Non-Interventional Study Protocol
A3921313

IDentification of Factors predictive of Tofacitinib's survival in patient with rheumatoid arthritis in routine practice and impact of patients' behavioural strategies on clinical parameters: the DeFacTo study

Statistical Analysis Plan
(SAP)

Version: 3

Author: PPD (Version 1)
PPD (Version 2)
PPD (Version 3)

Date: 26-Apr-2021

TABLE OF CONTENTS

LIST OF ABBREVATIONS

Abbreviation	Definition
ACPA	Anti-citrullinated Protein Antibody
ACR	American College of Rheumatology
AE	Adverse Event
CDAI	Clinical Disease Activity Index
CI	Confidence Interval
CRP	C Reactive Protein
csDMARD	Conventional Synthetic DMARD
CSQ	Coping Strategies Questionnaire
DAS28	Disease Activity Index 28
DMARD	Disease Modifying Anti-Rheumatic Drug
eCRF	Electronic Case Report Form
EMA	European Medical Agency
ESR	Erythrocyte Sedimentation Rate
EULAR	European League Against Rheumatism
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy Fatigue
FAS	Full Analysis Set
FiRST	Fibromyalgia Rapid Screening Tool
HR	Hazard Ratio
JAK-Inhibitor	Janus Kinase Inhibitors
KM	Kaplan-Meier
LDA	Low Disease Activity
MCS	Mental Component Summary
MTX	Methotrexate
OOB	Out-Of-Bag (Bootstrapping)
OR	Odds Ratio
PCS	Pain Catastrophising Scale
PCS_{SF12}	Physical Component Summary
PH	Proportional Hazards
PhGA	Physician Global Assessment
PtGA	Patient Global Assessment
QoL	Quality of Life
RA	Rheumatoid Arthritis
SAP	Statisticial Analysis Plan
SDAI	Simplified Disease Activity Index
SF-12	Short Form-12 Health Survey
SJC	Swollen Joint Count
SFR	French Society of Rheumatology
TEAE	Treatment Emergent Adverse Event
TJC	Tender Joint Count
Ts-DMARD	Targeted synthetic-DMARD
VAS	Visual Analogue Scale

1 AMENDMENTS FROM PREVIOUS VERSION(S)

1.1 VERSIONS

Version	Effective Date	Change Type (New, Revise, Admin)	Summary of Revisions
1	11-Dec-2019	New	
2	25-Feb-2020	Revise	To add subgroup analyses and to include comments from the scientific committee.
3	26-Apr-2021	Revise	To be in line with protocol amendment 2 To modify the definition of the Full Analysis Set To assess the impact of Covid-19 on study results

1.2 AMENDMENTS FROM THE PROTOCOL

The handling of visit windows has been amended in the SAP compared to the protocol. The Protocol states that:

In order to collect comparable study data, visits 2 to 9 occurring +/- 14 days of the scheduled visit date will be used for data analysis.

As the visits will be scheduled according to clinical practice there are likely to be many visits outside of the windows defined in the protocol. To ensure the primary analysis incorporates all visits at which a patient might experience a treatment escalation all visits will be included for the primary analysis, based on the windowing in Table 9 of Section 11.1.

2 INTRODUCTION

Note: in this document any text taken directly from the non-interventional (NI) study protocol is ***italicised***. This SAP is based on A3921313 NIS DeFacTo Final Protocol, dated 09 November 2018.

Rheumatoid arthritis (RA) is an auto-immune disease characterised by inflammation and joint destruction which is at the origin of progressive incapacity and negative psychological effects [1, 2]. RA is a problem of public health, whether on the level of the individual or on society, in particular, regarding incapacity for work and resultant loss of productivity [3, 4].

According to the recent European League Against Rheumatism (EULAR) recommendations [5] and the current recommendations of the French Society of Rheumatology (SFR) [1], earlier diagnosis is leading to earlier therapeutic intervention

with the goal of disease activity control, maintenance of physical function, optimization of quality of life and improvement of patient mindset [1]. "Treat-to-target" is defined as the basic therapeutic principle in order to reach a goal of sustained remission or low disease activity in every patient [1, 5].

Xeljanz® (Tofacitinib) is an orally targeted synthetic DMARDs (ts-DMARDs) of the new class called Janus kinase inhibitors (JAK-inhibitors). In combination with methotrexate (MTX), Tofacitinib has been approved in 2017 by the European Medicine Agency (EMA) [6] for the treatment of moderate-to-severe active rheumatoid arthritis in adult patient who have responded inadequately to, or who are intolerant to one or more disease modifying anti-rheumatic drugs (DMARDs). Furthermore, Tofacitinib can also be administered as monotherapy in case of intolerance to MTX or when treatment with MTX is inappropriate [7, 8]

Tofacitinib efficacy and safety in the treatment of rheumatoid arthritis has been proven and tried not only across a wide clinical development plan including 7 phase III studies and 2 long-term extension studies [7, 9-15, 16], but also under real life conditions of use since its first approval in the United States in 2012 and subsequent years in other countries worldwide [17].

Although the efficacy and safety profile of Tofacitinib is currently well-documented in the clinical setting [7, 9-15, 16], many fields of research remain to be investigated regarding this drug of a novel therapeutic class, including outcomes in the real world setting. For example, nowadays it is recognized that treatment retention is an overall marker of treatment success depending on multiple factors including efficacy, safety, patients' and disease characteristics at baseline, patients' and physicians' confidence with the treatment [1, 5]. Survival rate is a standardized method for identifying predictive factors of drug survival. Existing data with biologic DMARDs in rheumatoid arthritis have identified a number of independent predictors for treatment continuation [18-23]. Regarding Tofacitinib, newly introduced on the market, few data on drug retention [24] and no data on the factors predictive of Tofacitinib drug survival in patients with RA are available.

Therefore, the primary objective of the DeFacTo study will be to identify the factors predictive of Tofacitinib drug survival in patients with RA.

Moreover, rheumatoid arthritis substantially affects quality of life. Different factors such as pain, disability or fatigue which are directly or indirectly the result of inflammation can have a negative impact on quality of life [1, 25]. Nevertheless, such quality of life is not governed solely by symptoms of the disease, but also by the behaviour of the patient, as well as by behavioural strategies which he/she evidences with regard to the disease. We often refer to catastrophisation [26-28] and coping [29-32] to evaluate these behavioural strategies.If catastrophisation is described as a distortion of the perception of pain involving a both emotional and cognitive component, pushing the patient to see only the worst [26-28], coping involves adaptive strategies by which the patient attempts to find solutions in order to better cope with his/her disease [29-32]. It has been demonstrated that

such behavioural strategies can influence directly or indirectly the intensity of symptoms [33].

As secondary objectives, the impact of behavioural strategies on drug survival and other clinical parameters as well as Tofacitinib effectiveness and tolerability will be studied under real-life conditions of use in French patients with RA.

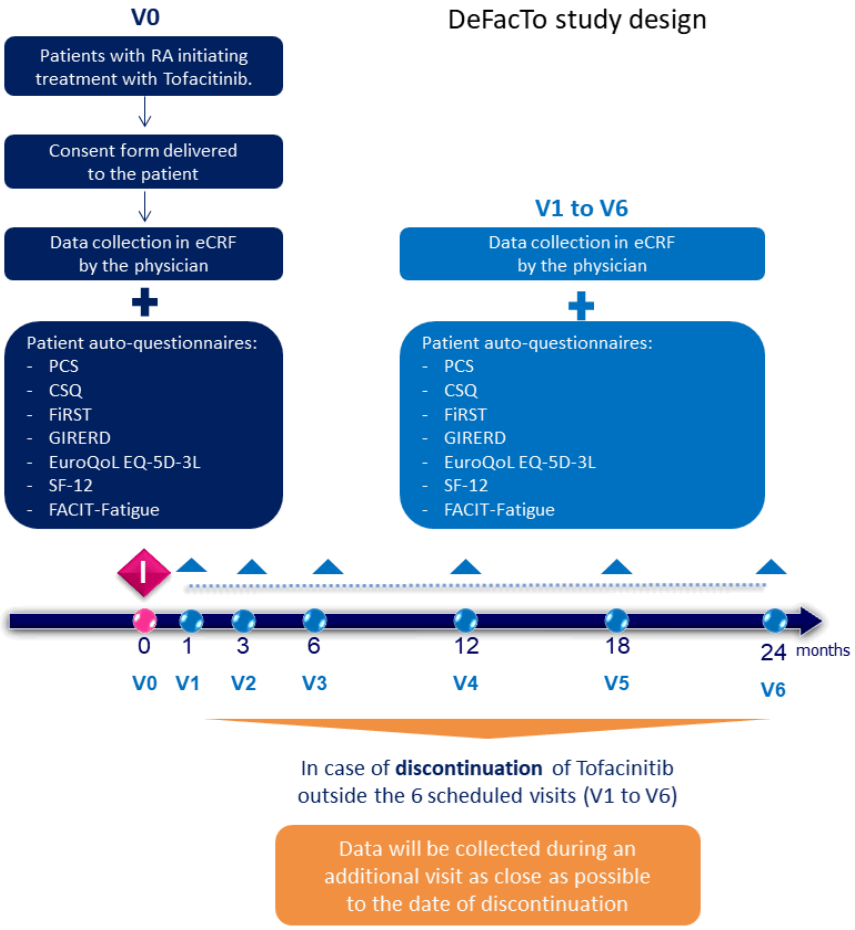
2.1 STUDY DESIGN

This is an observational, open-label, prospective, multi-centre, national study designed to evaluate the factors predictive of Tofacitinib's survival in patients with rheumatoid arthritis.

The study will consist of 500 patients recruited from approximately 100 centres. *The duration of this study will be approximately 48 months including a 24-month recruitment period and a 24-month patient follow-up period. Patients will be followed prospectively and follow-up visits will be conducted after the initial consultations. No visit or additional test is required by the protocol, since the study is observational: the modalities for follow-up and treatment will be left up to the sole judgement of the participating doctor.* Figure 1 shows the study schema.

Data from France will be compared and also pooled with data from other countries. Details will be included in a separate analysis plan for a pooled analysis.

Figure 1: Study schema



Eligible patients will be followed starting from the date of the first prescription of Tofacitinib during 24 months, whether they continue treatment with Tofacitinib or not.

Table 1 shows the schedule of activities.

Table 1 Schedule of Activities

	Inclusion visit Initiation of treatment (Target Day: 1)	Follow-up visit						Additional visit in case of discontinuation of Tofacitinib outside the 6 scheduled visits
		At 1 month (V1) (Target Day: 30)	At 3 months (V2) (Target Day: 91)	At 6 months (V3) (Target Day: 183)	At 12 months (V4) (Target Day: 365)	At 18 months (V5) (Target Day: 548)	At 24 months (V6) (Target Day: 731)	
Patient consent	X							
Verification of eligibility criteria	X							
Demographic characteristics								
Age, gender	X							
Weight, height, BMI	X							
Smoking status	X							
Alcohol intake (excessive use M>3 glasses wine/day, F>2 glasses wine/day)	X							
socio-educational status lives alone (Y/N)	X							
Status with respect to hepatitis B Status with respect to hepatitis C Status with respect to tuberculosis	X							
Vaccine status - Pneumococcus - Herpes zoster History of chickenpox (Y/N)	X							
Biological parameters at inclusion								
Haemoglobin (g/dl)	X							
Mean corpuscular volume (μ^3)	X							
Lymphocyte absolute count (cells/mm^3)	X							
Neutrophil absolute count (cells/mm^3)	X							

	Inclusion visit Initiation of treatment (Target Day: 1)	Follow-up visit						Additional visit in case of discontinuation of Tofacitinib outside the 6 scheduled visits
		At 1 month (V1) (Target Day: 30)	At 3 months (V2) (Target Day: 91)	At 6 months (V3) (Target Day: 183)	At 12 months (V4) (Target Day: 365)	At 18 months (V5) (Target Day: 548)	At 24 months (V6) (Target Day: 731)	
Characteristics of RA								
Date of initial diagnosis	X							
Age at time of diagnosis of disease	X							
Erosions at time of diagnosis (Y/N)	X							
Rheumatoid nodules (previous or current): Y/N	X							
Secondary Sjögren Syndrome (Y/N)	X							
RF (negative, positive < 3x N, positive + ≥ 3x N)	X							
ACPA (negative, positive < 3x N, positive + ≥ 3x N)	X							
Previous treatments of RA (treatment line)	X							
If MTX as 1st line treatment (just before prescription of Tofacitinib), dosage and route of administration (PO /SC)	X							
Evaluation of RA								
Existence of comorbidities: - CV disease - Cancers - Infection - Gastro-intestinal disease - Osteoporosis - Depression - Diabetes	X							

	Inclusion visit Initiation of treatment (Target Day: 1)	Follow-up visit						Additional visit in case of discontinuation of Tofacitinib outside the 6 scheduled visits
		At 1 month (V1) (Target Day: 30)	At 3 months (V2) (Target Day: 91)	At 6 months (V3) (Target Day: 183)	At 12 months (V4) (Target Day: 365)	At 18 months (V5) (Target Day: 548)	At 24 months (V6) (Target Day: 731)	
Chronic low back pain (Y/N)	X							
Existence of fibromyalgia (FiRST questionnaire)	X	X	X	X	X	X	X	X
ESR (min/hour)	X	X	X	X	X	X	X	X
CRP (mg/mL)	X	X	X	X	X	X	X	X
PtGA (patient)	X	X	X	X	X	X	X	X
PhGA (physician)	X	X	X	X	X	X	X	X
Evaluation of pain by the patient	X	X	X	X	X	X	X	X
Duration of morning stiffness	X	X	X	X	X	X	X	X
TJC (28)	X	X	X	X	X	X	X	X
SJC (28)	X	X	X	X	X	X	X	X
PCS Questionnaire	X	X	X	X	X	X	X	X
CSQ Questionnaire	X	X	X	X	X	X	X	X
GIRERD Questionnaire	X	X	X	X	X	X	X	X
Euro QoL EQ-5D-3L Questionnaire	X	X	X	X	X	X	X	X
SF-12 Questionnaire	X	X	X	X	X	X	X	X
FACIT-fatigue Questionnaire	X	X	X	X	X	X	X	X
Administration of Tofacitinib								
Reason for initiation of Tofacitinib (problem of tolerability to previous treatment, primary failure (absence of significant improvement), secondary failure (significant improvement and then loss of efficacy), patient choice, other	X							
Dosage	X	X	X	X	X	X	X	X
Concomitant treatments								

	Inclusion visit Initiation of treatment (Target Day: 1)	Follow-up visit						Additional visit in case of discontinuation of Tofacitinib outside the 6 scheduled visits
		At 1 month (V1) (Target Day: 30)	At 3 months (V2) (Target Day: 91)	At 6 months (V3) (Target Day: 183)	At 12 months (V4) (Target Day: 365)	At 18 months (V5) (Target Day: 548)	At 24 months (V6) (Target Day: 731)	
Conventional synthetic DMARD: INN, dosage, date of start	X	X	X	X	X	X	X	X
Corticosteroid therapy: dose, p.o./bolus dose	X	X	X	X	X	X	X	X
Psychotropic treatment (antidepressant, BZD, BZD hypnotic)	X	X	X	X	X	X	X	X
Changes to treatment with Tofacitinib								
- Y/N - If yes: date of change, dosage, reasons for change		X	X	X	X	X	X	X
Discontinuation of treatment with Tofacitinib								
- Y/N, - If yes: date of discontinuation and reasons for discontinuation - Prescription of a biologic treatment or tsDMARD: Y/N. If yes: INN, dosage and date of start		X	X	X	X	X	X	X
Safety								
Collection of AEs		X	X	X	X	X	X	X

2.2 STUDY POPULATION

Patients eligible for this study can be included by the physician if they are starting a treatment with Tofacitinib for RA according to usual clinical practice and satisfy all of the criteria for inclusion. Patient treatment is independent from patient enrolment into the study.

*Data will be recorded during the 24 months participation of each patient.*Visits, target dates and windowing is described in Table 9, Section 11.1.

Patients will be followed prospectively and follow-up visits will be conducted after the initial consultations. No visit or additional test is required by the protocol, since the study is observational: the modalities for follow-up and treatment will be left up to the sole judgement of the participating doctor.

In case of discontinuation of treatment with Tofacitinib, for whatever reason, clinical data will be collected as close as possible to the date of such a discontinuation. Likewise, in case of Tofacitinib discontinuation outside the 6 scheduled visits, an additional visit to collect clinical data as close as possible to the date of this discontinuation should be organised.

Eligible patients will be followed starting from the date of the first prescription of Tofacitinib during 24 months, whether they continue treatment with Tofacitinib or not.

In order to ensure expected recruitment, the study will be conducted by about 100 rheumatologists. Participating physicians will be selected from among those all over France out of the total number of French health institutions which manage patients with rheumatoid arthritis.

In order to reach an objective of participation of 100 centres desired, the study will be offered to about 550 centres.

2.3 TREATMENT LABELS

The treatment labels for analyses are as follows:

Treatment	**Treatment Label**
Tofacitinib as monotherapy	Tofacitinib as Monotherapy
Tofacitinib as combination (with MTX or other csDMARDs)	Tofacitinib as Combination
Tofacitinib in total	Total

Each treatment group (Tofacitinib as Monotherapy or Tofacitinib as Combination) will correspond to patients who received a given treatment for at least 80% of their follow-up. The total group will include all patients enrolled in this study and took at least one dose of Tofacitinib. Note: due to the criteria for each treatment group the total column may not

equal the number in Tofacitinib as Monotherapy plus the number in Tofacitinib as Combination.

2.4 STUDY OBJECTIVES

The primary objective of this study is to identify factors predictive of drug survival with Tofacitinib in RA patients. To assess this the time (in days) which a patient is being treated with Tofacitinib, up to permanent discontinuation, will be modelled using Cox proportional hazard (PH) models and predictive factors will be investigated to assess their impact on the models.

The secondary objectives are:

- *To evaluate the impact of behavioural strategies in patients with rheumatoid arthritis on Tofacitinib drug survival* by including the PCS (total score) and CSQ (5 domain scores) in a Cox PH model
- *To study the correlation* (by considering relevant correlation coefficients) *between behavioural strategies and effectiveness*
- *To evaluate the effectiveness of Tofacitinib*
- *To evaluate the tolerability of Tofacitinib*

Unless specified otherwise, secondary endpoints will be assessed using summaries only.

3 INTERIM ANALYSES

Interim analyses will be planned each year. These interim analyses will be performed on the available data. They will be purely descriptive and will include the analyses planned in this SAP except multivariate analyses and correlation assessments. Outputs required will be highlighted in the study List of Tables.

4 HYPOTHESES AND DECISION RULES

The primary and secondary analyses will use predictive models, described in Section 8.2.3 and Section 8.2.5.1, to assess Tofacitinib drug survival. All other summaries will analyse data descriptively.

There are no formal statistical hypotheses for this study.

5 ANALYSIS SETS

5.1 FULL ANALYSIS SET

The Full Analysis Set (FAS) will include patients who meet the eligibility criteria (See section 9.2 in the protocol), receive at least one dose of Tofacitinib and have at least one set of post-baseline measurements.

All primary and secondary analyses will use FAS.

5.2 SAFETY ANALYSIS SET

The Safety Analysis Set will include all patients who have taken at least one dose of Tofacitinib.

All safety analyses will use the Safety Analysis Set.

5.3 MONTH 3 ANALYSIS SET

The Month 3 Analysis Set will include patients who receive at least one dose of Tofacitinib, have at least one set of post-baseline measurements and who have at least one non-missing month 3 measurement.

The sensitivity analyses where DAS28-4 CRP measured at month 3 is included as a potential factor will use the Month 3 analysis set.

5.4 SUBGROUPS

The relationship between behavioural strategies and the following adverse event subgroups will be investigated

- Patients with and without Herpes Zoster
- Patients with and without Serious Infection (Y/N)

Subgroup analyses will be performed by sex and according to patient age:

- < 65 years
- ≥ 65 years

6 ENDPOINTS AND COVARIATES

6.1 EFFICACY/EFFECTIVENESS ENDPOINT(S)

All study data will be collected by the participating physician directly in the eCRF.

Baseline measurements for all endpoints are the last measurement before Tofacitinib treatment, up to 28 days before starting Tofacitinib. Change from baseline at a timepoint will be calculated as:

Measurement at visit – Measurement at baseline

6.1.1 Efficacy/Effectiveness Endpoints

6.1.1.1 Primary Endpoint

The primary endpoint of this study is the duration of Tofacitinib drug survival. The duration of Tofacitinib drug survival (in days) is calculated as:

Date of permanent drug discontinuation – Date of first taking the drug + 1

If a patient does not present with the event of interest, that is they do not have a permanent drug discontinuation record during the study, then they will be censored at the time of their last visit.

6.1.1.2 Secondary Endpoints

There are multiple secondary endpoints in this study, described in the following subsections.

For the secondary endpoints which consider duration of time on the study, duration of time on the study will be calculated as:

Date of Last Visit – Date of Baseline Visit + 1

6.1.1.2.1 PCS – Pain Catastrophising Scale

The Pain Catastrophising Scale (PCS) is a 13 item questionnaire around thoughts and emotions experienced during pain, with each question on a 5 point scale (0 = 'Not at all' to 4 = 'All the time'). There are three subscales of PCS: Rumination (thinking deeply), exaggeration and helplessness. Table 2 shows an overview of the items which belong to each domain.

Table 2 Overview of Domains for PCS (5)

Domain	Items
Rumination	8, 9, 10, 11
Exaggeration	6, 7, 13
Helplessness	1, 2, 3, 4, 5, 12

Total scores and domain scores are derived by summing the scores of the individual items. The highest possible score of PCS is 52, and a total score of ≥ 20 indicates a clinically relevant level of catastrophisation. PCS will be considered both in terms of individual domain scores and total scores. PCS total score will be analysed as a continuous variable as well as a dichotomous variable (<20 *vs* ≥ 20).

For the handling of missing data when deriving PCS scores, see Section 7.4.

PCS is collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.2 CSQ – Coping Strategies Questionnaire

The Coping Strategies Questionnaire (CSQ) evaluates cognitive coping and, in its French version, is split into 5 domains: "Distraction", "Catastrophising" , "Distancing from Pain", "Ignoring Pain Sensations" and "Praying". The CSQ consists of 21 items in total and for each item, subjects rate the frequency of their use of each coping strategy on a 4-point Likert-type scale. CSQ is collected on the CRF as the following: 1: Never, 2: Sometimes, 3: Often and 4: Very often. In order to consider values comparable to the standard CSQ French Version, the following mapping will be done.

CSQ Text Value	CRF Value	Map to standard value
Never	1	0
Sometimes	2	1
Often	3	2
Very Often	4	3

This endpoint is considered in terms of the 5 domains, and a total score is not considered for CSQ. Domain scores are derived by summing the individual items scores that make-up the domain. Details of the 5 domains are provided below.

Domain	Items as numbered on CRF	Domain Score Range
Distraction	2, 10, 11, 19, 20	0 – 15
Catastrophising	3, 9, 15, 18	0 -12
Distancing from Pain	1, 5, 13, 21	0 – 12
Ignoring Pain Sensations	6, 7, 8, 14, 16	0 – 15
Praying	4, 12, 17	0 - 9

CSQ is collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

For the handling of missing CSQ data, see Section 7.5.

6.1.1.2.3 LDA

Low Disease Activity (LDA) is a binary endpoint and can be assessed in four different ways using DAS28-4 ESR, DAS28-4 CRP, SDAI and CDAI. The individual criteria for the four different ways of measuring LDA are:

- DAS28-4 ESR ≤ 3.2
- DAS28-4 CRP ≤ 3.2
- SDAI ≤ 11
- CDAI ≤ 10

Each criteria will be analysed separately.

All of the LDA factor individual components are collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.4 DAS28-4 ESR

The 28-joint Disease Activity Score erthrocyte sedimentation rate (DAS28-4 ESR) is a composite endpoint, calculated using 4 variables (represented by '-4' in the name). The calculation of DAS28-4 ESR is as follows:

$$0.56 * \text{sqrt}(TJC) + 0.28 * \text{sqrt}(SJC) + 0.70 * \ln(ESR) + 0.014*PtGA$$

With ESR in mm/h, ln = nautral logarithm and PtGA is patient global assessment in mm.

DAS28-4 ESR contributes to the assessment of patient LDA and patient remission, and is also summarised in it's own right.

The individual components of DAS28-4 ESR are collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.5 DAS28-4 CRP

The 28-joint Disease Activity Score C Reactive Protein with 4 variables (DAS28-4 CRP) is, similarly to DAS28-4 ESR, a composite endpoint calculated using 4 variables. However, instead of ESR, CRP is used and the calculation varies slightly. The calculation of DAS28-4 CRP is as follows:

$$0.56 * \text{sqrt}(TJC) + 0.28 * \text{sqrt}(SJC) + 0.36 * \ln(CRP+1) + 0.014*PtGA + 0.96$$

With CRP in mg/l, ln = nautral logarithm and PtGA in mm.

DAS28-4 CRP contributes to the assessment of patient LDA and patient remission, and is also summarised in it's own right.

The individual components of DAS28-4 CRP are collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.6 SDAI

The Simplified Disease Activity Index (SDAI) is a composite end point, calculated as

TJC + SJC + PtGA + PhGA + CRP

Where TJC is tender joint count (28), SJC is swollen joint count (28), PtGA is patient global assessment in cm and PhGA is physician global assessment in cm, and CRP is C-reactive protein in mg/dL.

The individual components of SDAI are collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.7 CDAI

The Clinical Disease Activity Index (CDAI) is a composite end point, calculated as

TJC + SJC + PtGA + PhGA

Where TJC is tender joint count (28), SJC is swollen joint count (28), PtGA is patient global assessment in cm and PhGA is physician global assessment in cm.

The individual components of CDAI are collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.8 Patient Remission

Patient remission will be calculated using 5 separate definitions:

- DAS28-4 ESR < 2.6
- DAS28-4 CRP < 2.6
- SDAI ≤ 3.3
- CDAI ≤ 2.8
- ACR-EULAR 2011 Boolean criteria should be met

The ACR-EULAR 2011 Boolean criteria is defined in more detail below, in Section 6.1.1.2.9.

6.1.1.2.9 ACR-EULAR 2011 Boolean Criteria

The America College of Rheumatology-European League Against Rheumatism (ACR-EULAR) 2001 Boolean criteria is derived as:

ACR_Remission = 1, if SJC ≤ 1, TJC ≤ 1, CRP ≤ 1 mg/dL, and PtGA ≤ 1cm

ACR_Remission = 0, Otherwise.

Where CRP is measured in mg/dL.

The proportion of patients which meet the ACR-EULAR 2011 Boolean criteria at Visit X is then calculated as:

> No. of patients with ACR-EULAR met at Visit X / Total Number of patients with non-missing ACR-Eular at Month X.

The individual factors of ACR-EULAR 2011 Boolean Criteria are collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.10 DAS28 (EULAR Criteria)

The DAS28 EULAR criteria to evaluate patient response to treatment is categorised by good responder patient, moderate responder patient, non-responder patient and non-evaluable patient based on patients DAS28-4 CRP score and the decrease in DAS28-4 CRP compared to Visit 0. Table 3 below gives further detail of the categorisation.

Table 3 Evaluation of patient response to treatment using DAS28 (EULAR Criteria)

DAS28 at visit	Decrease in DAS28 compared to V0		
	>1.2	>0.6 and ≤1.2	≤0.6
≤3.2	Good responder	Moderate responder	Non-responder
>3.2 and ≤5.1	Moderate responder	Moderate responder	Non-responder
>5.1	Moderate responder	Non-responder	Non-responder

The DAS-28 EULAR Criteria will also be derived based on the DAS28-4 ESR score using the same approach as above.

The individual components of DAS28-4 CRP and DAS28-4 ESR, which are the composite endpoints used in this criteria, are collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.11 TJC

Tender Joint Count (TJC) (28) is the count of the total number of tender joints out of a possible 28 joints which are checked at the visit.

TJC is collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.12 SJC

Swollen Joint Count (SJC) (28) is the count of the total number of swollen joints out of a possible 28 joints which are checked at the visit.

SJC 28 is collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.13 Morning Stiffness

The duration of morning stiffness is measured in minutes.

Duration of morning stiffness is collected at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.14 EuroQoL EQ-5D-3L

The European Quality of Life (Euro-QoL) developed the EQ-5D-3L questionnaire. It is a standardised, self-reported instrument used to evaluate health-related quality of life. It is copyrighted and consists of 5 domains of questions:

1. Mobility
2. Self-care
3. Usual activities
4. Pain/discomfort
5. Anxiety/depression

Each question domain has 3 levels, of which a patient will choose the most appropriate level for each, see Table 4.

Table 4 EQ-5D-3L Scores

Dimension	Response	Score
Mobility	I have no problems in walking about	1
	I have some problems in walking about	2
	I am confined to bed	3
Self-care	I have no problems with self-care	1
	I have some problems washing or dressing myself	2
	I am unable to wash or dress myself	3
Usual activities (e.g. work, study, housework, family or leisure activities)	I have no problems with performing my usual activities	1
	I have some problems with performing my usual activities	2
	I am unable to perform my usual activities	3
Pain/discomfort	I have no pain or discomfort	1
	I have moderate pain or discomfort	2

	I have extreme pain or discomfort	3
Anxiety/depression	I am not anxious or depressed	1
	I am moderately anxious or depressed	2
	I am extremely anxious or depressed	3

The score for each dimension is weighted in accordance with Table 5.

Table 5 EQ-5D-3L Weightings (10)

EQ-5D Dimension	Score = 1	Score = 2	Score = 3
Mobility	0	0.155	0.372
Self-Care	0	0.212	0.326
Usual Activities	0	0.156	0.189
Pain/Discomfort	0	0.112	0.265
Anxiety/Depression	0	0.090	0.204

The following algorithm is then applied to calculate the EQ-5D Total Score:

1. If all five EQ-5D dimensions have a score of 1 then the EQ-5D Total Score is 1.

2. If any of the five EQ-5D dimensions have a score of 3, then the EQ-5D Total Score is:

$$1 - \left(\sum_{1}^{5} weighted\ dimension\ score \right) - 0.174$$

3. If none of the five EQ-5D dimensions has a score of 3, then the EQ-5D Total Score is:

$$1 - \left(\sum_{1}^{5} weighted\ dimension\ score \right)$$

For the handling of missing weighted dimension scores, see Section 7.6.

The Euro QoL EQ-5D-3L Questionnaire is completed at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.15 SF-12v2 Questionnaire

The short form-12 version 2 (SF-12v2) health survey consists of 12 questions, which evaluate 8 aspects of quality of life (QoL). All items, domain scores and summary measures are scored so that a higher score indicates a better health state. The scores will

be derived based on the guidelines provided in User's Manual for the SF-12v2 Health Survey,Third Edition (4).

Table 6 shows an overview of the SF-12v2 questionnaire domains.

Table 6 Overview of Domains for SF-12v2 questionnaire

Domain	Item numbers
Physical Functioning (PF)	2a, 2b
Role-Physical (RP)	3a, 3b
Bodily Pain (BP)	5
General Health (GH)	1
Vitality (VT)	6b
Social Functioning (SF)	7
Role-Emotional (RE)	4a, 4b
Mental Health (MH)	6a, 6c

SF-12v12 will be considered in terms of the physical and mental component summary measures (PCS_{SF12} and MCS) as well as the 8 health domains seen in Table 6.

The SF-12v2 Questionnaire is completed at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.16 FACIT-Fatigue Questionnaire

The Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) questionnaire consists of 13 self evaluated questions. The questions consider the patients functionality over the 7 days prior to completing the questionnaire. Each question has a response of values 0 to 4. The FACIT-Fatigue total score is derived by summing the response to each question which gives a value between 0 and 52. Higher scores represent better patient status (less fatigue).

For some of the questions, the scale indicates a higher score is better and for others a lower score is better. Therefore, in order to correct this and ensure that higher scores represent better patient status, the items marked with a '-' in the

Table 7 will be reversed by subtracting the response from '4' before summing all subscale items.

Table 7 FACIT-Fatigue Item subscale

Item No.	Question	+/-
HI7	I feel fatigued	-
HI12	I feel weak all over	-
An1	I feel listless ("washed out")	-
An2	I feel tired	-
An3	I have trouble starting things because I am tired	-

An4	I have trouble finishing things because I am tired	-
An5	I have energy	+
An7	I am able to do my usual activities	+
An8	I need to sleep during the day	-
An12	I am too tired to eat	-
An14	I need help doing my usual activities	-
An15	I am frustrated by being too tired to do the things I want to do	-
An16	I have to limit my social activity because I am tired	-

For handling of missing data, see Section 0.

The FACIT-Fatigue Questionnaire is completed at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.17 PtGA

Patient Global Assessment (PtGA) is measured using a visual analogue scale (VAS) on a 0-100 mm scale, where 0 mm = no symptoms of disease and 100 mm = maximum seriousness which a patient can imagine.

PtGA is measured at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.18 PhGA

Physician Global Assessment (PhGA) is measured using a VAS on a 0-100 mm scale, where 0 = no disease activity and 100 = maximum disease activity.

PhGA is measured at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.1.2.19 Pain

Pain is measured by the patient using a VAS on a 0-100 mm scale, where 0 = no pain and 100 = maximum pain.
Pain is measured at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.2 Other Endpoints

6.1.2.1 FiRST Questionnaire

Fibromyalgia Rapid Screening Tool (FiRST), used to detect fibromyalgia in patients with diffuse rheumatic pain, is a questionnaire which contains 6 items:

1. Diffuse pain
2. Painful symptoms
3. Fatigue
4. Sleep and cognitive disorders
5. Nonpainful abnormal sensations
6. Functional somatic symptoms

Each of the 6 items contain a yes or no response, and a positive answer to 5 out of the 6 answers can detect fibromyalgia.

The FiRST Questionnaire is completed at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.1.2.2 GIRERD Questionnaire

The GIRERD Questionnaire is used to evaluate treatment compliance, and is made up of 6 yes or no questions, such that:

- *If the patient responds "no" to all questions, he is considered as a good complier.*
- *If the patient responds "yes" once or twice, he is considered as a minor complier.*
- *If the patient responds "yes" three times or more, he is considered as a non-complier.*

The GIRERD Questionnaire is completed at all visits (Inclusion visit, V1, V2, V3, V4, V5, V6 and any additional visits in the case of discontinuation of Tofacitinib).

6.2 SAFETY ENDPOINTS

The reporting of safety data will be in accordance with Pfizer Data Standards.
In addition, endpoints involving the occurrence of the following adverse events of special interest will also be investigated:

1. Herpes Zoster.
2. Serious Infections.

A serious infection is any infection (viral, bacterial, and fungal) that requires hospitalization for treatment or requires parenteral antimicrobial therapy or meets other criteria that require it to be classified as a serious adverse event (results in death, is life-threatening, causes substantial disability or incapity, or results in a congenital anomaly or birth defect).

The number of AE's adjusted by a patients duration on the study, where duration of time on the study is calculated as per Section **Error! Reference source not found.**, will be calculated as:

$$\frac{\text{Number of Adverse Events}}{\text{Duration of time on study}}$$

7 HANDLING OF MISSING VALUES

Missing data indicators method (see Section 7.1) will be used for the primary analysis and sensitivity analyses, where specified. If there are more than 10% missing data in baseline covariates then a sensivity analysis using complete case analysis will also be produced. The sensitivity analysis will check for bias using the missing data indicators methods, which is understood to exhibit bias in the occurrence of a large proportion of missing data. (1) If a large difference is seen between the primary analysis and the sensitivity analysis then other missing data methods may be used.

Furthermore, random survival forests methods will be produced as an additional sensitivity analysis to complement the primary model. Missing data methods for random survival forests are described below in Section 7.3.

Missing data methods for questionnaire data are described in Sections 7.4 to Section 0.

7.1 MISSING DATA INDICATORS

Missing data indicators is not a form of imputing, and is used to check if data is missing at random (2). For each factor with missing data in the primary analysis, an indicator variable will be created as

1 = Missing
0 = Otherwise

The indicator variables will be added to the multivariate Cox PH model, following backwards selection, described in Section 8.1.1.1 Cox Proportional Hazards. The model will have a missing data indicator for each variable in the model with missing data.

7.2 COMPLETE CASE ANALYSIS

If there is more than 10% missing data then complete analysis will be used as a sensitivity analysis. For a complete case analysis, if a patient has missing data for any factor within the model then this patient will be removed from the analysis.

7.3 RANDOM SURVIVAL FORESTS MISSING DATA

As part of the sensitivity analyses random survival forests will be produced, described in Section 8.1.1.2 Random Survival Forests, which will use the R package 'randomForestSRC'. For our predictive model, a call to rfsrc() will be made, and in the presence of missing data the option 'n.action = "na.impute"' will be used to impute missing data. The missing data algorithm which will be envoked can be summarized as (6):

- Variables with missing data are imputed by randomly drawing from the in-bag non-missing data. Split statistics can be calculated from the imputed data so that cases are assigned to daughter nodes.
- Following the node split, imputed data is set back to missing.
- Repeat the first two steps until the terminal node is reached. Missing data within the terminal nodes are imputed using in-bag, non-missing terminal node data.
- Records where all outcome and variable information is missing are removed. Variables which have only missing values are removed.

7.4 PCS MISSING DATA

No imputation will be used for missing PCS data.

7.5 CSQ MISSING DATA

No imputation will be used for missing CSQ data. In the event of missing data the corresponding domain score will not be calculated.

7.6 EQ-5D-3L MISSING DATA

Missing weighted dimension scores are replaced by the mean of the non-missing weighted dimension scores. If a weighted score is missing and replaced in this way by a mean weighted score of zero, step 3 of the algorithm described in Section 6.1.1.2.14 is applied.

7.7 SF-12V2 QUESTIONNAIRE MISSING DATA

Rules suggested by the producer of this instrument will be followed in calculating the values. If these rules are not enough to impute a value, and a patient still has a missing value at a visit, the missing values will be assumed to be missing at random.

7.8 FACIT-FATIGUE QUESTIONNAIRE MISSING DATA

If any items are not answered the following approach will be taken:

If more than 50% of the items are non-missing, a prorated subscale will be derived as

$$\text{Prorated Subscale Score} = \frac{\text{Sum of indivudal answered items} * 13}{\text{Number of answered questions}}$$

If 50% or more of the items are missing the score will be set to missing.

8 STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1 STATISTICAL METHODS

Visits, defined by Table 9 of Section 11.1, will be used for the analyses.

8.1.1 Analysis of Time to Event Data

8.1.1.1 Cox Proportional Hazards

Cox proportional hazards (PH) models will be used to analyse time to event data, where the event of interest is duration of Tofacitinib drug survival described in Section **Error! Reference source not found.**. If a patient is not recorded as having permanently discontinued Tofacitinib, which is the event of interest, then they will have their time of follow-up censored at the date of their last visit.

For the primary analysis, univariate Cox PH models will be created for all of the potential predictive factors, see Figure 2 in Section 8.2.3, that is a separate Cox PH model will be created for each factor. If the variable significantly impacts Tofacitinib drug survival at the 10% level then the factor will be put forward into a multivariate model.

Once the univariate model selection is complete, backwards selection will be used to remove factors from the multivariate model which do not have a significant p value at the 5% level. The covariate with the highest p value will be removed in turn until all factors have a p value less than 0.05. If a significant interaction is found then the individual factors of this interaction will be kept in the model, regardless of their p value. A significant interaction is tested for between all variables in the multivariate model.

Furthermore, for the primary analysis a table showing all of the 95% confidence intervals and p values for the “raw” hazard ratios coming from the univariate models will be created. This will be for all of the factors in Figure 2, regardless of whether they’re included in the multivariate model.

For the secondary analysis, a multivariate Cox PH model will be created including the scores from the 5 domains of the CSQ at baseline and the total PCS score at baseline included as covariates.

A further two Cox PH models will be produced for the secondary analysis, for CSQ and PCS: the first including the last evaluation available before discontinuation (or last visit if

patient does not discontinue during the study) and the second including change from baseline to the last recorded measurement.

The SAS procedure PROC PHREG will be used to fit any Cox PH models and check the assumption of proportional hazards. If the assumption of proportional hazards is violated then alternative models will be considered.

Hazard ratios (HRs) and the 95% confidence intervals (CIs) will be created and displayed, along with the associated p value from the models. All statistical tests will be two-sided with alpha of 0.05, apart from during variable selection using univariate models for the primary analysis, where alpha will be considered as 0.1. Two-sided CI's will be presented whenever it is appropriate. No adjustment for multiplicity will be used.

Drug survival with Tofacitinib will be described with the Kaplan-Meier (KM) method, using the SAS procedure PROC LIFETEST. Confidence intervals will be displayed, using the Klein and Moeschberger method. An estimate for drug survival at 12 and 24 months will be presented. A KM plot will be created to display the survival function of the Tofacitinib Total group. As a sensitivity comparison a KM plot will also be presented showing the Tofacitinib as Monotherapy group and the Tofacitinib as Combination Therapy group.

8.1.1.2 Random Survival Forests

Random survival forests will produced as part of the sensitivity analyses, and will analyse time to Tofacitinib discontinuation. It will consider all of the potential factors seen in Figure 2 in Section 8.2.3. The program R will be used for this, using the package 'randomForestSRC'. For survival data, this package uses splits with either the default rule as log-rank splitting (9, 7) or log-rank score splitting(8). We will primarily use the default but a further sensitivity analysis, will also be performed using the log-rank score splitting and compared with the results when using log-rank splitting.

Suggested steps of random survival forests are as follows (6):

1. Draw ntree (1500 in our case) bootstrap samples from the original data. (Each bootstrap excludes a percentage of the data called out-of-bag (OOB) data which will be used for testing accuracy)

2. Grow a tree for each bootstrapped data set. At each node of the tree randomly select mtry predictors (covariates) for splitting on. Split on a predictor using a survival splitting criterion. A node is split on that predictor which maximizes survival differences across daughter nodes. We will use cross validation to determine the best mtry.

3. Grow the tree to full size under the constraint that a terminal node should have no less than nodesize unique failures.

4. Calculate an ensemble cumulative hazard estimate by combining information from the ntree trees. One estimate for each individual in the data is calculated.

5. Compute an OOB error rate for the ensemble derived using the first b trees, where b = 1, . . . , ntree. Estimate the prediction error for the ensemble cumulative hazard function.

Variable importance will also be compiled and presented in an output.

Ensemble estimation
The randomForestSRC package produces an ensemble estimate for the cumulative hazard function and this will be presented in a table. Error rate performance is calculated based on this value, which will also be presented.

8.1.2 Correlation Analysis

The correlation between behavioural strategies (measured by PCS total score and 5 domains of the CSQ) *and effectiveness will be studied with Pearson's or Spearman's coefficient according to the nature of variables compared.* This will be calculated for all effectiveness endpoints. To investigate the relationship between a categorical variable and a continuous variable, boxplots for each category will be produced and a visual comparison will be made (there will be no correlation co-efficient for these comparisons).

8.1.3 Logistic Regression

Logistic regression will be used to analyse the relationship between binary endpoints and behavioural strategies. The binary endpoints will be the response variables at Months 3, 6, 12 and 24 and PSC total score and the 5 domains of the CSQ will be fitted as a continuous, independent variables in separate univariate models. Estimates for the odds ratio's (OR's), confidence intervals (CI's) and p-values will be presented.

Reference code for logistic regression:

```
ODS OUTPUT  oddsratios = _out1 parameterestimates = _out2;
PROC LOGISTIC DATA = <data>;
MODEL <binary endpoint at visit> = <behavioural endpoint at
visit>;
RUN;
ODS OUTPUT CLOSE;
```

Where ODS OUTPUT option odds ratios selects the OR's and CI's, and parameter estimates can be used to select the p-value.

The association between baseline PCS total score (<20 *vs* ≥ 20) and baseline characteristics will also be explored using a logistic regression. A univariate analysis will first be performed with a significance level set at 10%. All significant characteristics will be put forward into a multivariate model. A backward selection will be used to remove

characteristics from the multivariate model which do not have a significant p value at the 5% level.

8.1.4 Analysis of Continuous Data

Continuous data will be summarised in terms of mean, standard deviation (SD), median, first and third quartiles, and minimum and maximum for each visit. Additionally the number of patients with evaluable records for each visit will be presented. Change from baseline tables will be produced for continuous data for each post-baseline timepoint, and will be calculated as per Section 6.1.

8.1.5 Analysis of Categorical Data

Categorical data will be presented as the number of patients and the percentage of patients within a group, based on the number of patients with non-missing data at each visit. These will be presented for each visit.

8.1.6 Analysis of Binary Endpoints

Binary data will be presented using counts and percentages for each visit. Percentages will be calculated based on the number of subjects with a non-missing data at the visit.

8.1.7 Area Under the Curve

For PCS total score, the 3 domains of PCS and the 5 domains of the CSQ, the Area under the curve (AUC), for each score over time, will be calculated from baseline to last follow-up visit.

As patients will not be on the study for the same length of time, AUC will then be adjusted by the duration which each patient is in the trial, where duration is calculated as per Section **Error! Reference source not found.**.

Thus, adjusted AUC will be calculated as:

$$\frac{\text{Area under the Curve}}{\text{Duration of time on study}}$$

8.2 STATISTICAL ANALYSES

All analyses will be performed in the statistical software SAS 9.3 (or higher), unless otherwise specified.

8.2.1 Background, Demographic Characteristics and Study Disposition

The demographic and baseline summaries will use the Full Analysis Set (FAS). Continuous variables will be summarised as per Section **Error! Reference source not**

found., the categorical variables will be summarised as per Section 8.1.5 and the binary variables will be summarised as per Section **Error! Reference source not found.**.

These summaries will include age, sex, height, weight, smoking status, severity of disease, duration of disease, anti-citrullinated protein antibodies (ACPA), rheumatoid factor (RF), vaccination status, presence of poor prognostic factors, comorbidities, co-medication and concomitant medication.

The number and percentage of patients at each time point, described in Table 9 in Section 11.1, will be provided along with reasons for withdrawal from the study. For patients who have discontinued Tofacitinib treatment but remain in the study information on the treatments prescribed following study treatment (Tofacitinib) will be listed.

8.2.2 Safety Analyses

All safety analyses will use the Safety Analysis Set. All safety analyses will present a single Tofacitinib Total column using the treatment label seen in Section 2.3.

8.2.2.1 Adverse Events

Adverse events (AE's) *will be summarized according to Pfizer standards,* presenting AE's, Serious AE's and non-serious AE's.

Two sets of AE's will be summarised as 'All Casuality' and 'Treatment Emergent'. For the treatment emergence table, a drug lag of 28 days will be applied following cessation of Tofacitinib.

The AE listings will list the individual drugs that were beging taken at the time of the AE.

Adverse events which occurred under another DMARD, following cessation of Tofacitinib, will be presented in separate listing. Note: A drug lag of 0 days will be applied following the cessation of Tofacitinib, for AE's occurring under another DMARD, e.g. if an AE is within 28 days of Tofacitinib and started after the cessation of Tofacitinib then it will be seen within both listings (for Treatment-Emergent and occurred under another DMARD following cessation of Tofacitinib).

8.2.2.2 Exposure Data

Exposure data will be summarised including dose, dose change and reason for change.

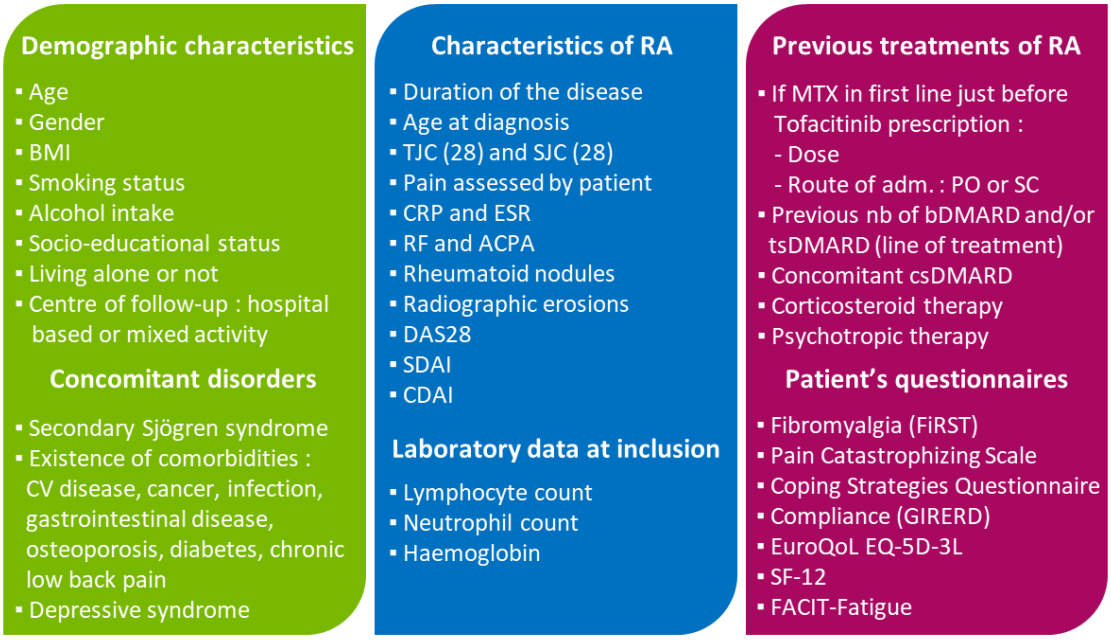
8.2.3 Primary Analyses

The primary endpoint, duration of Tofacitinib drug survival, will be analysed using the FAS and the methods described in Section 8.1.1.1 Cox Proportional Hazards. The Tofacitinib drug survival tables will present a single Tofacitinib Total column, using the treatment label seen in Section 2.3. A Kaplan-Meier plot will present the survival function for Tofacitinib Total group. As a comparison, a second KM plot will present the

survival functions for both Tofacitinib as Monotherapy and Tofacitinib as Combination groups. The treatment labels seen in Section 2.3 will be used.

The list of potential predictive factors to include in the model for the primary analysis is seen in Figure 2. Note: the factors seen in Figure 2 are all baseline predictors. Where SF-12 is considered within Figure 2, the 8 health domains as well as PCS_{SF12} and MCS will each be considered as potential predictive factors. As PCS_{SF12} and MCS are linear combinations of the 8 health domains, collinearities are anticipated, however considering all of them as potential predictors will allow us to investigate which domains contribute the most to PCS_{SF12} and/or MCS. If necessary, transitional models will be investigated including separately component summary measures (PCS_{SF12} and MCS) and the 8 health domains.

Figure 2: Potential predictive factors of drug survival



In the event of missing data, see Section 7.1. Missing data methods will be performed on the final multivariate Cox PH model, and not for the individual univariate models.

8.2.4 Sensitivity Analyses

Sensitivity analyses, except where specified, will use the FAS. All sensitivity analyses tables will present a single Tofacitinib Total column, using the treatment label seen in Section 2.3. All sensitivity analyses plots will present Tofacitinib as Monotherapy and Tofacitinib as Combination, using the treatment labels seen in Section 2.3.

Complete Case Analysis

In the event of more than 10% missing data in baseline covariates then a complete case analysis will take place, using the methods described in Section 7.2.

Random Survival Forests

Random survival forests will be created as a sensitivity analysis to complement the standard Cox PH model, as described in Section 8.1.1.2 Random Survival Forests, and will consider the potential predictive factors seen in Figure 2 in Section 8.2.3. In the event of missing data, see Section 7.3.

DAS28-4 CRP at Month 3 as a Potential Predictive Factor

8.2.5 As part of the sensitivity analyses, DAS28-4 CRP at month 3 will be considered as an additional potential predictive factor along with the factors seen in Figure 2 in Section 8.2.3, and will follow the methods described in Section Analysis of Time to Event Data

8.1.1.1 Cox Proportional Hazards. This will use the Month 3 Analysis Set. In the event of missing data, follow Section 7.1.

The methods for random survival forests, described in Section 8.1.1.2 Random Survival Forests, will also be repeated including DAS28-4 CRP at month 3 as a potential predictive factor, along with the factors in Figure 2 in Section 8.2.3. This will use the Month 3 Analysis Set. In the event of missing data, follow Section 7.3.

8.2.6 Secondary Analyses

All secondary analyses will use the FAS. Secondary analyses will consider Tofacibitab Total as the treatment group. Summaries of secondary endpoints will be presented for Tofacitinib total, Tofacitinib as monotherapy and Tofacitinib as combination, using the treatment labels seen in Section 2.3.

Each endpoint will apply the specific rules for missing data as described in Section 7 but if there are missing data points after this, no missing data methods will be used to impute data for the secondary analyses.

All the analyses described hereafter in this section will be performed on all assessments collected before permanent Tofacitinib discontinuation (On-Treatment Analysis). In addition, analyses described in sub-sections 8.2.6.2, 8.2.6.3, 8.2.6.4 and 8.2.6.5 will be re-run on all available data including retrieved dropout data (Analysis on All Available Data).

8.2.6.1 Behavioural Strategies – PCS and CSQ

The impact of behavioural strategies (PCS total score and 5 domains of the CSQ) on Tofacitinib drug survival will be analysed, using three multivariate Cox PH models as described in Section 8.1.1.1 Cox Proportional Hazards.

PCS (individual domain scores and total scores) and CSQ (individual domain scores) will also be summarised as continuous variables as per Section **Error! Reference source not found.**. PCS total score will also be summarised as a dichotomous variable (<20 *vs* ≥ 20) as per Section **Error! Reference source not found.**.

Correlations between the behavioural strategies (PCS total score and 5 domains of the CSQ) and each of the endpoints listed below will be investigated as outlined in Section **Error! Reference source not found.**, for the endpoints seen in Figure 3:

- DAS28-4 ESR
- DAS28-4 CRP
- TJC
- SJC
- Duration of morning stiffness
- GIRERD Questionnaire
- EuroQoL EQ-5D-3L
- SF-12v2 Questionnaire
- FACIT-Fatigue Questionnaire
- DAS28 (EULAR Criteria)
- Pain

Figure 3 Timepoints for correlation analyses

Behavioural Strategies Timepoint	Effectiveness Endpoint Timepoint
M0 (Baseline)	M3
M0 (Baseline)	CFB to M3 (M3-M0)
M3	M3
M3	CFB to M3 (M3-M0)
CFB to M3 (M3-M0)	M3
CFB to M3 (M3-M0)	CFB to M3 (M3-M0)
M0 (Baseline)	M6
M0 (Baseline)	CFB to M6 (M6-M0)
M6	M6
M6	CFB to M6 (M6-M0)
CFB to M6 (M6-M0)	M6
CFB to M6 (M6-M0)	CFB to M6 (M6-M0)
M0 (Baseline)	M12
M0 (Baseline)	CFB to M12 (M12-M0)
M12	M12
M12	CFB to M12 (M12-M0)
CFB to M12 (M12-M0)	M12
CFB to M12 (M12-M0)	CFB to M12 (M12-M0)
M0 (Baseline)	M24
M0 (Baseline)	CFB to M24 (M24-M0)
M24	M24
M24	CFB to M24 (M24-M0)
CFB to M24 (M24-M0)	M24
CFB to M24 (M24-M0)	CFB to M24 (M24-M0)

M = Month. CFB = Change from baseline.

Logistic regression will be considered to assess the relationship between behavioural strategies (PCS total score and 5 domains of the CSQ) and each of the binary endpoints listed below as outlined in Section **Error! Reference source not found.**, with a separate model created for each of the timepoints seen in Figure 4:

- LDA
- Patients in remission
- FiRST Questionnaire

Figure 4 Timepoints for logistic regression

Behavioural Strategies Timepoint	Effectiveness Endpoint Timepoint
M0 (Baseline)	M3
M3	M3
CFB to M3 (M3–M0)	M3
M0 (Baseline)	M6
M6	M6
CFB to M6 (M6–M0)	M6
M0 (Baseline)	M12
M12	M12
CFB to M12 (M12–M0)	M12
M0 (Baseline)	M24
M24	M24
CFB to M24 (M24-M0)	M24

M = Month. CFB = Change from baseline.

Further summaries will be produced to investigate the relationship between behavioural strategies and the adverse event (AE) groups Herpes Zoster and Serious Infections. Serious infections are as defined in Section 6.2. The summaries will be presented separately for the two groups, and are as follows:

- The baseline behavioural strategies (PCS total score, PCS (3 domains) and CSQ (5 domains)) statistics, summarized as per Section **Error! Reference source not found.**, will be presented for patients who experienced each AE of interest (Herpes Zoster/Serious Infection) and for patients who did not experience the AE of interest.
- The last behavioural strategies (PCS total score, PCS (3 domains) and CSQ (5 domains)) statistics, summarized as per Section **Error! Reference source not found.**, for measurements prior to the first occurrence of the AE of interest (for patients who experienced the AE of interest) will be presented. The last behavioural strategies measurements in the study for patients who did not experience the AE of interest will also be presented.
- Correlation between the number of AE's of interest and behavioural strategies (PCS total score, PCS (3 domains) and CSQ (5 domains)) at baseline will be

examined using the methods described in Section **Error! Reference source not found.**.

- Correlation between the number of AE's of interest and the last behavioural strategies (PCS total score, PCS (3 domains) and CSQ (5 domains)) prior to the first occurrence of an AE will be examined using the methods described in Section **Error! Reference source not found.**.
- The correlation between the adjusted number of AE's (see Section 6.2) and the adjusted AUC for PCS and CSQ (see Section **Error! Reference source not found.**) will be calculated and presented, as per Section **Error! Reference source not found.**. This will be done for:
  - PCS total score.
  - The 3 domains of the PCS.
  - The 5 domains of the CSQ.
- The change in behavioural strategies (PCS total score, PCS (3 domains) and CSQ (5 domains)) before and after start of the first occurrence of the AE of interest will be presented for the subset of patients who experience the AE of interest, summarized as per Section **Error! Reference source not found.**.

8.2.6.2 Analysis of Continuous Variables

Continuous secondary endpoints will be presented as per Section **Error! Reference source not found.**. The continuous secondary endpoints are:

- DAS28-4 ESR
- DAS28-4 CRP
- TJC
- SJC
- Duration of morning stiffness
- EuroQoL EQ-5D-3L
- SF-12v2 Questionnaire (PCS_{SF12}, MCS and 8 health domains)
- FACIT-Fatigue Questionnaire
- Pain

8.2.6.3 Analysis of Categorical Variables

Categorical secondary endpoints will be presented as per Section 8.1.5. The categorical secondary endpoints are:

- DAS28 (EULAR Criteria)
- GIRERD Questionnaire

8.2.6.4 Analysis of Binary Variables

Binary secondary endpoints will be presented as per Section **Error! Reference source not found.**. The binary secondary endpoints are:

- LDA
- Patients in remission
- FiRST Questionnaire

8.2.6.5 Further Analyses

Summaries will be produced for SDAI, CDAI, PhGA, PtGA, with summarises as per Section **Error! Reference source not found.**. The ACR-EULAR 2011 Boolean criteria will be summarised as per Section **Error! Reference source not found.**.

The association between baseline PCS total score (<20 *vs* ≥ 20) and baseline characteristics will be explored using univariate and multivariate logistic regressions. Baseline characteristics are listed hereafter:

- Sex (Female/Male)
- Age (years)
- BMI (kg/m²)
- Time between initial diagnosis of rheumatoid arthritis and inclusion (years)
- Number of paintful joints
- Number of swollen joints
- Evaluation of pain (mm)
- Erosions (No/Yes)
- RF (-/+)
- ACPA (-/+)
- CRP (mg/ml)
- ESR (mm/hour)
- DAS28-4 CRP
- DAS28-4 ESR
- FACIT-Fatigue
- SF-12 Physical Component Summary
- SF-12 Mental Component Summary
- EuroQoL EQ-5D-3L
- Established diagnosis of a depressive syndrome (No/Yes)
- Previously treated by csDMARD (No/Yes)
- Previously treated by biological treatments or tsDMARD (No/Yes)
- Number of previous biological treatments or tsDMARD by patient
- Concomitant conventional synthetic DMARD (No/Yes)

8.2.7 Subgroup Analyses

All efficacy variables, TEAEs and AEs of interest (Herpes Zoster and Serious Infections) will be described by sex and patient age (< 65 years, ≥65 years).

8.2.8 Impact of Covid-19 on study results

The impact of Covid-19 (if any) on the study results will be assessed. The number of delayed/missing visits during the lockdown period in France (from the 17th march 2020 to the 10th may 2020) will be provided.

8.2.9 Summary of Analyses

Table 8 Summary of Main Analyses

Outcome	Analysis Set	Supports Protocol Objective Number	Subgroups	Statistical Method	Covariates/ Strata	Missing Data
Time to discontinuation of Tofacitinib	Full Analysis Set (FAS)	1	None	Cox PH model	Univariate model selection, followed by backwards selection on the variables in Figure 2.	Missing data indicators
Sensitivity Analysis – Only to be used in the event of >10% missing data: Time to discontinuation of Tofacitinib	FAS	1 – **Sensitivity Analysis**	None	Cox PH model	Univariate model selection, followed by backwards selection on the variables in Figure 2.	Complete Case Analysis
Sensitivity Analysis: Time to discontinuation of Tofacitinib	FAS	1 – **Sensitivity Analysis**	None	Random Survival Forests	Variables in Figure 2, following methods seen in Section 8.1.1.2 Random Survival Forests.	Random Survival Forests missing data algorithm.
Sensitivity Analysis: Time to discontinuation of Tofacitinib	Month 3 Analysis Set	1 – **Sensitivity Analysis**	None	Cox PH model	Univariate model selection, followed by backwards selection on the variables in Figure 2 and Das28-4 CRP at Baseline.	Missing data indicators
Sensitivity Analysis: Time to discontinuation of Tofacitinib	Month 3 Analysis Set	1 – **Sensitivity Analysis**	None	Random Survival Forests	Variables in Figure 2 and DAS28-4 CRP at Month 3, following the methods seen in Section 8.1.1.2 Random Survival Forests.	Random Survival Forests missing data algorithm.
Time to discontinuation of Tofacitinib	FAS	2	None	Cox PH model	Baseline CSQ and baseline PCS	No missing data methods used.
Time to discontinuation of Tofacitinib	FAS	2	None	Cox PH model	Last recorded CSQ and PCS before discontinuation.	No missing data methods used.
Time to discontinuation of Tofacitinib	FAS	2	None	Cox PH model	Change from baseline to last recorded CSQ and PCS before discontinuation.	No missing data methods used.

9 LIST OF TABLES AND TABLE SHELLS

To be presented in a separate document.

10 REFERENCES

1. *A Critical Look at Methods for Handling Missing Covariates in Epidemiology Regression Analyses.* **Finkle, William D and Greenland, Sander.** 1995, American Journal of Epidemiology, pp. 1255 - 1264.
2. *Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression.* **Jones, Michael P.** 433, s.l. : Journal of the Americsan Statistical Association, 1996, Vol. 91.
3. *Random Survival Forests for R.* **Kogalur, Hemant Ishwaran and Udaya B.** 2007, R News, pp. 25-31.
4. **Maruish, Mark E.** *User's Manual for the SF-12v2 Health Survery, Third Edition.* Lincoln, RI : QualityMetric Incorporated, 2012.
5. **Sullivan, Michael JL.** *The Pain Catastrophizing Scale User Manual.* Montreal, Quebec : School of Physical and Occupational Therapy, 2009.
6. *Random Survival Forests.* **Hemant Ishwaran, Udaya B Kogalur et al.** 3, s.l. : The Annals of Applied Statistics, 2008, Vol. 2.
7. *Relative risk trees for censored survival data.* **Crowley, M. LeBalnc and J.** s.l. : Biometrics, 1992, Vols. 48: 411-425.
8. *On the exact distribution of maximally selected rank statistics.* **Lausen, Hothorn and B.** s.l. : Comput. Statist. Data Analysis, 2003, Vols. 43:121-137.
9. *Regression trees for censored data.* **Segal, M. R.** s.l. : Biometrics, 1988, Vols. 44:35-47.
10. *Valuing EQ-5D using Time Trade-Off in France.* **Chevalier, Julie and de Pouvourville, Gérard**. 2013, Eur J Health Econ, Vol 14: 57-66.

11 APPENDICES

11.1 APPENDIX 1: DATA DERIVATION DETAILS

Table 9 Study Visits and Visit Windowing

Visit	Target Date	Target Day	Visit Window
Visit 0	Inclusion	1	Day 1 or before
Visit 1	1 month	30	Day 2 to Day 61
Visit 2	3 months	91	Day 62 to Day 137
Visit 3	6 months	183	Day 138 to Day 274
Visit 4	12 months	365	Day 275 to Day 457
Visit 5	18 months	548	Day 458 to Day 639
Visit 6	24 months	731	Day 640 onwards

In case of multiple observations falling within a given window, the observations selected for analysis will be identified as follows:

1. The observation closest to the target day will be used.
2. If the observations are at equal distance from the target day in absolute value, the one with a correct nominal visit label will be used.
3. If neither (1) nor (2) can be used to identify the observation windowing, then the latest observation within the analysis window will be used.