NCT03449758

STATISTICAL ANALYSIS PLAN

Effects of Sarilumab on patient-reported outcomes in patients with moderately to severely active rheumatoid arthritis and with inadequate response or intolerance to current conventional synthetic DMARDs or tumor necrosis factor inhibitors

STUDY NUMBER: SARILL08755

STUDY NAME: SariPRO

COMPOUND: SAR153191 (Sarilumab)

STATISTICIAN: [redacted] **(ITEC Services)**

DATE OF ISSUE: 19 Nov-2019

VERSION: Version 1.1

Total number of pages: 53

SIGNATURE PAGE

SariPRO

SARILL08755

EFFECT OF SARILUMAB ON PATIENT-REPORTED OUTCOMES IN PATIENTS WITH MODERATELY TO SEVERELY ACTIVE RHEUMATOID ARTHRITIS AND WITH INADEQUATE RESPONSE OR INTOLERANCE TO CURRENT CONVENTIONAL SYNTHETIC DMARDS OR TUMOR NECROSIS FACTOR INHIBITORS

SANOFI

Name	Position	Date	Signature
[redacted]	Epidemiologist-Methodologist	[redacted]	
	Clinical Project Leader		
	Medical Manager		

ITEC Services

Name	Position	Date	Signature
[redacted]	Statistician	[redacted]	[redacted]

TABLE OF CONTENTS

STATISTICAL ANALYSIS PLAN 1

TABLE OF CONTENTS 3

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS 6

1. OVERVIEW AND STUDY PLAN 8

1.1. STUDY DESIGN 8

1.2. OBJECTIVES 9

1.2.1. Primary objectives (6-months core period) 9

1.2.2. Secondary objectives (6-months core period) 9

1.2.3. Exploratory objectives 9

1.3. ENROLMENT OF CENTRES AND PATIENTS 10

1.4. DETERMINATION OF SAMPLE SIZE 12

1.5. STUDY PLAN 12

1.6. MODIFICATIONS FROM THE PROTOCOL 13

1.7. MODIFICATIONS FROM THE APPROVED STATISTICAL ANALYSIS PLAN 14

2. COLLECTED DATA 15

2.1. SITE QUESTIONNAIRE 15

2.2. SCREENING LOG 15

2.3. PATIENT DATA 15

3. GENERAL STATISTICAL APPROACH 16

3.1. DESCRIPTIVE STATISTICS 16

3.2. ANALYSIS SETS 16

4. ANALYSIS OF SITE QUESTIONNAIRE 17

5. ANALYSIS OF PATIENT DATA 18

5.1. ANALYSIS VARIABLES 18

5.1.1. Patients' characteristics 18

5.1.2. Primary endpoint (6-month core period) 19

5.1.3. Secondary endpoints (6-month core period) 19

5.1.4. Safety endpoints 21
5.1.5. Exploratory endpoints 23
5.1.5.1. Associations between changes in RAID score and composite disease activity scores (DAS28-ESR^{DD} and CDAI^{DD}). 23
5.1.5.2. Associations between changes in Hospital Anxiety and Depression Scale (HADS) scores and in composite disease activity scores (DAS28-ESR^{DD} and CDAI^{DD})....23
5.1.5.3. To explore patients' profiles 23
5.2. ANALYSIS POPULATIONS....24
5.3. STATISTICAL METHODS 24
5.3.1. Disposition of patients....24
5.3.2. Analyses of evaluation variables....25
5.3.2.1. Descriptive analysis 25
5.3.2.2. Primary efficacy analysis 25
5.3.2.3. Secondary efficacy analyses 25
5.3.2.4. Safety analyses....26
5.3.2.5. Exploratory analysis....26
6. DATA HANDLING CONVENTIONS....29
6.1. DATA HANDLING CONVENTIONS FOR SITE QUESTIONNAIRE....29
6.2. DATA HANDLING CONVENTIONS FOR PATIENT DATA....29
6.3. DATA HANDLING CONVENTIONS FOR PATIENT REPORTED OUTCOMES 29
6.4. MISSING DATA 29
6.5. DEFINITION OF REGIONS 29
6.6. WINDOWS FOR TIME POINTS 29
6.7. STATISTICAL TECHNICAL ISSUES....30
7. INTERIM ANALYSIS 31
8. SOFTWARE DOCUMENTATION 32
8.3. VALIDATION OF PROGRAMS....32
8.4. RESTITUTION OF THE PROGRAMS....32
9. LIST OF APPENDICES....33
APPENDIX A: DERIVED DATA 34
APPENDIX B: DECISION TABLES....40
APPENDIX C: RHEUMATOID ARTHRITIS IMPACT OF DISEASE - RAID QUESTIONNAIRE 43
APPENDIX D: HOSPITAL ANXIETY AND DEPRESSION SCALE (HADS)....44

APPENDIX E: FACIT FATIGUE SCALE (VERSION 4) 45

APPENDIX F: HEALTH ASSESSMENT QUESTIONNAIRE – DISABILITY INDEX (HAQ-DI) 46

APPENDIX G: INTERNATIONAL PHYSICAL ACTIVITY QUESTIONNAIRE (IPAQ) 49

APPENDIX H: MULTIDIMENSIONAL ASSESSMENT OF THYMIC STATES (MATHYS).....53

APPENDIX I: GLOBAL ASSESSMENT OF DISEASE ACTIVITY 55

APPENDIX J: POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES (PCSA) CRITERIA.....56

10. REFERENCES.....58

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

- ACR: American College of Rheumatology
- AE: Adverse Event
- AESI: Adverse Event of Special Interest
- ATC: Therapeutic Class
- BMI: Body Mass Index
- CDAI: Clinical Disease Activity Index
- CI: Confidence Interval
- COX-2: Cyclo-Oxygenase-2 Inhibitors
- csDMARD: Conventional synthetic Disease-Modifying Anti-Rheumatic Drug
- DAS28: Disease Activity Score in 28 joints
- DD: Derived Data
- DMARD: Disease-Modifying Anti-Rheumatic Drug
- D*x* Day *x*
- ECG: Electrocardiogram
- ESR: Erythrocyte Sedimentation Rate
- EULAR: European League Against Rheumatism
- FACIT-F: Functional Assessment of Chronic Illness Therapy – Fatigue
- HADS: Hospital Anxiety and Depression Scale
- HAQ-DI: Stanford Health Assessment Questionnaire - Disability Index
- HBsAg: Hepatitis B surface antigen
- HCV: Hepatitis C Virus
- ICF: Informed Consent Form
- IMP: Investigational Medicinal Product
- IPAQ : International Physical Activity Questionnaire
- ITT: Intent-To Treat

- LTE: Long-Term Extension
- MAThys: Multidimensional Assessment of Thymic States
- MedDRA: Medical Dictionary for Regulatory Activities
- NSAID: Non-Steroidal Anti-Inflammatory Drugs
- PPS: Per Protocole Set
- PRO: Patient-Reported Outcomes
- PT: Preferred Term
- RA: Rheumatoid Arthritis
- RAID: Rheumatoid Arthritis Impact of Disease
- SAE: Serious Adverse Event
- SAF: Safety Population
- SAP: Statistical Analysis Plan
- SD: Standard Deviation
- SJC: Swollen Joints
- SOC: System Organ Class
- TB: Tuberculosis
- TJC: Tender Joints
- TNF: Tumor Necrosis Factor
- ULN: Upper Limit of Normal
- VAS: Visual Analogue Scale
- V*x*: Visit *x*
- WHO - DD: World Health Organization - Drug Dictionary
- W*x*: Week *x*

1. OVERVIEW AND STUDY PLAN

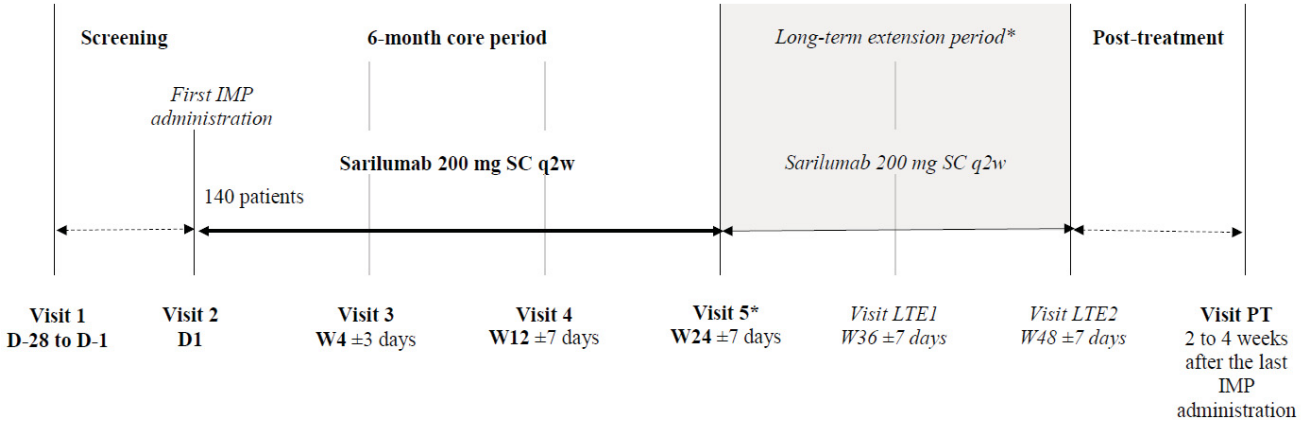
This Statistical Analysis Plan (SAP) provides a comprehensive and detailed description of strategy and statistical technique to be used to realize the analysis of data for study SariPRO SARILLL08755.

1.1. STUDY DESIGN

SariPRO is a national, multicenter, open-label, single-arm phase IV study. The total duration of the core study period per patient expected to be approximately 32 weeks:

Four study periods are defined for this study:

- Screening period: within 28 days prior to study treatment start (Baseline visit, Day 1 (D1)).
- 6-month core period: 3 visits are scheduled (W4, W12 and W24).
- Long-Term Extension (LTE) period: for all the patients who will complete the 6-month-core-period (allowed to continue the treatment with sarilumab). Patient visits are scheduled every 12 weeks over the LTE period (at W36 and W48).
- Post-treatment follow-up: each patient will have one follow-up visit, between 2 and 4 weeks after his/her last administration of sarilumab.



** Patients who will complete the 6-month-core-period will be allowed to continue the treatment with sarilumab, every 2 weeks for a maximum of another 24 weeks or until the commercial availability of sarilumab, depending whichever comes first.*

Figure 1: Graphical study design

1.2. OBJECTIVES

1.2.1. Primary objectives (6-months core period)

The primary objective of this study is to describe the evolution of the perception of patients regarding the effect of sarilumab in combination with conventional synthetic Disease-Modifying Anti-Rheumatic Drug (csDMARD) and/or monotherapy on patient-reported impact of disease, using the rheumatoid arthritis impact of disease (RAID) questionnaire, in patients with moderately to severely active rheumatoid arthritis (RA) and inadequate response or intolerance to current csDMARD or tumor necrosis factor (TNF) inhibitors.

1.2.2. Secondary objectives (6-months core period)

The secondary objectives of this study are:

- To assess the change of the RAID score from baseline (to Week 4 (W4) and W12) in patients with moderately to severely active RA and inadequate response or intolerance to current csDMARD or TNF inhibitors, treated with sarilumab in combination with csDMARD and/or monotherapy.
- To assess the effect of sarilumab in combination with csDMARD and/or monotherapy on other patient-reported outcomes (global assessment of disease activity, disability, morning stiffness, fatigue, anxiety/ depression, mood disorders, and physical activities), in patients with moderately to severely active RA and inadequate response or intolerance to current csDMARD or TNF inhibitors.
- To assess the efficacy of sarilumab in combination with csDMARD and/or monotherapy using Disease Activity Score in 28 joints-Erythrocyte Sedimentation Rate (DAS28-ESR) and Clinical Disease Activity Index (CDAI) in patients with moderately to severely active RA and inadequate response or intolerance to current csDMARD or TNF inhibitors.
- To assess the safety of sarilumab in combination with csDMARD and/or monotherapy in patients with moderately to severely active RA and inadequate response or intolerance to current csDMARD or TNF inhibitors.

1.2.3. Exploratory objectives

The exploratory objectives of this study are:

Over the 6-month core period

- To describe the associations between changes in RAID score and composite disease activity scores (DAS28-ESR and CDAI).
- To describe the associations between changes in Hospital Anxiety and Depression Scale (HADS) score and in composite disease activity scores (DAS28-ESR and CDAI)
- To explore patient profiles

Over the long-term extension (LTE) period (maximum 24 weeks)

- To assess the safety of sarilumab in combination with csDMARD and/or monotherapy in patients with moderately to severely active RA and inadequate response or intolerance to current csDMARD or TNF inhibitors over the LTE period. This objective will be considered as a **safety objective** for the rest of the document.

1.3. ENROLMENT OF CENTRES AND PATIENTS

Inclusion criteria:

I 01. Patient with moderately to severely active RA according to the 2010 American College of Rheumatology (ACR)/European League against Rheumatism (EULAR) criteria

I 02. Patient with moderate to severe disease activity defined as DAS28-ESR>3.2 at Screening

I 03. Patients with inadequate response within at least the last 3 months or intolerance to current csDMARD or to at least one anti-TNF therapy (as defined by the investigator)

I 04. Oral corticosteroids (≤15 mg/day prednisone or equivalent) and non-steroidal anti-inflammatory drugs (NSAID) or cyclo-oxygenase-2 inhibitors (COX-2) (up to the maximum recommended dose) are allowed if at a stable dose for at least 4 weeks prior to Baseline

I 05. Permitted csDMARDs are allowed if at a stable dose for at least 4 weeks prior to Baseline

I 06. Patients able and willing to give written informed consent and to comply with the requirements of the study protocol.

Exclusion criteria:

E 01. <18 years of age

E 02. Patient unable to understand and write adequately to complete the study Patient Reported Outcomes (PRO) assessments

E 03. Exposure to sarilumab at any time prior to Baseline

E 04. Use of intra-articular or parenteral corticosteroids within 4 weeks prior to Baseline

E 05. Treatment with any investigational agent within the 4 weeks of Screening

E 06. Last RA treatment prior to inclusion with any antiJAK or biologic DMARD other than anti-TNF

E 07. Patients treated with anti-TNF (e.i. adalimumab, infliximab, certolizumab, golimumab, etanercept) before the screening period, which are maintained within the 4 weeks before the inclusion (i.e.the first injection of sarilumab)

E 08. Rheumatic autoimmune disease other than RA or prior history or current inflammatory joint disease other than RA

E 09. Evidence of active malignant disease, malignancies diagnosed within the previous 10 years (except basal and squamous cell carcinoma of the skin or carcinoma in situ of the cervix uteri previously excised and cured)

E 10. Patient who is institutionalized due to regulatory or legal order or patient who is mentally disabled or educationally disadvantaged

E 11. Pregnant or breastfeeding woman

E 12. Women of childbearing potential not protected by highly-effective contraceptive method(s) of birth control (as defined in the informed consent form and/or in a local protocol addendum/amendment) over the study period and for at least 3 months following the last dose of sarilumab, and/or who are unwilling or unable to be tested for pregnancy

E 13. History of severe allergic or anaphylactic reactions to human, humanized, or murine monoclonal antibodies (or to any of the excipients associated to sarilumab)

E 14. Immunization with a live/attenuated vaccine within 4 weeks prior to Baseline

E 15. Stage III or IV cardiac failure according to the New York Heart Association classification

E 16. History of previous gastrointestinal perforation or diverticulitis

E 17. Known active current/ recurrent infections (including but not limited to active tuberculosis (TB) or history of incompletely treated TB and atypical mycobacterial disease, hepatitis B and C, and herpes zoster)

NOTE: in case of latent TB infection the patient may be included if a subsequent appropriate anti TB treatment is initiated since at least 3 weeks

E 18. Positive hepatitis B surface antigen (HBsAg), and/or positive total hepatitis B core antibody, and/or positive hepatitis C antibody (HCV) at the Screening visit

E 19. Evidence of serious uncontrolled concomitant disease, including severe uncontrolled hypercholesterolemia or hypertriglyceridemia

E 20. Patients with any of the following laboratory abnormalities at the Screening or Baseline visit:

- Hemoglobin <8.5 g/dL
- White blood cells <3000/mm³
- Absolute neutrophil count <2000/mm³
- Absolute lymphocyte count <500/mm³
- Platelet count <150 000 cells/mm³
- AST or ALT >1.5 x Upper Limit of Normal (ULN)
- Bilirubin (total) >ULN, unless Gilbert's disease has been determined by genetic testing and has been documented
- Total fasting cholesterol >3.50 g/L [9.1 mmol/L] or triglycerides >5.00 g/L [5.6 mmol/L]

Selection criteria (LTE period):

All the patients who will complete the 6-month-core period will be allowed to enter the LTE period, if sarilumab is not yet commercially available in France (as monotherapy or in combination with csDMARDs).

1.4. DETERMINATION OF SAMPLE SIZE

The primary study objective is to assess the change in RAID score from baseline to W24. A sufficient precision (half-length of the 95% confidence interval (CI)) is required. According to the literature [1,2,3] and findings from the MONARCH and TARGET studies [4,5], the standard deviation (SD) of the RAID score could be estimated at 2.5 points. On this basis, a sample size of 140 patients (122 evaluable, if 20% non-evaluable patients) produces a precision, two-sided 95% confidence interval (CI), with a distance from the mean to the limits that will be less than 0.5 when the estimated SD is 2.5. Table below gives the 95% CI of change from baseline to W24 of RAID score with this precision (0.44), assuming that the change from baseline will be from 30% to 50% (MCII defined by Dougados [6] in 2012) of baseline score.

1.5. STUDY PLAN

The data are recorded prospectively during one baseline visit, one follow-up visit at week 4 ± 3 days, a second follow-up visit at 12 ± 1 week, a third follow-up visit at 24 ± 1 week.

For the LTE period, data are recorded prospectively during one follow-up visit at 36 ± 1 week and a second follow-up visit at 48 ± 1 week.

	Screening	Treatment						Post-treatment observations	Early withdrawal
		6-month core period				*Long-term extension period*			
VISIT	V1	V2 (Baseline)	V3	V4	V5	*V LTE1*	*V LTE2*	V PT	
DAY / WEEK	D-28 to D-1	D1	W4 ±3D	W12 ±7D	W24 ±7D	*W36 ±7D*	*W48 ±7D*	2 to 4 weeks after the last IMP administration	
Eligibility									
Written Informed Consent	X								
Inclusion/Exclusion criteria	X	X							
Patient demography [a]	X								
Medical/surgical history	X								
Prior medication history	X								
Physical examination [b]	X				X				X
12-lead ECG and Chest X-ray [c]	X								
Treatment									
IMP administration [d]		◄	———	———	———	►			
Concomitant medication review		X	X	X	X	*X*	*X*	X	X
Compliance / review patient diary			X	X	X	*X*	*X*	X	X
Patient Reported Outcomes									
RAID, HADS, MAThyS, FACIT-Fatigue, HAQ-DI, duration of morning stiffness, IPAQ, and patient global assessment of disease activity (VAS)		X	X	X	X				X
Efficacy									
ESR, DAS28-ESR, CDAI, and joint counts (TJC and SJC on 28 joints) [e]	X	X	X	X	X				X
Size of the largest rheumatoid nodule (if applicable)		X			X				X
Safety									
AE/SAE recording [f]		◄	———	———	———	———	———	———	►
Vital signs [g]	X	X	X	X	X	*X*	*X*	X	X

Pregnancy test (for women of childbearing potential) [h]	X	X	X	X	X	X	X	X	X
Hematology [i]	D-7 to D-1		X	X	X	X	X		X
Serum chemistry [j]	D-7 to D-1		X	X	X	X	X		X
Serum iron, 25OHD [k]	X				X				X
Fasting lipids [l]	D-7 to D-1		X	X	X	X	X		X
Serology [m]	X								
Tuberculosis assessment [n]	X								

a) Gender and age.
b) A general physical examination should be performed during the Screening period (between D-28 and D-1). Body weight and waist circumference are to be recorded only at V2 and V5 (and at early withdrawal if applicable).
c) If at Baseline a chest X-ray has been taken within 90 days that shows no clinically significant abnormality, a further chest X-ray is not required. A 12-lead ECG should be performed during the Screening period.
d) Sarilumab 200 mg SC q2w, as monotherapy or in combination with a csDMARD (at Investigator's discretion). The first SC injection will be administered at the hospital site under supervision of the Investigator. For patients and/or non-professional caregivers able to perform SC injections at home, the investigator or delegate will provide training regarding the preparation and injections (at Baseline). For patients and/or non-professional caregivers completely unable to perform such injections at home, a home nursing will be needed.
e) Self-reported questionnaires should be completed by the patient before all other assessments performed during the study visits (and befoire the IMP administration if performed at this time).
f) All adverse events (date of onset, seriousness, maximum intensity, causal relationship with sarilumab, events leading to treatment modification or discontinuation, outcome of events).
g) Pulse rate, systolic and diastolic blood pressure (while seated).
h) Pregnancy tests will be locally conducted for female patients of child-bearing potential. Pregnancy test at Screening will be a serum test. All following pregnancy tests will be urine tests. If a urine test is positive it must be confirmed by a serum pregnancy test.
i) Hemoglobin, complete blood count (CBC), platelet count, and erythrocyte sedimentation rate (ESR) (blood should be drawn before drug administration). First tests will be performed during the 7-day period before inclusion. All tests will be performed locally.
j) Creatinine clearance, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin (blood should be drawn before drug administration). First tests have to be performed during the 7-day period before inclusion. All tests will be performed locally.
k) Blood should be drawn before drug administration. First tests will be performed during the Screening period. All tests will be performed locally.
l) Total fasting cholesterol, high density lipoprotein (HDL) cholesterol, low density lipoprotein (LDL) cholesterol, triglycerides (blood should be drawn before drug administration). First tests have to be performed during the 7-day period before inclusion. All tests will be performed locally.
m) HBsAg, hepatitis B virus antibody and hepatitis C virus antibody will be performed during the Screening period only, unless clinically indicated during the trial.
n) If at Baseline a tuberculosis assessment (QUANTIFERON) has been taken within 90 days that shows no abnormality, a further QUANTIFERON test is not required. Further tuberculosis assessments are not required unless clinically indicated during the trial.

Figure 2: Study flow-chart

1.6. MODIFICATIONS FROM THE PROTOCOL

Analysis sets definition have been reviewed:

Population	Protocol definition	Statistical analysis plan definition
Enrolled patients	No definition	All patients screened, who have signed the Informed Consent Form (ICF) and who never took the study treatment at any time before baseline.
Intent-to-treat (ITT) population	All patients who met all the eligibility criteria and who are actually treated.	All patients included (i.e. for whom the inclusion decision "Has the decision to include the patient been taken" is 'Yes') and who took the treatment at least once
Per-protocol population	Patients of ITT population without major protocol deviations. Major deviations will be defined in the statistical analyses plan before database frozen.	Patients of ITT population without major protocol deviations as defined during the Data Review meeting and who are evaluable for the main criterion

Data will be presented according to a subgroup not defined in the protocol. This subgroup has 3 categories[DD]:

- Patients treated with csDMARDs for whom Sarilumab is prescribed in addition to csDMARDs at baseline.

- Patients treated with biological therapy for whom Sarilumab is prescribed in addition to csDMARDs at baseline.

- Patients previously treated with csDMARDs or biological therapy for whom only Sarilumab in monotherapy is prescribed at baseline.

Also, an analysis has been performed with all screened patients who have met all the eligibility criteria and who have received the treatment at baseline visit (D1). Only data from the baseline visit have been analysed.

In the protocol section 11.4.1.3, compliance should be evaluated according to the planned dose required by the protocol. No planned dose is required by the protocol. Hence, compliance will be evaluated according to the number of injections during the study (see section 5.1.4).

Adverse events will not be described by SOC, HLGT, HLT and PT but only by SOC and PT.

1.7. MODIFICATIONS FROM THE APPROVED STATISTICAL ANALYSIS PLAN

Approved SAP (V1.0 of 23OCT2019)	**Modification**
Section 1.7: “an interim analysis has been performed with all screened patients who have met all the eligibility criteria and who have received the treatment at baseline visit (D1).”	**“**an ~~interim~~ analysis has been performed with all screened patients who have met all the eligibility criteria and who have received the treatment at baseline visit (D1)”
Section 5.1.5.3: “Patients’ profiles will be explored using the same analyses performed for the interim analysis”	“Patients’ profiles will be explored using the same analyses performed for the ~~interim~~ analysis of baseline data”
Section 7 : ” Interim analysis details have been presented in a separate document (SANOFI - SARIPRO- Statistical Analysis Plan - INTERIM ANALYSIS Version 2.0 20190611).”	~~Interim analysis details~~ An analysis of baseline data have been presented in a separate document (SANOFI - SARIPRO- Statistical Analysis Plan - INTERIM ANALYSIS Version 2.0 20190611).

[D] See Appendix A: Derived data

2. COLLECTED DATA

2.1. SITE QUESTIONNAIRE

Not applicable.

2.2. SCREENING LOG

A screening log form has been provided for each site. All screened patients have been registered on the screening log form with the following information:

- Date of visit (DD/MM/YYYY)
- Patient number (if enrolled)
- Reason for non-enrolment into the registry
    - Non-compliance with inclusion/ exclusion criteria
    - Patient's refusal of consent.

2.3. PATIENT DATA

Data collected on patients during the study are about:

- Eligibility criteria
- Demography
- Vital signs and physical exam
- Medical (Medical and surgical history in relation with RA, medical history and associated diseases)
- Characteristics of RA
- RA activity
- Laboratory tests
- Investigational Medicinal Product (IMP) administration
- Concomitant treatments
- PROs questionnaires
- Adverse Events (AEs)

Collected data will be presented more precisely below.

3. GENERAL STATISTICAL APPROACH

3.1. DESCRIPTIVE STATISTICS

Each modality of qualitative data will be presented with their number, percentage and 95% CI will also be presented.

For quantitative data, mean, median, minimum and maximum as well as standard deviation , interquartile range and 95% CI will be described.

For all the variables, number of available and missing observations will be specified.

Data will be presented using an appropriate number of decimal places (i.e. the number of decimal places used does not imply undue precision):

- Raw data : same number of decimal as collected,
- Derived data: The appropriate number of decimal places will be determined by general practice, mathematical rationale or scientific rationale (e.g. age should be presented in whole numbers),
- Mean, median and standard deviation : reported to one decimal place greater than the raw/derived data that they summarise,
- Minimum and maximum : same precision as the raw data,
- Percentage : one decimal place,
- P-values : four decimal places (e.g.: p=0.0037), after rounding. P-values which are less than 0.0001 will be presented as '<0.0001'.

3.2. ANALYSIS SETS

3.2.1. Enrolled patients

All patients screened, who have signed the Informed Consent Form (ICF) and who never took the study treatment at any time before baseline.

3.2.2. Intent-to-treat population (ITT)

All patients included (i.e. for whom the inclusion decision "Has the decision to include the patient been taken" is 'Yes') and who took the treatment at least once.

3.2.3. Per Protocol Set (PPS)

3.2.4. Patients of ITT population without major protocol deviations as defined during the Data Review meeting and who are evaluable for the main criterion.**Safety population (SAF)**

All patients who have signed the Informed Consent Form (ICF) and have been exposed to at least one dose or part of a dose of investigational medicine product (IMP).

4. ANALYSIS OF SITE QUESTIONNAIRE

Not applicable.

5. ANALYSIS OF PATIENT DATA

5.1. ANALYSIS VARIABLES

5.1.1. Patients' characteristics

- **Demography**

- Age at baseline (years)
- Sex (Female / Male)

- **Physical examination and vital signs at baseline**

- Height (cm)
- Weight (kg)
- Body Mass Index (BMI)[DD] (kg/m^2)
    - As quantitative
    - In classes:
        - Underweight (<18.5)
        - Normal (healthy) weight ([18.5; 25[)
        - Overweight ([25.0; 30[)
        - Obese (≥30)
- Abdominal girth (cm)
- Blood pressure
    - Systolic (mmHg)
    - Diastolic (mmHg)
- Heart rate (beats/min)

- **Medical and surgical history**

➢ <u>Medication history for rheumatoid arthritis</u>

- Time from RA diagnosis (years) [DD]
- Patient previously treated with conventional synthetic DMARDs (Yes/No)
- Patients previously treated with biological therapies and targeted synthetic DMARDs at baseline (Yes/No)
- Patients previously treated with corticosteroids (Yes/No)
- Patient currently treated with NSAIDs, analgesics (Yes/No)
- Patient currently receiving other treatments for their RA (Yes/No)
- Patient treated with other treatment (other than those for RA) (Yes/No)

Conventional synthetic DMARDs, biological therapies and targeted synthetic DMARDs, corticosteroids and NSAIDs, analgesics will be presented according to their status at baseline: ongoing or ended at baseline.

➢ <u>Medical history and associated diseases</u>

- Patient having experienced any past or concomitant diseases (clinically significant) (Yes/No)

If yes, past or concomitant diseases will be presented by System Organ Class (SOC) and Preferred Term (PT).

➢ Surgical history in relation with rheumatoid arthritis

- Patient having any surgical history in relation with rheumatoid arthritis (Yes/No)

If yes, surgical history will be presented by SOC and PT.

- **Characteristics of rheumatoid arthritis**

- Anti-CCP (Positive/Negative/Not performed)
- Rheumatoid factor (Positive/Negative//Not performed)
- Signs of structural joint involvement (Yes/No)
- At least one rheumatoid nodule (Yes/No)

  If yes:
  - Size of the largest nodule at screening and at the last measure (mm)
  - Changes from baseline[DD]
  - Time between measures of the rheumatoid nodule (months)[DD]
- Chest X-Ray (Normal or with a non-clinically-significant abnormality / Abnormal with a clinically-significant abnormality / not performed)

- **Evaluation of rheumatoid arthritis activity**

- Visual Analogue Scale (VAS) Global Assessment of disease activity (mm) by physician (0-no pain to 100-maxima pain imaginable)

- **Evaluation of the patient's feeling**

- Emotional items on a 5-Likert scale (from 0=never to 4=constantly):
  - Sadness
  - Joy
  - Irritability
  - Panic
  - Anxiety
  - Anger
  - Exaltation

- **Prior and concomitant treatments at baseline**

Only treatments with a start date prior to date of baseline will be presented.

- Subjects with at least one prior treatment (ended) at baseline
- Subjects with at least one prior treatment ongoing at baseline

Prior and concomitant treatments will be coded using World Health Organization - Drug Dictionary (WHO-DD) (current version at the time of encoding).

5.1.2. Primary endpoint (6-month core period)

Primary endpoint is the change in RAID score from baseline to W24. The RAID score ranging from 0 (best) to 10 (worst).

5.1.3. Secondary endpoints (6-month core period)

- **PROs questionnaires**

- RAID score per item and total score[DD]

The RAID score gives a number between 0 and 10. A higher score on the RAID indicates more impact of the disease.
A score $\leq$ 2 out of 10 is considered a patient-acceptable status.

- Hospital Anxiety and Depression Scale (HADS):
    - Per item
    - Anxiety calculated score as quantitative[DD] and in classes
        - Normal (0-7)
        - Borderline abnormal (8-10)
        - Abnormal (11-21)
    - Depression calculated score as quantitative[DD] and in classes:
        - Normal (0-7)
        - Borderline abnormal (8-10)
        - Abnormal (11-21)
- Multidimensional Assessment of Thymic States (MAThys) [DD]
    - Per item
    - 5-dimension scores:
        - Emotion
        - Cognition
        - Psychomotor function
        - Motivation
        - Sensory perception
    - MAThys total score

- Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-F) per item and total score

- Health Assessment Questionnaire - Disability Index (HAQ-DI) per domain (dressing & grooming, arising, eating, walking, hygiene, reach, grip and activities) and total score.
- International Physical Activity Questionnaire (IPAQ) classification per domain and total score as quantitative and in classes:
    - Low physical activity
    - Moderate physical activity
    - High physical activity

- Patient global assessment of disease activity using a 100 mm Visual Analogue Scale (VAS)

- Duration of morning stiffness (minutes)

- **Evaluation of RA activity**

- Erythrocyte sedimentation rate (ESR) (1st hour) (mm/h)
- Number of swollen joints (SJC)
- Number of tender joints (TJC)
- DAS28-ESR[DD]
    - Low Disease Activity (LDA) (Yes/No)
    - Remission (Yes/No)
- CDAI[DD]
    - Low Disease Activity (LDA) (Yes/No)
    - Remission (Yes/No)

5.1.4. Safety endpoints

Variables used to assess the safety of Sarilumab are as follows:

- **Adverse events (AEs)**

All AEs recorded during the study period, including serious Adverse Events (SAEs) Adverse Events of Special Interest (AESIs), Treatment-Emergent Adverse Event (TEAE) and AE leading to death will be considered. Adverse events will be coded with MedDRA Dictionary (current version at the time of encoding).

- **Laboratory safety variables:**

- Chemistry Blood: data will be presented as quantitative variable
  - Creatinine clearance (Cockroft-Gault equation) (mL/min)
  - Serum iron (µg/L)
  - 25 (OH) D (ng/mL)
  - Total bilirubin (mg/L)
  - AST (SGOT) (IU/L)
  - ALT (SGPT) (IU/L)
- Lipid panel: data will be presented as quantitative variable
  - Total cholesterol (g/L)
  - HDL cholesterol (g/L)
  - LDL cholesterol (g/L)
  - Triglycerides (g/L)
- Hematology: data will be presented as quantitative variable and as qualitative variable (Normal values / lower than normal range / upper than normal range)
  - Erythrocytes ($\times 10^9$/L)
  - White blood cells (g/dL)
  - Hemoglobin ($\times 10^9$/L)
  - Absolute neutrophil count ($\times 10^9$/L)
  - Absolute eosinophil count ($\times 10^9$/L)
  - Absolute basophil count ($\times 10^9$/L)
  - Absolute lymphocyte count ($\times 10^9$/L)
  - Monocytes ($\times 10^9$/L)
  - Platelet ($\times 10^9$/L)
- Serology
  - HBs antigen (Positive/Negative)
  - Anti-HBc antibodies (Positive/Negative)
  - Anti-HCV antbodies (Positive/Negative)
- Tuberculosis test (Positive/Negative/Not Performed)

- **Treatment with Sarilumab**

- Duration of IMP exposure (week) [DD]
- Number of injections performed per patient (as qualitative and quantitative)
- IMP Interruption
  - Patient with at least one temporary interruption (Yes/No)
    If yes:
    - Patient with at least one temporary interruption due to an adverse event (Yes/No)

        - Patient with at least one temporary interruption due to forgetting (Yes/No)
        - Patient with at least one temporary interruption due to another reason (Yes/No)
          A listing for other reason why not performed will be provided.
    - Patient with definitive interruption (Yes/No). A listing of patients with definitive interruption will be provided.

- Dose
    - Average dose per patient (mg)
    - Patient with at least one dose modification (Yes/No)
      If yes:
        - Patient with at least one dose increase (Yes/No)
        - Patient with at least one dose decrease (Yes/No)
- Injection site
    - Patient with at least one injection performed in the abdomen (Yes/No)
    - Patient with at least one injection performed in the thigh (Yes/No)
    - Patient with at least one injection performed in another injection site (Yes/No)
      A listing of other injection site will be provided.
- Place of injection
    - Patient with at least one injection performed on site (Yes/No)
    - Patient with at least one injection performed at home (Yes/No)
- Person who performed the injection
    - Patient with at least one injection performed by himself (Yes/No)
    - Patient with at least one injection performed by a healthcare professional (Yes/No)
    - Patient with at least one injection performed by another person (Yes/No)
- Problem encountered during injection
    - Patient who received at least once a smaller volume that that in the syringe (Yes/ No)
    - Patient with at least one injection done incorrectly (Yes/No)
    - Patient with at least one other problem encountered during injection (Yes/No)
      A listing will be provided for other problem encountered during injection.
- Treatment compliance
    - Injection performed (Yes/No)

  If yes:
        - Dose (150mg/200mg)
        - Injection site (Abdomen/Thigh/Other)
        - Place of injection (On site/At home)
        - Person who performed the injection (Patient/Healthcare professional/Other)
        - Problem encountered during injection (None/Volume injected below that of the syringe/ Injection done incorrectly/Other)

  If no:
        - Reason (Adverse Event/ Forgetting/ Other reason)

- **Vital signs and physical examination**

- Weight (kg)
- BMI[DD] (kg/m^2)
    - As quantitative
    - In classes:
        - Underweight (<18.5)
        - Normal (healthy) weight ([18.5; 25[)
        - Overweight ([25.0; 30[)
        - Obese (≥30)
- Blood pressure
    - Systolic (mmHg)
    - Diastolic (mmHg)
- Heart rate (beats/min)
- Abdominal girth (cm)
- Electrocardiogram (ECG) result (Normal or with a non-clinically-significant abnormality / Abnormal with a clinically-significant abnormality / not performed)

- **Concomitant treatments**

Prior and concomitant treatments will be presented and coded using World Health Organization - Drug Dictionary (WHO-Drug Dictionary) (current version at the time of encoding).

5.1.5. Exploratory endpoints

5.1.5.1. Associations between changes in RAID score and composite disease activity scores (DAS28-ESR[DD] and CDAI[DD]).

At each visit, all the efficacy parameters will be assessed.
- DAS28-ESR [DD] score
- CDAI [DD] score
- RAID total score

5.1.5.2. Associations between changes in Hospital Anxiety and Depression Scale (HADS) scores and in composite disease activity scores (DAS28-ESR[DD] and CDAI[DD])

At each visit, all the efficacy parameters will be assessed.
- DAS28-ESR [DD] score
- CDAI [DD] score
- HADS scores:
    - Anxiety calculated score
    - Depression calculated score

5.1.5.3. To explore patients' profiles

Patients' profiles will be explored using the same analyses performed for the analysis of baseline data. Association between cognitive troubles and patients' characteristics will be tested. Cognitive troubles will be evaluated by the MAThyS cognition score as qualitative variable (using the tertiles: <1st tertile, [1st tertile; 2nd tertile[, ≥ 2nd tertile).

Characteristics (at baseline) to be tested are:

- Age in quantitative or in classes (using quartiles) if linearity is not verified
- Sex (Female / Male)
- BMI in classes (Normal/Abnormal i.e Underweight or Overweight or Obese)
- Time between baseline and initial diagnosis (years) in quantitative or in classes (using quartiles), if linearity is not verified
- DAS28-ESR in quantitative or in classes (using quartiles), if linearity is not verified
- Patient previously treated with biological therapies and targeted synthetic DMARDs (Yes/No)
- Patient currently treated with corticosteroids for his/her RA (Yes/No)
- Patient currently treated with NSAIDs, analgesics (Yes/No)
- Patient having experienced any past or concomitant diseases (clinically significant) (Yes/No)
- Depression evaluated by HADS (Yes/No). If HADS depression = 'Normal' then Depression='No'. Else if HADS depression = 'Boderline abnormal' or 'Abnormal' then Depression='Yes'
- Anxiety evaluated by HADS (Yes/No). If HADS anxiety = 'Normal' then Anxiety='No'. Else if HADS anxiety = 'Boderline abnormal' or 'Abnormal' then Anxiety='Yes'
- Fatigue (FACIT-F total score) in quantitative or in classes (using quartiles), if linearity is not verified
- Global Assessment of disease activity (mm) by patient (0-no pain to 100-maxima pain imaginable) in quantitative or in classes (using quartiles), if linearity is not verified
- Global Assessment of disease activity (mm) by physician (0-no pain to 100-maxima pain imaginable) in quantitative or in classes (using quartiles), if linearity is not verified
- MAThyS global score in quantitative (only descriptive, not in the model due to collinearity with cognition score)
- Positive Anti-CCP (Yes/No)
- Positive rheumatoid factor (Yes/No)

5.2. ANALYSIS POPULATIONS

Descriptive analyses will be performed on the ITT population. Primary and secondary analyses will be done on ITT population.

For the primary analysis, the sensitivity analyses will be conducted in reproducing the same analyses, using the PPS.

Secondary efficacy analyses will be performed on the PPS, only if PPS represents less than 90% of the ITT population.

Exploratory analyses to search some patients' profiles at baseline will be performed on the ITT population. Safety analyses will be performed on the safety set (SAF).

5.3. STATISTICAL METHODS

5.3.1. Disposition of patients

Key information of the study will be presented (First patient first visit, last patient last visit, first patient last visit, last patient last visit, number of centers with at least one patient included, enrollment duration (months) [DD]).
Disposition of patients will be depicted for patient status and patient analysis populations. The total number of patients for each of the following categories will be described:

- Enrolled patients

- Patients failed to be enrolled and the reason for non-enrollment
- Safety population
- ITT population
- Per-Protocol population.

Summary tables and flowcharts will present the number and percentage of subjects who not entered in the study and reason for non-enrolment into the registry will be presented overall (non-compliance with inclusion / exclusion criteria, patient's refusal of consent).

Summary tables and flowcharts will present the number of subjects at each assessment and the number of subjects by population. Follow-up duration[DD] (weeks) will be calculated overall. The number and percentage of subjects who withdrew from the study and the reasons for premature withdrawal will be presented overall (adverse event, death, treatment inefficacy, major protocol deviations, consent withdrawal, lost to follow-up or other reason). Time between baseline and premature withdrawal[DD] (weeks) will also be presented.

5.3.2. Analyses of evaluation variables

5.3.2.1. Descriptive analysis

Descriptive statistics of all data collected in the CRF and which are not considered as primary, secondary or exploratory endpoint will be provide. These data are defined in the section 5.1.1. A frequency table of the number and percentage of subjects will be provided for medical and surgical history by System Organ Class (SOC) and Preferred Term (PT) overall. A frequency table of the number and percentage of subjects will be provided for prior and concomitant treatments at baseline by therapeutic class (ATC).

5.3.2.2. Primary efficacy analysis

The primary study objective is to describe the effect of sarilumab on the variation of the total RAID score from baseline to V5 (W24). Mean change from baseline (V2) of the total RAID score to W24 (V5) will be provided with its bilateral 95% CI.

Change from baseline in total RAID score after 24 weeks will be provided using a mixed model repeated measurement. This model includes total RAID score at baseline as covariate, and interactions between the covariate and visit.

SAS code:

```
Proc MIXED;
Class Visit;
Model RAID_score_change= Visit + Raid_score_baseline*Visit;
Repeated Visit / type=UN sub=id;
Run;
```

5.3.2.3. Secondary efficacy analyses

For each of the secondary endpoints, values at each assessment visit over the treatment period and changes from baseline for each PROs questionnaires total score to evaluate the RA activity will be presented (absolute and relative variation). 95% CI will be provided.

5.3.2.4. Safety analyses

All AEs (Serious AE (SAE), AE of special interest (AESI), Treatment-Emergent AE (TEAE), AE leading to death, TEAE leading to death) will be presented by System Organ Class (SOC) sorted by internationally agreed order[1] and Preferred Term (PT). A TEAE is defined as an AE occurring in the time from the first injection of IMP up to 14 days after the last injection. In case of AE/TEAE leading to death, cause of death will be provided.

Laboratory safety variables, vital signs and ECG values will be presented at each assessment.

A frequency table of the number and percentage of subjects will be provided for prior and concomitant therapies by therapeutic class (ATC). The table will be sorted by decreasing frequency of anatomic category followed by all other therapeutic classes. In case of equal frequency regarding anatomic categories, alphabetical order will be used.

The following definitions will be applied to laboratory parameters and vital signs
• 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 (PCSA list provided in Appendix J).

• PCSA criteria will determine which patients had at least 1 PCSA during the on-treatment period, taking into account all evaluations performed during the on-treatment period, including unscheduled or repeated evaluations. The number of all such patients will be the numerator for the on-treatment PCSA percentage.

5.3.2.5. Exploratory analysis

5.3.2.5.1. Associations between changes in RAID score and composite disease activity scores (DAS28-ESR^{DD} and CDAI^{DD}).

The first exploratory study objective is to describe the associations between changes in RAID score and composite disease activity scores (DAS28-ESR^{DD} and CDAI^{DD}). Change from baseline in total RAID score after 24 weeks will be provided using a mixed model repeated measurement. This model includes total RAID score at baseline and DAS28-ESR^{DD} (respectively CDAI^{DD}) as covariates, and interactions between the covariates and visit. The same model will be used with the CDAI^{DD} instead of DAS28-ESR^{DD}.

SAS code:

```
Proc MIXED;
Class Visit;
Model RAID_score_change= DAS28-ESR + Visit + DAS28-ESR*Visit + Raid_score_base*Visit;
Repeated Visit / type=UN sub=id;Run;
```

5.3.2.5.2. Associations between changes in Hospital Anxiety and Depression Scale (HADS) scores and in composite disease activity scores (DAS28-ESR^{DD} and CDAI^{DD})

The second exploratory study objective is similar to the first. It is to describe the associations between changes in HADS scores and composite disease activity scores (DAS28-ESR^{DD} and CDAI^{DD}). Change from baseline in total HADS scores after 24 weeks will

[1] Internationally agreed order as defined in table 3-2 of Introductory Guide MedDRA Version 22.1 (available on https://www.meddra.org/how-to-use/support-documentation/english)

be provided using a mixed model repeated measurement. This model includes total HADS score at baseline and DAS28-ESR^{DD} (respectively CDAI^{DD}) as covariates, and interactions between the covariates and visit. The same SAS code will be considered, using HADS scores instead of RAID score for both composite disease activity scores.

5.3.2.5.3. To explore patients' profiles

To assess the association between cognitive troubles and patient's characteristics, MAThyS cognition score will be tested. Factors (patient's characteristics) associated with cognitive troubles will be identified using:

- 1 multidimensional analysis (Factorial Analysis for Mixed Data [FAMD])
- 1 logistic regression for cognition score using tertiles (adjacent-category logits model with second tertile as reference)

In a preliminary step, patient's characteristics will be described according to the variable to explain.

FAMD analysis:

This method will allow to establish subjects' profiles, dealing with the small expected number of subjects compared to the high number of variables.
Due to lack of robustness to missing data with this method, subjects with at least one missing value will be excluded from the analysis. Their characteristics (all characteristics listed above) will be described in a summary table.
The FAMD will be conducted as follows:

1. All characteristics and MAThys cognition score (in 3 classes according to the tertiles) will be entered in the FAMD model.
2. The choice of the number of axes will be determined graphically (percentage of inertia). The raw eigen values histogram will be provided.

   Note: As the number of variables impacts the inertia of the analysis (higher number of variables are, higher inertia is.
3. Graphics of variables on factorial axes will be provided to analyse the correlations between variables. Contribution, quality of representation (cos^2) and v-test will be studied to explain the axes.
4. Graphics of subjects on factorial axes will be provided to analyse the profiles of subjects
5. An Ascending Hierarchical Clustering (AHC) will also be performed after FAMD to create clusters of subjects.
6. A summary table of each cluster will be provided.

Logistic regression analyses:

For the analysis, identification of factors will be conducted in two steps:

1. A univariate analysis of each of the baseline parameter listed above separately. If some factors appear to be associated (i.e., statistically significant at level α= 0.20) with outcome, they will be gathered into a multivariate logistic model.
2. A multivariate logistic regression analysis: a stepwise selection of all factors selected in the second step described above, including their interaction, if relevant (significant at p<0.10), will be performed. The significance level will be set at 5%.

SAS code (varA is qualitative, varB is quantitative):

```
proc logistic data= table order=data;

class varA (param=ref ref='reference category');

model OUTCOME=varA varB /link=alogit rsquare maxiter=25;

run;
```

6. DATA HANDLING CONVENTIONS

6.1. DATA HANDLING CONVENTIONS FOR SITE QUESTIONNAIRE

Not applicable.

6.2. DATA HANDLING CONVENTIONS FOR PATIENT DATA OUTLIERS

Any outlier identified prior to database lock which is impossible/implausible will be excluded from the analysis. A search of outliers should be performed before the database lock and actions with the sponsor should be defined.

6.3. DATA HANDLING CONVENTIONS FOR PATIENT REPORTED OUTCOMES

The management of missing data for PRO is described in Appendix A.

6.4. MISSING DATA

For the primary endpoint, in case of treatment discontinuation due to lack of efficacy (as defined in Protocol Version 2 section 11.4.2.1) or adverse event (except for AEs due to NSAID), the last available endpoint before the considered event will be taken in account (for RAID score for primary endpoint and for all other secondary endpoints). For all other cases, missing data will not be replaced.

MISSING OR INCOMPLETE DATES

For all dates (except adverse events and concomitant treatments):

- If only day is missing, it will be replaced by ‘15’
- If day and month are missing, day will be replaced by 1 and month will be replaced by July ‘7’
- In case of missing year or complete missing date, no replacement will be performed

For dates relative to adverse events (AE) and concomitant treatments, if a date is completely or partially missing, date will not be replaced but:

- If day is missing and start date/end date of AE/treatment occurred before or the same month as the end of study, AE/treatment will be considered as “treatment-emergent” by default.
- If day and month are missing and start date/end date of AE/treatment occurred the same year as the end of study, AE/treatment will be considered as “treatment-emergent” by default.
- In case of missing year or complete missing date, start date/end date of AE/treatment will be considered as “treatment-emergent” by default.

6.5. DEFINITION OF REGIONS

Not applicable.

6.6. WINDOWS FOR TIME POINTS

For each visit, a window of time for analysis endpoints may need to be defined:

- Week 4 (V3) may be defined as between 25 and 31 days.
- Week 12 (V4) may be defined as between 77 and 91 days.
- Week 24 (V5) may be defined as between 161 and 175 days.
- Visit LTE1 (W36) may be defined as between 245 and 259 days.
- Visit LTE2 (W48) may be defined as between 329 and 343 days.
- Premature withdrawal visit will be allocated to a study visit according to the delay between baseline and premature withdrawal, if and only if this study visit has not been attended yet. For example, if a patient withdrew after V3 and time to withdrawal between 77 and 91 days, then premature withdrawal data collected will be considered as V4 data. If time to withdrawal is below 77 days, data will not be considered in the analysis.

See Graphical study design (Figure 1) in the Section 1.1.

6.7. STATISTICAL TECHNICAL ISSUES

Not applicable.

7. ANALYSIS OF BASELINE DATA

Analysis of baseline data details have been presented in a separate document (SANOFI - SARIPRO- Statistical Analysis Plan - INTERIM ANALYSIS Version 2.0 20190611).

8. SOFTWARE DOCUMENTATION

8.1. HARDWARE

The statistical analysis will be performed using Windows 7 professional and a 64-bit operating system.

8.2. SOFTWARE

All tables, listings and figures will be produced and statistical analysis performed using SAS® version 9.3 or above. All outputs will be edited in Microsoft Word Format.

8.3. VALIDATION OF PROGRAMS

Validations will be done according to the procedure of the CRO ITEC Services.

The program reviewer is responsible for reviewing each project program and output provided to the sponsor. Program logs are reviewed for logical, syntax and fatal errors. The review in SAS® includes, but is not limited to, all ERRORS, WARNINGS, BY-VALUE merge messages, NOTES, and UNINITIALIZED variables. Program logs are also reviewed for accurate and consistent variable and observation counts following each procedure and data step.

The reviewing statistician is responsible for checking and reviewing the work produced using whatever method he/she feels is appropriate (e.g., SAS® code review, hand calculation, etc.) to reassure of the quality of the output.

Outputs are reviewed for typographical errors, misspellings and nonsensical values or results and to check the consistency with the statistical and analysis plan. Outputs are cross-checked against each other for accuracy and consistency. For statistical tables, listings, appendix listings, and figures, this procedure includes comparison of subject group numbers, counts of subjects at each observation point, and consistency of results for variables between outputs.

Findings of the quality control reviews are communicated to the party responsible for making necessary changes. The programs will be retested after modifications.

After final review, and when no further changes are required to produce the deliverable, the program reviewer and the statistician in charge of the study need to complete and sign a validation checklist, to indicate that they have successfully performed all of their responsibilities.

Copies of the internal quality control forms produced for the validation process and the sign-off forms will be provided to the sponsor to support the validation if asked.

8.4. RESTITUTION OF THE PROGRAMS

All programs (including Macros and SAS® analysis datasets) producing the tables, listings and statistical output along with associated logs should be given to the sponsor when the tables, listings, figures and statistical analysis has been finalised.

9. LIST OF APPENDICES

Appendix A: Derived data **34**
Appendix B: Decision tables **40**
Appendix C: Rheumatoid Arthritis Impact of Disease - RAID questionnaire **43**
Appendix D: Hospital Anxiety and Depression Scale (HADS) **44**
Appendix E: FACIT Fatigue Scale (Version 4) **45**
Appendix F: Health Assessment Questionnaire – Disability Index (HAQ-DI) **46**
Appendix G: International Physical Activity Questionnaire (IPAQ) **49**
Appendix H: Multidimensional Assessment of Thymic States (MAThys) **53**
Appendix I: Global assessment of disease activity **55**
Appendix J: Potentially clinically significant abnormalities (PCSA) criteria **56**

APPENDIX A: DERIVED DATA

The derived data are variables which are calculated from the raw data in the CRF/eCRF and not included in the database (e.g.: AUC, total score, duration, summary value from repeated observations, study exposure, study drug exposure, study drug cumulative dose, etc.).

The formula used for derived data should be provided and examples may be necessary to clarify the calculation. The strategy for missing data must be specified in the calculation of derived data (e.g.: the minimum number of observations required to calculate an average score, handling incomplete dates).

All references used to calculate the derived data should be mentioned below.

The following derived data will be calculated and included in the listings:

(1) SUBGROUPS DECISION

Treatment started before screening visit and ongoing at screening visit	**Prescribed treatment at baseline**
csDMARds	Sarilumab + csDMARDs
Biological therapy	Sarilumab + csDMARDs
csDMARDs or biological therapy	Sarilumab in monotherapy

(2) DAS28-ESR

DAS28-ESR=0.56×√(TJC28)+0.28×√(SJC28)+ln(ESR)+0.014×PtGA (in mm)

Where:

- ln is the natural logarithm
- SJC is the swollen joints
- TJC is the tender joints
- PtGA is the patient's global assessment of disease activity (using a 100 mm VAS)
- ESR is the erythrocyte sedimentation rate (mm/h)

(3) CDAI

CDAI=TJC28+SJC28+PtGA(in cm)+PhGA(in cm)

Where:

- SJC is the swollen joints
- TJC is the tender joints
- PhGA is the physician's global assessment of disease activity (using a 100 mm VAS)
- PtGA is the patient's global assessment of disease activity (using a 100 mm VAS)

(4) BMI

BMI (kg/m^2) will be derived as Weight (kg)/[Height(cm)/100]^2 and rounded to the nearest decimal.

(5) Time between baseline and initial diagnosis (years)

Time between baseline and initial diagnosis in years will be calculated as (initial diagnosis date-baseline visit date+ 1) / 365.25.

(6) **Changes from baseline (V2)**

Absolute variation (points) from baseline will be calculated as a difference from baseline (e.g. assessment at the visit – assessment at baseline).
Relative variation (%) from baseline will be calculated as: ((assessment at the visit – assessment at baseline)/ assessment at baseline) * 100.
If necessary, evolution will be presented according to decision tables from Appendix B.

(7) **Duration of IMP exposure (week)**

The duration of IMP exposure is defined as follows: (last dose date – first dose date + 14 days)/7 regardless of unplanned intermittent discontinuations.

(8) **Enrollment duration (months)**

Enrollment duration (months) will be calculated as: ((date of last patient last visit-date of first patient first visit) + 1) / 30.4375.

(9) **Follow-up duration (weeks)**

Follow-up duration in weeks will be calculated as: ((date of last contact with the patient – date of baseline) + 1) / 7.

(10) **Time between baseline and premature withdrawal (weeks)**

Time between baseline and premature withdrawal in weeks will be calculated as ((withdrawal date – baseline visit date) + 1) / 7.

(11) **Time between measures of the rheumatoid nodule (months)**

Time between measures of the rheumatoid nodule (months) will be calculated as: (date of the last measure of the rheumatoid nodule - date of screening)+1) / 30.4375.

(12) **RAID**

The RAID is calculated based on 7 Numerical rating scales (NRS) questions (see Appendix C). Each NRS is assessed as a number between 0 and 10. The 7 NRS correspond to pain, function, fatigue, sleep, emotional well-being, physical well-being, and coping/self-efficacy.

Calculation: RAID final value = (pain NRS value (range 0-10) x 0.21) + (function NRS value (range 0-10) x 0.16) + (fatigue NRS value (range 0-10) x 0.15) + (phys well being NRS value (range 0- 10) x 0.12) + (sleep NRS value (range 0-10) x 0.12) + (emotional well being NRS value (range 0-10) x 0.12) + (coping NRS value (range 0-10) x 0.12).

Thus, the range of the final RAID value is 0-10 where higher figures indicate worse status. Missing data imputation: If one of the 7 NRS values composing the RAID is missing, the imputation is as follows:

a. calculate the mean value of the 6 other (non-missing) NRS (range, 0-10)

b. impute this value for the missing NRS

c. Then, calculate the RAID as explained above. If 2 or more of the NRS are missing, the RAID is considered as missing value (no imputation).

See: http://pitie-salpetriere.aphp.fr/psaid/raid_psaid_quest_home.php

(13) Hospital Depression Scale (issue from HADS)

Total score of depression is the sum of tick boxes of columns D (see Appendix D). The range of the total score of depression value is 0-21.

(14) Hospital Anxiety Scale (issue from HADS)

Total score of anxiety is the sum of tick boxes of columns A (see Appendix D). The range of the total score of anxiety value is 0-21.

(15) Multidimensional Assessment of Thymic States (MAThys):

The MAThyS is a visual analog scale consisting of 20 items variing from 0 to 10 (see Appendix H). These items corresponded to five dimensions: emotional, cognition, psychomotor function, motivation and sensory perception.

Total score is the sum of the 20 items. The measure is the number of centimetres from the left hand anchor. Items measured from 0 to 10 are: 1; 2; 3; 4; 11; 12; 13; 14; 15; 16; 19; 20 and items measured from 10 to 0 are : 5; 6; 7; 8; 9; 10; 17; 18. A score of 0 corresponds to inhibition of the state evaluated by the item. A score of 5 indicates no change from the patient's usual state and a score of 10 corresponds to excitation for the evaluated state. An overall score of between 0 and 200 is thus obtained. This scale is not devoted to make a diagnosis of mood state but allow to determine the general level of inhibition/activation processes (lower scores indicate general inhibition and higher scores indicate general excitation).

The subscore for **emotional** is obtained by the sum of items: 3; 7; 10; 18. This range is 0 (hyporeactive)-40 (herperreactive).

The subscore for **cognition** is obtained by the sum of items: 5; 9; 12; 14. This range is 0 (Retardation)-40 (Acceleration).

The subscore for **psychomotor function** is obtained by the sum of items: 2; 11; 19. This range is 0 (Retardation)-30 (Acceleration).

The subscore for **motivation** is obtained by the sum of items: 4; 15; 16; 17. This range is 0 (Decrease)-40 (Increase).

The subscore for **sensory perception** is obtained by the sum of items: 1; 6; 8; 13; 20. This range is 0 (Decrease)-50 (Increase).

See: Chantal Henry, Katia M'Bailara, Flavie Mathieu, Rollon Poinsot, Bruno Falissard Construction and validation of a dimensional scale exploring mood disorders: MAThyS (Multidimensional Assessment of Thymic States). BMC Psychiatry. 2008; 8: 82.

(16) Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-F) per item and total score

The FACIT measurement system is a comprehensive compilation of questions that measure health related QOL in patients with chronic illnesses. The FACIT-F includes the 13 items.

FACIT-F score is the sum of items, this range is 0 to 52. High FACIT-F scores represent less fatigue.

See: CELLA, D, YOUNT, S, SORENSEN, M, CHARTASH, E, SENGUPTA, N ET GROBER, J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. J Rheumatol. 2005: 32: 811-9.

(17) The Health Assessment Questionnaire - Disability Index (HAQ-DI)

There are 8 sections: dressing, arising, eating, walking, hygiene, reach, grip, and activities. There are 2 or 3 questions for each section. Scoring within each section is from 0 (without any difficulty) to 3 (unable to do). For each section the score given to that section is the worst score within the section, i.e. if one question is scored 1 and another 2, then the score for the section is 2. In addition, if an aide or device is used or if help is required from another individual, then the minimum score for that section is 2. If the section score is already 2 or more then no modification is made.

The 8 scores of the 8 sections are summed and divided by 8. The result is the disability index. In the event that one section is not completed by a subject then the summed score would be divided by 7. In the event that two section is not completed by a subject then the summed score would be divided by 6 etc. The score range is 0 to 3.

See: Bruce, Bonnie, et James F Fries. « The Stanford Health Assessment Questionnaire: Dimensions and Practical Applications ». *Health and Quality of Life Outcomes* 1, nº 1 (2003): 20. https://doi.org/10.1186/1477-7525-1-20.

(18) International Physical Activity Questionnaire (IPAQ)

IPAQ assesses physical activity undertaken across a comprehensive set of domains including: leisure time physical activity, domestic and gardening (yard) activities, work-related physical activity and transport-related physical activity.

Results can be reported in categories (low activity levels, moderate activity levels or high activity levels) or as a continuous variable (MET minutes a week). MET minutes represent the amount of energy expended carrying out physical activity. A MET is a multiple of your estimated resting energy expenditure. One MET is what you expend when you are at rest. To get a continuous variable score from the IPAQ (MET minutes a week) we will consider walking to be 3.3 METS, moderate physical activity to be 4 METS and vigorous physical activity to be 8 METS.

Walking MET-minutes/week = 3.3 * walking minutes * walking days

Moderate MET-minutes/week = 4.0 * moderate-intensity activity minutes * moderate days

Vigorous MET-minutes/week = 8.0 * vigorous-intensity activity minutes * vigorous-intensity days

Total physical activity MET-minutes/week = sum of Walking + Moderate + Vigorous METminutes/week scores.

If ‘don’t know’ or ‘refused ‘ or data are missing for time or days then that case is removed from analysis.

The items in the short IPAQ form were structured to provide separate scores on walking, moderate-intensity and vigorous-intensity activity. Computation of the total score for the short form requires summation of the duration (in minutes) and frequency (days) of walking, moderate-intensity and vigorous-intensity activities. Domain specific estimates cannot be estimated. Below, physical activity categories are defined:

- Category 1 Low

This is the lowest level of physical activity. Those individuals who not meet criteria for Categories 2 or 3 are considered to have a 'low' physical activity level.

- Category 2 Moderate

The pattern of activity to be classified as 'moderate' is either of the following criteria:

a) 3 or more days of vigorous-intensity activity of at least 20 minutes per day

OR

b) 5 or more days of moderate-intensity activity and/or walking of at least 30 minutes per day

OR

c) 5 or more days of any combination of walking, moderate-intensity or vigorous intensity activities achieving a minimum Total physical activity of at least 600 MET-minutes/week.

Individuals meeting at least one of the above criteria would be defined as accumulating a minimum level of activity and therefore be classified as 'moderate'.

- Category 3 High

A separate category labelled 'high' can be computed to describe higher levels of participation. The two criteria for classification as 'high' are:

a) vigorous-intensity activity on at least 3 days achieving a minimum Total physical activity of at least 1500 MET-minutes/week

OR

b) 7 or more days of any combination of walking, moderate-intensity or vigorous-intensity activities achieving a minimum Total physical activity of at least 3000 MET-minutes/week.

Rules to handle inconsistencies: if any inconsistency is detected (for example, an activity 11 days per week, an activity ticked 'No activity' but a time activity completed…), the data will be considered as missing.

See: Guidelines of IPAQ (2005)

(19) Conversion method for laboratory tests

a) Lipid panel:

Value (mmol/L) = 5.5 × Value (g/L)

b) Hematology:

Value ($\times 10^{12}$/L) = 1000 × Value ($\times 10^{9}$/L)

Value (g/L) = 0.1 × Value (g/dL)

(20) **Normal ranges for laboratory tests (hematology)**

	Man	Women	RANGE
Erythrocytes	4,6 à 6,2 millions/mm3	4,2 -5,4	<4 et >6
Hemoglobin	13.5 – 17 g/dl	12 – 15,5 g/dl	< 12 et > 18
White Blood Cells	3.5 – 10x10⁹/L		<3.5 et > 10x10⁹/L
Neutrophils	3000 – 5800 x10⁶/L		<3000 et >6000 x10⁶/L
Lymphocytes	1500-3000 x10⁶/L		<1500 et >3000 x10⁶/L
Monocytes	300-500 x10⁶/L		<300 et >500 x10⁶/L
Eosinophils	50-250 x10⁶/L		<50 et 250 x10⁶/L
Basophils	15- 50 x10⁶/L		<15 et >50 x10⁶/L
Platelets	150 000 -400 000		<150 000 et > 400 000

APPENDIX B: DECISION TABLES

To evaluate the variation of each total score of the PROs in classes, decision tables below will be used.

Total RAID score		**At V*x***	
		Acceptable status	*Non acceptable status*
At baseline	*Acceptable status*	Unchanged (Acceptable status)	Worsening
	Non acceptable status	Improvement	Unchanged (Non acceptable status)

Figure 3: Decision table for the variation of the total RAID score in classes

HADS scores		**At V*x***		
		Normal	*Borderline Abnormal*	*Abnormal*
At baseline	*Normal*	Unchanged (Normal)	Worsening	Worsening
	Borderline Abnormal	Improvement	Unchanged (Borderline Abnormal)	Worsening
	Abnormal	Improvement	Improvement	Unchanged (Abnormal)

Figure 4: Decision table for the variation of the HADS scores (anxiety and depression) in classes

HAQ-DI scores		**At Vx**			
		Without any difficulty	*With some difficulty*	*With much difficulty*	*Unable to do*
At baseline	*Without any difficulty*	Unchanged (Without any difficulty)	Worsening	Worsening	Worsening
	With some difficulty	Improvement	Unchanged (With some difficulty)	Worsening	Worsening
	With much difficulty	Improvement	Improvement	Unchanged (With much difficulty)	Worsening
	Unable to do	Improvement	Improvement	Improvement	Unchanged (Unable to do)

Figure 5: Decision table for the variation of the HAQ-DI domain scores in classes

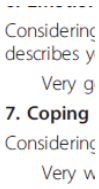
IPAQ total score		**At V*x***		
		Low physical activity	*Moderate physical activity*	*High physical activity*
At baseline	*Low physical activity*	Unchanged (Low physical activity)	Improvement	Improvement
	Moderate physical activity	Worsening	Unchanged (Moderate physical activity)	Improvement
	High physical activity	Worsening	Worsening	Unchanged (High physical activity)

Figure 6: Decision table for the variation of the IPAQ classification in classes

Comparison	Classes
Value at Vx < Value at baseline	Decrease
Value at baseline = Value at Vx	Unchanged
Value at baseline < Value at Vx	Increase

Figure 7: Decision table for the size of the rheumatoid nodule evolution in classes

APPENDIX C: RHEUMATOID ARTHRITIS IMPACT OF DISEASE - RAID QUESTIONNAIRE


Considering
describes y
Very g
7. Coping
Considering
Very w

APPENDIX D: HOSPITAL ANXIETY AND DEPRESSION SCALE (HADS)

Plea

APPENDIX E: FACIT FATIGUE SCALE (VERSION 4)

APPENDIX F: HEALTH ASSESSMENT QUESTIONNAIRE – DISABILITY INDEX (HAQ-DI)

☐ Dr

☐ Hy

repres

APPENDIX G: INTERNATIONAL PHYSICAL ACTIVITY QUESTIONNAIRE (IPAQ)

T

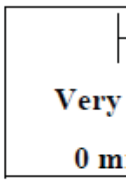
APPENDIX H: MULTIDIMENSIONAL ASSESSMENT OF THYMIC STATES (MATHYS)

environn

APPENDIX I: GLOBAL ASSESSMENT OF DISEASE ACTIVITY

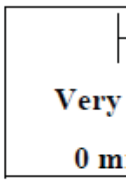
Patient global assessment of disease activity

Considering all the ways that your rheumatoid arthritis affects you, rate how you are doing on the following scale by placing a vertical mark on the line:



Physician global assessment of disease activity

Place a vertical mark on the line for how you would assess your patient's current activity:



APPENDIX J: POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES (PCSA) CRITERIA

Parameter	PCSA	Comments
Chemistry Blood		
Creatinine clearance (Cockroft-Gault equation) (mL/min)	<15 (end stage renal disease) ≥ 15 - < 30 (severe decrease in Glomerular Filtration Rate (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 ondosing and labeling
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.
AST (SGOT)	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.
ALT (SGPT)	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.
ALT (SGPT) and Total Bilirubin	ALT>3 ULN and Total Bilirubin > 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.
Lipid panel		
Total Cholesterol	≥7.74 mmol/L ≥6.2 mmol/L	Threshold for therapeutic intervention.
LDL Cholesterol	≥4.1 mmol/L ≥4.9 mmol/L	
Triglycerides	≥4.6 mmol/L ≥5.6 mmol/L	Threshold for therapeutic intervention.
Hematology		
Erytrocytes	≥ 6 $x10^{12}$/L	
White blood cells	≥ 16.0 $x10^{9}$/L	

Hemoglobin	≤115 g/L (Male); ≤95 g/L (Female) ≥185 g/L (Male); ≥165 g/L (Female) Decrease from screening visit ≥20 g/L	
Absolute neutrophil count	< 1.0 $x10^9$/L	International Consensus meeting on drug-induced blood cytopenias, 1991. FDA criteria.
Absolute eosinophil count	> 0.5 $x10^9$/L or > ULN (if ULN ≥ 0.5 $x10^9$/L)	Harrison- Principles of internal Medicine 17th Ed., 2008.
Absolute basophil count	> 0.1 $x10^9$/L	
Absolute lymphocyte count	< 0.5 $x10^9$/L ≥ 0.5 $x10^9$/L – LLN > 4.0 $x10^9$/L	
Monocytes	> 0.7 $x10^9$/L	
Platelets	< 50 $x10^9$/L ≥ 50 - 100 $x10^9$/L ≥ 700 $x10^9$/L	
Vital signs		
Systolic blood pressure	≤ 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.
Diastolic blood pressure	≤ 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.
Heart rate	≤ 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.
Physical examination		
Weight	≥ 5% increase from baseline ≥ 5% decrease from baseline	FDA Feb 2007.

10. REFERENCES

[1]. Heiberg T, Austad C, Kvien TK, et al. Performance of the Rheumatoid Arthritis Impact of Disease (RAID) score in relation to other patient-reported outcomes in a register of patients with rheumatoid arthritis. Ann Rheum Dis. 2011; 70(6): 1080-2.

[2]. Austad C, Kvien TK, Olsen IC, et al. Sleep disturbance in patients with rheumatoid arthritis is related to fatigue, disease activity, and other patient-reported outcomes. Scand J Rheumatol. 2016; 20: 1-9.

[3]. Medrare L, Allali F, Ngueleu A, et al. Factors influencing the Rheumatoid Arthritis Impact of Disease (RAID) score for Moroccan patients with rheumatoid arthritis (RA) IJIAS 2016; 15(2): 437-42.

[4]. Burmester GR, Lin Y, Patel R, et al. Efficacy and safety of sarilumab monotherapy versus adalimumab monotherapy for the treatment of patients with active rheumatoid arthritis (MONARCH): a randomised, double-blind, parallel-group phase III trial; Ann Rheum Dis. 2016; 76: 840-7.

[5]. Strand V, Reaney M; Chen CI, et al. Sarilumab improves patient-reported outcomes in rheumatoid arthritis patients with inadequate response/ intolerance to tumour necrosis factor inhibitors. RMD Open 2017; 3: e000416.

[6]. Dougados M, Brault Y, Logeart I, et al. Defining cut-off values for disease activity states and improvement scores for patient-reported outcomes: the example of the Rheumatoid Arthritis Impact of Disease (RAID). Arthritis