

Protocol A3921192

A PHASE 3B/4 RANDOMIZED DOUBLE BLIND PLACEBO CONTROLLED STUDY
OF METHOTREXATE (MTX) WITHDRAWAL IN SUBJECTS WITH RHEUMATOID
ARTHRITIS (RA) TREATED WITH TOFACITINIB 11MG MODIFIED RELEASE (MR)
FORMULATION

Statistical Analysis Plan
(SAP)

Version: 4.0

Date: 09-Nov-2018

TABLE OF CONTENTS

LIST OF TABLES 4

LIST OF FIGURES 4

1. VERSION HISTORY 5

2. INTRODUCTION 7

 2.1. Study Objectives 7

 2.1.1. Primary Objectives 7

 2.1.2. Secondary Objectives 7

 2.1.3. Other Objectives 8

 2.1.4. Exploratory Objectives 8

 2.2. Study Design 8

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS 9

 3.1. Primary Efficacy Endpoint(s)..... 9

 3.2. Secondary Efficacy Endpoint(s)..... 9

 3.3. Other Efficacy Endpoints 10

 3.4. Exploratory Endpoints..... 10

 3.5. Baseline Variables 10

 3.6. Safety Endpoints 12

 3.6.1. Adverse Events 12

 3.6.2. Laboratory Data 12

4. ANALYSIS SETS 12

 4.1. Full Analysis Set 12

 4.2. Per Protocol Analysis Set..... 13

 4.3. Safety Analysis Set..... 13

 4.4. Other Analysis Sets 13

5. GENERAL METHODOLOGY AND CONVENTIONS..... 13

 5.1. Hypotheses and Decision Rules 13

 5.2. General Methods 14

 5.2.1. Analyses for Binary Data..... 14

 5.2.2. Analyses for Continuous Data 14

 5.2.3. Analyses for Categorical Data 14

 5.3. Methods to Manage Missing Data 15

5.3.1. Joint Counts	15
5.3.2. DAS28-4(ESR), DAS28-4(CRP), CDAI, SDAI	15
5.3.3. LDA and Remission	16
5.3.4. ACR20, ACR50, ACR70.....	16
5.3.5. Patient Reported Outcomes	17
6. ANALYSES AND SUMMARIES	17
6.1. Primary Endpoint(s).....	17
6.1.1. Change in DAS28-4 (ESR) Score from Randomization (at Week 24) to the End of Double-Blind MTX Withdrawal Phase (at Week 48).....	17
6.1.1.1. Primary Analysis	17
6.1.1.2. Sensitivity/Robustness Analyses	18
6.2. Secondary Endpoint(s) and Other Endpoints in the Double-Blind Phase.....	19
6.2.1. Change in the DAS28-4(ESR) from Week 24 to Week 36	19
6.2.2. Changes in the DAS28-4 (CRP) from Week 24 to Week 48 and from Week 24 to Week 36.....	19
6.2.3. Other Continuous Endpoints in the Double-Blind Phase	20
6.2.4. LDA as assessed by DAS28-4(ESR) \leq 3.2 at Weeks 48 and 36	21
6.2.5. Other Binary Endpoints in the Double-Blind Phase.....	21
6.2.6. ACR 20, 50, 70 at Weeks 48 and 36.....	21
6.3. Other Endpoint(s).....	22
6.3.1. Changes in DAS28-4 (ESR) from Day 1 to Week 12 and from Day 1 to Week 24	22
6.3.2. Other Continuous Endpoints in Open-Label Run-in Phase.....	22
6.3.3. LDA as assessed by DAS28-4(ESR) \leq 3.2 at Weeks 12 and 24	23
6.3.4. Other Binary Endpoint in Open-Label Run-in Phase	23
6.3.5. ACR 20, 50, and 70 at Weeks 12 and 24.....	23
6.4. Exploratory Endpoints.....	24
6.5. Subset Analyses.....	24
6.6. Baseline and Other Summaries and Analyses	24
6.6.1. Baseline Summaries.....	24
6.6.2. Study Conduct and Subject Disposition	24
6.6.3. Study Treatment Exposure	24
6.6.4. Concomitant Medications and Non-Drug Treatments.....	24

6.7. Safety Summaries and Analyses	25
6.7.1. Adverse Events	25
6.7.2. Laboratory Data	25
6.7.3. Vital Signs	25
6.7.4. Physical Examination	25
6.7.5. ECG	25
CCI	25
8. REFERENCES	26
9. APPENDICES	27

LIST OF TABLES

Table 1. Summary of Major Changes in SAP Amendments	5
-----------------------------------------------------------	---

LIST OF FIGURES

Figure 1. Study Design.....	9
-----------------------------	---

APPENDICES

Appendix 1. DATA DERIVATION DETAILS	27
Appendix 1.1. Endpoints Definition and Derivation	27
Appendix 1.2. Definition and Use of Visit Windows in Reporting.....	28

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study A3921192 is based on Protocol Amendment 2 dated 04-Dec-2017.

Table 1. Summary of Major Changes in SAP Amendments

SAP Version	Change	Rationale
1	Not Applicable	Not Applicable
2	<ul style="list-style-type: none"> a. Section 7 was changed for analysis of openlabel phase b. Window of open-label Week 24 was changed as “Days 127 – Discontinuation Day /randomization (early one)” 	Based on Protocol Administrative Change Letter 05-Jul-2017
3	<ul style="list-style-type: none"> a. In Section 2.2, based on recent calculations, the number of subjects was increased from 580 to 680 for the open-label phase in order to achieve at least 232 subjects in the double-blind phase. b. The LDA assessment was changed from DAS28-4(ESR) < 3.2 and DAS28-4(CRP) < 3.2 to DAS28-4(ESR) ≤ 3.2 and DAS28-4(CRP) ≤ 3.2, in Sections 3.2, 3.3, 5.3.3, 6.2.4, 6.2.5, 6.3.3, 6.3.4 and Appendix 1.1. c. Change in tender/painful joint counts, swollen joint counts, subject assessment of arthritis pain, subject global assessment of arthritis, physician global assessment of arthritis, and CRP were added as other efficacy endpoints in Sections 3.3, 6.2.3 and 6.3.2. d. Analysis for exploratory endpoint RAPID3 was added in Section 6.4. e. Baseline Disease Characteristics of 	Based on Protocol Amendment 2 04-Dec-2017

	<p>prior use of tsDMARD, baseline MTX dose, baseline steroid dose, ESR, CRP, tender joint counts, swollen joint counts, pain VAS, subject global assessment, physician global assessment, DAS28-4(ESR), DAS28-4(CRP), HAQ-DI, CDAI, and SDAI were added in Section 3.5.</p> <p>f. “Line plots of the LS means and 95% CI” was changed to “Line plots of the LS means (+/- SE)” in Section 6.</p> <p>g. “Line plots of the percentage and 95% CI” was changed to “Line plots of the percentage (+/- SE)” in Section 6.</p> <p>h. Observed cases were added for analyses in Section 6.</p> <p>i. Summary of ECG was added in Section 6.7.5.</p> <p>j. Analysis window for Week 6 was added in Appendix 1.2.</p> <p>k. Updated analysis window for Baseline in Appendix 1.2.</p>	
4	<p>a. Added details for MMRM when the model with unstructured covariance structure doesn’t converge in Section 5.2.2.</p> <p>b. Added details for NRI in Section 5.3.3 and Section 5.3.5.</p> <p>c. Removed “Canada” in the region definition in Section 6.5.</p> <p>d. Added details for subgroup MMRM model in Section 6.5.</p> <p>e. Added details for windowing in</p>	To add analysis details based on BDR1 discussions

	<p>Appendix 1.2.</p> <p>f. Updated the window for Randomization in Appendix 1.2</p> <p>g. Corrected typos in the document.</p>	
--	--------------------------------------------------------------------------------------------------------------------------------	--

2. INTRODUCTION

The tofacitinib Immediate Release (IR) 5mg twice daily (BID) formulation has demonstrated efficacy both as a monotherapy and in combination with methotrexate (MTX) in previous trials. However, it remains unknown whether the tofacitinib efficacy will be sustained after the withdrawal of MTX among RA patients who achieve low disease activity following treatment with MTX plus tofacitinib. Results of the current study will inform the value of the continued MTX treatment among subjects who have responded to combined therapy of tofacitinib and MTX. A modified release (MR) 11mg once daily (QD) regimen has the potential to offer an additional dosing option to subjects, and based on the equivalence of key pharmacokinetic parameters, efficacy and safety are expected to be consistent with that demonstrated for tofacitinib IR 5 mg BID. Therefore, the current study will evaluate the sustained efficacy and safety of the tofacitinib MR 11 mg QD regimen among rheumatoid arthritis (RA) patients undergoing blinded withdrawal of MTX.

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study A3921192. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives

2.1.1. Primary Objectives

To compare the efficacy of tofacitinib MR 11mg monotherapy to tofacitinib MR 11mg with continued MTX, as measured by the change in the Disease Activity Score utilizing 4 components including erythrocyte sedimentation rate (DAS28-4 (ESR)) from randomization (at Week 24) to the end of the double-blind MTX withdrawal phase (at Week 48).

2.1.2. Secondary Objectives

- *To compare the efficacy of tofacitinib MR 11mg monotherapy to tofacitinib MR 11mg with continued MTX in the double-blind MTX withdrawal phase (at Week 36), as measured by DAS28-4 (ESR).*
- *To compare the efficacy of tofacitinib MR 11mg monotherapy to tofacitinib MR 11mg with continued MTX in the double-blind MTX withdrawal phase (at Weeks 48 and 36), as measured by Disease Activity Score 28-4 (C reactive protein) (DAS28-4 (CRP)), Clinical Disease Activity Index (CDAI), Simplified Disease Activity Index (SDAI), Low Disease Activity (LDA), remission, ACR20, ACR 50, and ACR 70.*

- *To compare effects of tofacitinib MR 11mg monotherapy to tofacitinib MR 11mg with continued MTX on health outcome measures in the double-blind MTX withdrawal phase (at Weeks 48 and 36), as measured by Health Assessment Disability Questionnaire-Disability Index (HAQ-DI), Medical Outcomes Survey Short Form -36 (SF-36), Work Productivity and Activity Impairment (WPAI), European Quality of Life 5 dimensions questionnaire (EuroQol EQ 5D), and Functional Assessment of Chronic Illness Therapy (FACIT)-Fatigue scale.*
- *To evaluate the safety and tolerability of tofacitinib MR 11mg monotherapy versus tofacitinib MR 11mg with continued MTX.*

2.1.3. Other Objectives

- *To describe efficacy of treatment with tofacitinib 11mg with MTX in the open-label run-in phase (at Weeks 12 and 24), as measured by DAS28-4 (ESR), DAS28-4 (CRP), CDAI, SDAI, LDA, remission, ACR20, ACR 50 and ACR70.*
- *To describe effects of treatment with tofacitinib MR 11mg with MTX on health outcome measures in the open-label run-in phase (at Weeks 12 and 24), as measured by HAQ-DI, SF-36, WPAI, EuroQol EQ-5D, and FACIT-Fatigue scale.*

2.1.4. Exploratory Objectives

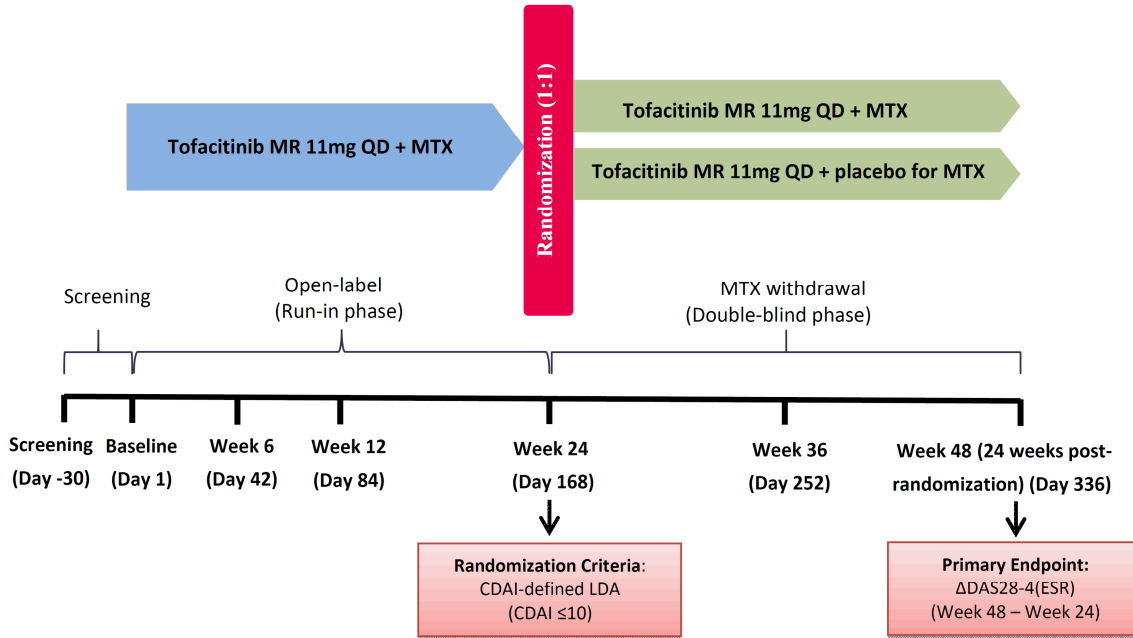
- *To collect exploratory biomarker/genomics samples for bio-banking.*
- *US and Canada ONLY: To explore effects of treatment with tofacitinib MR 11mg with/without MTX on health outcome measure in the open-label run-in phase (at Weeks 12 and 24) and the double-blind MTX withdrawal phase (at Weeks 48 and 36), as measured by Routine Assessment of Patient Index Data 3 (RAPID3).*

2.2. Study Design

Study A3921192 is a 12-month randomized, double-blind, placebo-controlled withdrawal study in a total of approximately 680 subjects with moderate to severe RA (defined as CDAI >10 and DAS28 (ESR) \geq 3.2) who are insufficiently responding to their stable dose of MTX treatment. In addition, the proportion of subjects who have had an insufficient response to one or more biologics prior to the enrollment should be \leq 25% (ie, \leq 145 subjects). Following enrollment, all subjects will receive open-label tofacitinib MR 11mg QD added to MTX at their previously stabilized dose for 24 weeks (run-in phase). At the end of the run-in phase (Week 24), only subjects who achieve low disease activity (LDA) as assessed by CDAI (ie, CDAI \leq 10) (approximately 232 subjects) will be randomized into the 24-week double-blind, placebo-controlled MTX withdrawal phase. Subjects who do not achieve LDA (defined by CDAI >10) at this time point will be discontinued from the study. Subjects entering the double-blind phase will be randomized in a 1:1 ratio to either continue the run-in phase treatment regimen (tofacitinib MR 11mg QD with blinded MTX) or to receive tofacitinib monotherapy (tofacitinib MR 11mg QD with blinded matching placebo for MTX). Randomization will be stratified by subject's prior exposure to a biologic DMARD (bDMARD). Randomized subjects will be followed up to the end of Week 48. The primary

efficacy endpoint is the change in DAS28-4 (ESR) from randomization (at Week 24) to the end of double-blind MTX withdrawal phase (at Week 48).

Figure 1. Study Design



3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Efficacy Endpoint(s)

Change in DAS28-4 (ESR) score from randomization (at Week 24) to the end of double-blind MTX withdrawal phase (at Week 48).

3.2. Secondary Efficacy Endpoint(s)

- Change in the DAS28-4(ESR) from Week 24 to Week 36;
- Changes in the DAS28-4 (CRP), CDAI and SDAI, respectively, from Week 24 to Week 48 and from Week 24 to Week 36;
- LDA as assessed by DAS28-4(ESR) ≤ 3.2, DAS28-4(CRP) ≤ 3.2, CDAI ≤ 10 and SDAI ≤ 11, respectively, at Weeks 48 and 36;
- Remission as assessed by ACR-EULAR Boolean remission criteria, DAS28-4 (ESR) < 2.6, DAS28-4 (CRP) < 2.6, CDAI ≤ 2.8 and SDAI ≤ 3.3, respectively, at Weeks 48 and 36;
- ACR20, ACR 50, and ACR70 responses, respectively, at Weeks 48 and 36;

- Change in the HAQ-DI, the SF-36 (8 domain scores and 2 component scores), WPAI, EuroQol EQ-5D and the FACIT-Fatigue scale score, respectively, from Week 24 to Week 48 and from Week 24 to Week 36;
- HAQ-DI response (ie, decrease of at least 0.22) at Weeks 48 and 36.
- Note:
- For subjects entering the double-blind phase, the value at randomization (Week 24) for each endpoint is the last nonmissing observation before the first dose of study drug in double-blind phase.
- The baseline (Day 1) value for each endpoint is the last nonmissing observation before the first dose of study drug in open-label run-in phase.
- ACR20, ACR 50, ACR70, and HAQ-DI responses are relative to baseline (Day 1).

3.3. Other Efficacy Endpoints

- Changes in DAS28-4 (ESR), DAS28-4 (CRP), CDAI, SDAI, HAQ-DI, SF-36 (8 domain scores and 2 component scores), WPAI, EuroQol EQ-5D score and FACIT-Fatigue scale score, respectively, from day 1 to Week 12 and from Day 1 to Week 24;
- LDA as assessed by $\text{DAS28-4(ESR)} \leq 3.2$, $\text{DAS28-4(CRP)} \leq 3.2$, $\text{CDAI} \leq 10$ and $\text{SDAI} \leq 11$, respectively, at Weeks 12 and 24;
- Remission as assessed by ACR-EULAR Boolean remission criteria, $\text{DAS28-4 (ESR)} < 2.6$, $\text{DAS28-4 (CRP)} < 2.6$, $\text{CDAI} \leq 2.8$ and $\text{SDAI} \leq 3.3$, respectively, at Weeks 12 and 24;
- ACR20, ACR 50, and ACR70 responses, respectively, at Weeks 12 and 24;
- HAQ-DI response (ie, decrease of at least 0.22) at Weeks 12 and 24.
- Change in tender/painful joint counts, swollen joint counts, subject assessment of arthritis pain, subject global assessment of arthritis, physician global assessment of arthritis, and CRP, respectively, from Day 1 to Week 12 and from Day 1 to Week 24, from Week 24 to Week 48 and from Week 24 to Week 36.

3.4. Exploratory Endpoints

- *US and Canada ONLY: RAPID3 summary scores at Weeks 12, 24, 36 and 48.*

3.5. Baseline Variables

Demographic and baseline characteristics include:

- Age (in years);

- Sex;
- Race;
- Region;
- Height (in cms);
- Weight (in kg);
- Body Mass Index;
- Duration of disease (in years);
- Rheumatoid Factor (positive/negative);
- Anti-cyclic citrullinated peptide (Anti-CCP) antibody (positive/negative);
- Prior use of bDMARD (yes/no);
- Prior use of TNFi (yes/no);
- Prior use of MTX (yes/no);
- Prior use of csDMARD (yes/no);
- Prior use of tsDMARD (yes/no);
- Prior use of corticosteroid (yes/no);
- Baseline MTX dose;
- Baseline Steroid Dose;
- Disease activity (moderate or severe based on DAS28-4(ESR) at baseline);
- DAS28-4(ESR);
- DAS28-4(CRP);
- ESR;
- CRP;
- Tender Joint Counts;
- Swollen joint Counts;
- Subject global assessment;
- Physician global assessment;
- Pain VAS;
- HAQ-DI;
- CDAI;
- SDAI.

Note: Prior use of bDMARD will be used in the mixed-effect model of repeated measures (MMRM) model as a covariate. It will be based on the CRF data.

3.6. Safety Endpoints

3.6.1. Adverse Events

- *All adverse events (AEs), including serious adverse events (SAEs).*
- Adjudicated safety events

An AE is considered treatment emergent relative to a given treatment if:

- the event occurs for the first time during the effective duration of treatment and was not seen prior to the start of treatment, or
- the event was seen prior to the start of treatment but increased in severity during treatment.

The effective duration of treatment is determined by the lag time. Any event occurring within 28 days, whether this occurs during a break in treatment or at the end of treatment, is attributed to the corresponding treatment period.

3.6.2. Laboratory Data

- *Clinically significant abnormal laboratory parameters.*
- Clinical laboratory values and change from baseline at Weeks 12 and 24.
- Clinical laboratory values and change from baseline and randomization at Weeks 36 and 48.

4. ANALYSIS SETS

4.1. Full Analysis Set

There are two full analysis sets, one is for open-label run-in phase and one is for double-blind MTX withdrawal phase.

- **Open-Label Period Full Analysis Set (FAS-OL)**

FAS-OL includes all subjects who received at least one dose of tofacitinib MR 11mg plus MTX during open-label phase.

- **Double-Blind Period Full Analysis Set (FAS-DB)**

FAS-DB includes all subjects who received at least one dose of tofacitinib MR 11mg plus MTX during the open-label run-in phase, and were randomized and received at least one dose of the randomized investigational drug regimen (tofacitinib MR 11mg with MTX or tofacitinib MR 11mg with placebo for MTX) during the blinded MTX withdrawal phase.

A randomized but not treated subject, or treated but not randomized subject during MTX withdrawal phase will be excluded from FAS-DB.

A subjects who was randomized but took incorrect treatment during MTX withdrawal phase will be reported under the randomized treatment group for all efficacy analyses.

4.2. Per Protocol Analysis Set

The Per Protocol Analysis Set (PPAS) will be a subset of subjects from FAS-DB. Subjects who had a protocol deviation during the blinded MTX withdrawal period thought to affect the efficacy analysis will be excluded from the Per Protocol efficacy analysis. The PPAS will be defined for double-blind period only. Note that the protocol deviation may have occurred during open-label period. Protocol deviations will be assessed by the study team prior to unblinding the study. This list of subjects along with exclusions from per protocol analyses will be put into the trial master file.

4.3. Safety Analysis Set

There are three safety analysis sets, two for two study phases (open-label run-in phase and double-blind MTX withdrawal phase) and one for overall study period.

- **Open-Label Run-in Phase Safety Analysis Set (Safety-OL)**

Safety-OL will include all subjects who received at least one dose of tofacitinib MR 11mg QD plus MTX during open-label run-in period. It is same as FAS-OL.

- **Double-Blind MTX Withdrawal Phase Safety Analysis Set (Safety-DB)**

Safety-DB will include all subjects who received at least one dose of study drug (tofacitinib MR 11mg QD with blinded MTX or tofacitinib MR 11mg QD with blinded matching placebo for MTX) during double-blind MTX withdrawal phase.

- **Overall Study Safety Analysis Set (Safety-All)**

Safety-All consists of all subjects who received at least one dose of study drug during the study.

4.4. Other Analysis Sets

N/A.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

The primary analysis will be a non-inferiority test to show that tofacitinib MR 11mg QD monotherapy is not less effective (non-inferior) than tofacitinib MR 11mg QD with MTX within the margin of non-inferiority set at 0.6 (ie, the difference in mean change from randomization in DAS28-4 (ESR) between the two arms). If the upper bound of the 95% two-sided confidence interval (CI) for the difference (tofacitinib monotherapy arm minus tofacitinib with MTX arm) in change from randomization to Week 48 in DAS 28-4 (ESR)

between the two arms is lower than 0.6, then tofacitinib monotherapy will be declared to be non-inferior to the tofacitinib with continued MTX treatment.

5.2. General Methods

In general, number and percent will be presented for categorical 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 will be used to present the data.

The data will be described and analyzed by study phase.

Note that the primary statistical hypotheses are based on data solely from blinded MTX withdrawal phase. Analyses of endpoints during open-label run-in phase are intended to be descriptive in nature.

5.2.1. Analyses for Binary Data

Binary endpoints in the blinded MTX withdrawal phase will be analyzed using the normal approximation for the difference in binomial proportions. The estimated treatment difference, the associated 95% CI and p-value will be presented.

Binary endpoints in the open-label run-in phase will be summarized using number, percentage and 95% CI of percentage.

5.2.2. Analyses for Continuous Data

The primary endpoint of change from randomization in DAS28-4(ESR) at Week 48 will be analyzed using a linear mixed-effect model of repeated measures (MMRM) that includes the fixed effects of treatment, visit, treatment-by-visit interaction, prior use of a biological DMARD, and baseline DAS28-4 (ESR) value as a covariate. Within-subject variability will be accounted for using a random effect with an unstructured covariance matrix. The least square (LS) mean of treatment, and LS mean difference and the associated 95% confidence interval (CI) will be presented.

If the model with an unstructured covariance matrix doesn't converge or is unstable, 1st order autoregressive will be fitted instead. If the convergence problem persists, compound symmetry will be used.

Continuous secondary endpoints in the blinded MTX withdrawal phase will be analyzed using similar methods as the primary endpoint.

Continuous endpoints in the open-label run-in phase will be descriptively summarized using number, mean, standard deviation (standard error of the mean), minimum, 1st, 2nd and 3rd quartiles and maximum.

5.2.3. Analyses for Categorical Data

Categorical endpoints will be summarized using number and percentage.

5.3. Methods to Manage Missing Data

For continuous endpoints, missing values will not be imputed for descriptive statistics. For continuous endpoints in the blinded MTX withdrawal phase, the missing values post-randomization will be handled in a linear mixed-effect model with repeated measures for this continuous variable, where the values are assumed to be missing at random.

Multiple Imputation will be used for sensitivity analysis for DAS28-4(ESR) change from randomization (Week 24) in the double-blind phase. Imputed values will be sampled (or randomly drawn) from the normal distribution based on the observed mean and SD of non-missing data by treatment and time point and the full analysis will be conducted on the 'filled-in' (observed and imputed together) data multiple times (500 times). The final results will be presented as a summary of these multiple analyses.

In addition, missing values for safety endpoints will not be imputed.

5.3.1. Joint Counts

- Tender joint count (TJC)

TJC is defined as the number of tender joints with abnormality. In case of missing joint assessment, the TJC will be prorated as follows:

$28 * (\text{number of joints with tender score as "Present"}) / (\text{number of non-missing tender joints})$.

Joints with a response of "Not Done" or "Not Applicable" will be handled as missing. If less than 20 joints have non-missing tender scores, then TJC will be defined as missing.

- Swollen joint count (SJC)

SJC is defined as the number of swollen joints with abnormality. In case of missing joint assessment, the SJC will be prorated as follows:

$28 * (\text{number of joints with swollen score as "Present"}) / (\text{number of non-missing swollen joints})$.

Joints with a response of "Not Done" or "Not Applicable" will be handled as missing. If less than 20 joints have non-missing swollen scores, then SJC will be defined as missing.

5.3.2. DAS28-4(ESR), DAS28-4(CRP), CDAI, SDAI

DAS28-4(ESR) is a derived variable from 4 components. If any component is missing, DAS28-4(ESR) will be missing.

Above approach will also be used for DAS28-4(CRP), CDAI, SDAI.

5.3.3. LDA and Remission

The binary variables derived from the continuous endpoints (such as DAS28-4 (ESR)) will be imputed using Non-Responder Imputation (NRI) method. Missing values due to a patient dropping from the study for any reason (eg, lack of efficacy or adverse event) will be handled by setting the the response value to nonresponsive from that visit onward. This also means that if a patient withdraws at a visit but had responsive values at that visit, the binary variable is still set to nonresponsive.

NRI approach will be used to impute the missing data of LDA as assessed by:

- DAS28-4(ESR) ≤ 3.2 ;
- DAS28-4(CRP) ≤ 3.2 ;
- CDAI ≤ 10 ;
- SDAI ≤ 11 .

and Remission as assessed by:

- ACR-EULAR Boolean remission criteria;
- DAS28-4 (ESR) < 2.6 ;
- DAS28-4 (CRP) < 2.6 ;
- CDAI ≤ 2.8 ; and
- SDAI ≤ 3.3 .

For subjects who don't enter the blinded MTX withdrawal phase, the imputation is only applied up to Week 24.

5.3.4. ACR20, ACR50, ACR70

Because the ACR20, ACR50, ACR70 variables are based on several component variables, it is possible that the values may still be calculated even if the component variables have some missing values. In this case, no imputation method is needed. If the ACR value is still missing, an imputation method will be applied.

If the ACR value is missing due to missing values in any of the components, while the patient 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 values of ACR20, ACR50 and ACR70 will be determined.

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 the LOCF imputation has been applied, missing values due to a patient dropping from the study for any reason (eg, lack of efficacy or adverse event) will be handled by setting the ACR value (ACR20, ACR50 and ACR70) to nonresponsive from that visit onward. This also means that if a patient withdraws at a visit but had responsive ACR values at that visit, the ACR value is still set to nonresponsive. This method also goes by the name Non-Responder Imputation, NRI.

It is also the case that LOCF method does not carry forward any baseline values, as a convention. That is, a value carried forward must be a post-baseline value. If there are no post-baseline values that are non missing, the component remains missing.

Regardless of any imputation or none, if for some reason the ACR value still can't be determined, including the case where baseline data is missing, then its value will be set to non-responder.

For subjects who don't enter the blinded MTX withdrawal phase, the imputation is only applied up to Week 24.

5.3.5. Patient Reported Outcomes

For the HAQ-DI, SF-36, WPAI, EuroQol EQ-5D and the FACIT-Fatigue, rules suggested by the producers of these will be followed in calculating the values. If these rules are not enough for imputing a value, then the value is missing.

NRI approach (see [Section 5.3.3](#)) will be used to impute the missing data of HAQ-DI response (ie, decrease of at least 0.22).

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint(s)

6.1.1. Change in DAS28-4 (ESR) Score from Randomization (at Week 24) to the End of Double-Blind MTX Withdrawal Phase (at Week 48)

6.1.1.1. Primary Analysis

- Analysis time points: Week 48
- Analysis population: FAS-DB
- Analysis methodology: Change from randomization (at Week 24) will be analyzed using the MMRM (specified in [Section 5.2.2](#)).
- Covariate: Baseline of DAS28-4(ESR) and prior use of a bDMARD.
- Imputation for missing values: Observed data will be included in the MMRM model, the missing values will be handled in the model, where the values are assumed to be missing at random.
- Supporting objective and Decision rule: Primary Objective, upper bound of 95% CI of treatment difference < 0.6.

Reporting results:

- Raw data: The sample size, mean, standard deviation, median, minimum and maximum at randomization (Week 24) and post-randomization visits will be presented for each treatment arm.
- Change from randomization: The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm. The LS means, 95% confidence interval for the LS means, difference between the LS means between treatment groups and the corresponding 95% CI, p-value will be presented.

Figures

- Line plots of the LS means (+/-SE) at Weeks 48 and 36 for two treatment groups.
- Plot of 95% CI of the LS mean difference between treatment groups at Weeks 48 and 36.

6.1.1.2. Sensitivity/Robustness Analyses

To support the interpretation of the primary analysis the following analyses will be performed:

6.1.1.2.1. Analysis for PPAS

- Dependent Variable: Change in the DAS28-4(ESR) from Week 24 to Week 48
- Analysis time points: Week 48
- Analysis population: PPAS
- Analysis methodology: Change from Week 24 will be analyzed using the MMRM (specified in [Section 5.2.2](#)).
- Covariate: Baseline of DAS28-4(ESR) and prior use of a bDMARD.
- Imputation for missing values: Observed data will be included in the MMRM model, the missing values will be handled in the model, where the values are assumed to be missing at random.
- Supporting objective and Decision rule: Primary Objective, upper bound of 95% CI of treatment difference <0.6.

Reporting results:

- Raw data: The sample size, mean, standard deviation, median, minimum and maximum at randomization (Week 24) and post-randomization visits will be presented for each treatment arm.

- Change from randomization: The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm. The LS means, 95% confidence interval for the LS means, difference between the LS means between treatment groups, the corresponding 95% CI and p-value will be presented.

Figures

- Line plots of the LS means (+/-SE) at Weeks 48 and 36 for two treatment groups.
- Plot of 95% CI of the LS mean difference between treatment groups at Weeks 48 and 36.

6.1.1.2.2. Multiple Imputations

- Dependence Variable: Change in the DAS28-4(ESR) from Week 24 to Week 48
- Analysis population: FAS-DB
- Analysis time point: Week 48
- Analysis methodology: Change from Week 24 will be analyzed using the MMRM (specified in [Section 5.2.2](#)).
- Covariate: Baseline of DAS28-4(ESR) and prior use of a biological DMARD
- Imputation for missing values: multiple imputation (see [Section 5.3](#)).

Reporting results:

- Change from Week 24: The LS means, 95% confidence interval for the LS means, difference between the LS means between treatment groups and the corresponding 95% CI will be presented.
- Plot of 95% CI of the LS mean difference between treatment groups at Weeks 48 and 36.

6.2. Secondary Endpoint(s) and Other Endpoints in the Double-Blind Phase

6.2.1. Change in the DAS28-4(ESR) from Week 24 to Week 36

All analyses performed for Week 48 will be performed for Week 36. See [Section 6.1.1](#).

6.2.2. Changes in the DAS28-4 (CRP) from Week 24 to Week 48 and from Week 24 to Week 36

- Analysis time points: Week 48 and Week 36
- Analysis population: FAS-DB
- Analysis methodology: Change from Week 24 will be analyzed using the MMRM (specified in [Section 5.2.2](#)).
- Covariate: Baseline of DAS28-4(CRP) and prior use of a bDMARD

- Imputation for missing values: observed data will be included in the MMRM model, the missing values will be handled in the model, where the values are assumed to be missing at random.

Reporting results:

- Raw data: The sample size, mean, standard deviation, median, minimum and maximum at randomization (Week 24) and post-randomization visits will be presented for each treatment arm.
- Change from Week 24: The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm. The LS means, 95% confidence interval for the LS means, difference between the LS means between treatment groups, the corresponding 95% CI and p-value will be presented.

Figures

- Line plots of the LS means (+/-SE) at Weeks 48 and 36 for two treatment groups.
- Plot of 95% CI of the LS mean difference between treatment groups at Weeks 48 and 36.

6.2.3. Other Continuous Endpoints in the Double-Blind Phase

The same analysis in [Section 6.2.2](#) will be performed for changes from Week 24 to Week 48 and from Week 24 to Week 36 in the:

- CDAI;
- SDAI;
- Tender/painful joint counts;
- Swollen joint counts;
- Subject assessment of arthritis pain VAS;
- Subject global assessment of arthritis;
- Physician global assessment of arthritis;
- CRP;
- HAQ-DI;
- SF-36 (8 domain scores and 2 component scores);
- WPAI;
- EuroQol EQ-5D;
- FACIT-Fatigue scale score;

- RAPID3 (US and Canada ONLY).

6.2.4. LDA as assessed by DAS28-4(ESR) \leq 3.2 at Weeks 48 and 36

- Analysis time points: Week 48 and Week 36;
- Analysis population: FAS-DB;
- Analysis methodology: normal approximation (specified in [Section 5.2.1](#));
- Imputation for missing values: NRI ([Section 5.3.3](#)), observed cases.

Reporting results:

- The sample size, frequency and percentage of responders at post-randomization visits will be presented for each treatment arm difference between treatment groups and the corresponding 95% CI and p-value will be presented.

Figures

- Line plots of the percentage (+/-SE) at Weeks 48 and 36 for two treatment groups.

6.2.5. Other Binary Endpoints in the Double-Blind Phase

The same analysis in [Section 6.2.4](#) will be performed at Weeks 48 and 36 for

- LDA as assessed by DAS28-4(CRP) \leq 3.2;
- LDA as assessed CDAI \leq 10;
- LDA as assessed SDAI \leq 11;
- Remission as assessed by ACR-EULAR Boolean remission criteria;
- Remission as assessed by DAS28-4 (ESR) $<$ 2.6;
- Remission as assessed by DAS28-4 (CRP) $<$ 2.6;
- Remission as assessed by CDAI \leq 2.8;
- Remission as assessed by SDAI \leq 3.3;
- HAQ-DI response (ie, decrease of at least 0.22).

6.2.6. ACR 20, 50, 70 at Weeks 48 and 36

- Analysis time points: Week 48 and Week 36;
- Analysis population: FAS-DB;
- Analysis methodology: normal approximation (specified in [Section 5.2.1](#));
- Imputation for missing values: NRI ([Section 5.3.4](#)), observed cases

Reporting results:

- The sample size, frequency and percentage of responders at post-randomization visits will be presented for each treatment arm difference between treatment groups and the corresponding 95% CI and p-value will be presented.

Figures

- Line plots of the percentage (+/-SE) at Weeks 48 and 36 for two treatment groups.

6.3. Other Endpoint(s)

6.3.1. Changes in DAS28-4 (ESR) from Day 1 to Week 12 and from Day 1 to Week 24

- Analysis time points: Week 12 and Week 24
- Analysis population: FAS-OL
- Analysis methodology: descriptive summary.
- Imputation for missing values: no imputation.

Reporting results:

- Raw data: The sample size, mean, standard deviation, median, minimum and maximum at baseline will be presented.
- Change from Day 1: The sample size, mean, standard deviation, median, minimum and maximum will be presented.

6.3.2. Other Continuous Endpoints in Open-Label Run-in Phase

The same analysis in [Section 6.3.1](#) will be performed for changes from Day 1 to Week 12 and from Day 1 to Week 24 in the

- DAS28-4 (CRP);
- CDAI;
- SDAI;
- Tender/painful joint counts;
- Swollen joint counts;
- Subject assessment of arthritis pain VAS;
- Subject global assessment of arthritis;
- Physician global assessment of arthritis;
- CRP;
- HAQ-DI;

- SF-36 (8 domain scores and 2 component scores);
- WPAI;
- EuroQol EQ-5D;
- FACIT-Fatigue scale score;
- RAPID3 (US ONLY).

6.3.3. LDA as assessed by DAS28-4(ESR) \leq 3.2 at Weeks 12 and 24

- Analysis time points: Week 12 and Week 24;
- Analysis population: FAS-OL;
- Analysis methodology: descriptive summary;
- Imputation for missing values: NRI ([Section 5.3.3](#)), observed cases.

Reporting results:

- The sample size, frequency, percentage of responders, and 95% CI of percentage will be presented.

6.3.4. Other Binary Endpoint in Open-Label Run-in Phase

The same analysis in [Section 6.3.3](#) will be performed at Weeks 12 and 24 for

- LDA as assessed by DAS28-4(CRP) \leq 3.2
- LDA as assessed CDAI \leq 10
- LDA as assessed SDAI \leq 11
- Remission as assessed by ACR-EULAR Boolean remission criteria
- Remission as assessed by DAS28-4 (ESR) $<$ 2.6
- Remission as assessed by DAS28-4 (CRP) $<$ 2.6
- Remission as assessed by CDAI \leq 2.8
- Remission as assessed by SDAI \leq 3.3
- HAQ-DI response (ie, decrease of at least 0.22)

6.3.5. ACR 20, 50, and 70 at Weeks 12 and 24

- Analysis time points: Week 12 and Week 24
- Analysis population: FAS-OL
- Analysis methodology: descriptive summary.

- Imputation for missing values: NRI ([Section 5.3.4](#)).

Reporting results:

- The sample size, frequency, percentage of responders and 95% CI of percentage will be presented.

6.4. Exploratory Endpoints

See [Section 6.2.3](#) and [Section 6.3.2](#) for the analysis of exploratory endpoint RAPID3.

6.5. Subset Analyses

To check the consistency of the primary endpoint across subgroups, change from randomization at Week 48 and Week 36 will be analyzed for below subgroups:

- Age group: 18-45, 46-64, ≥ 65 ;
- Geographic region: US, Europe, Latin America, Rest of the World;
- Race: White, Black, Asian, Other;
- Prior bDMARD: Yes, No;
- Disease severity: Moderate (DAS28-4 (ESR) ≥ 3.2 and ≤ 5.1), Severe (DAS28-4 (ESR) > 5.1).

Each subgroup factor along with its 2-way and 3-way interactions with treatment group and visit will be added to the MMRM model without imputation for missing data described in [Section 5.2.2](#), and outputs from this augmented model will be used for comparisons within each subgroup.

6.6. Baseline and Other Summaries and Analyses

6.6.1. Baseline Summaries

Demographic and baseline characteristics will be summarized for FAS-OL and FAS-DB according to Pfizer data standards.

6.6.2. Study Conduct and Subject Disposition

Subjects evaluation, disposition, discontinuation will be summarized for FAS-OL and FAS-DB according to Pfizer data standards.

6.6.3. Study Treatment Exposure

Treatment duration and number of subjects by treatment duration category will be summarized for FAS-OL and FAS-DB according to Pfizer data standards.

6.6.4. Concomitant Medications and Non-Drug Treatments

Prior drug and non-drug treatment for RA, concomitant drug and non-drug treatment for RA for FAS-OL and FAS-DB according to Pfizer data standards.

6.7. Safety Summaries and Analyses

6.7.1. Adverse Events

Adverse events will be summarized according to Pfizer data standards for each phase.

- TEAEs;
- Serious AEs;
- Serious infections;
- Adjudicated events.

AEs will not be double counted across phases. Counting of AEs will be based on start date of AE. If the start date of an AE is in the open-label phase, the end date of the AE in the double-blind phase and severity of AE remained the same or became lower in both phases, then the AE will be counted in open-label phase, not in the double-blind phase. However, if the same AE started in the open-label phase, and get worse in severity in the double-blind phase, then the AE will be counted in both phases.

In all these analyses attention will be restricted to assessments that occurred no more than 28 calendar days after the last dose of study drug.

6.7.2. Laboratory Data

Safety laboratory data will be summarized according to Pfizer standards for each phase.

6.7.3. Vital Signs

Descriptive summary tables will be provided in accordance with Pfizer reporting standards for each phase.

6.7.4. Physical Examination

Descriptive summary tables will be provided in accordance with Pfizer reporting standards for each phase.

6.7.5. ECG

ECG data at screening and Week 48/end of study, change from screening to Week 48/end of study, number (%) of subjects with any clinically significant changes will will be descriptively summarized by treatment group in double-blind phase and for FAS-DB.

CCI



8. REFERENCES

1. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis and rheumatism*. 2011;63(3):573-86. Epub 2011/02/05.
2. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *The Journal of rheumatology*. 1982;9(5):789-93. Epub 1982/09/01.
3. Ware JE KM, Dewey JE. . How to score Version 2 of the SF 36 Health Survey (Standard & Acute forms). In: *How to score Version 2 of the SF 36 Health Survey (Standard & Acute forms)*. Lincoln, Rhode Island: QualityMetric, Incorporated. 2000.
4. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *British journal of rheumatology*. 1997;36(5):551-9. Epub 1997/05/01.
5. Reilly MC, Zbrozek AS, Dukes E: The validity and reproducibility of a work productivity and activity impairment measure. *Pharmacoeconomics* 1993; 4(5):353-365.
6. Cella D, Lai JS, Chang CH, Peterman A, Slavin M. Fatigue in cancer patients compared with fatigue in the general United States population. *Cancer*. 2002;94(2):528-38. Epub 2002/03/20.
7. Pincus T, Yazici Y, Bergman MJ. RAPID3, an index to assess and monitor patients with rheumatoid arthritis, without formal joint counts: similar results to DAS28 and CDAI in clinical trials and clinical care. *Rheum Dis Clin North Am*. 2009 Nov;35(4):773-8.

9. APPENDICES

Appendix 1. DATA DERIVATION DETAILS

Appendix 1.1. Endpoints Definition and Derivation

Endpoint	Defination/Calculation
DAS28-4 (CRP)	$0.56*\sqrt{(TJC28)} + 0.28*\sqrt{(SJC28)} + 0.36*\ln(\text{CRP in mg/l} + 1) + 0.014*\text{PtGA in mm} + 0.96$
DAS28-4 (ESR)	$0.56*\sqrt{(TJC28)} + 0.28*\sqrt{(SJC28)} + 0.70*\ln(\text{ESR in mm/ hour}) + 0.014*\text{PtGA in mm}$
SDAI	$(28TJC) + (28SJC) + [\text{PhyGA in cm}] + [\text{PtGA in cm}] + [\text{CRP in mg/dL}]$
CDAI	$(28TJC) + (28SJC) + [\text{PhyGA in cm}] + [\text{PtGA in cm}]$
ACR20/50/70 responder	ACR20: at least 20% improvement in tender and swollen joint counts and at least 20% improvement in at least 3 of the 5 remaining ACR-core set measures: subject and physician global assessments, pain, HAQ-DI, and CRP. Similarly, ACR50 and 70 are calculated with the respective percent improvements.
ACR/EULAR Boolean-based definition of remission	all of the following satisfied: $TJC \leq 1$, $SJC \leq 1$, $CRP \leq 1$ mg/dL, $PtGA \leq 1$ on a 0-10 scale
LDA of DAS28-4 (CRP)	$DAS28-4 (CRP) \leq 3.2$
LDA of DAS28-4 (ESR)	$DAS28-4 (ESR) \leq 3.2$
LDA of SDAI	$SDAI \leq 11$
LDA of CDAI	$CDAI \leq 10$
Remission of DAS28-4 (CRP)	$DAS28-4 (CRP) < 2.6$
Remission of DAS28-4 (ESR)	$DAS28-4 (ESR) < 2.6$
Remission of SDAI	$SDAI \leq 3.3$
Remission of CDAI	$CDAI \leq 2.8$
HAQ-DI responder	Decrease ≥ 0.22
RAPID3	$(\text{HAQ} * 3.33 + \text{pain VAS in mm/10} + \text{PtGA in mm/10})/3$.

TJC = tender joint count; SJC = swollen joint count; CRP = C-reactive protein in mg/L; ESR = erythrocyte sedimentation rate in mm/first hour, PtGA = subject's global assessment of health; PhyGA = physician's global assessment of health, HAQ-DI = Health Assessment Disability Questionnaire-Disability Index; LDA = Low Disease Activity. RAPID3 = Routine Assessment of Patient Index Data 3.

Appendix 1.2. Definition and Use of Visit Windows in Reporting

Visit windows will be used for efficacy variables, and for any safety displays that display by week.

If more than one observation from the same subject falls into the same visit window, the value closest to the targeted day will be used as the observation for that week. If two visits are equal distant from the Targeted Day in absolute value, the later visit should be used. All observations will, however, be included in the listings.

For endpoints derived from multiple components, eg, DAS28-4(ESR), each component will be windowed to select the value for analysis at a specific visit, and the value of the endpoint will then be derived. This means the component values may be from different visit dates that fall within the analysis window.

Subjects’ observations will be excluded from visit windows if the data was collected more than 7 days off study drug. (Safety analysis may follow Pfizer standard.)

Visit Label	Targeted Day	Analysis window for data sets
Open-Label Run-In Phase		
Baseline	Day 1	Last Observation up to and including First Dosing Date (no earlier than Day -90)
Week 6	Day 42	Days 2 – 63
Week 12	Day 84	If there is Week 6 visit such as lab tests: Days 64 – 126; If no Week 6 visit: Days 2 – 126
Week 24	Day 168	Day 127 – Discontinuation Day /Randomization Day (the earlier one)
Double-Blind Phase		
Randomization (Week 24)	Randomization Day	Last Observation up to and including First Dosing Date in Double-Blind Phase (no earlier than Day 127)
Week 36	Randomization Day + 84	Randomization Day + 1 – Randomization Day + 126
Week 48	Randomization Day + 168	Randomization Day + 127 –