NCT01146652

STATISTICAL ANALYSIS PLAN

A multi-center, uncontrolled extension study evaluating efficacy and safety of sarilumab in patients with active Rheumatoid Arthritis (RA)

SAR153191-LTS11210

STATISTICIAN: [redacted]

DATE OF ISSUE: 22-Dec-2015

Total number of pages: 104

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

ACR	American College of Rheumatology
ADA	anti-drug antibody
AE(s)	adverse event(s)
AESI	adverse events of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANA	antinuclear antibody
ANC	absolute neutrophil count
AST	aspartate aminotransferase
ATC	Anatomic Therapeutic Class
BMI	body mass index
BUN	blood urea nitrogen
CAC	Cardiovascular adjudication committee
CCP	cyclic citrullinated peptides
CI	Confidence interval
CLcr	creatinine clearance
CPK	creatine phosphokinase
CRP	C-reactive protein
CRF	case report form
CSR	clinical study report
CV	coefficient of variation
CV event	cardiovascular event
DAS28	Disease Activity Score 28
DBP	diastolic blood pressure
DILI	drug induced liver injury
DMARD	disease modifying antirheumatic drug
DNA	deoxyribonucleic acid
ds-DNA	double-stranded DNA
ECG	electrocardiogram
EULAR	European League Against Rheumatism
HAQ-DI	Health Assessment Questionnaire-Disability Index
HbA1c	hemoglobin A1C
HDL	high density lipoprotein
HLGT	high level group term
HLT	high level term
HR	heart rate
hs-CRP	high sensitivity C-reactive protein
IL6-R	interleukin 6 receptor
IMP	investigational medicinal product
ITT	intent to treat

IVRS	Interactive Voice Response Systems
LDH	lactate dehydrogenase
LDL	low density lipoprotein
LLN	lower limit normal
LLT	lower level term
MACE	major adverse cardiovascular event
MedDRA	Medical Dictionary for Regulatory Activities
MTX	Methotrexate
PCSA	potentially clinically significant abnormalities
PFS-S	pre-filled syringe with safety system
PK	pharmacokinetic
PT	preferred term
PTC	product technical complaint
PTF	product technical failure
q2w	bi-weekly (every-other-week)
RA	rheumatoid arthritis
RBC	red blood cell
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
SD	standard deviation
SF-36 V2	Short Form 36 Version 2
SJC	swollen joint count
SMQ	standardized MedDRA query
SOC	system organ class
TEAE	treatment emergent adverse event
TJC	tender joint count
TNF	tumor necrosis factor
ULN	upper limit normal
VAS	Visual Analogue Scale
WBC	white blood cell
WHO-DD	World Health Organization – Drug Dictionary
WPAI	Work Productivity and Activity Impairment

1 OVERVIEW AND INVESTIGATIONAL PLAN

This document is the statistical analysis plan (SAP) for the final clinical study report (CSR) and guidance for interim analysis reports during the study.

1.1 STUDY DESIGN AND RANDOMIZATION

This is a multicenter, multinational open label long term study, for patients with RA who participated and completed EFC11072, ACT11575, EFC10832, SFY13370, and EFC13752 or for patients from 12 weeks onward with per protocol defined lack of efficacy from EFC10832. Patients, who were randomized in Part B of EFC11072 in a treatment arm subsequently not retained following pivotal dose selection, were also eligible for enrollment.

The 5 previous studies are described briefly below:

EFC11072: A randomized, double-blind, placebo-controlled, multi-centre, two-part, dose ranging (Part A) and confirmatory (Part B) study with an operationally seamless design, evaluating efficacy and safety of sarilumab added to methotrexate (MTX) in patients with active rheumatoid arthritis who are inadequate responders to MTX therapy. The study was completed. In the dose ranging part of the study (Part A), 5 doses (100 mg weekly [qw], 150 mg qw, 100 mg every other week [q2w], 150 mg q2w, and 200 mg q2w) and placebo were tested in combination with methotrexate (MTX). The confirmatory (Phase 3) portion (Part B) which initially had the same 6 arms (Cohort 1) and then after Phase 3 dose selection (150 mg q2w and 200 mg q2w) had 3 arms (Cohort 2: placebo, 150 mg q2w, 200 mg q2w) for 52-week treatment. In EFC11072 Part B, beginning at Week 16, patients with lack of efficacy (defined as less than 20% improvement from baseline on 2 consecutive visits in either TJC or SJC, or with any other clear lack of efficacy based on Investigator judgment) might have started OL rescue therapy with sarilumab 150 mg qw (150 mg qw prior to Phase 3 dose selection; 200 mg q2w after Phase 3 dose selection) in this study. These patients were also eligible to enter LTS11210 study if completed the treatment period. There were 1283 patients who rolled over to LTS11210.

ACT11575: A randomized, double-blind, parallel-group, placebo- and active calibrator-controlled study assessing the clinical benefit of sarilumab on top of MTX in patients with active RA who have failed previous TNF-α antagonists was terminated by the Sponsor due to administrative reasons with only 16 patients enrolled. There were 7 patients who rolled over to LTS11210.

EFC10832: A randomized, double-blind, parallel-group, placebo-controlled study assessing the efficacy and safety of sarilumab added to DMARD therapy in patients with active rheumatoid arthritis (RA) who are inadequate responders or intolerant to tumor necrosis factor-alpha (TNF-α) antagonists. Patients are randomized into 3 arms (placebo, 150 mg q2w, 200 mg q2w) with 24-week treatment. There were 456 patients who rolled over to LTS11210.

SFY13370: A randomized, double-blind, double-dummy study assessing the safety and tolerability of sarilumab and tocilizumab in patients with rheumatoid arthritis who are inadequate responders to or intolerant of TNF antagonists. Patients are randomized into 3 arms (Tocilizumab IV 4/8 mg/k q4w, 150 mg q2w, 200 mg q2w) with 24-week treatment. There were 168 patients who rolled over to LTS11210.

EFC13752: An Open-label, Randomized, Parallel Group Study Assessing the Immunogenicity and Safety of Sarilumab Administered as Monotherapy in Patients with Active Rheumatoid Arthritis. Patients are randomized into 2 arms (150 mg q2w, 200 mg q2w) with 24-week treatment. There were 111 patients who rolled over to LTS11210.

Patients may have been exposed with sarilumab for 12 weeks if they were initially randomized in EFC11072 Part A or ACT11575; between 2 and 52 weeks, if initially randomized in EFC11072 Part B; between 12 and 24 weeks, if initially randomized in EFC10832; or for 24 weeks, if initially randomized in SFY13370 and EFC13752.

Initially in LTS11210, patients received 150 mg of sarilumab SC weekly as the highest dose studied in EFC11072 Part A. The reduced dose (due to neutropenia, thrombocytopenia, or an increase in liver enzymes [ALT]), was 150 mg q2w. Once the dose regimens were selected for the Phase 3 studies (150 mg q2w and 200 mg q2w), patients already ongoing in the study at the highest dose were switched to the new highest sarilumab dose 200 mg q2w, as soon as permitted by administrative process. Patients previously assigned to the reduced dose of 150 mg q2w due to safety issue continued to receive 150 mg q2w.

The duration of the treatment period in the study will be at least 264 weeks, or up to a maximum of 516 weeks (*except in the UK where the duration of treatment will be 264 weeks*) descripted in Protocol Amendment 9. However, some patients may complete this study with 260 week treatment duration per Protocol Amendment 8 before the approval of Protocol Amendment 9 in their countries. The total maximum duration of study participation for a patient may be up to 523 weeks.

- 1-week screening period if applicable
- 516-week maximum treatment period in LTS11210 (for the earliest patients enrolled in LTS11210 in 2010 if sarilumab is not commercialized by 2020)
- 6-week follow up period as required per protocol

Patients will either enter a 1-week screening period after the EOT visit (Visit 1 in LTS11210) in initial studies (for patients from EFC11072 Part A or Part B1 and ACT11575), or enrol directly into the treatment period (for patients from EFC11072 Part B2, EFC10832, SFY13370, or EFC13752, the EOT visit in the initial study corresponds to the randomization visit, Visit 1/2 in LTS11210).

At Visit 2 or Visit 1/2 (Week 0) after confirmation of eligibility, patients will be administered the first open-label dose of sarilumab at the investigational site and will be observed for 30 minutes for any medical events. At dosing time points occurring outside site visits, sarilumab can be injected by the patients, or by trained caregivers.

Patients will return for the safety follow-up visit 6 weeks after the EOT visit (V49). In case of early treatment discontinuation, patients are required to complete the EOT visit (V49) and to return for the 6-week safety follow-up visit (Visit 50) as well.

The study will be considered completed for a patient at the time he/she completes all the scheduled procedures.

An optional 12-week sub-study aiming to evaluate the usability of pre-filled syringe with safety system (PFS-S) will be conducted in some countries. Patients enrolled in LTS11210 at selected sites in those countries may participate in the PFS-S sub-study from Week 24 of the main study after providing written informed consent for the sub-study. However, patients may continue to participate in the main trial without participating in the PFS-S sub-study if they wish. Full details of the PFS-S sub-study are presented in Appendix K.

1.2 OBJECTIVES

1.2.1 Primary objectives

The primary objective of the study is to evaluate the long term safety of sarilumab in patients with RA.

1.2.2 Secondary objectives

The secondary objective of the study is to evaluate the long term efficacy of sarilumab in patients with RA.

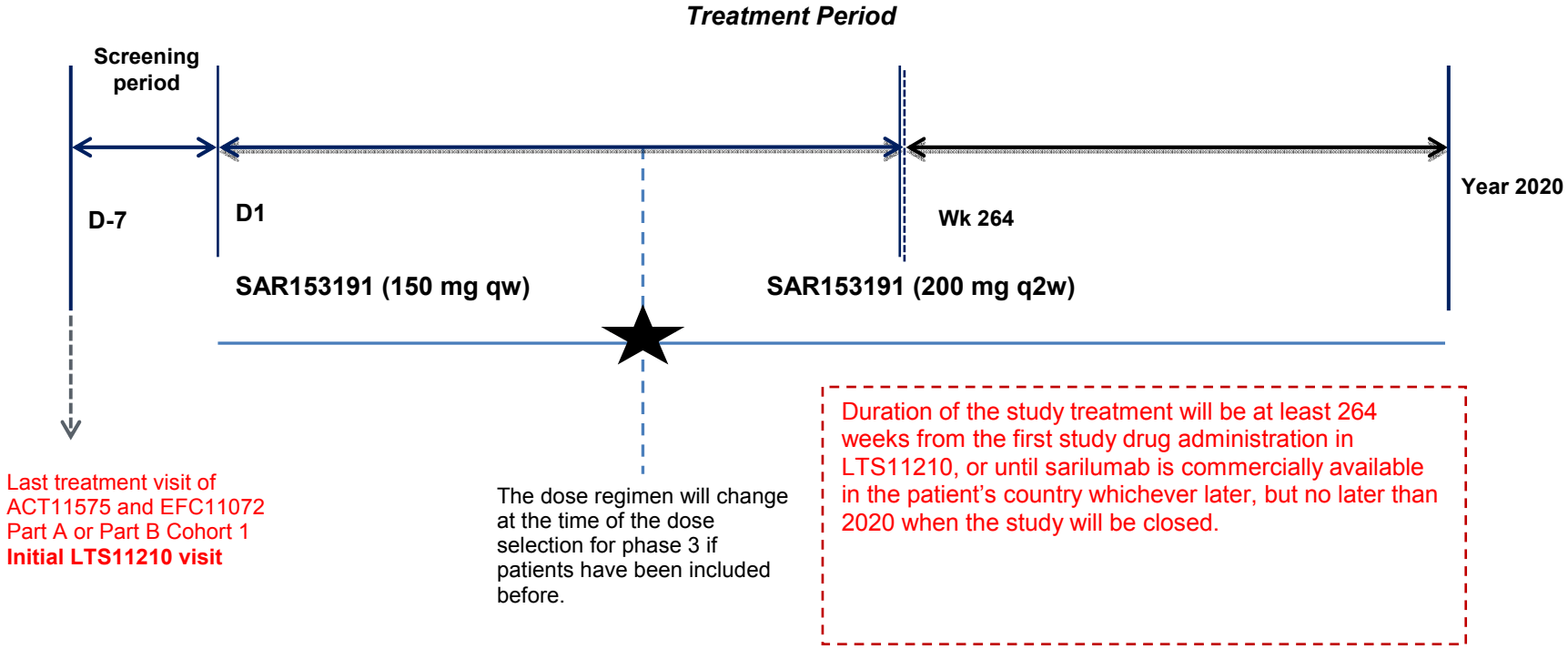
1.3 DETERMINATION OF SAMPLE SIZE

The number of patients to participate in this extension study is approximately 2000 from previous studies as follows:

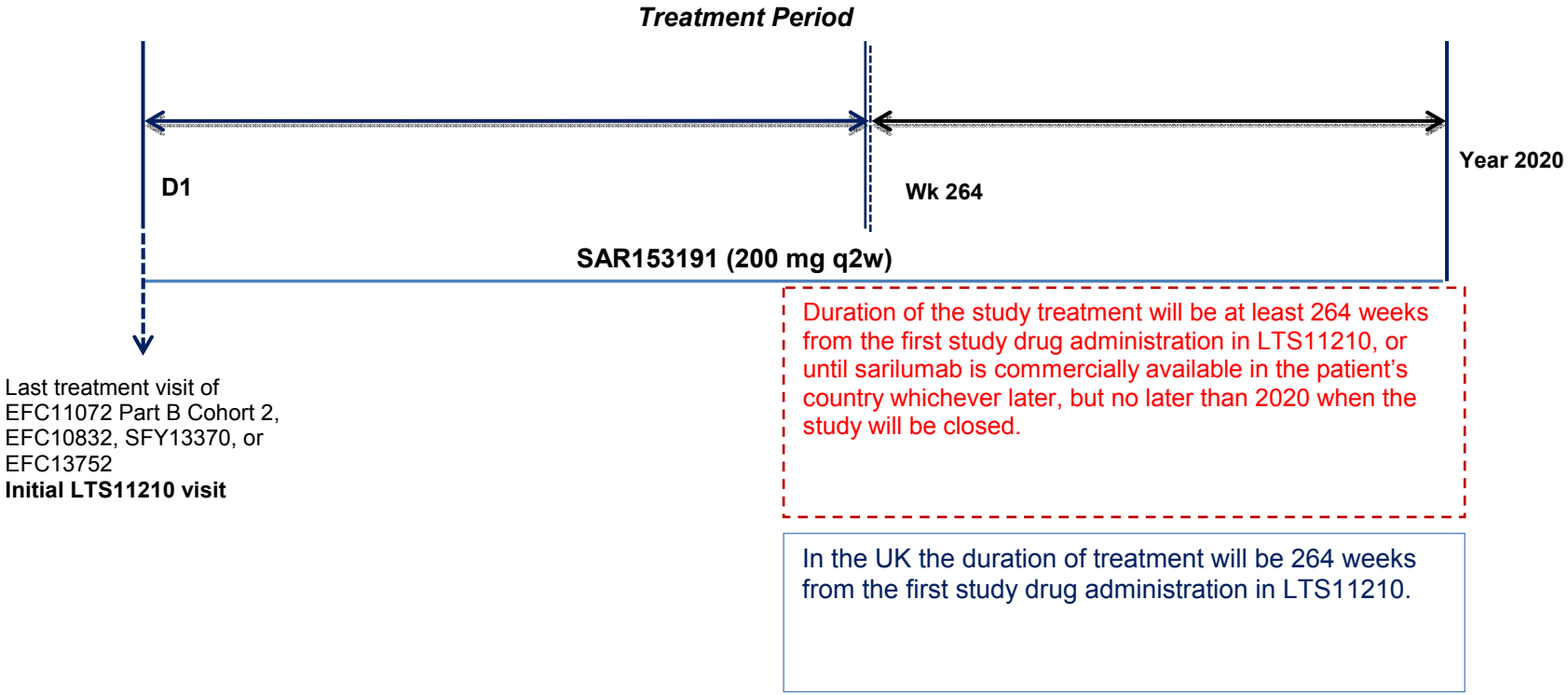
- EFC11072: 243 patients from Part A, 1040 patients from Part B
- ACT11575: 7 patients
- SFY13370: 168 patients
- EFC10832: 454 patients
- EFC13752: 111 patient

1.4 STUDY PLAN

1.4.1 Graphical Study Design for Patients in ACT11575 and EFC11072 Part A or Part B Cohort 1



1.4.2 Graphical Study Design for Patients in EFC11072 Part B Cohort 2, EFC10832, SFY13370, and EFC13752



1.4.3 Study Flowchart

1.4.3.1 Study Flow Chart for Patients Enrolling From ACT11575 and EFC11072 Part A or Part B Cohort 1

Evaluation	**Screening**	**Open Label Treatment**												**Post Treatment follow-up**
DAY **Week**	**Day -7 to D-1**	**Day 1** **Wk0**	**Day 15** **Wk 2**	**Day 29** **Wk 4**	**Day 43** **Wk 6**	**Day 57** **Wk 8**	**Day 71** **Wk 10**	**D85** **Wk 12**	**Wk 24-96 (every 12 wks) Wk 24, 36, 48, 60, 72, 84, 96**	**Wk 108- 492 IMP dispensing visits (every 24 weeks) Wk108, 132, 156, 180, 204, 228…492[a]**	**Wk120 – 264 (every 24 weeks) Wk120, 144, 168, 192, 216, 240, 264[a]**	**Wk288 – 504 (every 24 weeks) Wk 288, 312, …, 504[a]**	**EOT visit[a]**	**6 weeks after end of treatment[a] Post treatment follow-up**
Visit no.	**V 1 D-7 to D-1**	**V 2**	**HV 3[c] (± 3 days)**	**V 4 (±3 days)**	**HV 5[c] (±3 days)**	**V 6 (±3 days)**	**HV 6.1[c] (±3 days)**	**V 7 (±3 days)**	**V 8 to V 14 (±3 days)**	**V15, V17, V19, V21, V23, V25…V47(±3 days)**	**V16, V18, V20, V22, V24, V26 V28(±3 days)**	**V30, V32, V34, V36, V38, …, 48(±3 days)**	**V49 (±3 days)**	**V50(±3 days)**
Design														
Inclusion/exclusion criteria	X	X												
Previous medical/surgical history	X													
Informed consent	X													
Patient demography	X													
Prior medication history	X													
Smoking, alcohol, and illicit drug use history	X													

Evaluation	Screening	Open Label Treatment												Post Treatment follow-up
DAY Week	Day -7 to D-1	Day 1 Wk0	Day 15 Wk 2	Day 29 Wk 4	Day 43 Wk 6	Day 57 Wk 8	Day 71 Wk 10	D85 Wk 12	Wk 24-96 (every 12 wks) Wk 24, 36, 48, 60, 72, 84, 96	Wk 108- 492 IMP dispensing visits (every 24 weeks) Wk108, 132, 156, 180, 204, 228…492[a]	Wk120 – 264 (every 24 weeks) Wk120, 144, 168, 192, 216, 240, 264[a]	Wk288 – 504 (every 24 weeks) Wk 288, 312, …, 504[a]	EOT visit[a]	6 weeks after end of treatment[a] Post treatment follow-up
Visit no.	V 1 D-7 to D-1	V 2	HV 3[c] (± 3 days)	V 4 (±3 days)	HV 5[c] (±3 days)	V 6 (±3 days)	HV 6.1[c] (±3 days)	V 7 (±3 days)	V 8 to V 14 (±3 days)	V15, V17, V19, V21, V23, V25…V47(±3 days)	V16, V18, V20, V22, V24, V26 V28(±3 days)	V30, V32, V34, V36, V38, …, 48(±3 days)	V49 (±3 days)	V50(±3 days)
Detail history for tuberculosis (TB) and opportunistic infection	X													
Physical examination[e]	X[d]								X		X	X	X	
Confirm eligibility	X	X												
Treatment														
Study drug dispensing[b]		X		X		X		X	X	X	X	X		
Study drug compliance[b]				X		X		X	X	X	X	X	X	
Concomitant medications	X	X		X		X		X	X	X	X	X	X	X
Vital signs														
Temperature, heart rate, blood pressure	X[d]	X		X		X		X	X		X	X	X	X
Weight in Kg	X[d]	X		X		X		X	X		X	X	X	X

Evaluation	Screening	Open Label Treatment												Post Treatment follow-up
DAY Week	Day -7 to D-1	Day 1 Wk0	Day 15 Wk 2	Day 29 Wk 4	Day 43 Wk 6	Day 57 Wk 8	Day 71 Wk 10	D85 Wk 12	Wk 24-96 (every 12 wks) Wk 24, 36, 48, 60, 72, 84, 96	Wk 108- 492 IMP dispensing visits (every 24 weeks) Wk108, 132, 156, 180, 204, 228…492[a]	Wk120 – 264 (every 24 weeks) Wk120, 144, 168, 192, 216, 240, 264[a]	Wk288 – 504 (every 24 weeks) Wk 288, 312, …, 504[a]	EOT visit[a]	6 weeks after end of treatment[a] Post treatment follow-up
Visit no.	V 1 D-7 to D-1	V 2	HV 3[c] (± 3 days)	V 4 (±3 days)	HV 5[c] (±3 days)	V 6 (±3 days)	HV 6.1[c] (±3 days)	V 7 (±3 days)	V 8 to V 14 (±3 days)	V15, V17, V19, V21, V23, V25…V47(±3 days)	V16, V18, V20, V22, V24, V26 V28(±3 days)	V30, V32, V34, V36, V38, …, 48(±3 days)	V49 (±3 days)	V50(±3 days)
Efficacy														
ACR disease core set[h]	X[d]	X		X		X		X	X		X	X	X	
X-ray (hand, feet)[g]	X[d]								X		X		X	
Health Economic														
SF-36	X[d]							X	X		X		X[f]	
WPAI	X[d]							X	X		X		X[f]	
FACIT-Fatigue, Sleep questionnaire	X[d]							X	X		X		X[f]	
Safety														
Tuberculosis assessments	X[d]	X		X		X		X	X		X	X	X	X
AE/SAE recording (if any)	←													→
Laboratory Testing														
High sensitive- C-Reactive protein (hs-CRP)	X[d]	X		X		X		X	X		X	X	X	

Evaluation	Screening	Open Label Treatment												Post Treatment follow-up
DAY Week	Day -7 to D-1	Day 1 Wk0	Day 15 Wk 2	Day 29 Wk 4	Day 43 Wk 6	Day 57 Wk 8	Day 71 Wk 10	D85 Wk 12	Wk 24-96 (every 12 wks) Wk 24, 36, 48, 60, 72, 84, 96	Wk 108- 492 IMP dispensing visits (every 24 weeks) Wk108, 132, 156, 180, 204, 228…492[a]	Wk120 – 264 (every 24 weeks) Wk120, 144, 168, 192, 216, 240, 264[a]	Wk288 – 504 (every 24 weeks) Wk 288, 312, …, 504[a]	EOT visit[a]	6 weeks after end of treatment[a] Post treatment follow-up
Visit no.	V 1 D-7 to D-1	V 2	HV 3[c] (± 3 days)	V 4 (±3 days)	HV 5[c] (±3 days)	V 6 (±3 days)	HV 6.1[c] (±3 days)	V 7 (±3 days)	V 8 to V 14 (±3 days)	V15, V17, V19, V21, V23, V25…V47(±3 days)	V16, V18, V20, V22, V24, V26 V28(±3 days)	V30, V32, V34, V36, V38, …, 48(±3 days)	V49 (±3 days)	V50(±3 days)
Hematology: CBC and differential[i]	X[d]	X	X	X	X	X	X	X	X		X	X	X	
Liver Function Tests (LFTs)[j]	X[d]	X	X	X	X	X	X	X	X		X	X	X	
Lipids (fasting)[k]	X[d]			X		X		X	X		X	X	X	
Clinical chemistry (fasting)[l]	X[d]					X		X	X		X	X	X	
ANA/Anti-ds-DNA[m]	X[d]								X		X	X	X	
Dipstick urinalysis[n]	X[d]					X		X	X		X		X[f]	
Urine pregnancy test (for women of childbearing potential)[o]	X[d]			X		X		X	X		X	X	X	X
Dispense urine pregnancy kits[o]								X	X		X	X		
12-lead electrocardiogram[p]	X[d]								X		X		X[f]	
Other analysis														
Rheumatoid factor[q]		X							X		X		X[f]	
Serum IL-6[v]	X[d]			X				X	X					
Pharmacokinetics[r]	X[d]			X				X	X		X	X	X	X

Evaluation	Screening	Open Label Treatment												Post Treatment follow-up
DAY Week	Day -7 to D-1	Day 1 Wk0	Day 15 Wk 2	Day 29 Wk 4	Day 43 Wk 6	Day 57 Wk 8	Day 71 Wk 10	D85 Wk 12	Wk 24-96 (every 12 wks) Wk 24, 36, 48, 60, 72, 84, 96	Wk 108- 492 IMP dispensing visits (every 24 weeks) Wk108, 132, 156, 180, 204, 228…492[a]	Wk120 – 264 (every 24 weeks) Wk120, 144, 168, 192, 216, 240, 264[a]	Wk288 – 504 (every 24 weeks) Wk 288, 312, …, 504[a]	EOT visit[a]	6 weeks after end of treatment[a] Post treatment follow-up
Visit no.	V 1 D-7 to D-1	V 2	HV 3[c] (± 3 days)	V 4 (±3 days)	HV 5[c] (±3 days)	V 6 (±3 days)	HV 6.1[c] (±3 days)	V 7 (±3 days)	V 8 to V 14 (±3 days)	V15, V17, V19, V21, V23, V25…V47(±3 days)	V16, V18, V20, V22, V24, V26 V28(±3 days)	V30, V32, V34, V36, V38, …, 48(±3 days)	V49 (±3 days)	V50(±3 days)
Anti-sarilumab antibody	X[d]							X	X		X	X	X	X
Serum sample to be stored for future biomarkers[s]	X[d]		X					X	X					
Expression RNA[t]		X[d]	X											

a EOT visit and follow up visit 6 weeks later should be completed for all patients at the end of treatment (in patients who complete sarilumab treatment per protocol in countries where sarilumab is commercially available the post treatment follow-up visit may be cancelled if the patient initiates treatment with commercial sarilumab during the post treatment follow-up period; however, in this case, a telephone contact must be established with the patient to capture all AEs occurring up to the 1st administration of commercial sarilumab). EOT visit may occur when the patients complete 264 weeks of sarilumab treatment from the LTS11210, or at any visit thereafter when sarilumab is commercialized in their country, but no later than 2020 when the study will be closed. In case of permanent discontinuation of treatment, the patients will be assessed using the procedures normally planned for the EOT visit and the 6 week follow up visit 6 weeks later.

b From Week 24, patients at limited sites in selected countries may participate in PFS-S sub-study. If the patient enters the sub-study, then the sub-study IMP will be dispensed instead of the main study IMP at the entry into sub-study and IMP compliance with sub-study IMP will be reviewed at the end of sub-study. Full details of the sub-study are provided in Appendix R

c HV: Visit can be home visit or clinic visit to draw and collect the blood sample only for hematology and LFTs using the designated central laboratory.

d Last treatment visit from the study EFC11072 or ACT11575. Please note that the SF-36 must be completed at the screening visit of the LTS11210 study for patients rolling over from EFC11072 Part A and ACT11575 studies. Please note that the RNA sample must be collected at V2 of the LTS11210 study for patients rolling over from EFC11072.

e Physical examination to be done at Week 48, 96, 144, 192, 240, 264, 288, 336, 384, 432, 480 and EOT visit.

f Assessment is NOT required when EOT visit is after Week 264.

g X-ray of hand and feet (only for patients who completed Part B of the study EFC11072 (not performed for Part A patients, or patients from Part B non-selected dose arms) will be done at Weeks 48, 96, 144, 192 (±14 days for each assessment), and EOT; No X-ray evaluation will be performed after Week 192 including EOT if EOT is after Week 192. Required x-ray must be done after confirmation of negative urine pregnancy test in women of child bearing potential.

h After Week 264, only DAS28 components of ACR core set will be assessed (TJC & SJC for 28 joints, patient´s global assessment of disease activity).

i Hematology: Hemoglobin, hematocrit, red blood cell (RBC) morphology (if blood cell count is abnormal), white blood cell (WBC) with differential, platelets count. For all patients, a CBC test must be performed before or at Visit 3 (using either designated central lab or a local laboratory facility), but not earlier than the 12th day after the first dose of IMP administration in order to confirm that the neutrophil count and platelet count are not within the protocol-defined limits for temporary or permanent discontinuation of study drug.

j Liver Function Tests (LFTs): Prothrombin Time (PT), Albumin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), total bilirubin, conjugated bilirubin, unconjugated bilirubin. PT is NOT required at any visit after Week 264.

k Lipids (fasting): Total cholesterol (TC), High-Density Lipoprotein (HDL) cholesterol, Low-Density Lipoprotein (LDL) cholesterol, triglycerides, Apolipoprotein A, and Apolipoprotein B. After Week 264, lipids (excluding Apolipoprotein A, and Apolipoprotein B) will only be assessed at Weeks 288, 336, 384, 432, 480 (including EOT if EOT is after Week 264)

l Clinical chemistry (fasting): fasting glucose, total proteins, calcium, sodium, potassium, Lactate Dehydrogenase, (LDH), urea nitrogen and creatinine. Creatinine clearance will be calculated during the study if clinical indicated.

m ANA titer will be done at screening, at Week 48, 96, 144, 192, 240, 264, 288, 336, 384, 432, 480 and EOT visit or sooner if clinical indicated (Anti-ds-DNA only if ANA titer is >1:160)

n Dipstick urinalysis for: specific gravity, pH, glucose, blood, ketones, proteins, bilirubin, urobilinogen, nitrite, leukocytes. If any parameter is abnormal, a urinalysis sample should be sent to central laboratory for testing. If positive for proteins, microscopic analysis will be performed by central laboratory.

o After Visit 7 (Week 12) between study visits the patient will have a urine pregnancy test at home on a monthly basis. Patients will be given sufficient urine pregnancy kits to take home at each successive visit for monthly testing up until the final treatment visit. When the testing coincides with a clinical visit as indicated in the flow chart, the results should be reported in the eCRF. No pregnancy test kits will be dispensed at EOT.

p ECG to be done at Weeks 48, 96, 144, 192, 240, 264 or EOT; No ECG will be done after Week 264 including EOT if EOT is after Week 264.

q Rheumatoid factor only at baseline Visit 2 (Week 0) and at Weeks 48, 96, 144, 192, 240, 264 or EOT. No Rheumatoid factor will be assessed after Week 264 including EOT if EOT is after Week 264.

r If throughout the study a serious adverse event (SAE) occurs in a patient, blood samples should be collected for sarilumab at or near the onset and completion of the occurrence of the event, if possible. The exact date of sample collection and last dose must be recorded on the e-CRF.

s Serum sample for biomarkers to be collected at screening (ie, the last treatment visit from the EFC11072 or ACT11575), Week 2, Week 12, and Week 48.

t Blood sample for RNA to be collected at Week 0 (baseline, prior to administration of study drug) and Week 2 (at site visit only; cancel RNA collection if home visit).

u Deleted.

v Samples for IL-6 to be collected at Screening (ie the last treatment visit from EFC11072 or ACT11575) and at V4 (Week 4), V7 (Week 12), V8 (Week 24), V9 (Week 36), and V10 (Week 48).

1.4.3.2 Study Flow Chart for Patients Enrolling From EFC10832, EFC11072 Part B Cohort 2, SFY13370, and EFC13752

Evaluation	**Open Label Treatment**												**Post Treatment follow-up**
DAY and/or WEEK	**Day 1** **WK0**	**Day 15** **Wk 2**	**Day 29** **Wk 4**	**Day 43** **Wk 6**	**Day 57** **Wk 8**	**Day 71** **Wk 10**	**D85** **Wk 12**	**Wk 24-96** **(every 12 wks)** **Wk 24, 36, 48, 60, 72, 84, 96**	**Wk 108-492IMP dispensing visits** **(every 24 weeks)** **Wk108, 132, 156, 180, 204, 228… 492[a]**	**Wk120 – 264** **(every 24 weeks)** **Wk120, 144, 168, 192, 216, 240, 264[a]**	**Wk288 – 504** **(every 24 weeks)** **Wk288, 312,…, 504[a]**	**EOT visit[a]**	**6 weeks after end of treatment[a]** **Post treatment follow-up**
Visit no.	**V1/V 2**	**HV 3[c] (± 3 days)**	**V 4 (±3 days)**	**HV 5[c] (±3 days)**	**V 6 (±3 days)**	**HV 6.1[c] (±3 days)**	**V 7 (±3 days)**	**V 8 to V14 (±3 days)**	**V15, V17, V19, V21, V23, V25… V47 (±3 days)**	**V16, V18, V20, V22, V24, V26, V28 (±3 days)**	**V30, V32, V34, V36, V38, … , V48 (±3 days)**	**V49 (±3 days)**	**V50(±3 days)**
Design													
Inclusion/exclusion criteria	X												
Previous medical/surgical history	X												
Informed consent	X												
Patient demography	X												
Prior medication history	X												
Smoking, alcohol, and illicit drug use history	X												
Detail history for tuberculosis (TB) and opportunistic infection	X												
Physical examination[e]	X[d]							X		X	X	X	
Confirm eligibility	X												
Treatment													
Study drug dispensing[b]	X		X		X		X	X	X	X	X		

Evaluation	Open Label Treatment												Post Treatment follow-up
DAY and/or WEEK	**Day 1** **WK0**	**Day 15** **Wk 2**	**Day 29** **Wk 4**	**Day 43** **Wk 6**	**Day 57** **Wk 8**	**Day 71** **Wk 10**	**D85** **Wk 12**	**Wk 24-96** **(every 12 wks)** **Wk 24, 36, 48, 60, 72, 84, 96**	**Wk 108-492IMP dispensing visits** **(every 24 weeks)** **Wk108, 132, 156, 180, 204, 228… 492**[a]	**Wk120 – 264** **(every 24 weeks)** **Wk120, 144, 168, 192, 216, 240, 264**[a]	**Wk288 – 504** **(every 24 weeks)** **Wk288, 312,…, 504**[a]	**EOT visit**[a]	**6 weeks after end of treatment**[a] **Post treatment follow-up**
Visit no.	**V1/V 2**	**HV 3**[c] **(± 3 days)**	**V 4 (±3 days)**	**HV 5**[c] **(±3 days)**	**V 6 (±3 days)**	**HV 6.1**[c] **(±3 days)**	**V 7 (±3 days)**	**V 8 to V14 (±3 days)**	**V15, V17, V19, V21, V23, V25… V47 (±3 days)**	**V16, V18, V20, V22, V24, V26, V28 (±3 days)**	**V30, V32, V34, V36, V38, … , V48 (±3 days)**	**V49 (±3 days)**	**V50(±3 days)**
Study drug compliance[b]			X		X		X	X	X	X	X	X	
Concomitant medications	X		X		X		X	X	X	X	X	X	X
Vital signs													
Temperature, heart rate, blood pressure	X[d]		X		X		X	X		X	X	X	X
Weight in Kg	X[d]		X		X		X	X		X	X	X	X
Efficacy													
ACR disease core set[h]	X[d]		X		X		X	X		X	X	X	
X-ray (hand, feet)[g]	X[d]							X		X		X	
Health Economic													
SF-36 (EFC11072, EFC10832)	X[d]						X	X		X		X[f]	
WPAI (EFC11072)	X[d]						X	X		X		X[f]	
FACIT-Fatigue (EFC11072, EFC10832)	X[d]						X	X		X		X[f]	
Sleep questionnaire (EFC11072)	X[d]						X	X		X		X[f]	
WPS-RA (EFC10832)	X[d]						X	X		X		X[f]	

Evaluation	Open Label Treatment												Post Treatment follow-up
DAY and/or WEEK	**Day 1** **WK0**	**Day 15** **Wk 2**	**Day 29** **Wk 4**	**Day 43** **Wk 6**	**Day 57** **Wk 8**	**Day 71** **Wk 10**	**D85** **Wk 12**	**Wk 24-96** **(every 12 wks)** **Wk 24, 36, 48, 60, 72, 84, 96**	**Wk 108-492IMP dispensing visits** **(every 24 weeks)** **Wk108, 132, 156, 180, 204, 228… 492[a]**	**Wk120 – 264** **(every 24 weeks)** **Wk120, 144, 168, 192, 216, 240, 264[a]**	**Wk288 – 504** **(every 24 weeks)** **Wk288, 312,…, 504[a]**	**EOT visit[a]**	**6 weeks after end of treatment[a]** **Post treatment follow-up**
Visit no.	**V1/V 2**	**HV 3[c] (± 3 days)**	**V 4 (±3 days)**	**HV 5[c] (±3 days)**	**V 6 (±3 days)**	**HV 6.1[c] (±3 days)**	**V 7 (±3 days)**	**V 8 to V14 (±3 days)**	**V15, V17, V19, V21, V23, V25… V47 (±3 days)**	**V16, V18, V20, V22, V24, V26, V28 (±3 days)**	**V30, V32, V34, V36, V38, … , V48 (±3 days)**	**V49 (±3 days)**	**V50(±3 days)**
Safety													
Tuberculosis assessments	X[d]		X		X		X	X		X	X	X	X
AE/SAE recording (if any)	◄												►
Laboratory Testing													
High sensitive- C- Reactive protein (hs-CRP)	X[d]		X		X		X	X		X	X	X	
Hematology: CBC and differential[i]	X[d]	X	X	X	X	X	X	X		X	X	X	
Liver Function Tests (LFTs)[j]	X[d]	X	X	X	X	X	X	X		X	X	X	
Lipids (fasting)[k]	X[d]		X		X		X	X		X	X	X	
Clinical chemistry (fasting)[l]	X[d]				X		X	X		X	X	X	
ANA/Anti-ds-DNA [n]	X[d]							X		X	X	X	
Dipstick urinalysis[o]	X[d]				X		X	X		X		X[f]	
Urine pregnancy test (for women of childbearing potential)[p]	X[d]		X		X		X	X		X	X	X	X
Dispense urine pregnancy kits[p]							X	X		X	X		

Evaluation	Open Label Treatment												Post Treatment follow-up
DAY and/or WEEK	**Day 1** **WK0**	**Day 15** **Wk 2**	**Day 29** **Wk 4**	**Day 43** **Wk 6**	**Day 57** **Wk 8**	**Day 71** **Wk 10**	**D85** **Wk 12**	**Wk 24-96** **(every 12 wks)** **Wk 24, 36, 48, 60, 72, 84, 96**	**Wk 108-492IMP dispensing visits** **(every 24 weeks)** **Wk108, 132, 156, 180, 204, 228… 492[a]**	**Wk120 – 264** **(every 24 weeks)** **Wk120, 144, 168, 192, 216, 240, 264[a]**	**Wk288 – 504** **(every 24 weeks)** **Wk288, 312,…, 504[a]**	**EOT visit[a]**	**6 weeks after end of treatment[a]** **Post treatment follow-up**
Visit no.	**V1/V 2**	**HV 3[c] (± 3 days)**	**V 4 (±3 days)**	**HV 5[c] (±3 days)**	**V 6 (±3 days)**	**HV 6.1[c] (±3 days)**	**V 7 (±3 days)**	**V 8 to V14 (±3 days)**	**V15, V17, V19, V21, V23, V25… V47 (±3 days)**	**V16, V18, V20, V22, V24, V26, V28 (±3 days)**	**V30, V32, V34, V36, V38, … , V48 (±3 days)**	**V49 (±3 days)**	**V50(±3 days)**
12-lead electrocardiogram[q]	X[d]							X		X		X[f]	
Other analysis													
Rheumatoid factor[r]	X[d]							X		X		X[f]	
Serum IL-6[w]	X		X				X	X					
Pharmacokinetics[s]	X[d]		X				X	X		X	X	X	X
Anti-sarilumab antibody	X[d]						X	X		X	X	X	X
Serum sample to be stored for future biomarkers[t]	X[d]	X					X	X					
Expression RNA[u]	X	X											

a EOT visit and follow up visit 6 weeks later should be completed for all patients at the end of treatment (in patients who complete sarilumab treatment per protocol in countries where sarilumab is commercially available the post treatment follow-up visit may be cancelled if the patient initiates treatment with commercial sarilumab during the post treatment follow-up period; however, in this case, a telephone contact must be established with the patient to capture all AEs occurring up to the 1st administration of commercial sarilumab). EOT visit may occur when the patients complete 264 weeks of sarilumab treatment from LTS11210, or any visit thereafter when sarilumab is commercialized in their country, but no later than 2020 when the study will be closed. *(In UK, the duration of treatment will be 264 weeks from the first study drug administration in LTS11210).* In case of permanent discontinuation of treatment, the patients will be assessed using the procedures normally planned for the EOT visit and the 6 week follow up visit 6 weeks later.

b From Week 24, patients at limited sites in selected countries may participate in PFS-S sub-study. If the patient enters the sub-study, then the sub-study IMP will be dispensed instead of the main study IMP at the entry into sub-study and IMP compliance with sub-study IMP will be reviewed at the end of sub-study. Full details of the sub-study are provided in Appendix R

c HV: Visit can be home visit or clinic visit to draw and collect the blood sample only for hematology and LFTs using the designated central laboratory.

d Last treatment visit from the study EFC11072, EFC10832, SFY13370, or EFC13752.

e Physical examination to be done at Week 48, 96, 144, 192, 240, 264, 288, 336, 384, 432, 480 and EOT visit

f The assessment is NOT required when EOT visit is after Week 264.

g X-ray of hands and feet (only for patients who completed Part B of the study EFC11072 will be done at Weeks 48, 96, 144, 192 (±14 days for each assessment), and EOT; No X-ray evaluation will be performed after Week 192 including EOT if EOT is after Week 192. Required x-ray must be done after confirmation of negative urine pregnancy test in women of child bearing potential.

h After Week 264, only DAS28 components of ACR core set will be assessed (TJC & SJC for 28 joints, patient´s global assessment of disease activity).

i Hematology: Hemoglobin, hematocrit, red blood cell (RBC) morphology (if blood cell count is abnormal), white blood cell (WBC) with differential, platelets count. For all patients, a CBC test must be performed before or at Visit 3 (using either designated central or a local laboratory facility) but not earlier than the 12th day after the first dose of IMP administration in order to confirm that the neutrophil count and platelet count are not within the protocol-defined limits for temporary or permanent discontinuation of study drug.

j Liver Function Tests (LFTs): Prothrombin Time (PT), Albumin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), total bilirubin, conjugated bilirubin, unconjugated bilirubin. PT is NOT required at any visit after Week 264.

k Lipids (fasting): Total cholesterol (TC), High-Density Lipoprotein (HDL) cholesterol, Low-Density Lipoprotein (LDL) cholesterol, triglycerides, Apolipoprotein A, and Apolipoprotein B. After Week 264, lipids (excluding Apolipoprotein A, and Apolipoprotein B) will only be assessed at Weeks 288, 336, 384, 432, 480 (including EOT if EOT is after Week 264)

l Clinical chemistry (fasting): fasting glucose, total proteins, calcium, sodium, potassium, Lactate Dehydrogenase, (LDH), urea nitrogen and creatinine. Creatinine clearance will be calculated during the study if clinical indicated.

m Deleted.

n ANA titer will be done at V1/V2 (ie the last treatment visit from the initial study), at Week 48, 96, 144, 192, 240, 264, 288, 336, 384, 432, 480 and EOT visit , or sooner if clinical indicated (Anti-ds-DNA only if ANA titer is >1:160)

o Dipstick urinalysis for: specific gravity, pH, glucose, blood, ketones, proteins, bilirubin, urobilinogen, nitrite, leukocytes. If any parameter is abnormal, a urinalysis sample should be sent to central laboratory for testing. If positive for proteins, microscopic analysis will be performed by central laboratory.

p After Visit 7 (Week 12) between study visits the patient will have a urine pregnancy test at home on a monthly basis. Patients will be given sufficient urine pregnancy kits to take home at each successive visit for monthly testing up until the final treatment visit. When the testing coincides with a clinical visit as indicated in the flow chart, the results should be reported in the e-CRF. No pregnancy test kits will be dispensed at EOT.

q ECG to be done at Weeks 48, 96, 144, 192, 240, 264 or EOT; No ECG will be done after Week 264 including EOT if EOT is after Week 264

r Rheumatoid factor only at Weeks 48, 96, 144, 192, 240, 264 or EOT; No Rheumatoid factor will be assessed after Week 264 including EOT if EOT is after Week 264

s If throughout the study a serious adverse event (SAE) occurs in a patient, blood samples should be collected for sarilumab at or near the onset and completion of the occurrence of the event, if possible. The exact date of sample collection and last dose must be recorded on the eCRF.

t Serum sample for biomarkers to be collected for all patients except EFC13752: at V1/V2 (ie, the last treatment visit from the initial study), as well as Week 2, Week 12, and Week 48.

u Blood sample for RNA except for SFY13370 and EFC13752 patients: to be collected at V1/V2 (Week 0) (baseline prior to administration of study drug) and V3 (Week 2) at site visit only (cancel RNA collection if home visit).

v Deleted

w Samples for IL-6 to be collected at V1/V2 for EFC11072 patients only (ie the last treatment visit) and for all patients at V4 (Week 4), V7 (Week 12), V8 (Week 24), V9 (Week 36), and V10 (Week 48).

1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

This section summarizes major changes to the protocol statistical section with emphasis on changes after study start (after the first patient was enrolled).

A few modifications are made in the statistical section in the following amendments.

Amendment 1

The following cut-offs for age group and BMI group defined in the protocol:

- Age group (<45, ≥45 and < 65, ≥65 years old)
- BMI group (< 25, ≥25 kg/m^2)

were changed in Protocol Amendment 1to:

- Age group (< 65, ≥65 and <75, ≥75 years old)
- BMI group (<25, ≥25 and <30, ≥30 kg/m^2)

Amendment 3

- Updated the description of the study endpoints and statistical methodology for clarification purposes. The intent of the statistical analysis plan established in the original protocol is not changed.
- Incorporate the Rheumatoid Arthritis-Work Productivity Survey (WPS-RA) into the protocol for patients that completed the EFC10832 study. The WPS-RA is added into the LTS11210 study because the WPS-RA was included in the EFC10832
- Added x-ray data summary

Amendment 8

- Added: Data analysis results for patients who roll over from EFC13752 will be presented separately as sarilumab is administered as monotherapy in the EFC13752 and these patients will continue to be on sarilumab monotherapy in LTS11210
- Updated analysis treatment groups: The safety analyses will be presented by original actual treatment groups in the respective previous studies (placebo, 150 mg q2w, 200 mg q2w, other sarilumab, and active control treatment groups) as well as overall in LTS11210, based on the reported adverse events, clinical laboratory evaluations, vital signs, and 12-lead ECGs
- Updated analysis treatment groups: The efficacy analyses will be done descriptively on the safety population by the originally randomized treatment groups as assigned in the respective previous studies (placebo, 150 mg q2w, 200 mg q2w, other sarilumab, and active control treatment groups) as well as overall and by visit in observed case, as appropriate

- Add: Radiographic progression of the mTSS is defined as a change from baseline in the mTSS >0. A change from baseline in the mTSS of less than or equal to zero is considered as no progression. The event of missing a change from baseline in the mTSS will be considered as progression. Summary of incidence of radiographic progression of mTSS will be provided
- Deleted X-ray assessment at Week 192 as the time point is very close to EOT visit. X-rays performed at EOT will serve as source to provide X-ray data after 5 years of treatment with sarilumab.
- Changed Section 13.8 title “Database Lock” as “Interim Analysis” with sentence modification.

Amendment 9

- Adjust timepoints up to Week 264 for efficacy endpoints (except DAS28 and EULAR) and health economic variables
- The timepoints are up to Week 516 for DAS28 and EULAR endpoints
- Updated analysis treatment groups for patients from EFC11072, ACT11575, EFC10832 and SFY13370 with sarilumab + DMARD therapy in the initial and current studies , and for patients from EFC13752 with sarilumab monotherapy in the initial and current studies, per previous SAP version 2
- Clearly state serum concentrations of bound Sarilumab are not analyzed per sponsor’s decision after EFC11072 study
- Clearly state 3 campaigns of x-ray data (2 year data up to Week 48, 3 year data up to Week 96, and 5 year data up to Week 192 in LTS11210)
- Add the PFS-S sub-study (see Protocol Amendment 9 Appendix R)

1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN

The statistical analysis plan (SAP) Versions 1 and 2 were based on the original protocol dated 16 Febr2010, and Protocol Amendments 1-8, respectively. The current SAP is modified per Protocol Amendment 9 with the modification mentioned in Section 1.5

2 STATISTICAL AND ANALYTICAL PROCEDURES

Below are general rules for the analysis procedure.

For patients from EFC11072, ACT11575, EFC10832 and SFY13370 with sarilumab + DMARD therapy in the initial and current studies, all analyses will be performed based on previous treatment prior to enrollment into LTS11210 (placebo, sarilumab 150 mg q2w, sarilumab 200 mg q2w, other sarilumab (non-selected) doses, and active control treatment) as well as overall in LTS11210, unless otherwise specified. Interim analyses may be performed based on overall sarilumab +DMARD therapy only.

For patients from EFC13752 with sarilumab monotherapy in the initial and current studies, all analyses will be performed based on previous treatment prior to enrollment into LTS11210 (sarilumab 150 mg q2w, sarilumab 200 mg q2w) as well as overall in LTS11210, unless otherwise specified. Interim analyses may be performed based on overall sarilumab monotheraphy only.

Data analysis results will be presented by sarilumab +DMARD therapy and sarilumab monotherapy in LTS11210 separately. All analyses will be performed in observed cases unless otherwise specified.

Data analyses for the PFS-S sub-study will be described in Appendix K.

2.1 ANALYSIS ENDPOINTS

2.1.1 Demographic and baseline characteristics

For patients who roll over to LTS11210, new demographic data are collected in LTS11210, but some baseline characteristic information (such as height, time since diagnosis of RA, rheumatoid factor and prior biologic use) is not collected in LTS11210. This missed information will be captured from the previous study databases.

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

Demographic characteristics

The following demographic characteristics will be summarized:

- Gender (Male, Female)
- Race (Caucasian/white, Black, Asian/Oriental, other)
- Ethnicity (Hispanic, Not Hispanic)
- Region (Region1:Western countries, Region2:South America, Region3:Rest of the World)

- Age (years)
- Age group (< 65, ≥65 and <75, ≥75 years)
- Weight (kg)
- Height (cm)
- Body mass index (BMI, kg/m^2)
- BMI group (<25, ≥25 and < 30, ≥30 kg/m^2).

Medical or surgical history

The medical and surgical history data were collected in the previous studies.

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 MedDRA currently in effect at sanofi at the time of database lock.

Disease characteristics at baseline

The following baseline disease characteristics will be summarized:

- Time since diagnosis of RA (years)
- Rheumatoid factor (Positive: ≥1 IU/mL, negative: <1 IU/mL)
- Prior biologic DMARD use for RA
- Smoking history
- Alcohol use

In addition, the summary of baseline values for clinical signs and symptoms and quality of life assessments will be presented in the summary tables for these data.

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

2.1.2 Prior or concomitant medications

All medications taken at any time during LTS11210 period, from the enrollment to the end of the follow-up period, including those ongoing at study start, are to be reported in the CRF pages.

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

Concomitant medications are any treatments received by the patient concomitantly to the/any IMP, from the first IMP intake in LTS11210 to the end of the follow-up period.

Prior medications are those the patient used prior to first IMP intake in LTS11210. Prior medications can be discontinued before first dosing or can be ongoing during treatment phase. The data before enrollment were reported in the previous study reports. However the data from enrollment up to the time of first IMP intake will be reported in LTS11210 database.

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

2.1.3 Safety endpoints

The primary endpoint is the incidence of adverse events. The safety analysis will be based on the reported adverse events (AEs) and other safety information, such as clinical laboratory data, vital signs and ECG.

Observation period

The observation period will be divided into 4 epochs:

- The SCREENING epoch is defined as beginning with the last visit in the previous studies to the first dose of IMP in LTS11210
- The TREATMENT epoch is defined as the time from first dose of IMP in LTS11210 to last dose of IMP + 13 days (6 days for patients receiving weekly injections only)
- The FOLLOW-UP epoch is defined as the time from the end of TREATMENT epoch to last dose of IMP + 60 days
- The POST-STUDY epoch is defined as the time after the last dose of IMP + 60 days

The rational of having “last dose of IMP + 60 days” as the end of follow-up epoch is to capture the events observed within an approximately 5 half-lives of the last IMP. The treatment-emergent AE (TEAE) period will include both TREATMENT and FOLLOW-UP epochs.

Please note: Not all the patients will have all 4 epochs. For example, if a patient died on the 7th day after the last dose of IMP, his/her TREATMENT epoch will be from the first dose of IMP to his/her death date and will have no FOLLOW-UP or POST-STUDY epoch; if a patient had his last visit (i.e. follow-up visit 29) on the 50th day after the last dose of IMP and have no AE reported after this date, his/her FOLLOW-UP epoch will be ended on last dose of IMP+50 days and will have no POST-STUDY epoch.

2.1.3.1 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 dose of IMP in LTS11210. The ongoing AEs from previous studies filled out on AE switch CRF pages are also considered as pre-treatment AEs

- TEAEs are AEs that developed or worsened or became serious during the TEAE period
- Post-study AEs are AEs that developed or worsened or became serious during the post-study period

All AEs (including SAEs and AESIs and ongoing AEs from previous studies filled out on AE switch CRF pages) will be coded to a "Lowest Level Term (LLT)", "Preferred Term (PT)", "High Level term (HLT)", "High Group Level Term (HLGT)" and associated primary "System Organ Class (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 and ongoing AEs from previous studies filled out on AE switch CRF pages) will be recorded from the time of signed informed consent until the end of the study.

Adverse event of special interest (AESI)

AESIs will be flagged in the database using search criteria. All AEs captured on the general AE form, the ALT increase form, the suspected infection event form, the diverticulitis or gastrointestinal ulceration form, or the neurological event form will be searched. The list of the AESIs with the search criteria are provided below:

Table 1 - AESI and search criteria

AESI flag	Search[a] criteria
Leukopenia	SMQ: Haematopoietic leukopenia
Thrombocytopenia	SMQ: Haematopoietic thrombocytopenia
Infections (Opportunistic infections, Tuberculosis)	Primary SOC: Infections and infestations (CRF checkbox: Opportunistic infections; HLT Tuberculosis infections)
Hepatic disorders	SMQ : Drug-related hepatic disorders – comprehensive search
Diverticulitis/potential GI perforations[b]	SMQ: Gastrointestinal perforation and HLT: Diverticulum inflammations
GI ulcerations	SMQ: Gastrointestinal ulceration
Elevation in lipids	SMQ: Dyslipidemias
Anaphylaxis	SMQ: Anaphylactic reaction
Hypersensitivity	SMQ Hypersensitivity
Injection site reactions	HLT Injection site reactions
Malignancy	SMQ: Malignant or unspecified tumours Excluding non-melanoma skin cancer: SMQ Malignant or unspecified tumours excluding HLT of Skin neoplasms malignant and unspecified (excl melanoma)
Lupus-like syndrome	SMQ: Systemic lupus erythematosus
Demyelinating disorders	SMQ: Demyelination
Pregnancy[c]	CRF checkbox: Pregnancy
Overdose[c]	CRF checkbox: Symptomatic overdose

a All SMQs are narrow search
b These events will be further reviewed to identify cases of GI perforations
c Standard AESI applies to all Sanofi studies

Cardiovascular Adjudication Committee

All deaths and potential CV events reported by the investigator, identified by pre-specified search criteria described in the Cardiovascular Adjudication Committee (CAC) charter will be prospectively adjudicated in a blinded manner by the CAC. The committee will adjudicate qualifying events according to the CAC Charter.

The CV adjudication committee will adjudicate each event as: a CV event, a non-CV event or not evaluable (Appendix J provides further categorization of CV event, non-CV event).

Two composite Major Adverse Cardiac Events (MACE) endpoints are defined as:

- MACE (primary): CV Death, MI, Stroke, Hospitalization for unstable angina (UA) or Hospitalization for Transient Ischemic Attack (TIA).
- MACE (narrow): CV Death, MI, or Stroke.

2.1.3.2 Deaths

The deaths observation period are per the observation periods defined above:

- Death on-treatment: deaths occurring during the treatment period
- Death during follow-up: deaths occurring during the follow-up period
- Death post-study: deaths occurring during the post-study period

2.1.3.3 Laboratory safety variables

Clinical laboratory data consists of blood analysis, including 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.

Blood samples for clinical laboratories will be taken at Visits 1-49 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, RBC, RBC morphology (if RBC count is abnormal), and prothrombin time.
  - **White blood cells:** WBC, WBC differential (neutrophils/ANC, lymphocytes, monocytes, basophils, eosinophils)

- Clinical chemistry
  - **Metabolism:** total cholesterol, HDL, LDL, ratio of HDL versus LDL (HDL:LDL), triglycerides tests; glucose and total protein in fasting condition; HbA1c; Apolipoprotein A and B (determination is not performed in SFY13370 and results will not be available for the initial LTS11210 visit for SFY13370 patients); and acute phase reactants: hs-CRP.
  - **Electrolytes**: sodium, potassium, chloride, calcium, and bicarbonate.
  - **Renal function**: creatinine, CLcr, BUN, uric acid.
  - **Liver function**: ALT/SGPT, AST/SGOT, ALP, albumin, total bilirubin, conjugated and unconjugated bilirubin, and LDH.
  - **Antibody:** anti-CCP antibodies, RF, ANA/anti-ds DNA.
  - **Pregnancy test** : Serum β-hCG for all female patients.

Urine samples will be collected as follows (when urine dipstick is abnormal at site):

- **Urinalysis** - quantitative analyses: pH, specific gravity, ketone, proteins, blood, glucose, bilirubin, urobilinogen, nitrate and leukocytes.

Technical formulas are described in Section 2.5.1.

2.1.3.4 Vital signs variables

Vital Signs include: Heart Rate (HR), Systolic and Diastolic Blood Pressure (SBP and DBP) according to position (standing and supine), as well as orthostatic changes in blood pressure (standing – supine). It also includes weight and temperature.

2.1.3.5 Electrocardiogram variables

A standard 12-lead ECG will be performed and will be determined using centralized automatic and manual readings.

ECG parameters include Heart Rate, PR, QRS, QT as well as corrected QTc (according to Bazett/Fridericia), ST deviation, T-wave morphology, and U wave presence or absence.

2.1.4 Efficacy endpoints

Efficacy is a secondary endpoint of this study. The efficacy variables will include:

- ACR20 response over time up to Week 264
- ACR50 and ACR70 response over time up to Week 264
- Change from baseline in ACR components up to Week 264
- DAS28 –CRP response over time

- DAS28 remission over time
- EULAR response over time
- Change from baseline in the van der Heijde modified total Sharp score (for patients from EFC11072 Part B Cohort 2 and Cohort 1 selected dose arms with an end of treatment X-ray evaluation) at Weeks 48, 96, 144 and 192 or EOT if EOT is prior to Week 192
- Incidence of radiographic progression of the van der Heijde modified total sharp score at Weeks 48, 96, 144, and 192 or EOT if EOT is prior to Week 192
- Change from baseline in HAQ-DI over time up to Week 264.

Note that TJC and SJC (both based on 28 joints), patient's global assessment of disease activity and hs-CRP will still be assessed after Week 264 to calculate DAS28.

As mentioned before, the baseline value for efficacy parameters is the original baseline from the previous study databases for patients who rolled over to LTS11210.

A) ACR20 response over time

To be classified as an ACR20 Responder, a patient must achieve 20% improvement in the Tender joint count (68 joints) and the Swollen joint count (66 joints), as well as 20% improvement in at least 3 of the following 5 assessments:

- Levels of an acute phase reactant (CRP level)
- Patient's assessment of pain based on the 0 – 100mm Visual Analogue Scale (VAS)
- Patient's global assessment of disease activity based on the VAS
- Physician's global assessment of disease activity based on the VAS
- Patient's assessment of physical function based on the HAQ_DI

The 7 ACR components assessing the signs and symptoms of RA are measured as follows

a) Tender Joint Count (TJC)

A total of 68 joints will be assessed for tenderness. The 68 joints to be examined for tenderness are: temporomandibular (n=2), sternoclavicular (n=2), acromioclavicular (n=2), shoulder (n=2), elbow (n=2), wrist (n=2), metacarpophalageal (n=10), interphalangeal of thumb (n=2), distal interphalangeal (n=8), proximal interphalangeal (n=8), hip (n=2), knee (n=2), ankle mortise (n=2), ankle tarsus (n=2), metatarsophalangeal (n=10),interphalangeal of great toe (n=2), and proximal/distal interphalangeal of the toes (n=8).

A formal count of the joints will be performed by a trained independent assessor. Joint tenderness is defined as pain induced by the pressure of the joints, exerted by the assessor's thumb and index finger. The assessor will classify each joint as painful yes/no and swollen yes/no. A score of 0/1 will be given to each tender joint with 0 representing no pain and 1 representing pain. The tender joint count ranges from 0 to 68 where 0 is considered the best and 68 the worst.

b) Swollen Joint Count (SJC)

The 66 joints to be examined for swelling are the same as those examined for tenderness, except the hip joints are not included.

A formal count of the joints will be performed by a trained independent assessor. The assessor will classify each joint as swollen yes/no. A score of 0/1 will be given to each swollen joint with 0 representing no swollen and 1 representing a swollen joint. The swollen joint count ranges from 0 to 66 where 0 is considered the best and 66 the worst.

c) Physician's Global Assessments of Disease Activity

Physician's global assessments of the patient's current disease activity will be accessed 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).

d) 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 where 0 is considered the best disease activity (no disease activity) and 100 the worst (most disease activity).

e) Patient's Assessment of Pain

Patients will be requested to indicate their pain intensity due to their RA using a 100 mm horizontal VAS where 0 is considered "No pain" and 100 "the worst pain you can imagine".

f) Patient's Assessment of Physical function – Health Assessment Questionnaire Disease Index (HAQ-DI)

The HAQ-DI is a standardized questionnaire developed for use in RA. The HAQ-DI, with the past week as the time frame, focuses on whether the respondent "is able to…" do the activity and covers 8 categories in 20 items: dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities, for which there are at least 2 questions by category. The 4 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 8 category scores by using the highest sub-category score from each category.
   - For example, in the category ARISING there are 3 sub-category items. A patient who responds with a 1, 2, and 0, respectively, results in category score of 2.
2. Adjust for use of aids/devices and/or help from another person when indicated.
   - 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 2, it remains a 2.
   - 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 0-3 (3=worst functioning).

A HAQ-DI score cannot be calculated validly when there are scores for less than 6 of the 8 categories. HAQ-DI scoring ranges between 0 and 3. A high HAQ-DI score has been found to be a strong predictor of morbidity and mortality in RA.

g) The level of an acute phase reactant measured by CRP

High sensitivity CRP will be assessed centrally. Since CRP levels are directly correlated with Interleukin 6 receptor (IL6-R) activity, it is expected that active dose regimens will have a dramatic lowering effect on CRP levels.

The ACR components are further described in

Table 2 - ACR components

ACR component	Range	Direction
TJC	0-68	Lower is better
SJC	0-66	Lower is better
Pain VAS	0-100	Lower is better
Patient global VAS	0-100	Lower is better
Physician global VAS	0-100	Lower is better
HAQ-DI	0-3	Lower is better
CRP (mg/dL)	>0	Lower is better

TJC: Total Joint Counts, SJC: Swollen Joint Counts, VAS: Visual Analog Scale

B) ACR50 response over time

ACR50 is defined as the percentage of patients who achieve at least 50% improvement in both tender joint count and swollen joint count, and at least 50% improvement in at least 3 of the 5 remaining ACR components.

C) ACR70 response over time

ACR70 is defined as the percentage of patients who achieve at least 70% improvement in both tender joint count and swollen joint count, and at least 70% improvement in at least 3 of the 5 remaining ACR components.

D) Change in ACR components over time

The change in each ACR component will be calculated from baseline of the previous studies.

E) DAS28-CRP over time

DAS28-CRP is a composite score that includes 4 variables:

- Tender joints count (based on 28 joints: shoulder (n=2), elbow (n=2), wrist (n=2), metacarpophalageal (n=10), interphalangeal of thumb (n=2), proximal interphalangeal (n=8), knee (n=2))
- Swollen joints count (based on 28 joints: shoulder (n=2), elbow (n=2), wrist (n=2), metacarpophalageal (n=10), interphalangeal of thumb (n=2), proximal interphalangeal (n=8), knee (n=2))
- General health (GH) or patient's global assessment of disease activity based on VAS (mm) assessed from the ACR questionnaire
- Marker of inflammation here assessed by the hs-CRP (mg/L).

It is a continuous measure allowing for measurement of absolute change in disease burden and percentage improvement. It has been extensively validated for its use in clinical trials and accepted by health authorities. Using this data, the DAS28 can be calculated using the following formula:

$$\text{DAS28-CRP} = 0.56 \times \sqrt{28TJC} + 0.28 \times \sqrt{28SJC} + 0.36 \times \text{Log}(\text{CRP}+1) + 0.014 \times \text{GH} + 0.96,$$

Where 28TJC and 28SJC are the tender and swollen joint counts from 28 joints and GH is the patient's global assessment of disease activity and CRP is in mg/L.

DAS28-CRP provides a number on a scale from 0 to 10 indicating the current activity of the RA.

DAS28 –CRP will be considered as missing if one of the components is missing.

F) DAS28 remission over time

DAS28 Remission is the event of having a DAS28-CRP score <2.6.

G) European League Against Rheumatism (EULAR) response over time

The European League Against Rheumatism (EULAR) response criteria combine the DAS28 score at the time of evaluation with the change in DAS28 score between 2 time points. The EULAR response criteria are defined as:

- Good response = The event of having an improvement of >1.2 and a present DAS28-CRP score ≤3.2.
- Moderate response = The event of having either an improvement of >0.6 to ≤1.2 and a present DAS28-CRP score ≤5.1, OR, having an improvement of >1.2 and a present DAS28-CRP score >3.2.
- No response = The event of having either an improvement of ≤0.6, OR, having an improvement of >0.6 to ≤1.2 and a present DAS28-CRP score >5.1.

Table 3 - EULAR response categories

DAS28-CRP score	Change from baseline in DAS28-CRP score		
	< -1.2	< -0.6 to ≥ -1.2	≥ -0.6
≤ 3.2	Good response	Moderate response	No response
> 3.2 to ≤ 5.1	Moderate response	Moderate response	No response
> 5.1	Moderate response	No response	No response

H) Change in van der Heijde modified total Sharp score over time

X-ray data will be collected for patients from EFC11072 Part B Cohort 2 and Cohort 1 selected dose arms only. The degree of joint damage will be assessed using the van der Heijde modified total Sharp score (mTSS). This methodology quantifies the extent of bone erosions for 44 joints and joint space narrowing for 42 joints, with higher scores representing greater damage. The van der Heijde mTSS at a time point is the sum of the scores from both the erosion score and the joint space narrowing score, for a maximum score of 448.

Erosion score

The joint erosion score is a summary of erosion severity in 32 joints of the hands and 12 joints of the feet. Each joint is scored, according to the surface area involved, from 0 to 5 for hand joints and 0 to 10 for the joints of the foot. The maximum erosion score (5 for the hand and 10 for the foot) indicates extensive loss of bone from more than one half of the articulating bone. A score of 0 in either the hand or foot indicates no erosion. The maximum erosion score is 280 (160 in the hands and 120 in the feet) for a time point.

Joint space narrowing score

The joint space narrowing (JSN) score summarizes the severity of JSN in 30 joints of the hands and 12 joints of the feet. Assessment of JSN for each hand (15 joints per hand) and foot (6 joints per foot), including subluxation, is scored from 0 to 4, with 0 indicating no/normal JSN and 4 indicating complete loss of joint space, bony ankylosis or luxation. Thus, the maximum JSN score is 168 at a time point.

X-ray data reading

Campaign 1

X-rays read in Campaign 1 will include baseline and EOT (or last available visit) from EFC11072, Week 48 and any unscheduled visits between Week 0 and Week 48 in LTS11210. Two primary readers and an adjudicator will perform reading of data up to Week 48 in LTS11210 (see details in Appendix D). The X-rays at baseline and Year 1 (EOT or last available visit in EFC11072) will be reread.

The linear extrapolation method will be used to impute missing modified total Sharp score, erosion score, or joint space narrowing score at Week 48 in LTS11210 from Campaign 1 data for patients who have EOT (or last available data) in EFC11072 and any unscheduled visit data between Week 0 and Week 48 in LTS11210. In the linear extrapolation method, for example, the missing modified total Sharp score at Week 48 is imputed according to the following formula:

Week 48 modified total Sharp score = EOT score in EFC11072 + 336(modified total Sharp score of the last measurement in LTS11210 – EOT score in EFC11072)/(date of last measurement in LTS11210 – date of EOT visit in EFC11072).*

Campaign 2

X-rays in Campaign 2 will include baseline from EFC11072, Weeks 48 and 96, and any unscheduled visits between Week 48 and Week 96 in LTS11210. The linear extrapolation method will be used to impute missing the modified total Sharp score, erosion score, or joint space narrowing score at Week 96. Similarly, the missing modified total Sharp scores at Week 96 are imputed by the linear extrapolation method according to the following formula:

Week 96 modified total Sharp score = Week 48 score + 336(modified total Sharp score of the last measurement – Week 48 score)/(date of last measurement – date of Week 48 measurement),*

Campaign 3

X-rays in Campaign 3 will include baseline from EFC11072, Weeks 96, 144 and 192 (or EOT), and any unscheduled visits between Week 96 and Week 192 (or EOT) in LTS11210. The linear extrapolation method will be used to impute missing the modified total Sharp score, erosion score, or joint space narrowing score at Weeks 144 and 192 in LTS11210 based on Campaign 3 data. Similarly, the missing modified total Sharp scores at Week 144 and Week 192 in LTS11210 are imputed by the linear extrapolation method according to the following formula:

Week 144 modified total Sharp score = Week 96 score+ 336(modified total Sharp score of the last measurement – Week 96 score)/(date of last measurement – date of Week 96 measurement),*

Week 192 modified total Sharp score = Week 96 score + 672(modified total Sharp score of the last measurement – Week 96 score)/(date of last measurement – date of Week 96 measurement),*

Where EOT visit for X-ray in LTS11210 will be prior to Week 192 if a patient discontinues the study prior to Week 192.

I) Incidence of radiographic progression of the van der Heijde modified total Sharp score over time

Note that missing modified total Sharp scores are imputed by the linear extrapolation method. And then a change from the baseline in the modified total SHARP score will be calculated. A change from the baseline in the modified total SHARP scores of ≤ 0 is considered as no progression.

J) Change from baseline in HAQ-DI

The HAQ-DI is composed of 8 categories as follows: dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities, for which there are at least 2 questions by category. The patient will be asked to score how difficult he/she feels it is to perform such activities using a 0 to 3 scoring (0=without any difficulty, 1=with some difficulty, 2=with much difficulty and 3=unable to do). If the patient is using assistance for any of these activities, scoring may be adjusted. For each category, the highest score given for one of the question is attributed to the category. The total score is the sum of all categories' scores divided by the number of answered categories (at least 3 categories should be answered).

2.1.5 Pharmacokinetic and immunogenicity variables

Immunogenicity (anti-sarilumab antibody) and pharmacokinetic blood samples (trough concentrations) will be collected according to the study flow chart. Pharmacokinetic samples will also be collected at or near the onset and completion of the occurrence of a serious adverse event. Note that there are some missing data after Week 48 since pharmacokinetic evaluation were added after Week 48 per Protocol Amendment 8 and by that time some patients already reach more than 48 weeks of treatment.

At each sample time the result of the 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.

2.1.6 Pharmacodynamic/genomics endpoints

Serum concentrations of IL-6 will be assessed according to the study flow chart.

2.1.7 Quality-of-life endpoints

SF-36 V2 (for patients from EFC11072, ACT11575 and EFC10832 only)

The SF-36 version 2 is a multi-purpose, short-form health survey with 36 questions. It yields scores for 8 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 2 summary measures of physical and mental health: the Physical Component Summary and Mental Component Summary (1).

The scoring process is summarized below:

1. Enter item response data
2. Recode item response values
3. Determine health domain scale raw scores.
4. Transform health domain scale raw scores to 0-100 scores.

5. Transform health domain scale 0-100 to norm-based scores
6. 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	Physical Functioning (PF)	3a, 3b, 3c, 3d, 3e, 3f, 3g, 3h, 3i, 3j
Summary (PCS)*	Role-Physical (RP)	4a, 4b, 4c, 4d
	Bodily Pain (BP)	7, 8
	General Health (GH)	1, 11a, 11b, 11c, 11d
Mental Component	Vitality (VT)	9a, 9e, 9g, 9i
Summary (MCS)*	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 I provides the SAS code to calculate the 8 scales, the 2 summary measure scores and 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) will be analyzed.

2.1.8 Health economic endpoints

WPAI (for patients from EFC11072 and ACT11575 only)

Work Productivity and Activity Impairment (WPAI) will be assessed. WPAI outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity, ie., worse outcomes, as follows:

- Q1 = currently employed
- Q2 = hours missed due to specified problem (RA)
- Q3 = hours missed other reasons
- Q4 = hours actually worked
- Q5 = degree problem affected productivity while working
- Q6 = degree problem affected regular activities Scores

Scores will be expressed in percentages.

- Percent work time missed due to problem (RA): 100*Q2/(Q2+Q4)
- Percent impairment while working due to problem: 10*Q5
- Percent activity impairment due to problem: 10*Q6
- Percent overall work impairment due to problem (RA):

  100*Q2/(Q2+Q4)+100*[(1-Q2/(Q2+Q4))*(Q5/10)]

When deriving the 4 WPAI scores, missing data will be handled using the following rules.

- Percent work time missed due to problem (RA): The score will be missing if either Q2 or Q4 is missing.
- Percent impairment while working due to problem: The score will be missing if Q5 is missing.
- Percent activity impairment due to problems: The score will be missing if Q6 is missing. Percent overall work impairment due to problem (RA): The score will be missing if either percent work time missed due to problem (RA) is missing or percent impairment while working due to problem is missing.

FACIT-Fatigue (for patients from EFC11072, ACT11575 and EFC10832 only)

The Functional Assessment of Chronic Illness Therapy Fatigue scale (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 RA 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=some what, 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 7-8) 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 7-8)]/number of answered items*

Sleep questionnaire (for patients from EFC11072 and ACT11575 only)

RA like other chronic illness is associated with sleep disturbances. Sleep disturbances is linked to pain, mood and disease activity. The effect of sarilumab on pain will be assessed on a sleep questionnaire VAS scale ranges from 0 (sleep is not a problem) to 100 (sleep is a major problem).

WPS-RA (for patients from EFC10832 only)

The rheumatoid arthritis-work productivity survey (WPS-RA) is a validated questionnaire that evaluates productivity limitations within work and within home associated with RA over the previous month. The questionnaire is interviewer-administered and is based on patient self-report.

It contains 9 questions addressing employment status (1 item), productivity at work (3 items) and within and outside the home (5 items) (Appendix H).

2.2 DISPOSITION OF PATIENTS

This section describes patient disposition for both patient study status and the patient analysis populations for LTS11210 only, unless otherwise specified.

- All screened patients are defined as those patients who signed the informed consent.
- All enrolled patients are defined as all patients with signed informed consent who met the eligibility criteria to enter into LTS11210.
- The Safety population consists of all enrolled patients who received at least one dose of the LTS11210 study medication.

For patient study status, the total number of patients for each one of the following categories will be presented using a summary table:

- All screened patients (informed consent signed)
- Screen failure patients and reasons for screen failure (not eligible per eligible CRF page)
- Non-enrolled but treated patients
- All enrolled patients by previous studies
- Enrolled but not treated patients
- Enrolled and treated patients
- Patients who had dose reductions
- Patients who discontinued study treatment by main reason for permanent treatment discontinuation
- Status at last study contact

For all categories of patients (except for the screened and non-enrolled categories), percentages will be calculated using the total of all enrolled patients as the denominator. Reasons and time to treatment discontinuation will be supplied in tables giving numbers and percentages. This summary will be further sub-grouped by region and by prior biological use.

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

All major or critical deviations potentially impacting analyses, drug dispensing irregularities and other major or critical deviations will be summarized in tables giving numbers and percentages of deviations.

The analysis populations for safety, efficacy and PK will be summarized in a table by patient counts on the enrolled patients.

- Safety population (for safety and efficacy analyses),
- PK population.

2.2.1 Randomization and drug dispensing irregularities

There is no randomization procedure in LTS11210. Drug dispensing irregularities occur whenever:

- A patient is dispensed an IMP kit not allocated by the IVRS, such as, a) a patient at any time in the study is dispensed a different treatment kit than as assigned (which may or may not contain the correct assigned IMP), or b) a non-screened patient is treated with study medication reserved for enrolled patients.

Drug dispensing irregularities will be monitored throughout the study and reviewed on an ongoing basis.

All drug dispensing irregularities will be documented. If the number of irregularities is large enough to make a tabular summary useful, the irregularities will be categorized and summarized among all patients (number and percentages).

Drug dispensing irregularities to be prospectively identified include but are not limited to:

Table 5 - Drug Allocation Irregularities

Drug Allocation Irregularities
Kit dispensation without IVRS transaction
Erroneous kit dispensation
Kit not available
Subject switched to another site

2.3 ANALYSIS POPULATIONS

The primary analysis population is the safety population. All safety analyses and efficacy analyses will be performed based on the safety population.

The patients who enroll into LTS11210 study but are not treated will be reported separately, and these patients will not be in the safety population for this study.

2.3.1 Safety populations

Safety population is defined as all enrolled patients who have received at least one dose of the study treatment in LTS11210.

All analyses will be performed on the safety population based on previous treatment prior to enrollment into LTS11210. This will allow for assessment of any differences between treatment groups based on previous treatment.

2.3.2 Efficacy population

The efficacy analysis population is the same as safety population.

Efficacy analyses will be performed on the safety population based on previous treatment prior to enrollment into LTS11210. This will allow for assessment of the efficacy improvement if it is related to the different original treatments. Note that X-ray data will be collected for patients from EFC11072 Part B Cohort 2 and Cohort 1 selected dose arms only.

2.3.3 Pharmacokinetics and immunogenicity population

The PK population will consist of all patients in the safety population with at least one post-dose non-missing plasma 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.

2.4 STATISTICAL METHODS

2.4.1 Demographics and baseline characteristics

Only patients enrolled and treated in LTS11210 will be summarized on demographic and baseline characteristic data.

Continuous data will be summarized using number of available data, mean, SD, median, minimum, and maximum. Categorical and ordinal data will be summarized using number and percentage of patients.

Parameters described in Section 2.1.1 will be summarized based on the safety population using descriptive statistics.

Medical and surgical history will be summarized on the safety population by system organ class (SOC) and preferred term (PT) sorted by internationally agreed order of SOC and by the decreasing frequency of PT within SOC.

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

2.4.2 Prior or concomitant medications

Prior medications are those the patients used prior to first IMP intake in LTS11210. The data before enrollment were collected and reported in the previous studies. However the data from enrollment up to the time of first IMP intake will be reported in LTS11210 database.

Prior and concomitant medications in LTS11210 will be presented based on the enrolled population.

Medications will be summarized by treatment received according to the WHO-DD dictionary, considering the first digit of the ATC class (anatomic category) and the first three digits of the ATC class (therapeutic category). All ATC codes corresponding to a medication will be summarized. Patients will be counted once in each ATC categories (anatomic or therapeutic) linked to the medication, therefore, patients may be counted several times for the same medication. In addition, the summaries for concomitant DMARDs (MTX, non-MTX DMARDs), folic acid, NSAIDs, corticosteroids, and lipid modifying agents will also be provided.

The tables for concomitant medications will be sorted by decreasing frequency of anatomic category followed by all other therapeutic classes. In case of equal frequency regarding anatomic categories (respectively therapeutic categories), alphabetical order will be used.

2.4.3 Extent of investigational medicinal product exposure and compliance

The extent of IMP exposure and compliance will be presented on the safety population. In addition, the extent of IMP exposure will also be assessed and summarized by actual treatment received (150 mg qw, 200 mg q2w, 150 mg q2w) in LTS11210.

Dose reduction

Per study design, patients can reduce dose due to safety consideration. These patients who were switched to a lower dose regimen (ie. 150 mg q2w) due to safety consideration, not administration errors, will be listed and/or summarized separately. There are 3 type cases of dose reduction: initial 150 mg qw reduced to 150 mg q2w, initial 200 mg q2w reduced to 150 mg q2w, initial 150 mg qw switched to 200 mg q2w (per Protocol Amendment 3) then reduced to 150 mg q2w. The reasons for dose reduction are collected and the summary will be provided accordingly. In addition, time to first dose reduction will be analyzed using Kaplan-Meier method for each type of dose reduction by previous treatment prior to enrollment into LTS11210. The maximal grade of neutropenia and maximal grade of ALT elevation after dose reduction will be summarized for each type of dose reduction by dose reduction reason. Treatment discontinuation after dose reduction, such as discontinuation reason and time from dose reduction date to treatment discontinuation date, will also be analyzed by summary and Kaplan-Meier plot. Detailed analyses will be performed in the integrated summary of safety (ISS) and the final CSR.

2.4.3.1 Extent of investigational medicinal product exposure

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

Extend of IMP exposure

Due to a global switch to the pivotal dose or a dose reduction to a lower pivotal dose if it applies, each patient may receive multiple doses. Hence the duration of IMP exposure will be calculated under 'sarilumab' treatment regardless what doses the patient received during LTS11210.

For patients receiving weekly injections at the time of last injection, duration of IMP exposure is defined as: [Date of last injection of the study medication in LTS11210] – [Date of first injection of the study medication in LTS11210] + 7 day, regardless of unplanned intermittent discontinuations (see Section 2.5.3 for calculation in case of missing or incomplete data).

For patients receiving biweekly injections at the time of last injection, duration of IMP exposure is defined as: [*Date of last injection of the study medication in LTS11210*] – [*Date of first injection of the study medication in LTS11210]* + 14 *day*, regardless of unplanned intermittent discontinuations (see Section 2.5.3 for calculation in case of missing or incomplete data).

Duration of each dose

Period for each treatment dose received is calculated as the days from the starting (injection) date of that dose regimen to one day before the starting date of the following dose regimen or to the end (injection) date of the same dose + 14 days (7 days for patients receiving weekly injections only) if that is the last dose regimen in the study. Duration of 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 counts and percentages for each of the following categories and cumulatively according to these categories: ≤4 weeks, >4 and ≤12 weeks, >12 and ≤24 weeks, > 24 and ≤48 weeks, > 48 and ≤72 weeks, >72 and ≤96 weeks, and at 24-week intervals through the end of the study.

Additionally, the cumulative duration of treatment exposure, defined as the sum of patients' duration of treatment exposure and expressed in patient years, will be provided.

2.4.3.2 Compliance

A given administration will be considered as non-compliant 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 a patient will be defined as the number of administrations (injections) the patient was compliant divided by the total number of administrations (injections) the patient was planned to take during the treatment epoch (ie, from the 1st to the last administration).

Treatment compliance percentages will be summarized descriptively as quantitative variables (Number, Mean, SD, Median, Min, and Max). The percentage of patients whose compliance is <80% will be summarized. In addition, patients with the wrong dose or an overdose will be summarized.

Cases of overdose (administering 2 or more sarilumab doses in less than 11 calendar days) will be listed as such. More generally, dosing irregularities are listed in Section 2.2.1.

2.4.4 Analyses of safety data

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:

- The baseline value for safety parameters is the original baseline from the previous studies.
- The value at Week 0 is the last non˜ missing measurement in data collected at scheduled visits: visit1, visit2 and/or visit1/2 in LTS11210 or EOT visit of previous studies prior to first IMP in LTS11210 study design.
- 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 version in BTD-009536 version 3, see Appendix A).
- PCSA criteria will determine which patients had at least one PCSA during the TEAE period, taking into account all evaluations performed during the TEAE period, including non-scheduled 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 TEAE period by treatment received 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 received. 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.
- The analysis of the safety variables will be essentially descriptive and no systematic testing is planned.

2.4.4.1 Analyses of adverse events

Generalities

The primary focus of adverse event reporting will be on TEAEs. Pre-treatment and post-study 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. The algorithm for imputing date/time of onset will be conservative and will classify an AE as a treatment emergent unless there is definitive information to determine it is pre-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 system-organ-class (SOC), High level group term (HLGT), High level term (HLT) and preferred term (PT) sorted in alphabetical order by treatment received and overall, the number (n) and percentage (%) of patients experiencing an AE. Multiple occurrences of the same event in the same patient will be counted only once in the tables within a dose group. Overall counts will also be provided, counting each patient once and also counting the total number of events. The denominator for computation of percentages is the safety population within the treatment group.

The incidence and number of events per 100 patient-years (number of events adjusted for the total duration of exposure) will be provided for all the TEAE summaries. In addition, the number of patients with event(s) per 100 patient-years (number of patients adjusted for the exposure up to the first event or up to the end of the duration of exposure for patients with no event) will also be provided for serious TEAEs and for each type of adverse event of special interest (AESI).

Sorting within tables should ensure same presentation. For that purpose, the table of all TEAEs presented by SOC and PT sorted by internationally agreed order of SOC and decreasing frequency of PT within SOC will define the presentation order for all other tables unless otherwise specified. Sorting will be based on results for overall sarilumab group.

Listings will be provided for all pre-treatment AEs, all TEAEs, all SAEs, all CV events undergone adjudication, and TEAEs leading to treatment discontinuation with flags indicating on-treatment status

Analysis of all TEAE(s)

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, and 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.
- 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 (i.e., mild, moderate, or severe), sorted by sorting order defined above.

Analysis of all treatment emergent SAE(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, sorted by SOC internationally agreed order and decreasing incidence of PTs within SOC
- A separate listing of non-treatment emergent SAEs occurred in previous studies which are ongoing from the previous studies into LTS11210 study will be provided.

Analysis of all TEAE(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. The table summary will be presented only if the number of patients concerned is greater than 3.
- All TEAEs leading to treatment discontinuation, 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
- A separate listing of AEs (occurred in previous studies) leading to treatment discontinuation in LTS11210, which are ongoing from the previous studies into LTS11210 study, will be provided.

Analysis of adverse events of special interest (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

The number at risk and the number with events will be provided by visit and treatment group

Analysis of adjudicated CV events

The following CV event summaries will be generated for the safety population:

- MACE (primary) and MACE (narrow), showing number (%) of patients with at least one adjudicated treatment-emergent CV event and number of such events per 100 patient-years (number of events adjusted for the total duration of exposure).
- All adjudicated treatment-emergent CV events by CV events categories, showing number (%) of patients with at least one adjudicated treatment-emergent CV event and number of events per 100 patient-years, sorted by alphabetical order.
- All events undergone adjudication, all events adjudicated as non-CV event, and all events that are not evaluable, showing number (%) of events (patients).

Summary describing the investigator-reported treatment-emergent SAE and the associated CAC adjudication classification will be provided.

2.4.4.2 Deaths

The incidence of death and number of death per 100 patient-years (number of death adjusted for the total duration of exposure) will be generated on the safety population. The following deaths summaries will be generated on the safety population.

- Number (%) of patients who died by study period (on-treatment, during follow-up or post-study)
- TEAE leading to death (death as an outcome on the AE CRF page as reported by the Investigator) by primary SOC , HLGT, HLT and PT showing number (%) of patients sorted by internationally agreed order of SOC and alphabetic order of HLGT, HLT, and PT.

Listings will be provided for all deaths with flags indicating on-treatment, during follow-up or post-study status.

2.4.4.3 Analyses of laboratory variables

The summary statistics (including number, mean, median, standard deviation, 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, and endpoint) by treatment received. This section will be organized by biological function as specified in Section 2.1.3.3.

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

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

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 original baseline status categories:

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

In addition, for all laboratory parameters, the mean change from baseline at each scheduled visit during the double-blind treatment period will be plotted by treatment groups.

Time to onset of the first G3 or G4 neutropenia (absolute neutrophil count <1.0 Giga/L) in LTS11210, and time to onset of the first platelet count <100 Giga/L in LTS11210 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 previous treatment prior to enrollment into LTS11210. Further analyses will be in the report of integrated summary of safety.

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 neutrophils value 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

For patients with Grade 3 or 4 neutropenia, a listing with the individual neutropenia counts, WBC, platelet counts, lymphocytes and hemoglobin at each visit (including unscheduled visits for re-test) will be provided. In addition, the neutropenia counts at each scheduled visit during the study will be plotted by previous treatment prior to enrollment into LTS11210. In addition, summaries will include the following data: discontinuation, restart dosing, dosing delay, number of episodes, and normalization (to >LLN or return to baseline if baseline is <LLN). The labs measured during the post-study period will be considered in the analysis of normalization.

Summaries of neutropenia by maximal grade at dose regimens (150 mg qw, 150 mg q2w, 200 mg q2w) may be provided.

Liver function tests

The liver function tests, namely AST, ALT, alkaline phosphatase 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 3xULN for ALT and a horizontal line corresponding to 2xULN for total bilirubin.

The normalization (to $\leq$1 ULN) or return to baseline (in case of baseline is >ULN) of elevated liver function tests will be summarized by categories of elevation (>3 ULN, >5 ULN, >10 ULN, >20 ULN for ALT and AST; >1.5 ULN for ALP; and >1.5 ULN and >2 ULN for total bilirubin), with following categories of normalization: 1) normalized on-treatment; 2) normalized after last dose; 3) last value not normal. The laboratory values measured during the post-study period will be considered in the analysis of normalization. Note that a patient will be counted only under the maximum elevation category.

2.4.4.4 Analyses of vital sign variables

The summary statistics (including number, mean, median, standard deviation, 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, and 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.

2.4.4.5 Analyses of ECG variables

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

The incidence of PCSAs at any time during the TEAE period will be summarized by treatment received 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.

2.4.5 Analyses of efficacy endpoints

There will be no confirmatory analysis for the efficacy variables. All analysis will be done descriptively on the safety population by visit in observed case, as appropriate. In addition to the by-visit analysis, efficacy assessments summarized on a yearly basis will be provided.

The baseline value for efficacy parameters is the original baseline from the previous studies. The value at Week 0 is the last non˜mi ssing measurement in data collected at scheduled visits: visit1, visit2, and/or visit1/2 in LTS11210 or EOT visit of previous studies prior to first IMP in LTS11210 study.

For categorical variables, for example, percentage of patients who have maintained/reached ACR20 overtime, the number and percentage will be provided from all patients who have data available at that time point, and the 95% confidence intervals (CI) will be calculated if appropriate.

For continuous variables, descriptive statistics such as the mean, standard deviation, median, minimum and maximum will be provided. In addition, 95% confidence interval of the mean will be presented. This applies to the original values and change (or percent change) from baseline.

The 2-year (EFC11072 52 weeks + LTS11210 48 weeks) X-ray data from Campaign 1 will be described using the van der Heijde modified total Sharp score for the subset of patients who previously completed study EFC11072, Part B when the 2-year X-ray data are available for all these patients. The modified total Sharp score and the change from baseline in the modified total sharp score will be described using mean, standard deviation, median, minimum and maximum, along with 95% confidence intervals. There is no formal treatment comparison and no alpha adjustment will be made (further analyses will be in the report of integrated summary of efficacy). In addition, the incidence of radiographic progression will be summarized using descriptive statistics.

The above analyses will also be performed for X-ray data from Campaign 2 and Campaign 3 respectively once X-ray data are available.

2.4.5.1 Multiplicity issues

Not applicable.

2.4.6 Analyses of pharmacokinetic and pharmacodynamic variables

The baseline value for PK and PD variables is the original baseline from the previous studies. The value at Week 0 is the last non˜mi ssing measurement in data collected at scheduled visits: visit1, visit2, and/or visit1/2 in LTS11210 or EOT visit of previous studies prior to first IMP in LTS11210 study design.

No PK values at baseline and Week 0 for patients with placebo treatment in EFC10832 and patients with active control (Tocilizumab) treatment in SFY13370. No ADA sample values at baseline and Week 0 for patients with active control (Tocilizumab) treatment in SFY13370.

Pharmacokinetic variables

All the PK analyses will be performed using the PK population.

Serum samples will be analyzed for functional SAR153191 concentrations using validated enzyme-linked immunosorbent assay (ELISA) methods with a targeted lower limit of quantitation (LLQ) of 0.294 mg/L (*REGN88-AV-07026-SA-01V1*) or 0.3125 mg/L (*REGN88-AV-13131-VA-01V1*). Serum concentrations of functional sarilumab will be summarized using arithmetic and geometric means, standard deviation (SD), standard error of the mean (SEM), coefficient of variation (CV%), minimum, median and maximum by actual dose and visit. Concentrations below LLOQ were replaced by LLOQ/2.

After the completion of EFC11072 (Mobility) Part B study, the Sponsor made the decision to discontinue measurement of bound sarilumab concentration from ongoing studies globally. Serum concentrations of bound sarilumab (serum sarilumab-sIL-6Rα complex) will not be summarized in this study.

Pharmacodynamics variables

Serum concentrations of IL-6 will be summarized using descriptive statistics (including number, arithmetic and geometric means, standard deviation, standard error of the mean (SEM), coefficient of variation (CV%), minimum, median and maximum) by actual dose received at each visit.

2.4.7 Analyses of quality of life/health economics variables

The summary statistics will be provided for the 8 dimensions of SF-36 and the 2 summary measures of physical and mental health (raw value and change from baseline) at each visit or study assessment (baseline, endpoint) by treatment group. Summary statistics will also be produced for WPAI questionnaires, FACIT-Fatigue, WPS-RA and the sleep questionnaire scales.

2.4.8 Analysis of immunogenicity variables

The baseline value for ADA is the original baseline from the previous studies. The value at Week 0 is the last non˜missing measurement in data collected at scheduled visits: visit1, visit2, or visit1/2 in LTS11210 or EOT visit of previous studies prior to first IMP in LTS11210 study design.

ADA population will consist of all patients in the safety population with at least one post-dose evaluable ADA sample (positive, negative) in LTS11210. Patients will be analyzed according to the treatment actually received in LTS11210.

Definitions

ADA positive patient is defined as patient with at least 1 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 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.

ADA prevalence and titer

The following summary will be provided based on ADA population:

- 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 ADA population 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
- 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 patients with neutralizing antibody
  - Number (%) of patients with non-neutralizing antibody
- Number (%) of treatment-boosted ADA positive patients.
  - Number (%) of patients with neutralizing antibody
  - Number (%) of patients with non-neutralizing antibody

A plot of the log titers over time will be provided by treatment group. At each time point, a box plot representing the titer range, Q1, median (Q2), Q3 (excluding outliers) and outliers of the positive ADA samples will be provided. The number of ADA positive samples, the number of total subjects evaluable for immunogenicity and the proportion of ADA positive subjects at that time point will be also be provided along with the box plot.

ADA and clinical safety

The safety assessment will focus on the following events:

- Hypersensitivity (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) and previous treatment prior to enrollment into LTS11210 . The relationship between the timing and the titer of the positive ADA and events will also be explored as necessary.

ADA and clinical efficacy

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

- ACR20/50 response by visit
- Number (%) of patients with lack of efficacy or loss of efficacy
  - Lack of efficacy is defined as permanent treatment discontinuation due to lack of efficacy.
  - Loss of efficacy is defined as permanent treatment discontinuation due to lack of efficacy after achieving an ACR50 or EULAR Good response (at least one visit).

ADA and PK

By visit descriptive summary of serum concentration of sarilumab will be provided by ADA patient classifications (positive or negative) for each actual sarilumab dose (150 mg qw, 150 mg q2w, or 200 mg q2w).

Scatter plot of serum concentration versus visit will also be provided by ADA classifications (positive or negative) and each actual sarilumab dose (150 mg qw, 150 mg q2w, or 200 mg q2w).

2.5 DATA HANDLING CONVENTIONS

2.5.1 General conventions

The following formulas will be used for computations of parameters.

Demographic formulas

Age is calculated as follows:

$$Age = (informed\ consent\ date - birth\ date)/365.25$$

BMI is calculated as follows:

$$BMI = Weight\ in\ kg/(height^2\ in\ meters)$$

Renal function formulas

Creatinine clearance (CLcr) value will be derived using the equation of Cockroft and Gault:

$$CLcr\ (ml/min) = (140 - age) * weight\ (kg) * (1 - 0.15*sex\ (0\text{-}M,\ 1\text{-}F)) /(0.814*creatinine\ (\mu mol/l))$$

CLcr will be calculated using weight assessed at the same visit that creatinine was assessed and age at the lab sampling date. Age is calculated as follows:

Age = (laboratory sampling date – birth date)/365.25

Number of TEAE per 100 patient years

The number of TEAE per 100 patient years is defined as the total number of TEAEs divided by the total exposure (expressed under the unit of 100 patient years) of the corresponding treatment group. Specifically, it is calculated as follows for each treatment group:

Number of TEAE/100 patient years = (Total number of TEAEs) / total treatment duration (Unit: 100 patient years)

The total treatment duration is defined as [(end of TEAE period) – (date of first IMP) + 1], where the end of TEAE period is defined as minimum [(the last IMP + 60 days), (last contact date), (death date)], where the last contact date is defined as maximum [last complete AE start date, last visit date, last subject vital status date].

The number of PCSA per 100 patient years for the laboratory values, vital signs and ECGs will be defined similarly.

Number of patients with special event per 100 patient-years

The number of patients with at least one specific event per 100 patient-years will be calculated as the number of patients having a specific event in question divided by the total person-year among patients at risk of an initial occurrence of the specific event in question. In particular, it is computed as follows:

$$100 \times \frac{n}{\sum t_i}$$

Where n is the number of patients with the specific events; t_i is patient exposure time in person-year unit. For each of the specific events of interest, the exposure time for patients who have experienced the specific events will be defined as the time to first specific event in question, whereas the exposure time for those who have not had this specific event will be total duration of exposure as defined above. Please note, the time to first PCSA will be calculated using the midpoint of the time interval between the first assessment showing the PCSA and the previous assessment.

2.5.2 Data handling conventions for efficacy variables

The missing data handling rules are summarized in a table in Appendix C.

ACR20/ACR50/ACR70

In the primary method of missing data handling, the data collected after treatment discontinuation will be set to missing. No imputation of missing post-baseline values will be performed. Responder status will be determined if possible. With these rules patients will automatically become non-responders for all time points beyond the time point they discontinue study treatment.

Patients with insufficient data (ie, TJC, SJC, and at least 3 of the remaining 5 ACR components) at baseline and a time point will be non-responders in ACR20/ACR50/ACR70 at that time point.

ACR components - TJC/SJC

When calculating the TJC and SJC, with individual missing joint scores (the '"replaced or fused" joints are not taken into consideration for the swelling or tenderness) imputed as the mean of the scored joints, the TJC/SJC after imputation are as follows:

TJC/SJC = sum (scored tender/swollen joints)(number of joints in the full joint set / number of scored tender/swollen joints)*

The number of joints in the full joint set is defined as (68 – number of replaced or fused joints) for tenderness; (66 – number of replaced of fused joints) for swollenness; and the scored joints refer to those with an answer (0 – no pain, 1- pain).

ACR components - HAQ-DI

The highest score for any component question in a category determines the category score. If the patient is using assistance for any of these activities, scoring may be adjusted. The total score is the sum of all category scores divided by the number of answered categories. At least 6 categories should be answered in order to have a total score.

DAS28-CRP

When calculating DAS28-CRP, the following rules will be applied in order:

- When calculating the 28TJC and 28SJC, with individual missing joint scores (the "replaced or fused" joints are not taken into consideration for the swelling or tenderness) imputed as the mean of the scored joints, the tender/swollen joint counts after imputation are as follows:

  28TJC/28SJC = sum (scored tender/swollen joints)(number of joints in the full joint set / number of scored tender/swollen joints)*

The number of joints in the full joint set is defined as (28 – number of replaced or fused joints) and the scored joints refer to those with an answer (0 – no pain, 1- pain).

- No imputation will be performed.
- If any of the 4 variables: 28TJC, 28SJC, GH (patient global), CRP is missing, then DAS28-CRP will be set to missing.

SF-36

The score of each of the 36 items is collected on the CRF. Appendix I provides the SAS code to calculate the 8 scales, the 2 summary measure scores and 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 the items are available and valid within the corresponding scale. The missing items will be imputed by the mean of available items.

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 drug end of treatment date is missing

For the calculation of the treatment duration, the date of the last dose of IMP is equal to the date of last administration reported on the dosing CRF page. If this date is missing, the exposure duration should be kept 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, so it cannot be determined whether it was taken prior or concomitantly, it will be considered as 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 AE when date and time of first investigational drug date is missing

When the date and time of the first investigational drug is missing all adverse events that occurred after or on the day of enrolled should be considered as treatment-emergent adverse events. The exposure duration should be kept as missing.

Handling of missing relationship to investigational product of AEs

If the assessment of the relationship to investigational product is missing, then the relationship to investigational product 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, that patient will be grouped in the category "normal /missing at baseline".

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

For PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range are missing; e.g., 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 nor be taken into account in the computation of PCSA values.

2.5.4 Windows for time points

For efficacy parameters (except X-ray), a discontinuation visit (EOT, Visit 49) will be mapped to the scheduled visit for the analyses. If a discontinuation visit happens at a regular visit, then the visit number will be reassigned as the regular visit number where an assessment is planned. If a discontinuation visit happens between 2 regular visits, then the visit number will be reassigned to the next visit number where an assessment is planned. For example, if a discontinuation visit happens between Visits 9 and 10, then the visit number will be reassigned as 10. Similarly, the EOT visit (Visit 49) for patients who complete the study with 260-week treatment duration per Protocol Amendment 8, will be mapped to the Visit 28.

For X-ray data, the linear extrapolation method for the modified total Sharp score, the erosion score, and the joint space narrowing score will use true date to extrapolate missing values at Weeks 48, 96, 144, and 192 with no visit remapping.

All efficacy analyses will be performed by the nominal visit number and no visit window will be defined.

For safety assessments, the reference date for the derivation of relative days of events or findings will be the date of first drug intake as documented in the study drug administration page of the eCRF. Selected safety variables will be summarized using the visit window defined in Table 6 for the by-visit descriptive analysis. Note that only the laboratory measurements performed at the central laboratory will be used for the by-visit descriptive analysis. All available values obtained between two visits including unscheduled measurements will be assigned to the appropriate visit window. The following rule will be applied when mapping the measurements to the visit:

1. When a patient has more than one measurement on the same laboratory parameter (or vital sign or ECG) on the same date, then the one with the later/largest sample ID will be used.
2. For the same laboratory parameter (vital sign or ECG), 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. If there is no scheduled measurement within the visit window, the unscheduled measurement that is closest to the target date will be used.
3. When a patient has more than one measurement on the same laboratory parameter with the same distance from the target date, select the one with the latest date.

Table 6 - Study window (in days) for safety variables in LTS11210

Week (Visit)	Target day	Blood hematology Blood liver function	Blood chemistry[a] /Urinalysis	Vital signs	ECGs	PK/ADA/ Serum IL-6[b] sample	ANA/Anti-ds-DNA/ Rheumatoid factor[c]
Baseline[d] (-1)							
0[e] (1,2 or 1/2)	1						
2 (3)	15	2-21					
4 (4)	29	22-35		2-42		2-56	
6 (5)	43	36-49					
8 (6)	57	50-63	2-70	43-70			
10 (6.1)	71	64-77					
12 (7)	85	78-126	71-126	71-126		57-126	
24 (8)	169	127-210	127-210	127-210		127-210	
36 (9)	253	211-294	211-294	211-294		211-294	
48 (10)	337	295-378	295-378	295-378	2-504	295-378	2-504
60 (11)	421	379-462	379-462	379-462		379-462	
72 (12)	505	463-546	463-546	463-546		463-546	
84 (13)	589	547-630	547-630	547-630		547-630	
96 (14)	673	631-756	631-756	631-756	505-840	631-756	505-840
120 (16)	841	757-924	757-924	757-924		757-924	
144 (18)	1009	925-1092	925-1092	925-1092	841-1176	925-1092	841-1176
168 (20)	1177	1093-1260	1093-1260	1093-1260		1093-1260	
192 (22)	1345	1261-1428	1261-1428	1261-1428	1177-1512	1261-1428	1177-1512
216 (24)	1513	1429-1596	1429-1596	1429-1596		1429-1596	
240 (26)	1681	1597-1764	1597-1764	1597-1764	1513-1764	1597-1764	1513-1764
264 (28)	1849	1765-1932	1765-1932	1765-1932	≥1765	1765-1932	1765-1932
288 (30)	2017	1933-2100	1933-2100	1933-2100		1933-2100	1933-2184
312 (32)	2185	2101-2268	2101-2268	2101-2268		2101-2268	
336 (34)	2353	2269-2436	2269-2436	2269-2436		2269-2436	2185-2520
360 (36)	2521	2437-2604	2437-2604	2437-2604		2437-2604	
384 (38)	2689	2605-2772	2605-2772	2605-2772		2605-2772	2521-2856
408 (40)	2857	2773-2940	2773-2940	2773-2940		2773-2940	
432 (42)	3025	2941-3108	2941-3108	2941-3108		2941-3108	2857-3192
456 (44)	3193	3109-3276	3109-3276	3109-3276		3109-3276	
480 (46)	3361	3277-3444	3277-3444	3277-3444		3277-3444	3193-3486
504 (48)	3529	3445-3570	3445-3570	3445-3570		3445-3570	
516 (49)	3613	≥3571	≥3571	≥3571		≥3571	≥3487

a Lipids profile (metabolism) was also performed at Weeks 4 and 8 (Week 4: 2-42 days, Week 8: 43-70 days)

b Serum IL-6 sample is collected at Weeks 0, 4, 12, 24, 36, and 48 only; ADA at Week 12 (2-126 days) due to no ADA at Week 4.

c Anti-ds-DNA only if ANA titer is >1:160; Rheumatoid factor only at Week 0, 48, 96, 144, 192, 240, and 264 (or EOT).

d Baseline from previous studies

e The value at Week 0 is defined as the last non-missing value in the measurements collected at Visits 1, 2 and/or 1/2, and/or EOT visit in previous studies.

For PK, ADA and Vital signs, the follow-up (ie, 6-week post-treatment) measurement will be defined based on the sample collected at the follow-up Visit 50.

2.5.5 Unscheduled visits

In any efficacy analyses or summaries, unscheduled visit measurements of CRP will be included for computation of the baseline, but not other visits.

Laboratory, vital sign and ECG data from scheduled and unscheduled visits will be used in PCSA analysis no matter if the data are from the central laboratory or local laboratories. For the by-visit summary on the laboratory parameters, only the measurements performed at the central laboratory will be used.

Unscheduled visit measurements of vital signs and ECG will be included in the by-visit summaries using the visit window defined in Section 2.5.4.

2.5.6 Pooling of centers for statistical analyses

Countries are pooled by region: Western countries, South American and Rest of the world. These lists may be updated as the study is ongoing. Countries may be added and non-active countries will be removed. Countries will not be switched from region to region.

- Region 1 (Western countries): Austria , Australia , Belgium , Canada , Czech Republic, Finland, Germany, Greece, Hungary, Italy, Israel, The Netherlands, New Zealand , Norway, Portugal, Spain, Sweden, United Kingdom, USA
- Region 2 (South American): Argentina , Brazil, Chile, Colombia, Ecuador, Guatemala, Mexico, Peru
- Region 3 (Rest of the world): Belarus , China, Egypt, Estonia, India , Lithuania, Malaysia , Philippines , Poland, Romania, Russia, South Africa, South Korea, Ukraine, Taiwan, Thailand, Turkey

2.5.7 Statistical technical issues

None.

3 INTERIM ANALYSIS

This is an open-label study. Interim reports may be prepared for regulatory submissions or other purposes. However, no alpha adjustment is needed for the final CSR.

Interim reports for BLA submissions (FDA, EU MAA, etc)

In general, the analyses for these reports will be performed in the overall sarilumab group, and not according to previous treatment prior to enrollment into LTS11210 due to ongoing imbalanced enrollment from previous studies, different previous treatment duration and rescue sarilumab in EFC11072, except the disposition tables and some special safety analyses such as neutropenia. Only selected efficacy and safety analyses will be determined and performed for the interim reports. PK and ADA analyses will be performed on certain cut-off data in each interim report since not all time points are reached by patients or by study design. The analysis of patient reported outcomes may not be performed. However, the full analyses described in this SAP will be performed in the final CSR.

4 DATABASE LOCK

At the time of the initial sarilumab marketing authorization application for the treatment of RA, an interim database (soft) lock will be performed. Additional database snapshots may be performed prior to subsequent marketing authorization applications in major markets, eg. the EU. The final database lock is planned to occur 4 weeks after last patient last visit.

5 SOFTWARE DOCUMENTATION

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

6 REFERENCES

1. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA: The Health Institute, 1994.

7 LIST OF APPENDICES

Appendix A: Potentially clinically significant abnormalities (PCSA) criteria

Appendix B: Summary of statistical analyses

Appendix C: Summary of missing data imputation rules for efficacy variables

Appendix D: X-Ray missing data imputation process

Appendix E: SF-36 V2 Scoring

Appendix F: FACIT-Fatigue scoring guidelines (Version 4)

Appendix G: WPAI scoring

Appendix H: WPS-RA Questionnaire

Appendix I: SAS code for SF-36 scoring

Appendix J: Classification of events reviewed by the CV adjudication committee

Appendix K: Analysis Plan of the PFS-S Study

Appendix A Potentially clinically significant abnormalities (PCSA) criteria

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

(From QSD-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.
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.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

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

Parameter	PCSA	Comments
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.
CPK**	>3 ULN >10 ULN	FDA Feb 2005. 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) ≥30% change from baseline ≥100% change from baseline	Benichou C., 1994.
CLcr (mL/min) (Estimated creatinine clearance based on the Cokcroft-Gault equation)	<15 (end stage renal disease) ≥15 - <30 (severe decrease in GFR) ≥30 - < 60 (moderate decrease in GFR) ≥60 - <90 (mild decrease in GFR) ≥ 90 (normal GFR)	FDA draft Guidance 2010 Pharmacokinetics in patients with impaired renal function-study design, data analysis, and impact on dosing and labeling

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

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

Parameter	PCSA	Comments
eGFR** (mL/min/1.73m2) (Estimate of GFR based on an MDRD equation)	<15 (end stage renal disease) ≥15 - <30 (severe decrease in GFR) ≥30 - < 60 (moderate decrease in GFR) ≥60 - <90 (mild decrease in GFR) ≥ 90 (normal GFR)	FDA draft Guidance 2010 Pharmacokinetics in patients with impaired renal function-study design, data analysis, and impact on dosing and labeling
Uric Acid Hyperuricemia Hypouricemia	 >408 µmol/L <120 µmol/L	Harrison- Principles of internal Medicine 17th Ed., 2008.
Blood Urea Nitrogen	≥17 mmol/L	
Chloride	<80 mmol/L >115 mmol/L	
Sodium	≤129 mmol/L ≥160 mmol/L	
Potassium	<3 mmol/L ≥5.5 mmol/L	FDA Feb 2005.
Total Cholesterol	≥7.74 mmol/L **≥6.2 mmol/L***	Threshold for therapeutic intervention.
LDL	**≥4.1 mmol/L*** ≥4.9 mmol/L	
Triglycerides	≥4.6 mmol/L **≥5.6 mmol/L***	Threshold for therapeutic intervention.
Lipasemia**	≥3 ULN	
Amylasemia**	≥3 ULN	
Glucose Hypoglycaemia Hyperglycaemia	 ≤3.9 mmol/L and <LLN ≥11.1 mmol/L (unfasted); ≥7 mmol/L (fasted)	 ADA May 2005. ADA Jan 2008.
HbA1c	>8%	
Albumin	≤25 g/L	
CRP	>2 ULN or >10 mg/L (if ULN not provided)	FDA Sept 2005.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

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

Parameter	PCSA	Comments
Hematology		
WBC	<3.0 Giga/L (Non-Black); <2.0 Giga/L (Black) ≥16.0 Giga/L	Increase in WBC: not relevant. 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) **<1.0 Giga/L***	International Consensus meeting on drug-induced blood cytopenias, 1991. FDA criteria.
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.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

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

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

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

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

Parameter	PCSA	Comments
ECG		Ref.: ICH E14 guidance (2005) and E14 Q&A (2012), and Cardiac Safety Research Consortium White Paper on PR and QRS (Nada et al. Am Heart J. 2013; 165(4) : 489-500)
HR	<50 bpm <50 bpm and decrease from baseline ≥20 bpm <40 bpm <40 bpm and decrease from baseline ≥20 bpm <30 bpm <30 bpm and decrease from baseline ≥20 bpm	Categories are cumulative
	>90 bpm >90 bpm and increase from baseline ≥20bpm >100 bpm >100 bpm and increase from baseline ≥20bpm >120 bpm >120 bpm and increase from baseline ≥20 bpm	Categories are cumulative
PR	>200 ms >200 ms and increase from baseline ≥25% > 220 ms >220 ms and increase from baseline ≥25% > 240 ms > 240 ms and increase from baseline ≥25%	Categories are cumulative
QRS	>110 ms >110 msec and increase from baseline ≥25% >120 ms >120 ms and increase from baseline ≥25%	Categories are cumulative
QT	>500 ms	
QTc	Absolute values (ms) >450 ms >480 ms >500 ms Increase from baseline Increase from baseline]30-60] ms Increase from baseline >60 ms	To be applied to any kind of QT correction formula. Absolute values categories are cumulative QTc >480 ms and ΔQTc>60 ms are the 2 PCSA categories to be identified in individual subjects/patients listings.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for phase 2/3 studies (oncology excepted)

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

Parameter	PCSA	Comments

*: Project specific request

**: Not applicable for sarilumab project.

Appendix B Summary of statistical analyses

Endpoint	Analysis Population	Primary Analysis	Statistical Method	Supportive Analysis	Subgroup Analysis	Other Analyses
Efficacy						
ACR20, ACR50, ACR70	Safety	Response rate at all visits	Descriptive statistics	No	No	No
DAS28-CRP	Safety	Change from baseline at all visits	Descriptive statistics	No	No	No
DAS28 remission	Safety	Remission rate at all visits	Descriptive statistics	No	No	No
EULAR response	Safety	Response rate at all visits	Descriptive statistics	No	No	No
Each of the seven individual ACR components	Safety	Change from baseline at all visits	Descriptive statistics	No	No	No
Van der Heijde modified total Sharp score (for patients from EFC11072 Part B)	Safety	Change from baseline at all visits	Descriptive statistics	No	No	No
Incidence of radiographic progression (for patients from EFC11072 Part B)	Safety	Rate of radiographic progression at all visits	Descriptive statistics	No	No	No
HAQ-DI	Safety	Change from baseline at all visits	Descriptive statistics	No	No	No
Safety						
Adverse Events	Safety	Follow safety guidelines	Descriptive statistics	Exposure adjusted	No	No
Lab, vital signs, ECGs	Safety	Follow safety guidelines	Descriptive statistics	No	No	No

Appendix C Summary of missing data imputation rules for efficacy variables

Endpoints	Approach	Missing data imputation rules
ACR20, ACR50, ACR70	Primary	**Missing tender/swollen joints:** When calculating the TJC and SJC, with individual missing joint scores (the 'replaced or fused' joints are not taken into consideration for the swelling or tenderness) imputed as the mean of the scored joints, the TJC/SJC after imputation are as follows: TJC/SJC = sum (scored tender/swollen joints)*(number of joints in the full joint set / number of scored tender/swollen joints) The number of joints in the full joint set is defined as (68 – number of replaced or fused joints) for tenderness; (66 – number of replaced of fused joints) for swollenness; and the scored joints refer to those with an answer (0 – no pain, 1- pain). **Missing questions/categories for HAQ-DI:** The highest score for any component question in a category determines the category score. If the patient is using assistance for any of these activities, scoring may be adjusted. The total score is the sum of all categories' scores divided by the number of answered categories. At least six categories should be answered in order to have a total score. **Missing items within a measurement:** No imputation of missing post-baseline values will be performed. Responder status will be determined if possible. **Visit level:** Patients who discontinue treatment due to lack of efficacy will be considered as non-responders for all time points beyond the time they discontinue. Patients with insufficient data (ie, TJC, SJC, and at least three of the remaining five ACR components) at baseline and a time point will be considered as non-responders at that time point.
HAQ-DI	Primary	**Missing questions/categories for HAQ-DI:** The highest score for any component question in a category determines the category score. If the patient is using assistance for any of these activities, scoring may be adjusted. The total score is the sum of all categories' scores divided by the number of answered categories. At least six categories should be answered in order to have a total score. **Visit level:** No imputation will be performed.

Endpoints	Approach	Missing data imputation rules
Seven individual ACR components	Primary	**Missing tender/swollen joints:** When calculating the TJC and SJC, with individual missing joint scores (the 'replaced or fused' joints are not taken into consideration for the swelling or tenderness) imputed as the mean of the scored joints, the TJC/SJC after imputation are as follows: TJC/SJC = sum (scored tender/swollen joints)*(number of joints in the full joint set / number of scored tender/swollen joints) The number of joints in the full joint set is defined as (68 – number of replaced or fused joints for tenderness); (66 – number of replaced of fused joints) for swollenness; and the scored joints refer to those with an answer (0 – no pain, 1- pain). **Missing questions/categories for HAQ-DI:** The highest score for any component question in a category determines the category score. If the patient is using assistance for any of these activities, scoring may be adjusted. The total score is the sum of all categories' scores divided by the number of answered categories. At least six categories should be answered in order to have a total score. **Visit level:** No imputation of missing post-baseline values will be performed.
Modified total sharp score, erosion, joint space narrowing	Primary	**Missing joint scores within a segment:** Same approach as described in Appendix D. **Visit level:** The linear extrapolation method will be used to impute missing the modified total Sharp score, erosion score, or joint space narrowing score at Weeks 48, 96, 144 and 192.
Radiographic progression of modified total sharp score, erosion, joint space narrowing	Primary	**Missing joint scores within a segment:** Same approach as described in Appendix D. **Visit level:** The linear extrapolation method will be used to impute missing modified total Sharp score, erosion score, or joint space narrowing score at Weeks 48, 96, 144 and 192. The event of missing a change from baseline in the modified total Sharp score at a time pint will be considered as progression. The event of missing a change from baseline in the erosion score at a time pint will be considered as progression. The event of missing a change from baseline in the joint space narrowing score at a time pint will be considered as progression.
DAS28-CRP	Primary	**Missing tender/swollen joints:** When calculating the 28TJC and 28SJC, with individual missing joint scores (the 'replaced or fused' joints are not taken into consideration for the swelling or tenderness) imputed as the mean of the scored joints, the tender/swollen joint counts after imputation are as follows: 28TJC/28SJC = sum (scored tender/swollen joints)*(number of joints in the full joint set / number of scored tender/swollen joints) The number of joints in the full joint set is defined as (28 – number of replaced or fused joints) and the scored joints refer to those with an answer (0 – no pain, 1- pain). **Missing items within a measurement:**. No imputation will be performed. **Visit level:** • Patients with any of the four terms missing (28TJC, 28SJC, GH and CRP) will have a missing DAS28-CRP score at that time point.
DAS28 remission	Primary	DAS28 remission will be defined using the DAS28 –CRP scores as described above.
EULAR response	Primary	EULAR response will be defined using the DAS28-CRP scores as described above.

Endpoints	Approach	Missing data imputation rules
SF-36 components and domains, FACIT-Fatigue, Sleep VAS	Primary	No imputation will be performed.
WPS-RA , Four WPAI scores	Primary	No imputation will be performed.

Appendix D X-Ray missing data imputation process

Details of the radiographic score derivation process conducted by BioClinica, Inc. are described in the supplementary document entitled "X-Ray Score Derivations Process for Determination of Adjudication in Rheumatoid Arthritis Clinical Trials".

Handling of missing joint data

The handling of missing joint data and the logistics for being able to calculate mTSS change score are given below and outlined in the attached document:

- The first step is to determine the adequacy of the joints in each segment. In case of missing or inadequate (eg. confirmed surgery, technical reasons) scores for a joint at one particular time point, it will first be determined whether these missing joint scores qualify for imputation. On a per joint basis, eligibility for imputation depends on the number of missing or inadequate joints in a segment in relation to the number of adequate joints in that segment. If the number of adequate joints is below the adequacy threshold, imputation of change scores will not be done (see below).

Table 7 - Adequacy threshold for the 8 segments

Measure	Segment	Number of Joints n (L + R)	Minimum number of joints scored
Erosion	PIP	8	5
	MCP	12	7
	Wrist Images	12	7
	Foot Images	12	7
Joint Space Narrowing	PIP	8	5
	MCP	10	6
	Wrist Images	12	7
	Foot Images	12	7

PIP: proximal interphalangeal, MCP: metatarsophalangeal.

If, for example, 4 joints of the PIP segment (erosion) are missing or inadequate and 4 joints are adequate, that segment will be set to missing for that time point, implying that this segment cannot be imputed.

Change from baseline (change score) will be calculated for each joint, in each of the 8 segments, at each time point. For adequate segments with individual missing joint scores (missing due to definite surgery/joint replacement or radiographically inadequate), the mean change from baseline for the adequately scored joints within the corresponding segment (ie., left and right sides

together) will be calculated, and that mean change score will be used to impute the missing change score for the joints with missing joint score.

- The next step in the algorithm is to determine the adequacy of the time points per reader per patient.
  - Baseline time point: All 8 segments must be adequate at the baseline time point for the baseline time point to be considered adequate.
  - Follow-up time point: At least 5 segments must be adequate at a following-up time point for the time point to be considered adequate. Note that these 5 adequate segments can be composed of only hand segments if the foot segment is inadequate.

For adequate follow-up time points that have less than 4 inadequate segments, the missing joint change score within inadequate segment will be calculated by using the mean of the total mean change from baseline of adequate segments to impute the change score of every joint with a missing value within the inadequate segments.

- The third step is to determine the adequacy of the reads per patient.
  - If both readers consistently score X-rays of all time points as adequate or inadequate, then both reads are considered adequate.
  - If both readers score baseline as inadequate, then the patient will not proceed to adjudication even if a follow-up time point is scored by one reader as adequate and the other reader as inadequate.
  - If one reader scores a time point as inadequate while the second reader scores the same time point as adequate, then the reads are considered inadequate and the patient will proceed to adjudication (all time points).

Adjudication

A patient will be selected for adjudication based on the following criteria:

- Technical adequacy: One reader scores inadequate segments within a time point while the second reader scores adequate segments within the same time point, then the patient will proceed to adjudication (all time points).
- Total change score: The total change score for a time point is the sum of all individual joint change scores (erosion and JSN) from the 8 segments within the time point. If the difference between the two readers' change scores is $> \pm 10$ at a follow-up time point, this patient will be adjudicated (all time points).

At each time point, the mean (rounded to 1 decimal place) of the two primary readers' total change score will be calculated, and that mean will be the total change score in mTSS for each corresponding time point. The mTSS will be back calculated as the total change score + baseline. In cases of adjudication, the following rules will be applied:

- In adjudication cases where the differences between the two primary readers' change score is > ±10, the mean of the adjudicator and the primary reader with a closer total change score will be calculated, and that will be the total change score of the mTSS.
- In adjudication cases of technical inadequacy in one primary reader, the mean of the other primary reader and the adjudicator will be calculated, and that will be the total change score of the mTSS.

Handling of missing mTSS data:

The imputation methods for missing radiographic data in randomized clinical trials in rheumatoid arthritis are compared in Huang et al.'s article "Missing radiographic data handling in randomized clinical trials in rheumatoid arthritis" as in the supplementary document which will appear in the *Journal of Biopharmaceutical Statistics*.

Appendix E SF-36 V2 Scoring

General Scoring information

Items and scales are scored in 3 steps

- Step 1. Item recoding, for then 10 items that require recoding
- Step 2. Computing scale scores by summing across items in the same scale (raw scale scores)
- 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:

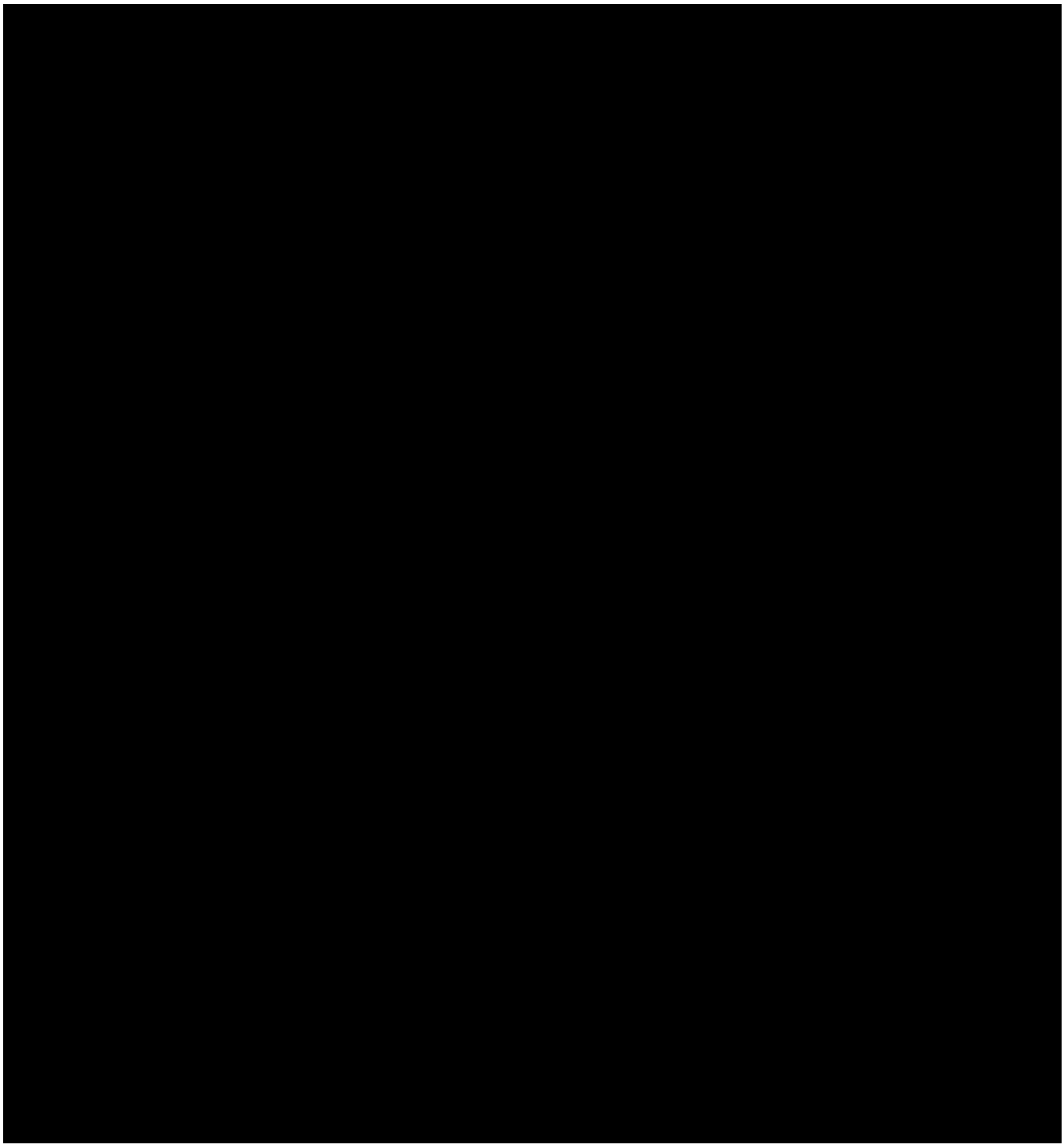
Transformed scale = [(actual raw score – lowest possible raw score) / possible raw score range] x 100

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

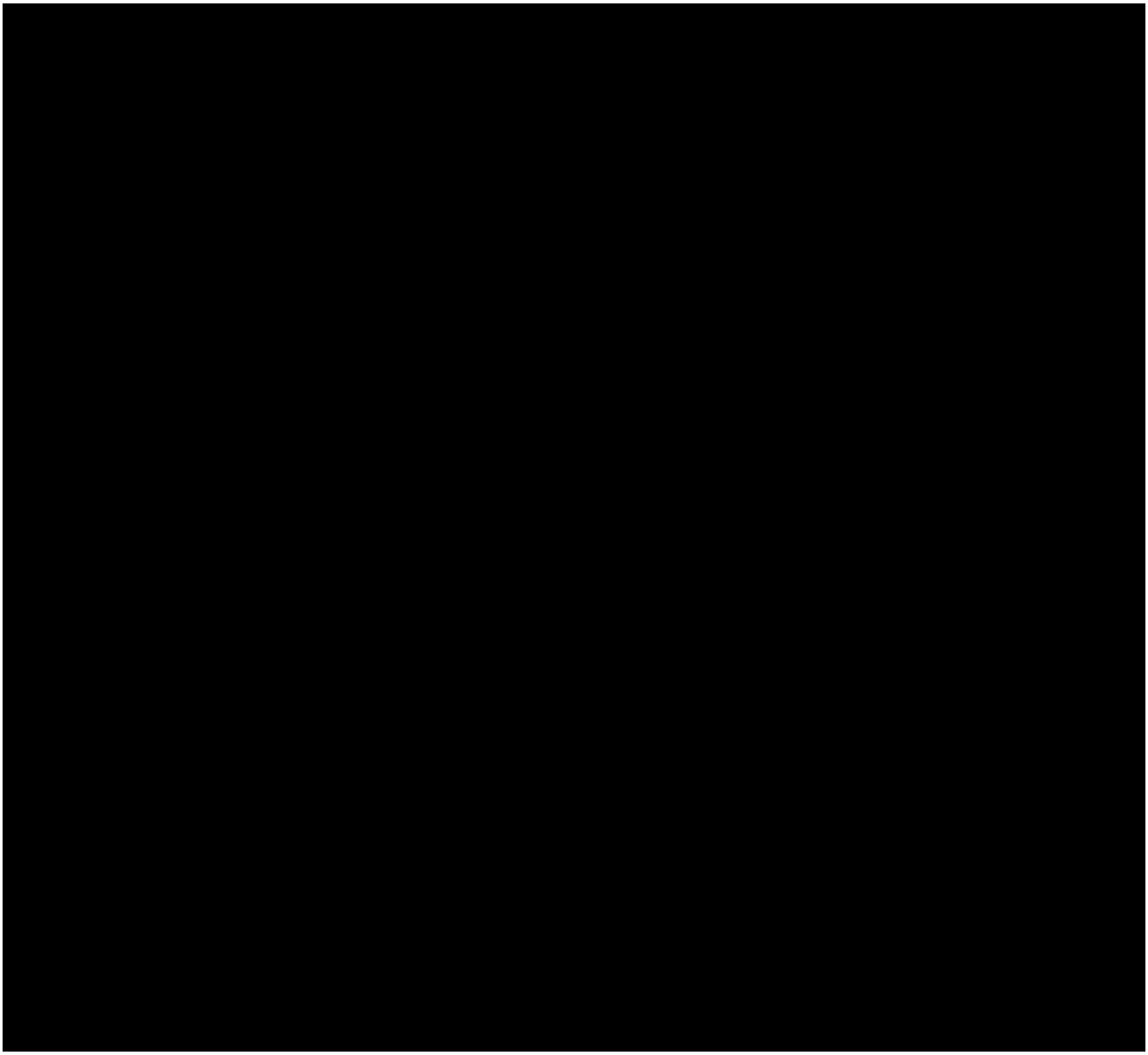
Table 8 - SF-36 V2 raw scores of eight domains

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

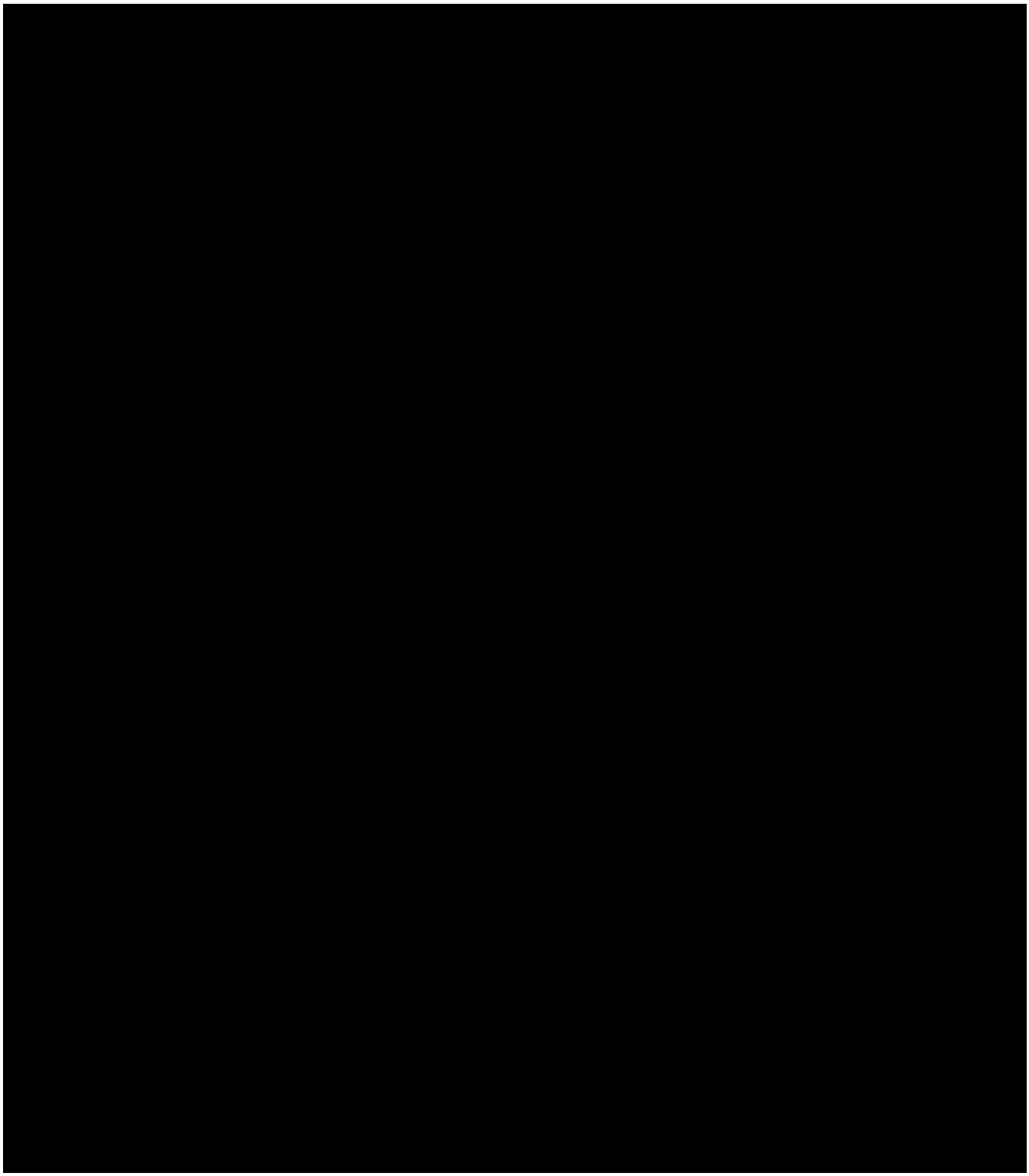
Appendix F FACIT-Fatigue scoring guidelines (Version 4)

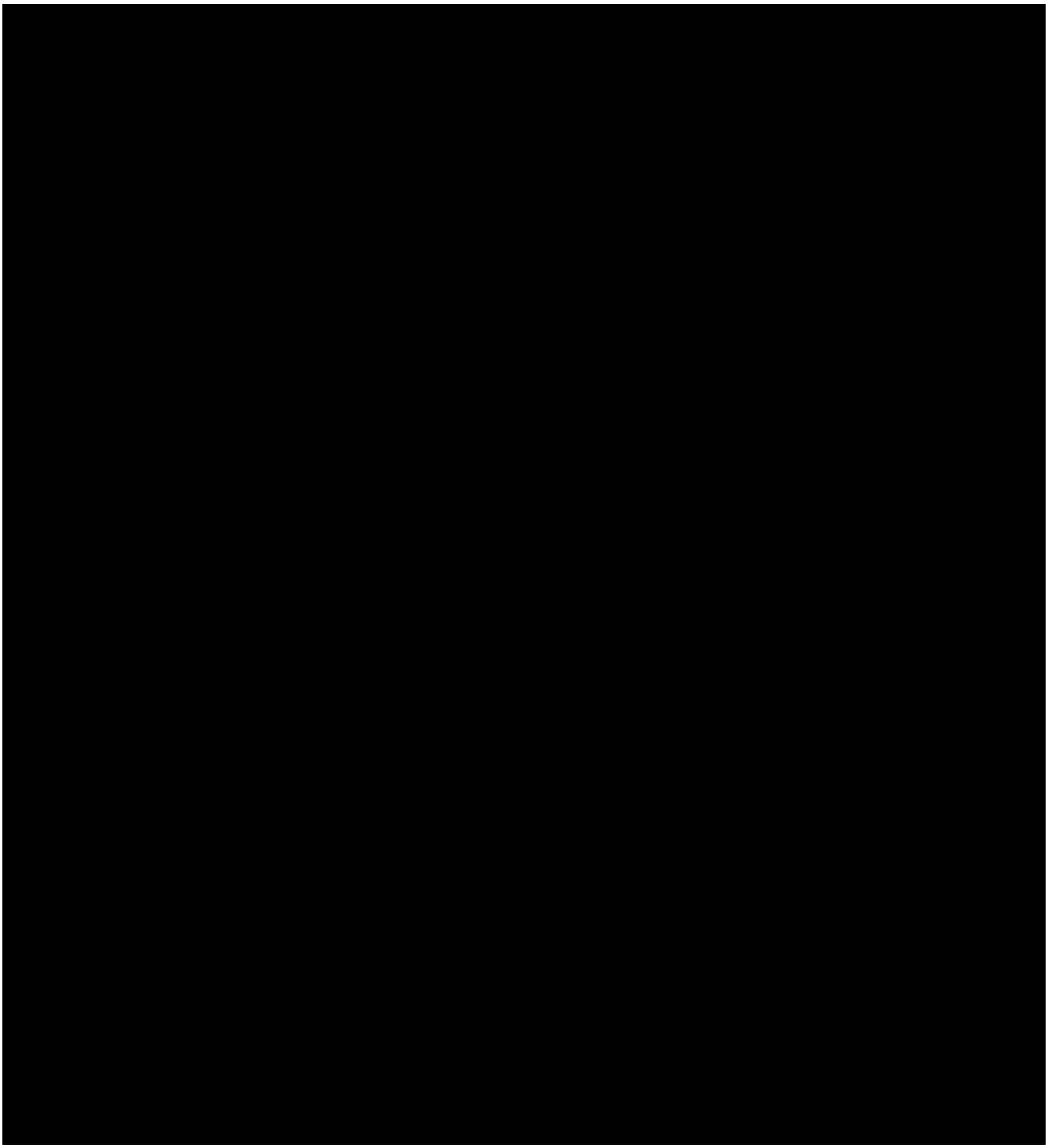


Appendix G WPAI scoring



Appendix H WPS-RA Questionnaire





Appendix I 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;
```

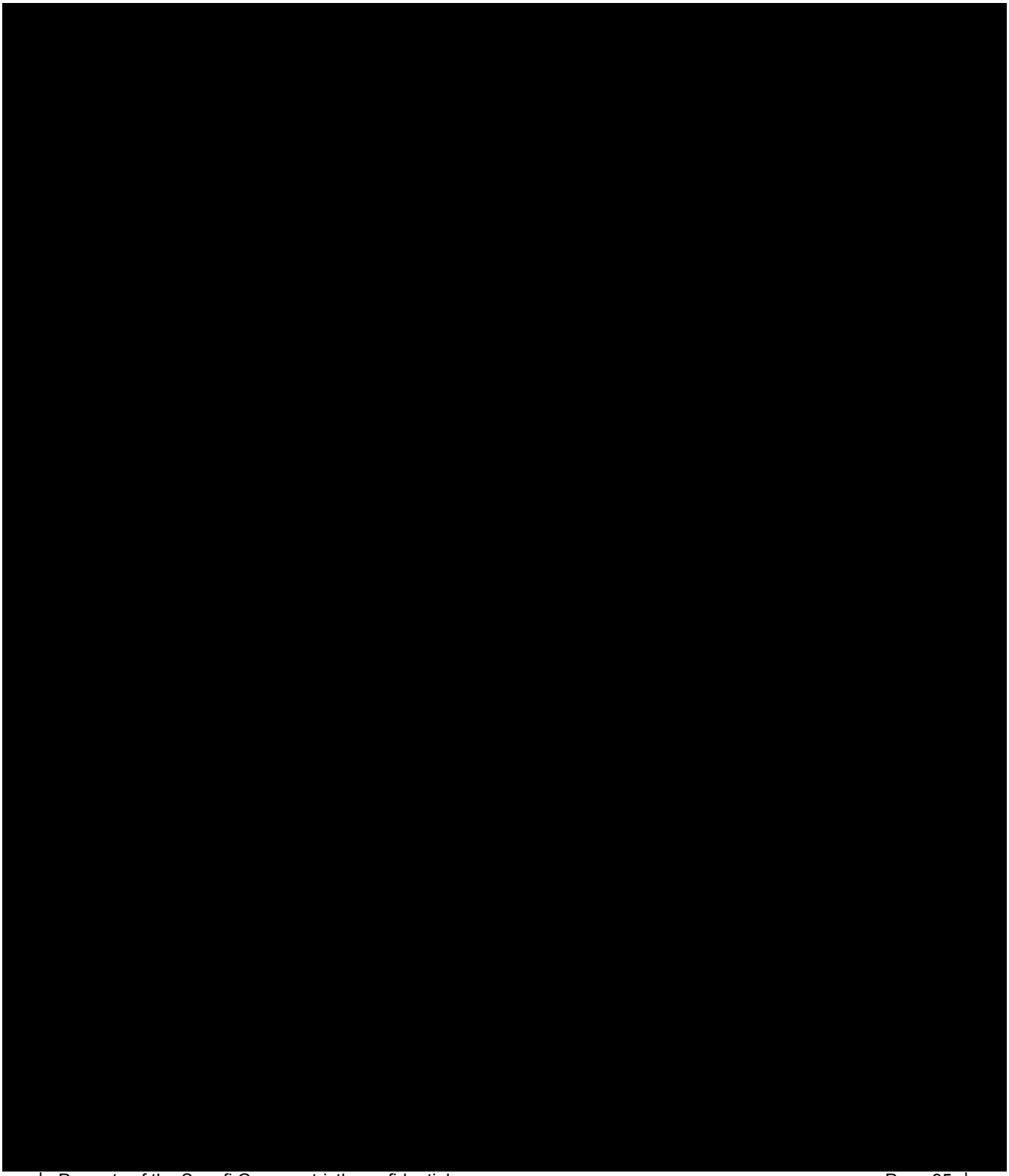
```
IF GH1 = 1 THEN RGH1 =   5;
IF GH1 = 2 THEN RGH1 = 4.4;
IF GH1 = 3 THEN RGH1 = 3.4;
IF GH1 = 4 THEN RGH1 =   2;
IF GH1 = 5 THEN RGH1 =   1;

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 (0-1
00 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 J Classification of Events Reviewed by the Committee



Appendix K Analysis Plan of the PFS-S Study

This appendix is the statistical analysis plan for the PFS-S sub-study.

Pre-filled syringe with safety system (PFS-S)

Sarilumab has been delivered via prefilled syringes (PFS) in most of the Phase 3 studies.

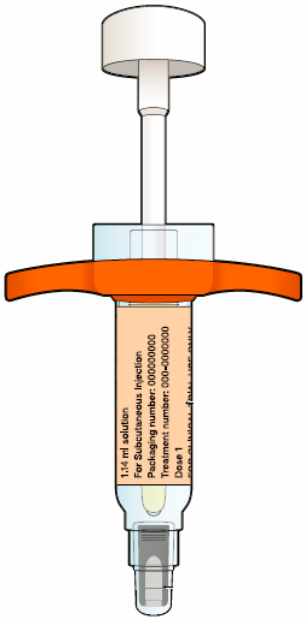
The pre-filled syringe with safety system (PFS-S) is a modified version of the prefilled syringe aimed at preventing needle stick injury by either the patient or by a caregiver (professional or non-professional) performing the administration. The "safety system" consists of a sharps injury prevention feature, which can be described as a sheath that automatically covers the needle once the injection is completed and thus prevents needle stick injury and re-use.

Compared to PFS, PFS-S also includes:

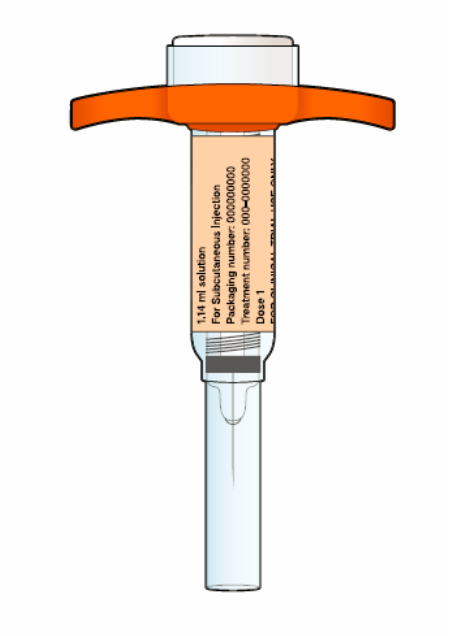
- a rigid needle shield that encloses the soft rubber needle shield covering the needle
- a finger flange with an ergonomic design and a large piston rod

The following figures show the unused and used PFS-S. At the end of the injection when the plunger rod has been pushed into the syringe as far as it will go, the needle safety shield is activated which can be seen as the syringe is pulled away from the skin.

Unused syringe



Used syringe with activated needle safety shield



Study design

This is a Phase 3, multicenter, worldwide, non-randomized, open-label, 1-arm, 12-week sub-study of the LTS11210/SARIL-RA-EXTEND study. The sub-study will only be performed in limited sites in selected countries.

Patients at participating sites who have been enrolled in the main study (LTS11210) for at least 24 weeks and are willing to participate in the sub-study will be enrolled in the sub-study. The entry into the sub-study (V101) will be at any scheduled visit in the main study from Visit 8 (Week 24) or after. Sarilumab 150 mg q2w PFS-S or sarilumab 200 mg q2w PFS-S will be used in the sub-study to replace the sarilumab 150 mg q2w PFS or 200 mg q2w PFS used in the main study. At the entry into the sub-study (V101), the patients/persons performing injections will be trained how to use the PFS-S, and the sub-study injection diary will be dispensed to replace the patient diary of the main study. All the procedures described in the sub-study are additional to the main study activities, unless specified to replace procedures in the main study.

The patient/person performing each sub-study injection will respond to 3 questions related to PFS-S use on the injection diary after each PFS-S injection. A “no” response to any of these 3 questions will be defined as a product technical complaint (PTC). For any PTC, the Investigator will complete a PTC form and will send this PTC form together with the PFS-S that triggered this PTC back to the Sponsor according to a predefined process. The patients should be instructed to keep any PFS-S associated with a PTC separately from other PFS-S, so it can be identified when returned to site at the end of sub-study. The sponsor will validate or invalidate the occurrence of a product technical failure (PTF), which is defined as a PTC that has a validated technical cause.

For patients receiving sarilumab 200 mg q2w PFS-S the dose can be reduced to 150 mg q2w PFS-S at any time during the sub-study duration if the defined criteria specified in the main protocol for dose reduction are fulfilled.

When the patients have completed the 12 weeks of the sub-study they will return to the clinic to complete Visit 102 (V102) in addition to the scheduled main study visit on that day. They will switch back to the main study and continue sarilumab treatment through PFS.

If the patient chooses to end the participation in the sub-study during the sub-study duration but to continue in the main study with IMP administration via PFS, the V102 should still be completed and the patient should switch back to take IMP administration through PFS on and after V102 per the patient’s IMP administration schedule.

If a patient discontinues Investigational Medicinal Product (IMP) permanently in LTS11210 during the sub-study, he/she should complete all procedures planned for V102 in the sub-study as well as the end of treatment (EOT) and post treatment follow-up visits in the main study (see flowchart in main study).

Study objective

Primary

To assess the usability of the sarilumab pre-filled syringe with safety system (PFS-S) when administered to patients with active Rheumatoid Arthritis (RA)

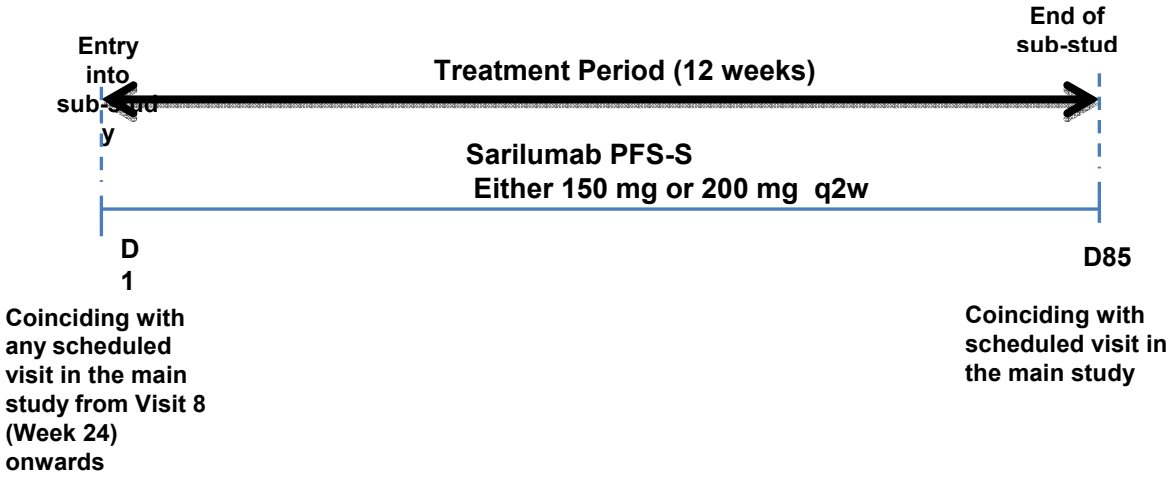
Secondary

To document the safety of sarilumab 200 mg q2w PFS-S and sarilumab 150 mg q2w PFS-S.

Determination of sample size

The sample size of this sub-study is based on empirical consideration. No formal sample size calculation is performed but an estimate of 120 patients included in this sub-study with up to 6 injections per patient will provide approximately 680 injections to evaluate PTF during the 12 weeks of the study assuming 5% drop-out rate.

Study plan



All assessments planned to be performed at the scheduled main study visit that coincides with the sub-study visit, are described in the main study flow charts. Below flowchart specifies only:

- those actions that are additional to the scheduled main study visit in LTS11210
- those actions that replace an action in the scheduled main study visit in LTS11210 (for example, study drug dispensing of the PFS-S IMP format, instead of the main study IMP in PFS)

	Sub-study assessments	
VISIT	V101	V102
DAY **Week**	D1 Wk0	D85 (+/- 3 days) Wk12
Eligibility	X	
Written Informed Consent	X	
IVRS call to start / end sub-study participation	X	X
Training on use of the PFS-S	X	
Dispense PFS-S[a]	X	
Dispense sub-study injection diary[b]	X	
Collect / review sub-study injection diary[b]		X
PTC reporting		X
AE related to a PTC reporting[c]		X
Serum Sarilumab (PK)[d]	X	X
ADA[d]	X	X

a PFS-S dispensing will replace the scheduled IMP dispensing at the main study visit

b Sub-Study Injection Diary dispensing and review will replace the dispensing and review of the main study Patient Diary in the scheduled main study visits

c Refer to Section 10.4.1 in Protocol Amendment 9 for instruction of reporting AE related to PTC

d If the sub-study visit coincides with a PK and ADA sampling time point in the main study, no additional samples are required for the sub-study.

Each patient will participate in the sub-study for a period of 12 weeks from the first sub-study visit (V101) to the final sub-study visit (V102). At the time of V101, eligible patients will start receiving sarilumab 200 mg q2w PFS-S or sarilumab 150 mg q2w PFS-S, if they reduced the dose in the main study per protocol, for 12 weeks as replacement of original sarilumab PFS in the main study in addition to background disease modifying anti-rheumatic drugs (DMARDs) therapy, if any. There is no post-treatment follow-up specific to the PFS-S sub-study, as all patients except patients who discontinue IMP permanently in LTS11210 during the sub-study will switch back to the main study at V102. For patients who discontinues IMP permanently in LTS11210 during the sub-study, they should complete all procedures planned for V102 in the sub-study as well as the end of treatment (EOT) and post treatment follow-up visits in the main study The last on-treatment sub-study visit for patients who complete the sub-study will occur at Day 85 (Visit 102). The sub-study will be considered completed when the last sub-study patient has completed V102.

Statistical analysis procedures and method

The phrase ‘treatment group’ used throughout this appendix refers to the following device/dose groups (PFS-S sarilumab 200 mg q2w, PFS-S sarilumab 150 mg q2w, PFS-S sarilumab 200 mg

q2w to 150 mg q2w). The baseline value is defined as the original baseline in initial/previous studies before LTS11210, as in the main study.

Disposition of patients

Below describes patient disposition for both patient study status and the patient analysis populations for the sub-study only, unless otherwise specified.

- All screened patients are defined as those patients who signed the informed consent for the sub-study.
- All enrolled patients are defined as all patients with signed informed consent who met the eligibility criteria to enter into the sub-study.
- The Safety population consists of all enrolled patients who received at least one dose of the IMP administered via PFS-S during this sub-study.
- The PK population consists of all enrolled patients on stable dose of sarilumab (receiving the same dose of sarilumab without treatment interruption) at least 24 weeks prior to V101 until V102 who receive at least 1 dose of the IMP and at least one non-missing serum concentration value during this sub-study.

For patient study status, the total number of patients for each one of the following categories will be presented using a summary table:

- All screened patients (informed consent signed)
- Screen failure patients and reasons for screen failure (not eligible per eligible CRF page)
- Non-enrolled but treated patients
- Enrolled but not treated patients
- Enrolled and treated patients
  - Patients who had stable dose at least 24 weeks prior V101
- Patients who had dose reductions during the duration of the sub-study
- Patients who discontinued the sub-study treatment by main reason for discontinuation
  - Patients who discontinued the sub-study treatment but continues in the main study
  - Patients who permanently discontinued the study treatment during participation in the sub-study
- Status for returning to the main study (Yes/No)

For all categories of patients (except for the screened and non-enrolled categories), percentages will be calculated using the total of all enrolled patients as the denominator. Reasons and time to treatment discontinuation will be supplied in tables giving numbers and percentages.

Demographics and baseline characteristics

Demographics and baseline characteristics will be presented for the enrolled patients in this sub-study.

Extent of investigational product exposure and compliance

The extent of study treatment exposure and compliance will be assessed and summarized by actual treatment received within the safety population.

The duration of IMP exposure is defined as: last dose date – first dose date at D1 + 14 days (regardless of unplanned intermittent discontinuations).

Treatment compliance to the IMP is defined as the actual number of doses received compared to the scheduled number of doses during the treatment period. Treatment compliance during the treatment period in the sub-study will be summarized descriptively.

Efficacy (Device) analysis

Data will be described for the 12 weeks of this sub-study using the safety population. The efficacy analyses will be performed using descriptive statistics and presented by treatment group.

Primary endpoint

The primary endpoint is the number of validated PFS-S-associated PTFs.

The injection diary will include specific questions about PTCs. The person performing each sub-study injection will record the answers to these questions in the diary each time he/she performs an injection with a PFS-S.

The questions include:

1. Were you able to remove the cap?
   - Yes
   - No

   If no, please describe / explain:
2. Was the needle safety system activated?
   - Yes
   - No

   If no, please describe / explain:
3. Did the safety system entirely cover the needle?
   - Yes

- No
  If no, please describe / explain:

At Week 12 (Visit 102), the Investigator will review the diary, confirm with the patient whether there are any problems associated with the use of PFS-S and record the responses in the e-CRF.

If the answer is “no” to any of the questions stated in the diary, the patients will be asked to provide further details of the complaint and a PTC form will be completed by the investigator. The PTC form and the PFS-S associated with PTC will be sent to the Sponsor following the pre-defined process. The sponsor will validate or invalidate the occurrence of a PTF. If the PTF is ruled out, the event is considered as a failed drug delivery (FDD) that is defined as patient’s failure to administer the full dose at a given attempt.

In addition, another question will be asked in the diary to confirm whether the person who performed the injection was trained:

4. Was the person who performed the injection the person who was trained by the site staff?
   - Yes
   - No

Secondary endpoints

- Number of PTCs
- Number and percentage of patients with PTFs
- Number and percentage of patients with PTCs
- Number and percentage of patients with FDDs

Safety analysis

Adverse events related to PTC per investigator’s judgment will be summarized by treatment group during the sub-study. If an adverse event reported during the sub-study is considered linked to a PTC by the investigator the potential link with the PTC should be specified in AE/SAE verbatim. If a laboratory or ECG abnormality or vital signs finding reported during the sub-study is considered clinically significant and linked to a PTC by the investigator an adverse event needs to be reported and the potential link with the PTC should be specified in AE/SAE verbatim.

Injection site reaction adverse events will be summarized during the sub-study period (that is, those events with start date from the date of first dose of the sub-study to the date of last sub-study dose+14 days, but no later than one day before the date of next non-sub study dose). The specific summaries to be performed are described for the main study.

Other adverse events, such as needle stick, skin irritation, etc, occurring during the sub-study period will be reviewed manually to evaluate if the events are related to the device. For the events related to the device but not PTC, a summary table will be provided or event description will be provided in details if only a few events reported.

Note: The safety analyses for the sub-study will be limited to the AEs noted above. Overall AEs, laboratory values, vital signs, and electrocardiograms (ECG) that occurs for patients participating in the sub-study will be described with the main LTS11210 study results.

PK analysis

Serum concentrations of functional sarilumab will be summarized using standard descriptive statistics by visit.

For PK parameters C_{trough}, estimates and 90% CIs for the ratio of geometric means (Day 85 versus Day 1) will be provided using a linear fixed effects model on the log-transformed data. Serum trough concentrations of functional sarilumab will be evaluated in patients on stable dose of sarilumab at V101 (D1) before using the PFS-S. Patients are considered to be on stable dose of sarilumab at V101 (D1) if they receive the same dose of sarilumab without treatment interruption for at least 24 weeks in LTS11210 prior to V101. Patients changing dose during the 12 week sub-study will be excluded from the C_{trough} analyses

Interim analysis

No interim analysis is planned for this sub-study.

LTS11210 16.1.9 Statistical Analysis Plan

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm)
[redacted]		