indicating the PCSA values.

2.4.6 Analyses of pharmacokinetic variables

All the PK analyses will be performed using the PK populations including only the sarilumab treated patients.

Serum concentrations of functional sarilumab (both pre-dose and post-dose) will be summarized using descriptive statistics (including number, arithmetic and geometric means, SD, standard error of the mean, coefficient of variation, minimum, median, and maximum) for each visit. The pre-dose samples will be considered non-eligible for these analyses if the previous dosing time is <11 days or >17 days before the sampling time for every other week regimens. Concentrations below the lower limit of quantification (LLOQ) will be set to zero for samples at baseline (i.e., pre-dose at Week 0). Other post-treatment concentrations below LLOQ will be replaced by LLOQ/2. In addition, serum concentrations of functional sarilumab will be summarized using descriptive statistics mentioned above by immunogenicity status for each visit.

2.4.7 Analyses of quality of life variables

Descriptive statistics including number of subjects, mean, median, SD, minimum, maximum will be provided by treatment groups for the value and change from baseline at each visit (baseline, Week 12, Week 24 and Week 52) for the 2 summary measures of SF-36 (physical component summary score and mental component summary score) and the 8 domains, the quantitative variables of EQ-5D-3L (visual analog scale and single index utility), FACIT-Fatigue, HAQ-DI (standardized score, pain score and patient global assessment) and MD-VAS. Change from baseline at each visit will also be plotted by treatment groups.

The change from baseline at Week 52 for the above measures will be analyzed with mixed model repeated measures (MMRM) approach. The repeated-measures analysis will be based on the

restricted maximum likelihood method assuming an unstructured covariance structure to model the within-subject errors. The model, including treatment, visit, and treatment -by -visit interaction, baseline and baseline -by -visit interaction as covariates, will be used to test the difference of least square means (LS means) between treatment groups in the change from baseline in each variable. Descriptive statistics including number of subjects, mean, standard error and LS means will be provided. In addition, difference in LS means between treatment groups, the corresponding 95% CI and the p-value will be provided.

2.4.8 Analysis of immunogenicity endpoints

The following summary will be provided separately for ADA positive patients, ADA negative patients, patients with persistent ADA response, and patients with transient ADA responses.

ADA prevalence and titer

The following summary will be provided based on the ADA populations:

- Number (%) of patients with an ADA positive sample at baseline:
  - Number (%) of neutralizing antibody.
  - Number (%) of non-neutralizing antibody.
  - The summary statistics (including number, median, Q1, Q3, minimum and maximum) of the titer for the baseline ADA positive patients.
- Number (%) of patients with an ADA negative sample at baseline.

ADA incidence and titer

The following summary will be provided based on the ADA populations during TEAE period:

- Number (%) of ADA-negative patients.
- Number (%) of ADA-positive patients:
  - Number (%) of patients with neutralizing antibody.
  - Number (%) of patients with non-neutralizing antibody.
  - The summary statistics (including number, median, Q1, Q3, minimum, and maximum) of the peak post-baseline titer for all ADA-positive patients.
- Number (%) of treatment-emergent ADA-positive patients:
  - The summary statistics (including number, median, Q1, Q3, minimum, and maximum) of the peak post-baseline titer for treatment-emergent ADA-positive patients.
  - Number (%) of transient treatment-emergent ADA-positive patients:
    - Number (%) of patients with neutralizing antibody.
    - Number (%) of patients with non-neutralizing antibody.

  - Number (%) of persistent treatment-emergent ADA positive patients:
    - Number (%) of patients with neutralizing antibody.
    - Number (%) of patients with non-neutralizing antibody.
  - Number (%) of persistent treatment-emergent ADA positive patients because last sample was positive.
- Number (%) of treatment-boosted ADA positive patients:
  - Number (%) of patients with neutralizing antibody.
  - Number (%) of patients with non-neutralizing antibody.
  - The summary statistics (including number, median, Q1, Q3, minimum, and maximum) of the peak post-baseline titer for all treatment-boosted ADA-positive patients.

In addition, number (%) of patients with ADA positive or negative response at each visit will be summarized by treatment group.

ADA and PK

By visit descriptive summary of functional sarilumab concentrations in serum will be provided by ADA patient classifications (positive or negative) for the sarilumab treatment groups.

ADA and clinical safety

The safety assessment will focus on the following events:

- Hypersensitivity (standardized MedDRA query (SMQ): Hypersensitivity [Narrow]).
- Anaphylaxis (SMQ: Anaphylaxis [Narrow]).
- Injection-site reactions (HLT: Injection site reactions).
- TEAEs leading to permanent treatment discontinuations.

Number (%) of patients with these events will be summarized by ADA patient classifications (positive or negative) for the sarilumab treatment group. The relationship between the timing and the titer of the positive ADA and events may also be explored as necessary.

ADA and clinical efficacy

The following efficacy endpoints will be analyzed by ADA patient classifications (positive or negative; persistent, transient; neutralizing, non-neutralizing):

- Proportion of patients achieving sustained remission at Week 52
- Number (%) of patients with lack of efficacy.

Lack of efficacy is defined as permanent treatment discontinuation due to lack of efficacy.

2.4.9 DNA/RNA

The DNA test is optional and patients need to sign a separate informed consent for DNA testing. Exploratory analysis of DNA/RNA will be addressed in a separate document by a separate biomarker analysis group.

2.5 DATA HANDLING CONVENTIONS

2.5.1 General conventions

The following formulas will be used for computations of parameters.

Demographic formulas:

BMI is calculated as follows:

$$\text{BMI} = \text{Weight in kg}/(\text{height}^2 \text{ in meters}).$$

2.5.2 Data handling conventions for secondary efficacy variables

GTI

General Considerations

For vital signs and laboratory data used together with the GTI CRF to derive the scoring the scheduled measurements at the time of the visit should generally be used, but if any of these are missing they can be replaced by the last available measurement in the corresponding time window for that parameter.

GTI missing domains

The missing lab measurement for any GTI domain at a visit will be imputed by last observation carried forward (LOCF) method, and missing medication information for Glucose, Blood Pressure and Hyperlipidemia domains will be imputed as ‘no change’ in order to calculate the composite and cumulative GTI scores. For patients who discontinued study treatment prematurely (prior to Week 52), domain scores after treatment discontinuation visit will be considered as ‘no change’ for the calculation of cumulative GTI scores at Week 52. No imputation will be performed for cumulative GTI scores at Week 52 for patients with no post-randomization measurements (i.e. will not carry forward from baseline).

2.5.3 Missing data

For categorical variables, patients with missing data are not included in calculations of percentages unless otherwise specified. When relevant, the number of patients with missing data is presented.

Handling of computation of treatment duration if investigational medicinal product end of treatment date is missingFor the calculation of the treatment duration, the date of the last dose of SC IMP is equal to the date of last administration reported on the SC IMP administration CRF page. If this date is missing, the exposure duration should be left as missing.

Handling of medication missing/partial dates

No imputation of medication start/end dates or times will be performed. If a medication date or time is missing or partially missing and it cannot be determined whether it was taken prior or concomitantly, it will be considered a prior and concomitant medication.

Handling of AEs with missing or partial date/time of onset

Missing or partial missing AE onset dates and times will be imputed so that if the partial AE onset date/time information does not indicate that the AE started prior to treatment or after the TEAE period, the AE will be classified as treatment-emergent. No imputation of AE end dates/times will be performed. These data imputations are for categorization purpose only and will not be used in listings. No imputation is planned for date/time of AE resolution.

Handling of AEs when date and time of first investigational medicinal product administration is missing

When the date and time of the first IMP administration is missing, all AEs that occurred on or after the day of randomization should be considered as TEAE. The exposure duration should be kept as missing.

Handling of missing relationship of adverse events to investigational medicinal product

If the assessment of the relationship to IMP is missing, then the relationship to IMP has to be assumed and the AE considered as such in the frequency tables of possibly related AEs, but no imputation should be done at the data level.

Handling of missing severity of AEs

If the severity is missing for one of the treatment-emergent occurrences of an AE, the maximal severity on the remaining occurrences will be considered. If the severity is missing for all the occurrences a “missing” category will be added in summary table.

Handling of PCSA

If a patient has a missing baseline he will be grouped in the category “normal/missing at baseline.”

For PCSAs with 2 conditions, one based on a change from baseline value or a normal range and the other on a threshold value, with the first condition being missing, the PCSA will be based only on the second condition.

For a PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range is missing; eg, for eosinophils the PCSA is > 0.5 GIGA/L or >ULN if ULN ≥0.5 GIGA/L. When ULN is missing, the value 0.5 should be used.

Measurements flagged as invalid by the laboratory will not be summarized or taken into account in the computation of PCSA values.

2.5.4 Windows for time points

Visit windows based on days relative to the first IMP dose will be used to map the efficacy, lab, vital signs, PK and ADA measurements to each scheduled visit (see Appendix I). The following rule will be applied when mapping the measurements to the visit:

- For the same parameter (i.e., efficacy, lab, or vital signs), if a patient has more than one measurement at different dates within the same visit window, the scheduled measurement that is closest to the target date will be used (in case of tie select the latest). If there is no scheduled measurement within the visit window, the unscheduled measurement that is closest to the target date will be used (in case of tie select the latest).
- When a patient has more than one measurement on the same parameter on the same date, then the one with the later/largest sample ID will be used.
- If there is no measurement for a given parameter in an analysis window, data will be considered missing for the corresponding visit.

2.5.5 Unscheduled visits

Unscheduled visit measurements of laboratory parameters and vital signs will be used for computation of baseline, the last on-treatment value, analysis according to PCSA analysis, and the shift summaries for safety. They will also be included in the by-visit summaries if they are re-allocated to scheduled visits.

2.5.6 Remote Assessment

Due to the nature of the COVID-19 pandemic, related travel restrictions, and on-site visits limitations, some efficacy assessments and laboratory parameters have been taken remotely or locally. All the measurements will be used in the analyses.

2.5.7 Pooling of centers for statistical analyses

Countries are pooled by region: Western countries, South American and Rest of the world.

- Region 1 (Western countries): Australia, Belgium, Canada, Estonia, France, Germany, Hungary, Italy, Netherlands, Spain, Switzerland, United Kingdom, United States.
- Region 2 (South American): Argentina.
- Region 3 (Rest of the world): Russia, Israel, Japan.

2.5.8 Statistical technical issues

Not applicable.

3 INTERIM ANALYSIS

No interim analysis is planned for this study.

4 DATABASE LOCK

The database lock will occur approximately 4 weeks after the last patient has completed the EOS Visit.

5 SOFTWARE DOCUMENTATION

All summaries and statistical analyses will be generated using SAS version 9.2 or higher.

6 REFERENCES

1. J.H. Stone, K. Tuckwell, S. Dimonaco, et al. Trial of Tocilizumab in Giant-Cell Arteritis. The New England Journal of Medicine. 2017;377(4):317–28

2. Eli M. Miloslavsky, Ray P. Naden, Johannes W.J. Bijlsma, et al. Development of a Glucocorticoid Toxicity Index (GTI) using multicriteria decision analysis. Annals of the rheumatic diseases. 2017;76(3):543–6

3. EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208

7 LIST OF APPENDICES

Appendix A Study flow chart

VISIT	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12 (EOT)	V13 (EOS)
DAY	D-28 to D-1	D1	D15 (±3)	D29 (±3)	D57 (±3)	D85 (±3)	D113 (±3)	D141 (±3)	D169 (±3)	D225 (±3)	D281 (±3)	D365 (±3)	D407 (±3)
WEEK		Week 0	Week 2	Week 4	Week 8	Week 12	Week 16	Week 20	Week 24	Week 32	Week 40	Week 52	Week 58
Eligibility													
Written informed consent	X												
Inclusion/exclusion criteria[t]	X	X											
Ultrasound central reading[v]	X												
Patient demography	X												
Medical/surgical/smoking-alcohol history	X												
Prior medication history	X												
Family cardiovascular history	X												
Full physical examination	X											X	
Targeted physical examination[a]		X				X			X		X		
Confirm eligibility		X											
Randomization		X											
Call IRT	X	X		X	X	X	X	X	X	X	X	X	X
Treatment													
Initial treatment kit assignment (IRT)		X											
IMP administration[b]		X[w]	X	X	X	X	X	X	X[s]	X[s]	X[s]		

VISIT	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12 (EOT)	V13 (EOS)
DAY	D-28 to D-1	D1	D15 (±3)	D29 (±3)	D57 (±3)	D85 (±3)	D113 (±3)	D141 (±3)	D169 (±3)	D225 (±3)	D281 (±3)	D365 (±3)	D407 (±3)
WEEK		Week 0	Week 2	Week 4	Week 8	Week 12	Week 16	Week 20	Week 24	Week 32	Week 40	Week 52	Week 58
Concomitant medication review	X[u]	X	X	X	X	X	X	X	X	X	X	X	X
Dispense patient diary		X	X	X	X	X	X	X	X	X	X		
Compliance/review patient diary			X	X	X	X	X	X	X	X	X	X	
Vital signs													
Temperature, heart rate, blood pressure	X	X	X	X	X	X	X	X	X	X	X	X	
Weight	X	X				X			X		X	X	
Height	X												
Efficacy													
PMR clinical assessments (including disease flare)	X	X	X	X	X	X	X	X	X	X	X	X	X
Physician global assessment (MD-VAS)		X				X			X			X	
Patient reported outcomes[c]		X				X			X			X	
Glucocorticoid toxicity index (excl. bone density assessment)		X				X			X		X	X	
Bone density assessment[d]		X										X	
Safety													
AE/SAE recording													
Tuberculosis assessment	X	X	X	X	X	X	X	X	X	X	X	X	X
QuantiFERON®	X												

VISIT	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12 (EOT)	V13 (EOS)
DAY	D-28 to D-1	D1	D15 (±3)	D29 (±3)	D57 (±3)	D85 (±3)	D113 (±3)	D141 (±3)	D169 (±3)	D225 (±3)	D281 (±3)	D365 (±3)	D407 (±3)
WEEK		Week 0	Week 2	Week 4	Week 8	Week 12	Week 16	Week 20	Week 24	Week 32	Week 40	Week 52	Week 58
Chest X-ray[e]	X												
Laboratory testing													
Hematology[f]	X	X		X		X			X		X	X	
Chemistry[g]	X	X		X		X			X		X	X	
ANA[h]		X										X	
Fasting lipids[i] and fasting glucose/insulin[i]	X			X		X			X		X	X	
HbA1c	X			X		X			X		X	X	
hs-CRP[k]	X	X	X	X	X	X	X	X	X	X	X	X	
ESR[k]	X	X	X	X	X	X	X	X	X	X	X	X	
Urinalysis[l]	X												
Virology[m]	X												
Serum pregnancy test[n]	X												
Urine pregnancy test[n]		X		X	X	X	X	X	X	X	X	X	
12-lead ECG	X												
Serum sarilumab[j]		X	X	X		X	X		X[r]			X	X
Antibodies to sarilumab/ADA[j]		X				X			X			X	X
Genotyping and biomarkers[j]													

VISIT	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12 (EOT)	V13 (EOS)
DAY	D-28 to D-1	D1	D15 (±3)	D29 (±3)	D57 (±3)	D85 (±3)	D113 (±3)	D141 (±3)	D169 (±3)	D225 (±3)	D281 (±3)	D365 (±3)	D407 (±3)
WEEK		Week 0	Week 2	Week 4	Week 8	Week 12	Week 16	Week 20	Week 24	Week 32	Week 40	Week 52	Week 58
Biomarkers-IL-6 and sIL-6R		X	X			X			X			X	
Immune Cell Phenotyping (Whole Blood)[o]		X		X					X				
Future Use Samples (Serum and Plasma)[p] – *Optional*		X	X			X			X			X	
DNA[q]- *Optional*		X											
RNA[q]- *Optional*		X	X										

AE = Adverse event; D = Day; DNA = Deoxyribonucleic acid; ECG = Electrocardiogram; EOT = End of treatment; EOS = End of study; EQ-5D = EuroQol; ESR = erythrocyte sedimentation rate; EUL = elevate upper limb; FACIT-Fatigue = Functional assessment of chronic illness therapy fatigue scale; GCA = Giant cell arteritis; GTI = Glucocorticoid toxicity index; HbA1c = Hemoglobin A1c; HbsAg = Hepatitis B surface antigen; Hbcore Ab = Hepatitis B core antibodies; HCV = Hepatitis C virus; hs-CRP = High-sensitivity C-reactive protein; IL = Interleukin; IMP = Investigational medicinal product; IRT = Interactive voice response system; MD-VAS = Physician global assessment of disease activity- visual analog scale; Pain VAS = Pain visual analog scale; Pt-VAS = Patient global assessment of disease activity– visual analog scale; RNA = Ribonucleic acid; SAE = Serious adverse event; SF-36v2 = Short form 36v2; V = Visit; Wk = Week.

a Targeted physical examination: head, eyes, ears, neck and throat, skin, respiratory, cardiovascular, neurologic, lymphatic examinations and abdominal examination.

b Last administration of sarilumab is at Week 50.

c Patient Reported Outcomes include EQ-5D-3L, FACIT-Fatigue, SF-36v2, HAQ-DI

d Bone Mineral Density assessment will be performed at the baseline (Visit 2) and Week 52 (Visit 12) using a DXA scan. The scan can be performed within ± 14 days of Visit 2 and within -14 days of Visit 12 and needs to include the lumbosacral and femoral neck regions.

e Chest X-ray is required during the screening period if no chest imaging (X-ray, CT, MRI) is available within the previous 12 weeks of V1 that clearly documents the exclusion of TB or if it does not follow the local guidelines and requirements for active screening of TB. In countries for which a specific approval procedure for the x-ray is required by a different committee than the local EC/IRB, a chest MRI between V1 and V2 can be performed.

f Hematology (blood should be drawn before drug administration): Hemoglobin, hematocrit, red blood cell (RBC) count and morphology (if RBC count is abnormal), white blood cell (WBC) differential, platelet count, absolute neutrophil count (ANC).

g Chemistry(blood should be drawn before drug administration): Sodium, potassium, chloride, bicarbonate, blood urea nitrogen (BUN), creatinine and creatinine clearance, calcium, phosphate, total protein, albumin, alanine aminotransferase (ALT) (SGPT), aspartate aminotransferase (AST) (SGOT), alkaline phosphatase (ALP), total bilirubin, conjugated bilirubin, unconjugated bilirubin, lactate dehydrogenase (LDH), uric acid

h Anti-nuclear antibody (ANA) will be collected at baseline, Visit 2 and EOT visits only.

i Lipids (blood should be drawn before drug administration): Triglycerides (TG), total cholesterol, high density lipoprotein (HDL) cholesterol, low density lipoprotein (LDL) cholesterol. Fasting is defined as having no food or liquid intake (except water/ice) for six hours or more.

j Blood should be drawn before drug administration. Fasting is defined as having no food or liquid intake (except water/ice) for six hours or more.

k CRP and ESR results will be blinded to both Investigator and Sponsor (except screening and baseline). ESR kits will be provided by the central laboratory while the test will be performed locally at the site; results will be blinded to Investigator and staff directly involved in management of study patient except the safety assessor.

l Urinalysis dipstick: specific gravity, pH, glucose, blood, protein, nitrites, leukocytes, ketones, urobilinogen and bilirubin (by dipstick) at screening visit only. If any parameter on the dipstick is abnormal, a urine sample should be sent to the central laboratory for testing. If positive for proteins, microscopic analysis is performed by central laboratory.

m Human immunodeficiency virus antibodies; Hepatitis B: Hep B surface antigen, total Hep B core antibody, Hep B surface antibody, and Hep B viral DNA (if necessary); Hepatitis C: HCV-antibody.

n In women of child-bearing potential.

o Immune Cell Phenotyping (Whole Blood): Approximately 40 patients from each of the two treatment arms will be selected for this whole blood draw and immune cell phenotyping analysis.

p A separate Future Use Samples Informed Consent has to be obtained before any sampling. Both serum and plasma will be drawn and the samples will be used for future analysis (e.g, circulating proteins).

q A separate Pharmacogenetic Research Informed Consent for collecting and sequencing DNA and RNA samples has to be obtained before any sampling. One DNA (at baseline or any treatment or follow up visit) and RNA sample for sequencing sampling time point at baseline and pre-dose (V3) are needed.

r Additional sample is to be drawn 4-7 days after Week 24 dosing.

s Since the visit interval exceeds 4 weeks, interim shipments of IMP to patients home may be performed using direct to patient (DTP) shipping in order to provide the patient with only 4 weeks IMP at a time in order to minimize compliance errors.

t If the ultrasound is employed in the diagnosis of PMR, then the ultrasound images need to be submitted to the central reader for confirmation that they fulfill the ultrasound part of the diagnostic criteria for PMR.

u For patients who are on >15mg/day (but not exceeding 20mg/day) of prednisone at screening and during the screening period, the Investigator should judiciously taper the prednisone down to 15mg/day prior to randomization in order to prevent a disease flare upon entering the study at 15mg/day of prednisone.

v If the ultrasound is being used as a diagnostic tool of PMR, the image needs to be submitted to the central reader for confirmation of eligibility.

w IMP training. Prior to the first dose of IMP, provide instructions on preparation and self-injection of the pre-filled syringes and the use of the weekly blister packs of prednisone. Document this training in the patients study file. Note: If the patient is unable or unwilling to perform the subcutaneous injections themselves, arrangements must be made for qualified site personnel and/or caregiver to administer study drug every 2 weeks for doses that are not scheduled to be given to the study site.

Appendix B Standardized corticosteroid-taper regimen during double-blind study treatment period

Week	Group 1-Daily PS dose14 weeks taper	Group 2-Daily PS dose 52 weeks taper
0 Day 1	15	15
1	15	15
2	14	14
3	12	12
4	10	12
5	9	12
6	8	10
7	7	10
8	6	9
9	5	9
10	4	9
11	3	9
12	2	8
13	1	8
14	PS placebo	8
15	PS placebo	8
16	PS placebo	7
17	PS placebo	7
18	PS placebo	7
19	PS placebo	7
20	PS placebo	6
21	PS placebo	6
22	PS placebo	6
23	PS placebo	6
24	PS placebo	5
25	PS placebo	5
26	PS placebo	5
27	PS placebo	5
28	PS placebo	4

Week	Group 1-Daily PS dose14 weeks taper	Group 2-Daily PS dose 52 weeks taper
29	PS placebo	4
30	PS placebo	4
31	PS placebo	4
32	PS placebo	3
33	PS placebo	3
34	PS placebo	3
35	PS placebo	3
36	PS placebo	2
37	PS placebo	2
38	PS placebo	2
39	PS placebo	2
40	PS placebo	2
41	PS placebo	2
42	PS placebo	2
43	PS placebo	2
44	PS placebo	1
45	PS placebo	1
46	PS placebo	1
47	PS placebo	1
48	PS placebo	1
49	PS placebo	1
50	PS placebo	1
51	PS placebo	1

Appendix C Potentially clinically significant abnormalities (PCSA) criteria

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

(From BTD-009536 " Analysis and reporting of safety data from clinical trials through the Clinical Study Report"– Version 3 – 21-MAY-2014)

Parameter	PCSA	Comments
Clinical Chemistry		
ALT	By distribution analysis: >3 ULN >5 ULN >10 ULN >20 ULN **Additional analysis*:** >1 – 1.5 ULN >1.5 – 3 ULN >3 – 5 ULN >5 – 8 ULN > 8 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. Categories are cumulative. First row is mandatory. Rows following one mentioning zero can be deleted.
AST	By distribution analysis: >3 ULN >5 ULN >10 ULN >20 ULN **Additional analysis*:** >1 – 1.5 ULN >1.5 – 3 ULN >3 – 5 ULN >5 – 8 ULN > 8 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. Categories are cumulative First row is mandatory. Rows following one mentioning zero can be deleted.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

(From BTD-009536 " Analysis and reporting of safety data from clinical trials through the Clinical Study Report"– Version 3 – 21-MAY-2014)

Parameter	PCSA	Comments
Alkaline Phosphatase	>1.5 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008.
Total Bilirubin	>1.5 ULN >2 ULN	Must be expressed in ULN, not in µmol/L or mg/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. Categories are cumulative. First row is mandatory. Rows following one mentioning zero can be deleted.
Conjugated Bilirubin	>35% Total Bilirubin and TBILI>1.5 ULN **Additional analysis*:** >1.5 ULN >2 ULN	Conjugated bilirubin dosed on a case-by-case basis. PCSA to be retrieved manually
Unconjugated bilirubin	**Additional analysis*:** >1.5 ULN >2 ULN	Must be expressed in ULN, not in µmol/L or mg/L. Categories are cumulative.
ALT and Total Bilirubin	ALT>3 ULN and TBILI>2 ULN	Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. To be counted within a same treatment phase, whatever the interval between measurement.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

(From BTD-009536 " Analysis and reporting of safety data from clinical trials through the Clinical Study Report"– Version 3 – 21-MAY-2014)

Parameter	PCSA	Comments
CPK**	>3 ULN	FDA Feb 2005.
	>10 ULN	Am J Cardiol April 2006.
		Categories are cumulative.
		First row is mandatory. Rows following one mentioning zero can be deleted.
Creatinine	≥150 µmol/L (Adults)	Benichou C., 1994.
	≥30% change from baseline	
	≥100% change from baseline	
CLcr (mL/min)	<15 (end stage renal disease)	FDA draft Guidance 2010
(Estimated creatinine clearance based on the Cokcroft-Gault equation)	≥15 - <30 (severe decrease in GFR)	Pharmacokinetics in patients with impaired renal function-study design, data analysis, and impact on dosing and labeling
	≥30 - < 60 (moderate decrease in GFR)	
	≥60 - <90 (mild decrease in GFR)	
	≥ 90 (normal GFR)	
eGFR** (mL/min/1.73m2)	<15 (end stage renal disease)	FDA draft Guidance 2010
(Estimate of GFR based on an MDRD equation)	≥15 - <30 (severe decrease in GFR)	Pharmacokinetics in patients with impaired renal function-study design, data analysis, and impact on dosing and labeling
	≥30 - < 60 (moderate decrease in GFR)	
	≥60 - <90 (mild decrease in GFR)	
	≥ 90 (normal GFR)	
Uric Acid		Harrison- Principles of internal Medicine 17th Ed., 2008.
Hyperuricemia	>408 µmol/L	
Hypouricemia	<120 µmol/L	
Blood Urea Nitrogen	≥17 mmol/L	
Chloride	<80 mmol/L	
	>115 mmol/L	
Sodium	≤129 mmol/L	
	≥160 mmol/L	

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

(From BTD-009536 " Analysis and reporting of safety data from clinical trials through the Clinical Study Report"– Version 3 – 21-MAY-2014)

Parameter	PCSA	Comments
Potassium	<3 mmol/L	FDA Feb 2005.
	≥5.5 mmol/L	
Total Cholesterol	≥7.74 mmol/L	Threshold for therapeutic intervention.
	≥6.2 mmol/L*	
LDL	**≥4.1 mmol/L***	
	≥4.9 mmol/L	
Triglycerides	≥4.6 mmol/L	Threshold for therapeutic intervention.
	≥5.6 mmol/L*	
Lipasemia**	≥3 ULN	
Amylasemia**	≥3 ULN	
Glucose		
Hypoglycaemia	≤3.9 mmol/L and <LLN	ADA May 2005.
Hyperglycaemia	≥11.1 mmol/L (unfasted); ≥7 mmol/L (fasted)	ADA Jan 2008.
HbA1c	>8%	
Albumin	≤25 g/L	
CRP	>2 ULN or >10 mg/L (if ULN not provided)	FDA Sept 2005.
Hematology		
WBC	<3.0 Giga/L (Non-Black); <2.0 Giga/L (Black)	Increase in WBC: not relevant.
	≥16.0 Giga/L	To be interpreted only if no differential count available.
Lymphocytes	**<0.5 Giga/L***	
	≥0.5 Giga/L – LLN*	
	>4.0 Giga/L	
Neutrophils	<1.5 Giga/L (Non-Black);<1.0 Giga/L (Black)	International Consensus meeting on drug-induced blood cytopenias, 1991.
	<1.0 Giga/L*	FDA criteria.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

(From BTD-009536 " Analysis and reporting of safety data from clinical trials through the Clinical Study Report"– Version 3 – 21-MAY-2014)

Parameter	PCSA	Comments
Monocytes	>0.7 Giga/L	
Basophils	>0.1 Giga/L	
Eosinophils	>0.5 Giga/L or >ULN (if ULN≥0.5 Giga/L)	Harrison- Principles of internal Medicine 17th Ed., 2008.
Hemoglobin	≤115 g/L (Male); ≤95 g/L (Female) ≥185 g/L (Male); ≥165 g/L (Female) Decrease from Baseline ≥20 g/L	Criteria based upon decrease from baseline are more relevant than based on absolute value. Other categories for decrease from baseline can be used (≥30 g/L, ≥40 g/L, ≥50 g/L).
Hematocrit	≤0.37 v/v (Male) ; ≤0.32 v/v (Female) ≥0.55 v/v (Male) ; ≥0.5 v/v (Female)	
RBC	≥6 Tera/L	Unless specifically required for particular drug development, the analysis is redundant with that of Hb. Otherwise, consider FDA criteria.
Platelets	**< 50 Giga/L*** **≥ 50 – 100 Giga/L*** ≥700 Giga/L	International Consensus meeting on drug-induced blood cytopenias, 1991.
Urinalysis		
pH	≤4.6 ≥8	
Vital signs		
HR	≤50 bpm and decrease from baseline ≥20 bpm ≥120 bpm and increase from baseline≥20 bpm	To be applied for all positions (including missing) except STANDING.
SBP	≤95 mmHg and decrease from baseline ≥20mmHg ≥160 mmHg and increase from baseline ≥20 mmHg	To be applied for all positions (including missing) except STANDING.
DBP	≤45 mmHg and decrease from baseline ≥10 mmHg ≥110 mmHg and increase from baseline ≥10 mmHg	To be applied for all positions (including missing) except STANDING.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

(From BTD-009536 " Analysis and reporting of safety data from clinical trials through the Clinical Study Report"– Version 3 – 21-MAY-2014)

Parameter	PCSA	Comments
Orthostatic Hypotension		
Orthostatic SDB		
Orthostatic DBP	≤-20 mmHg	
	≤-10 mmHg	
Weight	≥5% increase from baseline	FDA Feb 2007.
	≥5% decrease from baseline	

Appendix D Summary of statistical analyses

	Analysis		Statistical	Supportive	Subgroup	Other
Endpoint	**Population**	**Primary Analysis**	**Method**	**Analysis***	**Analysis***	**Analyses***
Primary Endpoint						
Patients who achieved sustained remission at Week 52	ITT	Proportion of patients achieving sustained remission at Week 52	Fisher's exact test	Tipping point analysis and primary endpoint excluding acute phase reactants and excluding incorrectly diagnosed PMR patients.	Yes Subgroups: Gender, race, region, age, baseline weight, BMI, baseline CRP, baseline ESR, baseline CS dose, prior relapse, COVID-19 impacted population.	No
Secondary Endpoints						
Patients who achieved of disease remission by Week 12	ITT	Proportion of patients who achieved of disease remission by Week 12	Descriptive	No	No	No
Patients who have absence of disease flare from Week 12 through Week 52	ITT	Proportion of patients who have absence of disease flare from Week 12 through Week 52	Descriptive	No	No	No
Patients who have normalization of CRP (decrease to <10 mg/L) with sustained normalization from Week 12 through Week 52	ITT	Proportion of patients who have normalization of CRP (decrease to <10 mg/L) with sustained normalization from Week 12 through Week 52	Descriptive	No	No	No
Patients who successfully adhere to the prednisone	ITT	Proportion of patients who successfully adhere to the prednisone taper from Week 12 through Week 52	Descriptive	No	No	No

Endpoint	Analysis Population	Primary Analysis	Statistical Method	Supportive Analysis*	Subgroup Analysis*	Other Analyses*
taper from Week 12 through Week 52						
Total cumulative prednisone (or equivalent) dose	ITT	Total cumulative prednisone (or equivalent) dose over 52-week treatment period	Wilcoxon rank-sum test	No	No	No
Time to the first PMR flare after clinical remission up to Week 52	ITT	Time (from randomization) to the first GCA flare after clinical remission up to Week 52	Kaplan-Meier estimates and Cox proportional hazards model	No	No	No
Total composite GTI total score	ITT	Cumulative GTI scores (CWS and AIS) at Week 52	Analysis of covariance (ANCOVA) model	No	No	Summaries of domain specific scores and specific list events at Week 12, 24, 40 and 52.
PMR activity score	ITT	PMR-AS at Week 52	MMRM model	No	No	Summaries of PMR-AS and components at baseline, Week 12, 24, and 52.
SF-36, EQ-5D-3L, FACIT-Fatigue, HAQ-DI, MD-VAS	ITT	Change from baseline at Week 52	MMRM model	No	No	No
Safety						
Adverse Events	Safety	Follow safety guidelines	Descriptive	No	No	No
Laboratory	Safety	Follow safety guidelines	Descriptive	No	No	No
Vital signs and ECGs	Safety	Follow safety guidelines	Descriptive	No	No	No
PK						
Sarilumab serum concentration	PK	By visit	Descriptive	No	No	No
Immunogenicity						

	Analysis			Statistical	Supportive	Subgroup	Other
Endpoint	**Population**		**Primary Analysis**	**Method**	**Analysis***	**Analysis***	**Analyses***
ADA positive patients	ADA	NA		Descriptive	No	No	No
ADA negative patients	ADA	NA		Descriptive	No	No	No
Persistent ADA response	ADA	NA		Descriptive	No	No	No
Transient ADA response	ADA	NA		Descriptive	No	No	No

Appendix E SF-36 V2 Scoring

General Scoring information

Items and scales are scored in 3 steps:

- Step 1. Item recoding, for the 10 items that require recoding,
- Step 2. Computing scale scores by summing across items in the same scale (raw scale scores); and,
- Step 3. Transforming raw scale scores to a 0-100 scale (transformed scale).

Item recoding

- All 36 items should be checked for out-of-range values prior to assigning the final item value. All out-of-range values should be recoded as missing data.
- The following tables show the recoding of response choice.

How to treat missing data

A scale score is calculated if a respondent answered at least half of the items in a multi-item scale (or half plus one in the case of scales with an odd number of items).

The recommended algorithm substitutes a person-specific estimate for any missing item when the respondent answered at least 50 percent of the items in a scale. A psychometrically sound estimate is the average score, across completed items in the same scale, for that respondent. For example, if a respondent leaves one item in the 5-item mental Health scale blank, substitute the respondent's average score (across the 4 completed mental health items) for that one item. When estimating the respondent's average score, use the respondent's final item values.

Computing raw scale scores

After item recoding, including handling of missing data, a raw score is computed for each scale. This score is a simple algebraic sum of responses for all items in that scale.

If the respondent answered at least 50% of the items in a multi-items scale, the score can be calculated. If the respondent did not answer at least 50% of the items, the score for that scale should be set to missing.

Transformation of the scale scores

The next step involves transforming each raw score to a 0 to 100 scale using the following formula:

Transformed scale = [(actual raw score – lowest possible raw score) / possible raw score range] x 100

This transformation converts the lowest and highest possible scores to zero and 100, respectively.

Table 6 - SF-36 V2 raw scores of eight domains

Scale	Lowest and highest possible raw scores	Possible raw score range
Physical Functioning	10,30	20
Role-Physical	4,20	16
Bodily pain	2,12	10
General health	5,25	20
Vitality	4,20	16
Social Functioning	2,10	8
Role-Emotional	3,15	12
Mental Health	5,25	20

Appendix F FACIT-fatigue scoring guidelines (version 4)

Instructions:* 1. Record answers in "item response" column. If missing, mark with an X

2. Perform reversals as indicated, and sum individual items to obtain a score.

3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.

4. The higher the score, the better the QOL.

Subscale	Item Code	Reverse item?		Item response	Item Score
FATIGUE	HI7	4	-	_____	=_____
SUBSCALE	HI12	4	-	_____	=_____
	An1	4	-	_____	=_____
	An2	4	-	_____	=_____
	An3	4	-	_____	=_____
	An4	4	-	_____	=_____
	An5	0	+	_____	=_____
	An7	0	+	_____	=_____
	An8	4	-	_____	=_____
	An12	4	-	_____	=_____
	An14	4	-	_____	=_____
	An15	4	-	_____	=_____
	An16	4	-	_____	=_____

Sum individual item scores: _____

Multiply by 13: _____

Divide by number of items answered: _____ **=Fatigue Subscale score**

*For guidelines on handling missing data and scoring options, please refer to the Administration and Scoring Guidelines in the manual or on-line at www.facit.org.

Appendix G SAS code for EQ-5D-3L index utility scoring

```
/*================================================================*/;
/* Aim : Derive the EQ5-D index (utility) */;
/* Source for the algorithm : scoring EQ-5D health states ( [redacted] */;
/* (cf. G:\_HE\PRO questionnaire\EQ5D\scoring\YorkTariffx.doc) */;
/* Note : UK based population (Dolan, 1997) */;
/* Author : [redacted] */;
/*================================================================*/;

data EUROQOL; set temp;
profil= (10000*eqq1cd)+(1000*eqq2cd)+(100*eqq3cd)+(10*eqq4cd)+eqq5cd;
eq5d=1;
**********Mobility*********;
if eqq1cd=2 then eq5d=eq5d-0.069;
if eqq1cd=3 then eq5d=eq5d-0.314;
********Self-care**********;
if eqq2cd=2 then eq5d=eq5d-0.104;
if eqq2cd=3 then eq5d=eq5d-0.214;
******Usual activities*****;
if eqq3cd=2 then eq5d=eq5d-0.036;
if eqq3cd=3 then eq5d=eq5d-0.094;
*****Pain/discomfort******;
if eqq4cd=2 then eq5d=eq5d-0.123;
if eqq4cd=3 then eq5d=eq5d-0.386;
*****Anxiety/depression*****;
if eqq5cd=2 then eq5d=eq5d-0.071;
if eqq5cd=3 then eq5d=eq5d-0.236;

if (eqq1cd ne 1 or eqq2cd ne 1 or eqq3cd ne 1 or eqq4cd ne 1 or eqq5cd ne 1)
then eq5d=eq5d-0.081;

if (eqq1cd=3 or eqq2cd=3 or eqq3cd=3 or eqq4cd=3 or eqq5cd=3)
then eq5d=eq5d-0.269;

if (eqq1cd=. or eqq2cd=. or eqq3cd=. or eqq4cd=. or eqq5cd=.)
then eq5d=.;

run;
```

Appendix H SAS code for SF-36 scoring

```
*****************************************************************
*  PROGRAM: SF36 V2                                             *
*  PURPOSE: SAS SCORING PROGRAM FOR THE SF-36 VERSION2          *
*                                                               *
*****************************************************************;

*****************************************************************;
***               STEP 1: INPUT DATA                          ***;
*****************************************************************;

DATA SF36DATA;
INFILE IN;
INPUT ID $ 1-3
      @ 5 (GH1 HT PF01-PF10 RP1-RP4 RE1-RE3 SF1
      BP1-BP2 VT1 MH1 MH2 MH3 VT2 MH4 VT3 MH5
      VT4 SF2 GH2 GH3 GH4 GH5) (1.);
RUN;


*****************************************************************;
***               STEP 2: SF-36 SCALE CONSTRUCTION            ***;
*****************************************************************;


*****************************************************************
*  USING THE SAS DATASET CREATED IN PART 1, CHANGE OUT-OF-RANGE
*  VALUES TO MISSING FOR EACH ITEM.  RECODE AND RECALIBRATE ITEMS
*  AS NEEDED.  AN 'R' PREFIX MEANS THE VARIABLE IS RECODED.
*****************************************************************;

DATA SF36SCAL;
  SET SF36DATA;


*****************************************************************
*  THE SF-36 PHYSICAL FUNCTIONING INDEX.
*  ALL ITEMS ARE POSITIVELY SCORED -- THE HIGHER THE ITEM
*  VALUE, THE BETTER THE PHYSICAL HEALTH.
*
*  THIS SCALE IS POSITIVELY SCORED.
*  THE HIGHER THE SCORE THE BETTER THE PHYSICAL FUNCTIONING.
*****************************************************************;

ARRAY PFI(10) PF01-PF10;

DO I = 1 TO 10;
   IF PFI(I) < 1 OR PFI(I) > 3 THEN PFI(I) = .;
END;

PFNUM = N(OF PF01-PF10);
PFMEAN = MEAN(OF PF01-PF10);

DO I = 1 TO 10;
  IF PFI(I)= . THEN PFI(I) = PFMEAN;
```

```
END;

IF PFNUM GE 5 THEN RAWPF = SUM(OF PF01-PF10);
PF = ((RAWPF - 10)/(30-10)) * 100;

LABEL  PF = 'SF-36 PHYSICAL FUNCTIONING (0-100)'
       RAWPF = 'RAW SF-36 PHYSICAL FUNCTIONING';


*****************************************************************
*  THE SF-36 ROLE-PHYSICAL INDEX.
*  ALL ITEMS ARE POSITIVELY SCORED -- THE HIGHER THE ITEM VALUE,
*  THE BETTER THE ROLE-PHYSICAL FUNCTIONING.
*
*  THIS SCALE IS POSITIVELY SCORED.
*  THE HIGHER THE SCORE THE BETTER THE ROLE-PHYSICAL.
*****************************************************************;

ARRAY RPA(4) RP1-RP4;

DO I = 1 TO 4;
   IF RPA(I) < 1 OR RPA(I) > 5 THEN RPA(I) = .;
END;

ROLPNUM = N(OF RP1-RP4);
ROLPMEAN = MEAN(OF RP1-RP4);

DO I = 1 TO 4;
   IF RPA(I) = . THEN RPA(I) = ROLPMEAN;
END;

IF ROLPNUM GE 2 THEN RAWRP = SUM(OF RP1-RP4);
RP = ((RAWRP - 4)/(20-4)) * 100;
LABEL  RP = 'SF-36 ROLE-PHYSICAL (0-100)'
       RAWRP = 'RAW SF-36 ROLE-PHYSICAL';


*****************************************************************
*  THE SF-36 PAIN ITEMS.
*  ITEM RECODING DEPENDS ON WHETHER BOTH PAIN1 AND PAIN2
*  ARE ANSWERED OR WHETHER ONE OF THE ITEMS HAS MISSING DATA.
*  AFTER RECODING, ALL ITEMS ARE POSITIVELY SCORED -- THE HIGHER
*  THE SCORE, THE LESS PAIN (OR THE MORE FREEDOM FROM PAIN).
*
*   THIS SCALE IS POSITIVELY SCORED.  THE HIGHER THE
*   SCORE THE LESS PAIN OR THE MORE FREEDOM FROM PAIN.
*****************************************************************;

IF BP1 < 1 OR BP1 > 6 THEN BP1 = .;
IF BP2 < 1 OR BP2 > 5 THEN BP2 = .;

*  RECODES IF NEITHER BP1 OR BP2 HAS A MISSING VALUE;

IF BP1 NE . AND BP2 NE . THEN DO;

   IF BP1 = 1 THEN RBP1 = 6;
```

```
   IF BP1 = 2 THEN RBP1 = 5.4;
   IF BP1 = 3 THEN RBP1 = 4.2;
   IF BP1 = 4 THEN RBP1 = 3.1;
   IF BP1 = 5 THEN RBP1 = 2.2;
   IF BP1 = 6 THEN RBP1 = 1;

   IF BP2 = 1  AND BP1 = 1 THEN RBP2 = 6;
   IF BP2 = 1  AND 2 LE BP1 LE 6 THEN RBP2 = 5;
   IF BP2 = 2  AND 1 LE BP1 LE 6 THEN RBP2  = 4;
   IF BP2 = 3  AND 1 LE BP1 LE 6 THEN RBP2  = 3;
   IF BP2 = 4  AND 1 LE BP1 LE 6 THEN RBP2  = 2;
   IF BP2 = 5  AND 1 LE BP1 LE 6 THEN RBP2  = 1;

END;

*  RECODES IF BP1 IS NOT MISSING AND BP2 IS MISSING;

IF BP1 NE . AND BP2 = . THEN DO;
   IF BP1 = 1 THEN RBP1 = 6;
   IF BP1 = 2 THEN RBP1 = 5.4;
   IF BP1 = 3 THEN RBP1 = 4.2;
   IF BP1 = 4 THEN RBP1 = 3.1;
   IF BP1 = 5 THEN RBP1 = 2.2;
   IF BP1 = 6 THEN RBP1 = 1;
   RBP2 = RBP1;

END;

*  RECODES IF BP1 IS MISSING AND BP2 IS NOT MISSING;

IF BP1 = . AND BP2 NE . THEN DO;

   IF BP2 = 1 THEN RBP2 = 6;
   IF BP2 = 2 THEN RBP2 = 4.75;
   IF BP2 = 3 THEN RBP2 = 3.5;
   IF BP2 = 4 THEN RBP2 = 2.25;
   IF BP2 = 5 THEN RBP2 = 1;
   RBP1 = RBP2;

END;

BPNUM = N(BP1,BP2);

IF BPNUM GE 1 THEN RAWBP = SUM(RBP1,RBP2);
BP = ((RAWBP - 2)/(12-2)) * 100;

LABEL  BP = 'SF-36 PAIN INDEX (0-100)'
       RAWBP = 'RAW SF-36 PAIN INDEX';


*****************************************************************
*  THE SF-36 GENERAL HEALTH PERCEPTIONS INDEX.
*  REVERSE TWO ITEMS AND RECALIBRATE ONE ITEM.  AFTER RECODING
*  AND RECALIBRATION, ALL ITEMS ARE POSITIVELY SCORED --  THE
*  HIGHER THE SCORE, THE BETTER THE PERCEIVED GENERAL HEALTH.
*
```

```
*  THIS SCALE IS POSITIVELY SCORED.
*  THE HIGHER THE SCORE THE BETTER THE HEALTH PERCEPTIONS.
*************************************************************;

ARRAY GHP(5) GH1-GH5;

DO I= 1 TO 5;
  IF GHP(I) < 1 OR GHP(I) > 5 THEN GHP(I) = .;
END;

IF GH1 = 1 THEN RGH1 =   5;
IF GH1 = 2 THEN RGH1 = 4.4;
IF GH1 = 3 THEN RGH1 = 3.4;
IF GH1 = 4 THEN RGH1 =   2;
IF GH1 = 5 THEN RGH1 =   1;

RGH3 = 6 - GH3;
RGH5 = 6 - GH5;

GHNUM = N(GH1,GH2,GH3,GH4,GH5);
GHMEAN = MEAN(RGH1,GH2,RGH3,GH4,RGH5);

ARRAY RGH(5) RGH1 GH2 RGH3 GH4 RGH5;

DO I= 1 TO 5;
  IF RGH(I) = . THEN RGH(I) = GHMEAN;
END;

IF GHNUM GE 3  THEN RAWGH = SUM(RGH1,GH2,RGH3,GH4,RGH5);
GH = ((RAWGH - 5)/(25-5)) * 100;

LABEL  GH = 'SF-36 GENERAL HEALTH PERCEPTIONS (0-100)'
       RAWGH = 'RAW SF-36 GENERAL HEALTH PERCEPTIONS';


*************************************************************
*  THE SF-36 VITALITY ITEMS.
*  REVERSE TWO ITEMS.  AFTER ITEM REVERSAL, ALL ITEMS ARE
*  POSITIVELY SCORED -- THE HIGHER THE SCORE, THE LESS THE FATIGUE
*  AND THE GREATER THE ENERGY.
*
*  THIS SCALE IS POSITIVELY SCORED.
*  THE HIGHER THE SCORE THE GREATER THE VITALITY.
*************************************************************;

ARRAY VI(4) VT1-VT4;

DO I = 1 TO 4;
  IF VI(I) < 1 OR VI(I) > 5 THEN VI(I) =.;
END;

RVT1 = 6-VT1;
RVT2 = 6-VT2;

VITNUM = N(VT1,VT2,VT3,VT4);
VITMEAN = MEAN(RVT1,RVT2,VT3,VT4);
```

```
ARRAY RVI(4) RVT1 RVT2 VT3 VT4;

DO I = 1 TO 4;
  IF RVI(I) = . THEN RVI(I) = VITMEAN;
END;

IF VITNUM GE 2 THEN RAWVT= SUM(RVT1,RVT2,VT3,VT4);
VT = ((RAWVT-4)/(20-4)) * 100;

LABEL  VT = 'SF-36 VITALITY (0-100)'
       RAWVT = 'RAW SF-36 VITALITY';


*************************************************************
*  THE SF-36 SOCIAL FUNCTIONING INDEX.
*  REVERSE ONE ITEM SO THAT BOTH ITEMS ARE POSITIVELY SCORED --
*  THE HIGHER THE SCORE, THE BETTER THE SOCIAL FUNCTIONING.
*
*  THIS SCALE IS POSITIVELY SCORED.
*  THE HIGHER THE SCORE THE BETTER THE SOCIAL FUNCTIONING.
*************************************************************;

ARRAY SOC(2) SF1-SF2;

DO I = 1 TO 2;
   IF SOC(I) < 1 OR SOC(I) > 5 THEN SOC(I) = .;
END;

RSF1 = 6 - SF1;
SFNUM = N(SF1,SF2);
SFMEAN = MEAN(RSF1,SF2);

ARRAY RSF(2) RSF1 SF2;

DO I = 1 TO 2;
  IF RSF(I) = . THEN RSF(I) = SFMEAN;
END;

IF SFNUM GE 1 THEN RAWSF = SUM(RSF1,SF2);
SF = ((RAWSF - 2)/(10-2)) * 100;

LABEL  SF = 'SF-36 SOCIAL FUNCTIONING (0-100)'
       RAWSF = 'RAW SF-36 SOCIAL FUNCTIONING';


*************************************************************
*  THE SF-36 ROLE-EMOTIONAL INDEX.
*  ALL ITEMS ARE POSITIVELY SCORED -- THE HIGHER THE ITEM VALUE,
*  THE BETTER THE ROLE-EMOTIONAL FUNCTIONING.
*
*  THIS SCALE IS POSITIVELY SCORED.
*  THE HIGHER THE SCORE, THE BETTER THE ROLE-EMOTIONAL.
*************************************************************;

ARRAY RM(3) RE1-RE3;
```

```
DO I = 1 TO 3;
   IF RM(I) < 1 OR RM(I) > 5 THEN RM(I) = .;
END;

ROLMNUM = N(OF RE1-RE3);
ROLMMEAN = MEAN(OF RE1-RE3);

DO I = 1 TO 3;
   IF RM(I) = . THEN RM(I) = ROLMMEAN;
END;

IF ROLMNUM GE 2 THEN RAWRE = SUM(OF RE1-RE3);
RE = ((RAWRE - 3)/(15-3)) * 100;

LABEL  RE = 'SF-36 ROLE-EMOTIONAL (0-100)'
       RAWRE = 'RAW SF-36 ROLE-EMOTIONAL';


*****************************************************************
*  THE SF-36 MENTAL HEALTH INDEX.
*  REVERSE TWO ITEMS.  AFTER ITEM REVERSAL, ALL ITEMS ARE
*  POSITIVELY SCORED -- THE HIGHER THE SCORE, THE BETTER THE
*  MENTAL HEALTH.
*
*  THIS SCALE IS POSITIVELY SCORED.
*  THE HIGHER THE SCORE THE BETTER THE MENTAL HEALTH.
*****************************************************************;

ARRAY MHI(5) MH1-MH5;

DO I = 1 TO 5;
   IF MHI(I) < 1 OR MHI(I) > 5 THEN MHI(I)=.;
END;

RMH3 = 6-MH3;
RMH5 = 6-MH5;

MHNUM=N(MH1,MH2,MH3,MH4,MH5);
MHMEAN=MEAN(MH1,MH2,RMH3,MH4,RMH5);

ARRAY RMH(5) MH1 MH2 RMH3 MH4 RMH5;

DO I = 1 TO 5;
   IF RMH(I) = . THEN RMH(I) = MHMEAN;
END;

IF MHNUM GE 3 THEN RAWMH = SUM(MH1,MH2,RMH3,MH4,RMH5);
MH = ((RAWMH-5)/(25-5)) * 100;

LABEL  MH = 'SF-36 MENTAL HEALTH INDEX (0-100)'
       RAWMH = 'RAW SF-36 MENTAL HEALTH INDEX';


*****************************************************************
*  THE SF-36 HEALTH TRANSITION ITEM.
```

```
*  THIS ITEM SHOULD BE ANALYZED AS CATEGORICAL DATA,
*  PENDING FURTHER RESEARCH.
************************************************************;

IF HT < 1 OR HT > 5 THEN HT = .;

LABEL  HT='RAW SF-36 HEALTH TRANSITION ITEM';
RUN;


************************************************************;
***               STEP 3: SF-36 SCALE CONSTRUCTION          ***;
************************************************************;



DATA SF36INDX;
  SET SF36SCAL;


************************************************************;
*  purpose: create physical and mental health index scores
*           standardized but not normalized
*           and standard deviations calculated with vardef=wdf
************************************************************;


/************************************************************/
/* NORM-BASED SCORING OF SF-36, STANDARD RECALL (4 WEEKS)     */
/* USING  WEIGHTED MEANS AND SD'S FROM 1998 GENERAL US POPULATION   */
/* FROM COMBINED SAMPLES                                      */
/************************************************************/

**********************************************************
**********************1998 US NORMS**********************
**********************************************************;

*CREATING SF-36 STANDARDIZED SCORES (0-100 Means/SD's from 1998 General
Population)*;

pf_z=(pf-83.29094)/23.75883;
rp_z=(rp-82.50964)/25.52028;
bp_z=(bp-71.32527)/23.66224;
gh_z=(gh-70.84570)/20.97821;
vt_z=(vt-58.31411)/20.01923;
sf_z=(sf-84.30250)/22.91921;
re_z=(re-87.39733)/21.43778;
mh_z=(mh-74.98685)/17.75604;


************************************************************;
*     COMPUTE SAMPLE RAW FACTOR SCORES                      *;
*     Z SCORES ARE FROM ABOVE                               *;
*     SCORING COEFFICIENTS ARE FROM U.S. GENERAL POPULATION  *;
*     FACTOR ANALYTIC SAMPLE N=2393: HAVE ALL EIGHT SCALES   *;
************************************************************;
```

```
praw=(pf_z * .42402)+(rp_z * .35119)+(bp_z * .31754)+(sf_z * -.00753)+
     (mh_z * -.22069)+(re_z * -.19206)+(vt_z * .02877)+(gh_z * .24954);

mraw=(pf_z * -.22999)+(rp_z * -.12329)+(bp_z * -.09731)+(sf_z * .26876)+
     (mh_z * .48581)+(re_z * .43407)+(vt_z * .23534)+(gh_z * -.01571);


*COMPUTE AGGREGATE STANDARDIZED SUMMARY SCORES*;

PCS = (praw*10) + 50;
MCS = (mraw*10) + 50;

label PCS='STANDARDIZED PHYSICAL COMPONENT SCALE-00'
      MCS='STANDARDIZED MENTAL COMPONENT SCALE-00';


run;
```

Appendix I Visit windows based on day ranges

Visit windows based on days relative to the first IMP dose will be used to map the efficacy, lab, vital signs, PK, and immunogenicity data to each scheduled visit (see **Table 7**).

Table 7 - Day range for each scheduled visit in the DB period for efficacy, lab, vital sign, PK and ADA summaries

Visit Label	Target Day	Schedule A[a]	Schedule B[b]	Schedule C[c]	Schedule D[d]	Schedule E[e]	Schedule F[f]	Schedule G[g]	Schedule H[h]	Schedule I[i]	Schedule J[j]
Double-Blind Period											
Baseline	1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1		≤ 1	≤ 1	≤ 1	≤ 1
Week 2	15	>1*-21							>1* - 21	>1* - 49	
Week 4	29	22-42				>1* - 56	>1* - 56	>1* - 42	22 - 56		>1* - 98
Week 8	57	43-70						43 - 70			
Week 12	85	71-98	>1* - 126	>1* - 126		57 - 126	57 - 126	71 - 98	57 - 98	50 - 126	
Week 16	113	99-126						99 - 126	99 - 140		
Week 20	141	127-154						127 - 154			
Week 24	169	155-196	127 - 224	127 - 266		127 - 224	127 - 224	155 - 196	141 - 266	127 - 266	≥ 99
Week 32	225	197-252						197 - 252			
Week 40	281	253-322	225 - 322			225 - 322	225 - 322	253 - 322			
Week 52	365	≥323	≥ 323	≥ 267	>1*	≥ 323	≥ 323	≥ 323	≥ 267	≥ 267	

a Includes temperature, heart rate, blood pressure, PMR clinical assessments (including disease flare), CRP, ESR

b Includes weight, glucocorticoid toxicity index (excluding bone density assessment)

c Includes physician global assessments (MD-VAS), patient reported outcomes and antibodies to sarilumab/ADA

d Includes bone density assessment, ANA

e Includes hematology: hemoglobin, hematocrit, red blood cell (RBC) count and morphology (if RBC count is abnormal), white blood cell (WBC) differential, platelet count, absolute neutrophil count (ANC) and chemistry: sodium, potassium, calcium, chloride, bicarbonate, total protein, creatinine and clearance, blood urea nitrogen (BUN), uric acid, lactate dehydrogenase (LDH), phosphate, albumin, ALT (SGPT), AST (SGOT), alkaline phosphatase (ALP), total bilirubin, conjugated bilirubin, and unconjugated bilirubin

f Includes HbA1c, fasting glucose, fasting lipids: triglycerides (TG), total cholesterol, HDL, LDL

g Includes urine pregnancy test

h Includes serum sarilumab

i Includes biomarkers-IL-6 and sIL-6R

j Includes immune cell phenotyping (whole blood)

* Includes value measured on day 1 but after the first IMP dose

For PK and ADA, the follow-up (i.e., 6-week post-treatment) measurement will be defined based on the sample collected at the follow-up visit (ie, Visit 13). The rest of PK and ADA measurements will be mapped to each visit using the visit window defined in **Table 7**.

Appendix J Baseline glucocorticoid toxicity

Approach to calculating the Baseline glucocorticoid toxicity score: Toxicities assigned weighted scores from the Composite Index to establish a baseline glucocorticoid toxicity score

Toxicity Domain		Points
Body Mass Index (BMI)	BMI < 27	0
	BMI ≥ 27 but < 30	21
	BMI ≥ 30	36
Glucose Metabolism	HgbA1c < 5.7	0
	HgbA1c < 5.7 but on medication	32
	Hgb A1c ≥ 5.7	32
	HgbA1c ≥ 5.7 and on medication	44
	Diabetic retinopathy, nephropathy, or neuropathy (count only one)	44
Blood pressure	Normotensive: systolic ≤ 120 and diastolic ≤ 85 no medications	0
	Systolic ≤ 120 and diastolic ≤ 85 but on medications	19
	Systolic > 120 and diastolic > 85 no medications	19
	Systolic ≤ 120 and diastolic > 85 no medications	44
	Hypertensive emergency or PRES (count only one)	44
Lipid Metabolism	LDL ≤ target[a]	0
	LDL ≤ target[a] but on medications	10
	LDL > target[a] on no medications	10
	LDL > target[a] on treatment	30
Bone/Tendon	Normal BMD or no known history of osteoporosis	0
	Osteoporosis	29
	Insufficiency fracture secondary to osteoporosis	29
	Osteonecrosis	29
	Tendon rupture while on corticosteroid	29
Glucocorticoid Myopathy	No myopathy	0
	Minor glucocorticoid myopathy	9
	Moderate glucocorticoid myopathy	63
	Severe glucocorticoid myopathy	63
Skin	No skin toxicity	0
	Minor skin toxicity (one or more than one minor skin item)	8
	Moderate skin toxicity (one or more than one moderate skin item)	26
	Severe skin toxicity (one or more than one moderate skin item)	26
Neuropsychiatric	No neuropsychiatric toxicity	0
	Minor (one or more than one minor NP item: insomnia, mania, depression, cognitive)	11
	Moderate (one or more than one moderate NP item: insomnia, mania, depression, cognitive)	74
	Severe (one or more than one severe NP item: insomnia, mania, depression, cognitive)	74
	Psychosis	74
	Glucocorticoid-induced violence	74

Infection	No GTI-relevant infections within the pre-baseline GTI interval of the study	0
	Oral or vaginal candidiasis or non-complicated zoster (<Grade3) within the pre-baseline GTI interval of the study	19
	Grade 3 or Grade 4 infection within the pre-baseline GTI interval of the study	93
Ocular	Increased IOP	33
	Posterior subcapsular cataract	33
	Central serous retinopathy	33
Gastrointestinal	GI perf absence of NSAIDs	33
	PUD without H. pylori	33
Endocrine	Adrenal insufficiency	33

[a] Target range of <70 mg/dL will be used for LDL domain.

Signature Page for VV-CLIN-0613825 v1.0
efc15160-16-1-9-sap

Approve & eSign	[redacted]
Approve & eSign	[redacted]