Protocol A3921133

PHASE 3B/4 RANDOMIZED SAFETY ENDPOINT STUDY OF 2 DOSES OF TOFACITINIB IN COMPARISON TO A TUMOR NECROSIS FACTOR (TNF) INHIBITOR IN SUBJECTS WITH RHEUMATOID ARTHRITIS

Statistical Analysis Plan (SAP)

Version: 5 (Amendment 4)

Date: 08-Jul-2020

TABLE OF CONTENTS

TABLE OF CONTENTS 2
LIST OF FIGURES 3
APPENDICES 3
LIST OF ABBREVIATIONS AND DEFINITION OF TERMS 4
1. AMENDMENTS FROM PREVIOUS VERSION(S) 6
2. INTRODUCTION 9
2.1. Study Design 10
SCHEDULE OF ACTIVITIES BY VISIT (SCREENING – MONTH 6) 12
SCHEDULE OF ACTIVITIES BY VISIT (MONTH 9- END OF STUDY) 13
2.2. Study Objectives 14
2.2.1. Safety Objective 14
2.2.2. Efficacy Objective 14
3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING 14
4. HYPOTHESES AND DECISION RULES 15
4.1. Statistical Hypotheses 15
4.2. Statistical Decision Rules 15
5. ANALYSIS SETS 15
5.1. Full Analysis Set 15
5.2. 'Per Protocol' Analysis Set 15
5.3. Safety Analysis Set 15
5.4. Treatment Misallocations 16
5.5. Protocol Deviations 16
5.5.1. Deviations Assessed Prior to Randomization 16
5.5.2. Deviations Assessed Post-Randomization 16
6. ENDPOINTS AND COVARIATES 17
6.1. Efficacy Endpoints 17
6.2. Safety Endpoints 17
6.3. Other Endpoints 22
6.3.1. Pharmacokinetics Endpoints 22
6.3.2. Pharmacodynamics Endpoints 22
6.3.3. Outcomes Research Endpoints 22
6.4. Covariates 23
7. HANDLING OF MISSING VALUES 23

7.1. Safety Data 23
7.1.1. Censoring Time for Time-to-event 24
7.2. Efficacy Data 25
7.2.1. Outcomes Research Endpoints 25
7.3. Missing Data due to COVID-19 Pandemic 26
8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES 26
8.1. Statistical Methods 26
8.1.1. Analyses of Time-to-Event Data 26
8.1.2. Analyses of Incidence Rates 26
8.1.3. Analyses of Continuous Efficacy Data 28
8.1.4. Analyses of Binary Efficacy Data 28
8.2. Statistical Analyses 28
8.2.1. Primary Analysis 29
8.2.2. Analyses of Other Safety Endpoints 30
8.2.3. Analyses of Efficacy Endpoints 31
8.2.4. Analyses of Outcome Research Endpoints 31
8.2.5. Subgroup Analyses by Age 32
9. REFERENCES 33
10. APPENDICES 34

LIST OF FIGURES

Figure 1. Study Design (minimum 1300 per arm, total N=4000) 11

APPENDICES

Appendix 1. DATA DERIVATION DETAILS 34
Appendix 1.1. Definition and Use of Visit Windows in Reporting 34
Appendix 1.2. Definition of Protocol Deviations that Relate to Statistical Analyses/Populations 36
Appendix 1.3. Definition of Analysis Populations/Sets 36
Appendix 1.4. Further Definition of Endpoints 36
Appendix 2. STATISTICAL METHODOLOGY DETAILS 37
Appendix 2.1. Further Details of Interim Analyses 37
Appendix 2.2. Further Details of the Statistical Methods 37

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation	Definition
ACR	American College of Rheumatology
AE	Adverse Event
ALT	Alanine Aminotransferase
AST	Aspartate Transaminase
ATE	Arterial Thromboembolism
BID	Twice Daily
BMI	Body Mass Index
CaPS	CDISC and Pfizer Standard
CDAI	Clinical Disease Activity Index
CI	Confidence Interval
CRF	Case Report Form
CRP	C-reactive Protein
CSR	Clinical Study Report
CXR	Chest X-Ray
DAS	Disease Activity Score
DILI	Drug-induced Liver Injury
DMARD	Disease Modifying Antirheumatic Drug
DMC	Data Monitoring Committee
DVT	Deep Vein Thrombosis
ECG	Electrocardiogram
EQ-5D/EuroQol - 5D	European Quality of Life - 5 Dimensions
EQ-VAS	EuroQol Visual Analog Scale
EULAR	European League against Rheumatism
FAS	Full Analysis Set
GI	Gastrointestinal
HAQ-DI	Health Assessment Questionnaire-Disability Index
HDL-c	High Density Lipoprotein Cholesterol
IEC	Independent Ethics Committee
INH	Isoniazid
IRB	Institutional Review Board
JTS	Joints
LDA	Low Disease Activity
LDL-c	Low Density Lipoprotein Cholesterol
LLA	Lipid Lowering Agent
LOCF	Last Observation Carried Forward
LSM	Least Squares Mean
MACE	Major Adverse Cardiovascular Event
MedDRA	Medical Dictionary for Regulatory Activities
MI	Myocardial Infraction
MMRM	Mixed Model for Repeated Measures
NMSC	Non-melanoma Skin Cancer

Abbreviation	Definition
NSAID	Nonsteroidal Anti-inflammatory Drug
OI	Opportunistic Infection
PASS	Post Authorization Safety Study
PD	Protocol Deviation
PE	Pulmonary Embolism
pIPD	Potentially Important Protocol Deviation
QFT	QuantiFERON-TB Gold
QOW	Every Other Week
RA	Rheumatoid Arthritis
SDAI	Simplified Disease Activity Index
SAE	Serious Adverse Event
SAS	Safety Analysis Set
SC	Subcutaneous Injection
SEER	Surveillance, Epidemiology, and End Results
SF-36	Short Form Health Survey - 36 Items
SIR	Standardized Incidence Ratio
SMQ	Standardized MedDRA Query
SOP	Standard Operation Procedure
TB	Tuberculosis
TEAE	Treatment Emergent Adverse Event
TNFi	Tumor Necrosis Factor Inhibitor
TST	Tuberculin Skin Test
ULN	Upper Limit of Normal
US FDA	United States Food and Dru Administration
VAS	Visual Analog Score
WPAI	Work Productivity Impairment

1. AMENDMENTS FROM PREVIOUS VERSION(S)

Document	Version Date	Summary of Changes
Original SAP (Version 1)	1-May-2013	N/A
Amendment 1 (Version 2)	26-Jan-2014	1. The study has not started to enroll patients. This amendment incorporates changes made to the original SAP, which are also aligned with the changes made in Amendments 1-3 to the protocol. Major changes are: 2. Eliminated the double-blind efficacy sub-study and related contents. 3. Removed MOS Sleep scale, FACIT-Fatigue, and RA Healthcare Resource Utililization Questionnaires from the endpoints. 4. Changed comparator from all adalimumab to TNFi (adalimumab in US, Puerto Rico and Canada, with etanercept in the rest of world). 5. Increased N from 3900 to approximately 4000. 6. Have all patients on background methotrexate. 7. Eliminated interim analysis. 8. Clearly stated statistical hypotheses for the primary objectives. 9. Updated statistical decision rules corresponding to the statistical hypotheses. 10. Changed Safety Analysis Set to be the set containing subjects randomized and had at least one dose of drug (the same as Full Analysis Set). It had all treated subjects with at least one dose of drug regardless of randomized or not. 11. Clarified the application for the analysis data sets: 12. Safety Analysis Set will be used for all safety endpoints; 13. Full Analysis Set will be used for all efficacy and outcome research endpoints; 14. Per Protocol Analysis Set will be used for robustness analysis of the primary safety endpoints.

Document	Version Date	Summary of Changes
		15. Added reference for endpoint definitions and clarification regarding endpoint composite for MACE in section 6.2. 16. Added gender as one of the covariates. 17. Changed the cutoff visit for missing data imputation from Month 6 (which corresponded to the deleted sub-efficacy study) to Month 36. 18. Clarified missing data handling for continuous endpoints: LOCF for component scores before subjects withdraw, and BOCF for component scores after subjects withdraw, to align with the approach for binary composite endpoints. 19. For continuous endpoints, changed MMRM model approach to t distribution based estimation for two-sample comparison. 20. For binary safety data, dropped logistic model approach from section 8.1. 21. Changed 3-month attribution window for AEs and malignancies to 60 days in the estimation of crude incidence rates in section 8.1.1 to align with the attribution window defined in time-to-event analyses. 22. Added definition of Tier-3 events in the 3-tier method. 23. Eliminated the day 14 visit, and modified cutoff days for the window definition for month 2, 3, 6, 9 visits in the appendices. 24. Typographical corrections and clarifications.
Amendment 2 (SAP version 3)	13-May-2019	1. Section 2.1: provided safety information based on protocol amendment 8 regarding subjects who randomized to tofacitinib 10 mg BID. 2. The definition of Per-Protocol Analysis Set was modified to be more clear. 3. Section 6.1: removed the reference statement about Pfizer rulebook. 4. Section 6.2: added following endpoints: DVT, PE, ATE, DVT or PE. Updated definition of thromboembolic events.

Document	Version Date	Summary of Changes
		5. Added Section 7.1.1 specified the rules of counting the time to event, the text was removed from Section 5.3 (Safety Analysis Set) to Section 7.1.1. 6. Section 7: specified that no imputation will be done for safety endpoints, and no imputation will be done for efficacy endpoints after the declared study completion day. 7. Section 8: specified that data collected after subject's discontinuation from the study will be summarized separately from the analyses described in this SAP. 8. Section 8.2: clarified that all analyses identified for the tofacitinib 10 mg BID treatment arm will be conducted as stated prior to protocol amendment 8. 9. Section 8.2.2: clarified how data collected under standard of care will be summarized for TEAE. 10. Section 8.2.2: removed 3-tier analysis of AEs. 11. Section 8.2.2: added TEAE definition. 12. Section 8.2.2: added endpoint of DVT, PE, ATE, DVT or PE to analysis. 13. Section 8.2.5: added for additional non-standard safety table details. 14. Appendix 1.1: added rules of visit windowing specific to SF-36, EQ-5D, and WPAI due to their different schedule of evaluations, as well as those specific to lipids (fasting) due to the same reason.
Amendment 3 (SAP version 4)	30-Jun-2020	1. Section 5: clarified FAS and SAS. 2. Section 6.2: added the word "adjudicated" for the adjudicated safety endpoints for clarity and internal consistency within the document. Clarified terminology for deaths (updated terms to: all-cause deaths and all-cause deaths associated with adjudicated primary endpoint criteria) for clarity and internal consistency within the document. 3. Section 7.2: for continuous endpoints, updated missing data handling by use of a MMRM for consistency with Sponsor's convention established across tofacitinib development programs. 4. Section 7.3: added a new section to describe missing data handling due to COVID-19 pandemic and associated summaries and data listings. 5. Section 8.1.2: exact 95% CI will be used instead of approximate 95% CI for the crude incidence rate for consistency with Sponsor's convention established across tofacitinib development

Document	Version Date	Summary of Changes
		programs. Clarified the calculation for SIR for malignancies. 6. Section 8.1.3: for analyzing continuous endpoints, t-test will be replaced by a MMRM for consistency with Sponsor's convention established across tofacitinib development programs. 7. "On-Treatment Time" will be renamed as "60-Day On-Treatment Time" for internal consistency throughout the document. 8. "28-day rule" will be renamed as "28-Day On-Treatment Time" for internal consistency throughout the document. 9. Section 8.2.5: subgroup analysis by age (<65 vs ≥65 years) for all-cause deaths will be added. 10. Appendix 1.1: clarified the baseline for ECG data. 11. Appendix 2.2: added algorithm of SIR and associated confidence interval. 12. Made minor updates for clarity and consistency throughout.
Amendment 4 (SAP version 5)	08JUL2020	Section 7.2.: Updated missing data handling for efficacy data to accommodate the trial design as a time-to-event safety study.

2. INTRODUCTION

Note: in this document any text taken directly from the protocol is ***italicized***.

As a commitment to the United States Food and Drug Administration (US FDA), this Post Authorization Safety Study (PASS) will evaluate the safety of tofacitinib 5 mg twice daily (BID) and tofacitinib 10 mg BID compared to a tumor necrosis factor inhibitor (TNFi).

The primary objective of this endpoint study is to evaluate the safety of tofacitinib at two doses versus a TNFi; the co-primary endpoints are adjudicated major adverse cardiovascular events (MACE) and adjudicated malignancies excluding non-melanoma skin cancers during study participation.

This document contains the plans for analyses in support of above objective.

2.1. Study Design

This is a Phase 3b/4 randomized, parallel arm, open-label safety endpoint study. All subjects will be randomized in a 1:1:1 ratio to one of the three treatment arms with approximately 1300 subjects in each treatment arm:

1. *Tofacitinib 5 mg BID (oral).*
2. *Tofacitinib 10 mg BID (oral).**
3. *TNFi: In the US, Puerto Rico and Canada, subjects randomized to TNFi will receive adalimumab 40 mg every other week (QOW) by subcutaneous injection (SC); in all other countries, subjects randomized to TNFi will receive etanercept 50 mg once weekly by SC injection.*

*** In response to new safety information about tofacitinib that was provided to Study A3921133 investigators on 19 February 2019, investigators were notified to verbally inform subjects randomized to tofacitinib 10 mg BID to reduce the dose of tofacitinib to 5 mg BID within 7 calendar days of this notification, secure their verbal agreement to continue in the study, and if they agreed, reduce their dose of tofacitinib to 5 mg BID for the remainder of the study. Tofacitinib is supplied as 5 mg tablets, which allowed subjects to decrease from their previous dose of 2 tablets twice daily to 1 tablet twice daily.*

During the study, subjects may require alternate therapies in addition to, or instead of, their randomized drug assignment. All subjects, regardless of their treatment regimen will participate in the study until study completion.

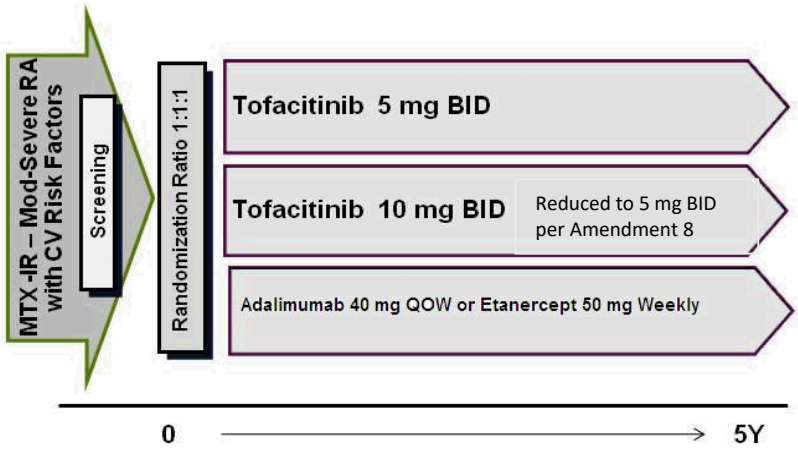
Study completion will be declared when all 3 of the following conditions are met:

1. *At least 1500 subjects have been followed for at least 3 years.*
2. *The targeted number of MACE are observed.*
3. *The targeted number of malignancies excluding non-melanoma skin cancers are observed.*

It is expected that approximately 4000 subjects will participate in the study and the expected duration of the study is approximately 5 years following randomization of the first subject. The exact number of subjects and duration of the study will be determined by the pre-specified rules outlined in the charter of the blinded Steering Committee.

The study design schematic for the study is represented in Figure 1.

Figure 1. Study Design (minimum 1300 per arm, total N=4000)



SCHEDULE OF ACTIVITIES BY VISIT (SCREENING – MONTH 6)

Protocol Activities – Screening to Month 6	***Screen***	***Visit 1*** *Day 1*	***Visit 2*** *Month[f] 2*	***Visit 3*** *Month[f] 3*	***Visit 4*** *Month[f] 6*
Visit Windows	*NA*	*See* Protocol *6.2.1*	*±2 weeks*	*±1.5 months*	*±1.5 months*
Informed Consent	*X*				
Medical History and Confirm Diagnosis	*X*				
Physical Examination, including height	*X*				
Assessment of New Physical Findings		*X*	*X*	*X*	*X*
Weight, Vital Signs	*X*	*X*	*X*	*X*	*X*
Inclusion/Exclusion criteria	*X*	*X*			
Blood Chemistry,[a] Hematology[a]	*X*	*X*	*X*	*X*	*X*
C-reactive Protein (CRP)	*X*	*X*	*X*	*X*	*X*
Banked Biospecimens		*X*			
HIV Serology, HBsAg, HBcAb, HBsAb[b]	*X*				
HCV Ab, HCV RNA[c]	*X*				
Lipid Profile (fasting)	*X*	*X*	*X*		
Pregnancy test[d]	*X*	*X*	*X*	*X*	*X*
QuantiFERON Gold®™ In-Tube Test	*X*				
Rheumatoid Factor	*X*				
Anti-cyclic citrullinated peptide (anti-CCP)	*X*				
Urinalysis	*X*				
Chest X-ray	*X*				
ECG	*X*				
Randomization		*X*			
Study Medication Dispensing		*X*	*X[g]*	*X*	*X*
Contraceptive Use Documentation	*X*	*X*	*X*	*X*	*X*
Joint Count (28)	*X*	*X*	*X*	*X*	*X*
Patient Assessment of Arthritis Pain		*X*	*X*	*X*	*X*
Patient Global Assessment of Arthritis		*X*	*X*	*X*	*X*
Physician Global Assessment of Arthritis		*X*	*X*	*X*	*X*
HAQ-DI		*X*	*X*	*X*	*X*
SF-36 (Version 2, Acute)		*X*		*X*	*X*
EuroQol EQ-5D		*X*		*X*	*X*
Work Productivity and Activity Impairment (WPAI)		*X*		*X*	*X*
Safety Assessment/AE Reporting		*X*	*X*	*X*	*X*
Concomitant Medication Review[e]	*X*	*X*	*X*	*X*	*X*
End of Study Assessment					

a. Re-testing of abnormal laboratories may be required (see Protocol *section 7.15).*

b. See Exclusion Criteria 4e and 4f for details (Protocol section 4.2).

c. See Exclusion Criteria 4g for details (Protocol section 4.2).

d. Pregnancy tests may also be repeated as per request of IRB/IECs or if required by local regulations.

e. History of concomitant medications at Screening includes a history of prior Disease Modifying Antirheumatic Drug (DMARD) treatment.

f. Calendar months.

g. Adalimumab or etanercept only.

SCHEDULE OF ACTIVITIES BY VISIT (MONTH 9- END OF STUDY)

Visits are outlined through Month 60, but will continue sequentially until the end of study is declared, at which time all subjects will be required to return for an End of Study Visit.

Protocol Activities – Months 9 to 60 and End of Study Follow-up Visit	***Visits 5, 9, 13, 17, 21*** *[Months[f] 9, 21, 33,45, 57]*	***Visit 6, 10, 14, 18, 22*** *[Months[f] 12, 24, 36, 48, 60]*	***Visits 7, 11, 15, 19*** *[Months[f] 15, 27, 39, 51]*	***Visits 8, 12, 16, 20*** *[Months[f] 18, 30, 42, 54]*	***End of Study Visit***	***End of Study Follow Up[g]***
Visit Windows	*±1.5 months*	*±1.5 months*	*±1.5 months*	*±1.5 months*	*±1.0 months*	*+0.5 months*
Assessment of New Physical Findings	*X*	*X*	*X*	*X*	*X*	*X*
Weight, Vital Signs	*X*	*X*	*X*	*X*	*X*	*X*
Blood Chemistry,[a] *Hematology*[a]	*X*	*X*	*X*	*X*	*X*	
C-reactive Protein (CRP)	*X*	*X*	*X*	*X*	*X*	
Banked Biospecimens		*X*[b]			*X*	
Lipid Profile (fasting)		*X*			*X*	
Pregnancy test[c]	*X*	*X*	*X*	*X*	*X*	*X*
QuantiFERON Gold®™ In-Tube Test		*X*[d]				
Chest X-ray		*X*[e]			*X*[e]	
ECG					*X*	
Study Medication Dispensing	*X*	*X*	*X*	*X*		
Contraceptive Use Documentation	*X*	*X*	*X*	*X*	*X*	*X*
Joint Count (28)	*X*	*X*	*X*	*X*	*X*	
Patient Assessment of Arthritis Pain	*X*	*X*	*X*	*X*	*X*	
Patient Global Assessment of Arthritis	*X*	*X*	*X*	*X*	*X*	
Physician Global Assessment of Arthritis	*X*	*X*	*X*	*X*	*X*	
HAQ-DI	*X*	*X*	*X*	*X*	*X*	
SF-36 (Version 2, Acute)		*X*			*X*	
EuroQol EQ-5D		*X*			*X*	
Work Productivity and Activity Impairment (WPAI) Questionnaire		*X*			*X*	
Safety Assessment/ AE Reporting	*X*	*X*	*X*	*X*	*X*	*X*
Concomitant Medication Review	*X*	*X*	*X*	*X*	*X*	*X*
End of Study Assessment					*X*	

a. Re-testing of abnormal laboratories may be required (see Protocol *section 7.15).*

b. Month 12 visit only.

c. Pregnancy tests may also be repeated as per request of IRB/IECs or if required by local regulations.

d. QFT is not performed in subjects who had a positive test previously and were adequately treated for latent TB, subjects newly testing positive for latent TB with negative CXR and no evidence of active disease are required to start a 9 month treatment course of isoniazid to continue in the study.

e. Only for subjects testing positive to QFT at annual visit.

f. Calendar Months.

g. Occurs 28 days after End of Study visit.

2.2. Study Objectives

2.2.1. Safety Objective

The primary objective of this endpoint study is to evaluate the safety of tofacitinib at two doses versus TNFi; the co-primary endpoints are adjudicated major adverse cardiovascular events (MACE) and adjudicated malignancies excluding non-melanoma skin cancers during study participation.

The following assessments will be used to evaluate safety in the entire study population:

- *Changes in physical assessments, including weight and vital signs.*
- *AE reporting, incidence and severity.*
- *Incidence and severity of safety laboratory abnormalities.*
- *Adherence to contraceptive use.*
- *Adherence to protocol requirements.*
- *Concomitant medication review.*
- *Review of eligibility to continue study drug.*

2.2.2. Efficacy Objective

The efficacy analysis is secondary objective for this study. The efficacy analysis is to estimate differences between each tofacitinib regimen and TNFi regimen, as well as between the two tofacitinib regimens. The following assessments related to the signs and symptoms of patient health and functional status will be captured directly in Case Report Form (CRF) and be used to evaluate efficacy in the study population:

- *28 joint counts;*
- *Patient Global Assessment;*
- *Patient Assessment of Pain;*
- *Physician Global Assessment;*
- *C-reactive Protein (CRP);*
- *A number of patient reported outcomes, including the Health Assessment Questionnaire – Disability Index (HAQ-DI).*

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

No interim analyses of the study data is planned for use outside the Data Monitoring Committee (DMC) and the Steering Committee. Any interim analyses planned for the DMC and the Steering Committee will be specified in their respective charters.

The final analysis will be performed using the complete, final, quality-controlled database generated from all visits and information generated as of the time of the end of active follow-up in the study.

Unblinding will not be needed as this is an open-label study.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

The hypothesis to be tested is, for the primary safety endpoints of malignancy and MACE, whether the two tofacitinib doses is non-inferior to the TNFi control and whether tofacitinib 10 mg BID is non-inferior to tofacitinib 5 mg BID.

4.2. Statistical Decision Rules

The hazard ratios relative to the TNFi control regimen for the tofacitinib regimens will be estimated together with 95% two-sided confidence intervals (CI). Non-inferiority of tofacitinib regimens to TNFi control regimen will be concluded if the upper confidence limits for both malignancy (Total Time Analysis) and MACE (60-Day On-Treatment Time Analysis), are less than 1.8.

Additionally, hazard ratios of tofacitinib 10 mg BID vs. tofacitinib 5 mg BID along with 95% two-sided CIs will be estimated. Non-inferiority of tofacitinib 10 mg BID to tofacitinib 5 mg BID will be concluded if the upper confidence limits for both malignancy (Total Time Analysis) and MACE (60-Day On-Treatment Time Analysis) are less than 2.0.

Adjustments for multiplicity will not be applied.

5. ANALYSIS SETS

The following analysis sets are defined.

5.1. Full Analysis Set

The term 'Full Analysis Set' (FAS) is used to describe the analysis set which is as complete as possible. It will include all subjects who were randomized in the study and received at least one dose of the randomized investigational drug (tofacitinib or TNFi). Subjects will be analyzed in the treatment groups as they are randomized.

5.2. ‘Per Protocol’ Analysis Set

The ‘Per Protocol’ analysis set will include all randomized subjects from FAS with no important protocol deviations that could impact the analysis, as determined by the potentially important protocol deviation (pIPD) process. The list of patients that will be excluded from the per-protocol analysis set will be finalized prior to the database release.

5.3. Safety Analysis Set

The Safety Analysis Set (SAS) will include all subjects who were randomized in the study and received at least one dose of the randomized investigational drug (tofacitinib or TNFi). Subjects will be analyzed in the treatment groups as they received.

5.4. Treatment Misallocations

For subjects with errors in treatment allocation, described below is the convention under which treatment groups they will be reported for efficacy and safety analyses:

If a subject was:

- Randomized but not treated, then the subject will be excluded from all analyses according to the definitions of FAS, Per Protocol, and Safety Analysis Sets given just above.

- Randomized but took incorrect treatment, then the subject will be reported under his/her randomized treatment group for all analyses based on FAS (or Safety Analysis Set), but will be omitted from analyses based on Per Protocol Analysis Set.

5.5. Protocol Deviations

Protocol deviation identification and documentation will be accomplished for this study according to the Sponsor's standard operating procedures (SOPs) that require periodic review of individual protocol deviations and periodic review of aggregated cumulative protocol deviations. Prior to initiation of the study, a list of potentially important protocol deviations (pIPDs) is developed in alignment with the requirements of the protocol. This list is updated, as necessary, to include other deviations identified during the course of the study.

Additionally, pIPDs are prospectively identified; pIPDs are used by the statistician and clinician in determining eligibility for inclusion in the statistical analysis of the study and are considered for inclusion in the body of the clinical study report.

Subjects with pIPDs that could impact endpoint analysis will be excluded from the Per Protocol Analysis Set, but included in the Full Analysis Set.

Protocol deviations classified as pIPDs are included in a Protocol Deviation Listing in the clinical study report.

5.5.1. Deviations Assessed Prior to Randomization

Granted exceptions or waivers to the inclusion or exclusion criteria are prohibited by the Sponsor's standard operating procedures and thus, are not expected to occur. Entry into the study by any subject who does not meet the study inclusion or exclusion criteria will be identified as protocol deviation.

5.5.2. Deviations Assessed Post-Randomization

The deviation resulting in exclusion from the safety analysis set and the full analysis set is failure to take any study medication.

6. ENDPOINTS AND COVARIATES

6.1. Efficacy Endpoints

Efficacy endpoints will be secondary endpoints in this safety study. The efficacy endpoints are:

- *Change from baseline to each post-baseline scheduled visit in DAS28-4 (CRP), Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI).*
- *Rate of remission at each post-baseline scheduled visit including:*
  - *ACR-EULAR Boolean remission (defined as the subject satisfying all of the following: tender joint count ≤1, swollen joint count ≤1, C-reactive protein ≤1 mg/dl, patient global assessment ≤1 on a 0-10 scale);*
  - *SDAI ≤3.3;*
  - *CDAI ≤2.8.*
- *Rate of low disease activity (LDA) at each post-baseline scheduled visit including:*
  - *SDAI ≤11;*
  - *CDAI ≤10;*
  - *Disease Activity Score (DAS)28-4(CRP) ≤3.2.*
- *ACR20, ACR50, and ACR70 response rate of at each post-baseline scheduled visit.*
- *Change from baseline to each post-baseline scheduled visit in the HAQ-DI.*

6.2. Safety Endpoints

Co-Primary

- *Malignancies, excluding non-melanoma skin cancers (adjudicated).*
- *Major adverse cardiovascular events (MACE) (adjudicated).*

  The definitions of MACE used in this study are consistent with those outlined in the Standardized Definitions for Cardiovascular and Stroke End Point Events in Clinical Trials[1] with the exclusion of cardiovascular death due to pulmonary embolism. The following events are included, as defined:

  - *Cardiovascular death:*
    - *Death due to acute myocardial infarction (MI);*
    - *Sudden cardiac death;*

- *Death due to heart failure;*
- *Death due to stroke;*
- *Death due to cardiovascular procedures;*
- *Death due to cardiovascular hemorrhage;*
- *Death due to other cardiovascular causes: peripheral artery disease.*

- *Non-fatal myocardial infarction (MI).*
- *Non-fatal stroke of any classification, including reversible focal neurologic defects with imaging evidence of a new cerebral lesion consistent with ischemia or hemorrhage.*

Secondary

- *Opportunistic infection events including tuberculosis (adjudicated):*
  - The opportunistic infection (OI) events by all OIs, tuberculosis adjudicated as OI, herpes zoster adjudicated as OI and all other OI;
  - Tuberculosis events separately from the evaluation of opportunistic infections:
    - Proportion of subjects with QuantiFERON-TB Gold (QFT) or Tuberculin Skin Test (TST) positive subjects with a history of TB by region;
    - QFT and TST newly positive by time in the study, newly positive treated with Isoniazid (INH) or other Tuberculosis (TB) prophylaxis, newly positive with positive Chest X-Ray (CXR), newly positive with negative CXR by region;
    - Patient Years exposure by treatment arm, with or without TB;
    - Proportion and Incidence Rate of Tuberculosis by time in the study;
    - Tuberculosis Severity and Seriousness by treatment arm.
- *Hepatic events (adjudicated) including:*
  - Drug-induced Liver Injury (DILI) - Probable, Highly Likely and Definite;
  - DILI – Listed separately;
  - DILI – Cases meeting classification and severity;
  - Subjects with elevations of Transaminase levels >1x Upper Limit of Normal (ULN), ≥3xULN, ≥5xULN (based on laboratory values).

- *Cardiovascular events other than MACE (adjudicated).*
- *All Adverse Events (AEs), including Serious Adverse Events (SAEs):*
  - Herpes zoster:
    - Baseline Characteristics with and without herpes zoster;
    - Patient Years of exposure with and without herpes zoster;
    - Proportion and Incidence Rate by Age Category, Time in study for all Herpes Zoster;
    - Proportion and Incidence Rate by Age Category and Time in Study for Herpes zoster by number of dermatomes;
    - List of subjects with Herpes Zoster Ophthalmic and Visceral cases;
    - List of subjects with serious Herpes Zoster with and without history of chicken pox (varicella); with and without history of chicken pox (varicella) vaccine; with and without history of herpes zoster vaccine (varicella); with and without history of history of shingles (herpes zoster);
    - List of subjects with non-serious Herpes Zoster with and without history of chicken pox (varicella); with and without history of chicken pox vaccine (varicella); with and without history of herpes zoster vaccine (varicella); with and without history of shingles (herpes zoster);
    - Risk factors associated with herpes zoster (cox regression analysis) including region, baseline corticosteroid use, age group, smoking status, treatment arm;
    - Time to resolution of all Herpes Zoster Cases, single dermatomes, 2 adjacent dermatomes, disseminated herpes zoster by treatment arm;
    - Severity and seriousness of Herpes Zoster by treatment arm;
    - Post Herpetic neuralgia after Herpes Zoster by treatment arm;
    - Permanent and temporary discontinuations due to Herpes Zoster by treatment arm;
    - Herpes Zoster considered OI by treatment arm.
  - Serious infections.
  - Gastrointestinal (GI) perforations (adjudicated):

- Incidence rate for all GI events adjudicated as meeting criteria by treatment arm;
- Listing of all GI events adjudicated, including location, risk factors and outcome, by treatment arm.

- Nonmelanoma skin cancers (adjudicated).
- Interstitial lung disease (adjudicated by a sponsor review committee, independent of the study team).
- Deep vein thrombosis (DVT) (adjudicated): subjects who experienced any DVT.
- Pulmonary embolism (PE) (adjudicated): subjects who experienced any PE.
- Arterial thromboembolism (ATE) (Standardized MedDRA Query (SMQ)): subjects who experienced any ATE.
- DVT or PE (adjudicated): subjects who experienced any DVT or PE.
- Thromboembolic events (TE) subjects who experienced any DVT (adjudicated), PE (adjudicated) or ATE (SMQ).
- SMQ of Hypertension.

- *Clinically significant abnormal laboratory parameters:*
  - Hematology, including hemoglobin, neutrophils and lymphocytes;
  - Lipids;
  - Liver tests including Alanine Aminotransferase (ALT), Aspartate Transaminase (AST) and total bilirubin.
- *All-cause mortality (adjudicated)* (also known as all-cause deaths associated with the adjudicated primary endpoint criteria under Bullet Deaths below).
- *Reasons for permanent or temporary discontinuation of study medication.*
- *Concomitant drugs by treatment group, in the following categories:*
  - *Lipid Lowering Agents (LLA) including Statins;*
  - *DMARDS;*
  - *Steroids;*
  - *Analgesics and* Nonsteroidal Anti-inflammatory Drug *(NSAIDS);*

- *INH and Rifampacin.*

- *Laboratory Data Descriptive Statistics for the following tests:*
  - *Hemoglobin;*
  - *Platelets;*
  - *Creatine Kinase;*
  - *Serum Creatinine;*
  - *Total Cholesterol;*
  - Low Density Lipoprotein Cholesterol *(LDL-c);*
  - High Density Lipoprotein Cholesterol *(HDL-c);*
  - *Total Cholesterol/HDL-c Ratio;*
  - *Total Cholesterol/LDL-c Ratio;*
  - *Triglycerides.*
- Vital Sign Descriptive Statistics Summaries, including BMI.
- Baseline Characteristics including:
  - Duration of Rheumatoid Arthritis (RA);
  - Cardiovascular Risk;
  - Subjects with Nodules, Sjoegren's Syndrome, Anemia of Chronic Disease, Pulmonary manifestations;
  - Joints (JTS), CDAI, SDAI, Smoking History, Alcohol use.
- Deaths:
  - List of all-cause deaths associated with the adjudicated primary endpoint criteria;
  - List of all-cause deaths;
  - Incidence rate (95% CI) for all-cause deaths associated with the adjudicated primary endpoint criteria;
  - Incidence rate (95% CI) for all-cause deaths.

- Patient Years exposure by treatment arm, including standard of care and by time point.

6.3. Other Endpoints

6.3.1. Pharmacokinetics Endpoints

None.

6.3.2. Pharmacodynamics Endpoints

None.

6.3.3. Outcomes Research Endpoints

Outcomes Research Endpoints, also known as Patient-Reported Outcomes, are:

Patient Global Assessment of Arthritis, a value on the scale of 0 to 100 (a visual-analog score, VAS).

Patient Global Assessment of Pain, a VAS, scale of 0 to 100.

Health Assessment Questionnaire-Disability Index (HAQ-DI) is a total score based on items that query the patient on how well the patient can dress and groom self, arise from bed, eat during the day, walk, perform hygiene, reach, grip, and perform common daily activities. The HAQ-DI alone is analyzed, and is on the scale of 0 to 3.

SF-36 Health Survey (Version 2, Acute) has 36 items that are summarized into eight domain scores measuring Physical Functioning, Role Physical (physical life), Social Functioning, Bodily Pain, Mental Health, Role Emotional (emotional life), Vitality, and General Health: these in turn are summarized into two component scores: Mental, and Physical. These 8 domains and 2 scores are individually analyzed. They are scaled to be centered at 50, with a standard deviation of 10. Each domain score can also be expressed on the scale of 0 to 100.

Work Productivity and Activity Impairment (WPAI) has 6 items that measure four concepts: Absenteeism (percentage of time missed from work); Presenteeism (percentage of impairment experienced while at work); Work productivity loss (overall work impairment - the combination of absenteeism and presenteeism); and Activity impairment (the percentage of impairment in daily activities). The four scores are individually analyzed.

EuroQoL EQ-5D has five items for each of the domains Unusual Activities; Anxiety/Depression; Pain/Discomfort; Mobility; and, Self Care. These five are in turn summarized into a single utility score. There is also a EuroQol visual-analog score (EQ-VAS, 0 -100) anchored with 0 as the worst health you can imagine and 100 as the best health you can imagine. The scores are individually analyzed, and the utility score is analyzed as well.

The endpoints of interest are the actual values and the change from baseline values at each post-baseline scheduled visit including:

- Patient Assessment of Arthritis;
- Patient Global Assessment of Pain;
- HAQ-DI;
- SF-36 (each of the 8 domain scores (norm-based) and 2 component summary scores);
- WPAI (each of the four scores);
- EQ-5D (five domain scores and Utility Score) and EQ-VAS.

6.4. Covariates

The following demographic variables will be included as covariates in the multivariate Cox proportional hazard model:

Geographic regions:

4 levels: USA/Canada, Europe, Latin America, and rest of world.

Age:

In years (to the nearest year) at entry to the study (categorized as 50 to <65, 65 to <75, ≥75).

Smoking:

Current or ex-smoker, lifetime nonsmoker.

Gender:

Male and Female.

7. HANDLING OF MISSING VALUES

7.1. Safety Data

Missing values for safety endpoints will not be imputed.

7.1.1. Censoring Time for Time-to-event

There are three ways in defining the time to an event:

- The "60-Day On-Treatment Time" is the time from first study dose to an event observed while a subject is exposed to the study treatment plus the applicable attribution window. The 60-Day On-Treatment Time analysis will censor subjects who cross over from tofacitinib or TNFi to standard of care at 60 days after the last dose of randomized study drug.

- The "Total Time" is the time from first study dose to an event regardless whether the event occurs during on-treatment period or after the subject discontinues the study treatment and is still being followed. The Total Time analysis will not censor subjects who cross over from tofacitinib or TNFi to standard of care and will ascribe events to randomized treatments.

- The "28-Day On-Treatment Time" is the time from first study dose to an event observed while a subject is exposed to the study treatment plus the applicable attribution window. The 28-Day On-Treatment Time analysis will censor subjects who cross over from tofacitinib or TNFi to standard of care at 28 days after the last dose of randomized study drug.

More specifically, in time-to-event analyses for a given event, a subject's time to event will be considered censored if any of the following conditions occurs prior to the subject experiencing the event in question:

- The subject is lost to follow-up (censoring time = study day of last on-study contact, eg, last visit, last lab test, and the like).

- The subject withdraws consent (censoring time = study day of last on-study contact).

- The subject is in the study under the protocol at the declared end of the study (censoring time = study day of last on-study contact).

- The subject dies without having experienced the event of interest (censoring time = study day of death).

For the **60-Day On-Treatment Time Analyses** of time-to-event, in addition to the foregoing, when a subject changes from the randomized study treatment to standard of care, the subject's time to event will be considered censored (censoring time = last dose day of randomized study regimen + the attribution window, or the study day of last on-study contact, whichever is earlier). It will be used for the primary safety endpoint MACE but excluding malignancy (excluding Non-melanoma Skin Cancer (NMSC)), and the attribution window will be 60 days. In other words, the censoring time will be the earliest of study day of last on-study contact, study day of death, and last dose day of randomized study regimen plus 60 days.

For the **Total Time Analyses** of time-to-event, the censoring time will follow the four foregoing listed cases, ie, the censoring time will not change when a subject changes from the randomized study treatment to standard of care. In other words, the censoring time will be the earlier of study day of last on-study contact and study day of death.

For the **28-Day On-Treatment Time Analyses** of time-to-event, in addition to the foregoing, when a subject changes from the randomized study treatment to standard of care, the subject's time to event will be considered censored (censoring time = last dose day of randomized study regimen + the attribution window, or the study day of last on-study contact, whichever is earlier). It will be used for all the secondary safety endpoints, and the attribution window will be 28 days. In other words, the censoring time will be the earliest of study day of last on-study contact, study day of death, and last dose day of randomized study regimen plus 28 days.

7.2. Efficacy Data

For binary variables, such as ACR20, remission with respect to SDAI≤3.3, if the value is missing due to missing values in any of the components, while the subject is still enrolled, the method of last observation carried forward (LOCF) will be used to carry forward any of the missing components, and from that mix of actual and carried forward values, the value of the binary composite variable will be determined. Note that, this approach should result in the same value for ACR20, ACR50 and ACR70 if their components are partially available but can be used to determine the ACR values without imputation (see definition for ACR response rate in Appendix 1.4).

> This type of LOCF method will be known as "LOCF mixed components", since it is based on calculating the composite value based on a mixture of values at a visit and values carried forward from previous visits. It can be justified as an attempt to carry forward as little information as possible.

After LOCF imputation has been applied, the composite binary variable can be calculated from the imputed components.

If baseline data is missing, then the post-baseline value will be set to non-responder.

DAS28-4(CRP), CDAI and SDAI are derived variables based on multiple components. If any component is missing, the endpoint will be considered as missing. A Mixed Model for Repeated Measures (MMRM) will be used to analyze continuous endpoints with repeated measures without explicit imputation for missing values.

7.2.1. Outcomes Research Endpoints

For the outcome research endpoints, rules for scoring and handling of missing data suggested by the developers will be followed.

7.3. Missing Data due to COVID-19 Pandemic

Because of the COVID-19 pandemic, some scheduled on-site visits may be substituted by remote visits such as telephone contacts. The data collected via the remote visits will be treated the same as the data collected by the on-site visits.

Missing data due to COVID-19 pandemic will be handled the same as that for other missing data.

In addition, the following summaries will be provided.

- Discontinuation due to COVID-19 impact will be reported in the disposition table.
- Missed visits due to COVID-19 impact are recorded as PD and will be in the specific PD tables in the Clinical Study Report (CSR).
- A set of summary/listing tables based on a subset of protocol deviations associated with COVID-19 will be provided. Study drug discontinuation, subject discontinuation due to COVID-19 and adverse events related to COVID-19 will be summarized as well. The source information will be captured in the protocol deviation.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

In general, number and percent will be presented for binary variables, number, mean, standard deviation (standard error of the mean), minimum, 1st, 2nd and 3rd quartiles and maximum will be presented for continuous variables. In addition, graphics may be used to present the data.

8.1. Statistical Methods

8.1.1. Analyses of Time-to-Event Data

Comparisons of hazard functions in time-to-event analyses will be carried out via (linear) Cox proportional hazard models. Censoring conventions are given in Section 7. A univariate analysis model containing only the single independent variable for treatment group will be calculated and reported.

In addition, the effects of adding other predictors to the Cox model, such as geographic region, age group, gender, and smoking status will be fully explored and reported.

8.1.2. Analyses of Incidence Rates

Crude incidence rates will be reported as the total number of subjects with admissible events divided by the total (for all qualifying subjects) time at risk for the cohort/treatment group of interest. For subjects with events, the time at risk will be the time to the first event, for subjects without events, it will be the censoring time defined in Section 7.1.1.

When subjects change regimens from study medication to standard of care then the following counting rules will be used to account for group changes in the estimation of crude incidence rates (where applicable):

- For MACE: there will be a 60-day risk window during which events will be attributed to both treatment groups, after which they will be assigned to the current treatment group.

- Malignancies (excluding NMSC): there will be a 60-day risk window during which events will only be attributed to the prior treatment group, after which they will be assigned to both treatment groups.

The 95% CIs for the crude incidence rate will be calculated for incidence rates based on the assumption that the actual count of cases arises from a Poisson distribution. The approach used in Daly et al.[2,3] describing the calculation of exact confidence limits will be used. If x denotes the actual number of subjects with events, $(1 - \alpha)$ denotes the confidence level, x_L denotes the lower confidence limit and x_U denotes the upper confidence limit, then the following formula yields values for x_L and x_U,

When $x > 0$,

$$x_L = 0.5 * \chi^2(2x;\ \alpha/2) \text{ and } x_U = 0.5 * \chi^2(2x+2;\ 1-\alpha/2),$$

When $x = 0$,

$$x_L = 0 \text{ and } x_U = 0.5 * \chi^2(2x+2;\ 1-\alpha/2).$$

In the above notation, $\chi^2(n; p)$ represents the p^{th} quantile of the chi-square distribution with n degrees of freedom.

Once x_L and x_U are obtained, they are to be adjusted for the at-risk population per 100 subject-years by multiplying by (100/total subject-years of event exposure).

Indirect standardization will also be used in a second analysis of malignancy data. With indirect standardization, age and sex-specific cancer rates in a standard population (ie, general US population via the Surveillance Epidemiology and End Results [SEER] registry) are used as weights for person-time of the age and sex-specific subgroups in the tofacitinib trial population. The resulting standardized incidence ratio (SIR) represents the ratio of the incident number of cases of cancer in the tofacitinib population to the incident number that would be expected if the study population had the same incidence rate as the US population. The SIR along with the 95% CI will be calculated for the adjudicated malignancies excluding NMSC and each cancer type (corresponding to the ones where the SEER incidence rates are available) for each randomized tofacitinib group (not for the TNFi group). For each cancer type and each treatment group, the expected number of cancer cases will be calculated as the SEER incidence rate (cancer type specific incidence rates) multiplied by the person-time (cancer type specific person-time) of each age (age categories according to SEER) and sex combination summed over all the age and sex combinations. The detailed calculation is provided in Appendix 2.2.

8.1.3. Analyses of Continuous Efficacy Data

Continuous data (ie, change from baseline) will be summarized.

Repeated measures data for continuous endpoints will be analyzed as change from baseline with a mixed model for repeated measures (MMRM) that includes fixed effects of treatment group, visit, treatment-group by visit interaction, baseline value, and baseline-value by visit interaction, without imputation for missing values. A common unstructured variance-covariance matrix will be used, provided the model converges, otherwise an alternative covariance structure, eg, heterogeneous compound symmetry (CSH), will be attempted. The Kenward-Roger degrees of freedom approximation will be used. Least squares means [LSM] of the treatments, LSM of the treatment difference, 2-sided p-value and 95% CI at each time point will be generated using this MMRM. If the Baseline is missing or if there are no post-baseline measurements, the subject will be excluded from this analysis.

8.1.4. Analyses of Binary Efficacy Data

Binary data will be summarized and compared between two treatment groups using the normal approximation of binomial proportions and 95% CI of treatment difference will be constructed. The normal-approximation to the test statistic for the difference in binomial random variables is calculated as:

$$Z_i = \frac{\hat{p}_i - \hat{p}_c}{\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n_i} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_c}}}$$

where $\hat{p}$ refers to the relative frequency, n to sample size, the subscripts i and c refer to two treatment group.

Two-sided 95% CIs are formed by:

$$(\hat{p}_i - \hat{p}_c) \pm 1.96\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n_i} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_c}}$$

8.2. Statistical Analyses

As of 19 February 2019, the tofacitinib 10 mg BID treatment arm reduced dosing to tofacitinib 5 mg BID; all analyses identified for the tofacitinib 10 mg BID treatment arm will be conducted as stated even though subjects in that treatment arm reduced their prescribed dose. Specifically, all references to the Tofacitinib 10 mg BID arm will include subjects randomized to tofacitinib 10 mg BID and discontinued study drug prior to Amendment 8 and those subjects who were randomized to tofacitinib 10 mg BID and had their dose of tofacitinib reduced to 5 mg BID with Protocol Amendment 8.

8.2.1. Primary Analysis

Safety endpoint analyses will be carried out using the data collected on the Safety Analysis Set.

Two methods in defining time to an event will be used for the co-primary time-to-event safety endpoints: 60-Day On-Treatment Time and Total Time (see Section 7.1.1 for definitions). The 60-Day On-Treatment Time definition will be used for primary MACE analysis. The Total Time definition will be used for primary malignancies (excluding NMSC).

This primary objective will be addressed via the analysis of the times to these events by proportional hazard (Cox) models via univariate model analysis. See details in Section 8.1.1.

In the time-to-event analyses using a Cox model, the treatment groups will be tofacitinib 5 mg BID, tofacitinib 10 mg BID, and TNFi regimen (adalimumab 40 mg QOW SC and etanercept 50 mg QW SC combined).

Ninety-five percent CIs for the hazard ratios for the combined tofacitinib regimens compared to the TNFi regimen as well as for the 3 pair-wise treatment comparisons, tofacitinib 10 mg BID vs. tofacitinib 5 mg BID, tofacitinib 5 mg BID vs. TNFi, tofacitinib 10 mg BID vs. TNFi, will be presented. For this analysis, only treatment group will be used as an independent variable in the model. For the Total Time analyses changes from original randomized treatment will be ignored. For the 60-Day On-Treatment Time analyses, rules for censoring times to event due to switching from randomized treatment to standard of care are given in Section 7.

Further, multivariate analysis and analysis by TNFi subset strata (ie, the sites/regions where either adalimumab or etanercept was used, in other words, USA/Puerto Rico/Canada as one stratum, the rest as the other stratum) will be conducted to assure consistency of results. No additional analyses comparing adalimumab vs. etanercept will be conducted for the purpose of reporting the primary endpoints of the study; when reporting the primary endpoint analyses, any difference in hazard ratio between the two TNFi controls will be attributed to geographic difference.

These analyses will be used to check the robustness and to assure consistency of the results for primary safety endpoints:

- Models containing independent variable of treatment group adjusted with geographic region, age group, gender, and smoking status (see covariates in Section 6.4) will be fitted using Safety Analysis Set.

- *An analysis by TNFi subset strata (ie, the sites/regions where either adalimumab or etanercept was used)* will be performed using Safety Analysis Set.

- The same analyses conducted based on Safety Analysis Set will be conducted using Per Protocol Analysis Set.

Unadjusted, or "crude" incidence rates by Total Time, 60-Day On-Treatment Time and 28-Day On-Treatment Time analysis will be calculated as well.

8.2.2. Analyses of Other Safety Endpoints

The analyses will be based on the Safety Analysis Set. The 28-Day On-Treatment Time will be applied to all the events of interest, see Section 7.1.1.

The times to the secondary endpoints of opportunistic infection events excluding tuberculosis, tuberculosis, hepatic events, cardiovascular events (MACE and others), serious infections, GI perforations, non-melanoma skin cancers, interstitial lung disease, DVT, PE, ATE, DVT or PE, thromboembolic events (TE) and all-cause mortality will be analyzed by Cox models via univariate model analysis. See Section 8.1.1. The 28-Day On-Treatment Time analysis will be conducted.

The times to herpes zoster events will be analyzed by a multivariate Cox model with region, baseline corticosteroid use, age group, smoking status, treatment as predictors. The 28-Day On-Treatment Time analysis will be conducted.

Unadjusted, or "crude" incidence rates will be calculated as well. Note that the ratios of these rates approximate, but are not exactly equal to, the hazard ratios.

A Surveillance, Epidemiology, and End Results (SEER) analysis of the malignancy endpoint will be conducted. In this analysis, standardized incidence ratio (SIR) for malignancies represents the ratio of the incident number of cases of cancer in the tofacitinib population to the incident number that would be expected if the study population had the same incidence rate as the US population. By way of interpretation, a SIR of more than 1.0 indicates that proportionally more cases of cancer have occurred in the tofacitinib population than the comparator, whereas a SIR of less than 1.0 indicates that proportionally fewer cancer cases have occurred.

The Sponsor has standard (ie, used for all its clinical trials) summary displays and listings for MedDRA-coded adverse events, demography, medical history, length of time in study and discontinuation, concomitant and concurrent medications, physical examination, ECG, laboratory tests, weight, pulse rate, and blood pressure. These standards, called *CaPS (CDISC and Pfizer Standards)*, will be utilized in the reporting of these routine safety variables. For laboratory tests total cholesterol, LDL-c, HDL-c, total cholesterol/HDL-c ratio, LDL-c/HDL-c ratio, Triglyceride, the summary will be produced by subjects on statins, subjects not on statins and subjects on any lower lipid agents including statins.

Only treatment emergent adverse events (TEAE) will be summarized. An event is considered TEAE if the event occurs for the first time during the effective duration of treatment and is not seen prior to the start of treatment or the event is seen prior to the start of treatment but increase in severity during treatment.

No inferential analysis is pre-planned for this set of variables.

These standard summaries of routine safety and subject characteristic data will be organized according to initially randomized group, ie, tofacitinib 5 mg BID, tofacitinib 10 mg BID, tofacitinib combined, and TNFi combined.

For the summaries of TEAE, the treatment emergence for the standard of care will be derived comparing events that happen after switching to the standard of care with events that happen before switching including those after the first dose of randomized treatment.

For subjects who switch from the randomized treatment to standard of care, a separate TEAE table will be presented for these subjects according to their initially randomized group.

8.2.3. Analyses of Efficacy Endpoints

All efficacy analyses will be based on the Full Analysis Set.

Descriptive statistics of efficacy endpoints for all subjects in the study will be provided from baseline to the declared end of the study. Differences between each tofacitinib regimen (5 mg BID, 10 mg BID, or combined) and the TNFi regimen as well as between tofacitinib regimens will be estimated. The difference estimates will be derived based on the following models:

- Analysis of continuous data, ie, change from baseline in DAS28-4(CRP), SDAI, CDAI, HAQ-DI will be analyzed using 95% CI of treatment difference based on MMRM analysis. (See discussions in Section 8.1.3).

- Dichotomized data will be summarized and compared between treatment groups using the normal approximation of binomial proportions and 95% CI of treatment difference (see discussions in Section 8.1.4). This approach will be used for the following efficacy variables:

  - Rate of remission: ACR-EULAR Boolean remission, SDAI ≤3.3, CDAI ≤2.8.

  - Rate of low disease activity (LDA): SDAI ≤ 11, CDAI ≤10, DAS28-4 (CRP) ≤3.2.

  - ACR20, ACR50, and ACR70 response rate.

For efficacy analyses, the treatment groups will be tofacitinib 5 mg BID, tofacitinib 10 mg BID, and TNFi regimen.

8.2.4. Analyses of Outcome Research Endpoints

All the outcome analyses are based on the Full Analysis Set.

Descriptive statistics at baseline and each scheduled post-baseline measure from the patient-reported outcome instruments will be calculated and reported. These outcome research endpoints are described in Section 6.3.3.

In general, the approach for the outcomes research endpoints will follow that of the other continuous endpoints. For the outcome analyses, the treatment groups will be tofacitinib 5 mg BID, tofacitinib 10 mg BID, and TNFi regimen.

8.2.5. Subgroup Analyses by Age

The following endpoints will be also analyzed using subjects with age ≥65 (vs <65) years based on the Safety Analysis Set.

- All infections;
- Serious infections;
- Herpes zoster;
- Opportunistic infections;
- Tuberculosis;
- All non-serious adverse events;
- All serious adverse events;
- All deaths;
- All other adjudicated events including endpoints.

9. REFERENCES

1. *Hicks KA, Hung J, Mahaffey KW, et al. Standardized Definitions for Cardiovascular and Stroke End Point Events in Clinical Trials. Draft Definitions for Testing, 2014. Available at the website:* https://pdfs.semanticscholar.org/5f9d/51b979ce9bd4da94dc4ce3a6689bfc9890dd.pdf.

2. Daly LB, Bourke GJ, Mcgilvray J. *Interpretation and use of medical statistics*, 4th Ed. Oxford: Blackwell Scientific Publications, 1991.

3. Daly LB. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. Comp. Biol. Med. 1992; 22:351-361.

4. Breslow NE, Day NE (1987). Statistical Methods in Cancer Research. Vol. II, The Design and Analysis of Cohort Studies (IARC Scientific Publication No. 82). Lyon, France: International Agency for Research on Cancer.

5. Sahai H, Khurshid A (1996). Statistics in Epidemiology: Methods, Techniques, and Applications. Boca Raton, FL: CRC Press, Inc.

6. Sahai H, Khurshid A (1993). Confidence Intervals for the Mean of a Poisson Distribution: A Review. Biometrical J, 35: 857-67.

10. APPENDICES

Appendix 1. DATA DERIVATION DETAILS

Appendix 1.1. Definition and Use of Visit Windows in Reporting

Visit Label	Month Label	Target Day	Definition [Day window]
Screening		0	30 days prior to Day 1
Visit 1	Month 0	Day 1, Randomization	Day 1
Visit 2	Month 2	60	Days 42 to 74
Visit 3	Month 3	90	Days 75 to 135
Visit 4	Month 6	180	Days 136 to 225
Visit 5	Month 9	270	Days 226 to 315
Visit 6	Month 12	360	Days 316 to 405
Visit 7	Month 15	450	Days 406 to 495
Visit 8	Month 18	540	Days 496 to 585
Visit 9	Month 21	630	Days 586 to 675
Visit 10	Month 24	720	Days 676 to 765
Visit 11	Month 27	810	Days 766 to 855
Visit 12	Month 30	900	Days 856 to 945
Visit 13	Month 33	990	Days 946 to 1035
Visit 14	Month 36	1080	Days 1036 to 1125
Visit 15	Month 39	1170	Days 1126 to 1215
Visit 16	Month 42	1260	Days 1216 to 1305
Visit 17	Month 45	1350	Days 1306 to 1395
Visit 18	Month 48	1440	Days 1396 to 1485
Visit 19	Month 51	1530	Days 1486 to 1575
Visit 20	Month 54	1620	Days 1576 to 1665
Visit 21	Month 57	1710	Days 1666 to 1755
Visit 22	Month 60	1800	Days 1756 to 1845
Visit 23	Month 63	1890	Days 1846 to 1935
Visit 24	Month 66	1980	Days 1936 to 2025
Visit 25	Month 69	2070	Days 2026 -

Additional visits will follow the same algorithm ie, Visit Month Label increases by 3 months; Target Day increases by 90 days and the window for each visit is 90 days from the previous visit window.

For the purposes of reporting, the convention is to refer to Day 1 as the day of randomization.

For the lab values, if the calculated study day for the labeled baseline visit is not study Day 1, but falls within 40 days before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

For ECG, if there are multiple visits prior to Day 1, the value collected at the latest visit will be used as baseline.

For the other values, if the calculated study day for the labeled baseline visit is not study Day 1, but falls within 30 days before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

If two or more visits fall into the same window, keep the one closest to the Target Day. If two visits are equal distant from the Target Day in absolute value, the later visit should be used.

Safety analysis to follow CaPS (CDISC and Pfizer Standards).

For outcome research endpoints collected on a different schedule, such as SF-36, EQ-5D, and WPAI, the following visit windows will be used:

Visit Label	**Month Label**	**Target Day**	**Definition [Day window]**
Visit 1	Month 0	Day 1, Randomization	Day 1
Visit 3	Month 3	90	Days 42 to 135
Visit 4	Month 6	180	Days 136 to 270
Visit 6	Month 12	360	Days 271 to 540
Visit 10	Month 24	720	Days 541 to 900
Visit 14	Month 36	1080	Days 901 to 1260
Visit 18	Month 48	1440	Days 1261 to 1620
Visit 22	Month 60	1800	Days 1621 -

Additional visits will follow the same algorithm ie, Visit Month Label increases by 12 months; Target Day increases by 360 days and the window for each visit is 360 days from the previous visit window.

For lipid profile (fasting), the following visit windows will be used:

Visit Label	**Month Label**	**Target Day**	**Definition [Day window]**
Visit 1	Month 0	Day 1, Randomization	Day 1
Visit 2	Month 2	60	Days 42 to 210
Visit 6	Month 12	360	Days 211 to 540
Visit 10	Month 24	720	Days 541 to 900
Visit 14	Month 36	1080	Days 901 to 1260
Visit 18	Month 48	1440	Days 1261 to 1620
Visit 22	Month 60	1800	Days 1621 -

Additional visits will follow the same algorithm ie, Visit Month Label increases by 12 months; Target Day increases by 360 days and the window for each visit is 360 days from the previous visit window.

Appendix 1.2. Definition of Protocol Deviations that Relate to Statistical Analyses/Populations

None.

Appendix 1.3. Definition of Analysis Populations/Sets

None.

Appendix 1.4. Further Definition of Endpoints

Indicator	Definition/Calculation
DAS28-4 (CRP)	DAS28-4(CRP) = 0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*ln(CRP+1) + 0.014* PtGA + 0.96
Simplified Disease Activity Index (SDAI)	(TJC28) + (SJC28) + PhyGA/10 + PtGA/10 + CRP/10
Clinical Disease Activity Index (CDAI)	(TJC28) + (SJC28) + PhyGA/10 + PtGA/10
American College of Rheumatology (ACR) Response Rates	The ACR's definition for calculating a 20% improvement in RA (ACR20) is as follows: a 20% improvement in tender and swollen joint counts and 20% improvement in 3 of the 5 remaining ACR-core set measures: patient and physician global assessments, pain, disability, and an acute-phase reactant (eg, CRP). Similarly, ACR50, 70 and 90 are calculated with the respective percent improvements.
ACR/EULAR Boolean-based definition of remission	subject must satisfy all of the following: TJC28 ≤1, SJC28 ≤1, CRP ≤10, PtGA ≤10

where TJC28 is number of painful joints out of 28 joints; SJC28 is number of swollen joints out of 28 joints;
PtGA is the patient's global assessment of disease activity on a 100 mm VAS;
PhyGA is the physician's global assessment of disease activity on a 100 mm VAS;
CRP = C-reactive protein in mg/L;
ln is the natural logarithm;
sqrt is square root

Patient's Global Assessment of Health: Subjects will answer the following question, "Considering all the ways your arthritis affects you, how are you feeling today?" The subject's response will be recorded using a 100 mm visual analog scale (VAS).

Physician's Global Assessment of Health: The investigator will assess how the subject's overall arthritis appears at the time of the visit. This is an evaluation based on the subject's disease signs, functional capacity and physical examination, and should be independent of the Patient's Global Assessment of Arthritis. The investigator's response will be recorded using a 100 mm visual analog scale (VAS).

Appendix 2. STATISTICAL METHODOLOGY DETAILS

None.

Appendix 2.1. Further Details of Interim Analyses

None.

Appendix 2.2. Further Details of the Statistical Methods

Definition of the SIR

Suppose the person-time from a treatment group is allocated among M subgroups defined by the cross-classification of sex and age group. Let t_k represent the person-time for that subgroup and D_k represent the observed number of subjects with events in the kth subgroup, and let λ_k^* represent the SEER incidence rate for the kth subgroup, where $k = 1, 2, ..., M$. Given this notation, the SIR is defined as:

$$\text{SIR} = \frac{\sum_{k=1}^{M} D_k}{\sum_{k=1}^{M} t_k \lambda_k^*} = \frac{D}{E^*}$$

where the total number of subjects with events observed in the treatment group is $D = \sum_{k=1}^{M} D_k$, and the expected total number of subjects with events is $E^* = \sum_{k=1}^{M} E_k^* = \sum_{k=1}^{M} t_k \lambda_k^*$ (Breslow and Day, 1987;[4] Sahai and Khurshid, 1996).[5]

Exact Confidence Limits for the True SIR

Sahai and Khurshid (1993,[6] 1996)[5] discuss the following method for obtaining the exact confidence limits, SIR_L and SIR_U, for the true SIR, ϕ. Assuming D is Poisson distributed with mean $\mu = E(D)$, confidence limits for μ are obtained using the relationship between the Poisson distribution and the chi-square distribution. Then these limits are divided by the expected total number of subjects with events, E^*, to obtain the limits.

$$\text{SIR}_L = \frac{\chi^2_{2D,\alpha/2}}{2E^*} \quad \text{and} \quad \text{SIR}_U = \frac{\chi^2_{2(D+1),1-\alpha/2}}{2E^*}$$

where $\chi^2_{v,\alpha}$ is the $100^*\alpha$ percentile of the chi-square distribution with v degrees of freedom.