

Protocol A3921092

**A LONG-TERM, OPEN-LABEL EXTENSION STUDY OF TOFACITINIB
(CP-690,550) FOR THE TREATMENT OF PSORIATIC ARTHRITIS**

**Statistical Analysis Plan
(SAP)
(Including SAP for the Methothrexate Withdrawal Sub-Study)**

Version: 4.0

Date: 12APR2018

TABLE OF CONTENTS

LIST OF TABLES	4
APPENDICES	4
1. AMENDMENTS FROM PREVIOUS VERSION(S)	6
2. INTRODUCTION	7
2.1. Study Design	8
2.2. Study Objectives	8
3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING	9
4. HYPOTHESES AND DECISION RULES	9
4.1. Statistical Hypotheses	9
4.2. Statistical Decision Rules	9
5. ANALYSIS SETS	9
5.1. Full Analysis Set	9
5.2. ‘Per Protocol’ Analysis Set	9
5.3. Safety Analysis Set	9
5.4. Other Analysis Sets	10
5.4.1. Endpoint Specific Analysis Sets	10
5.5. Treatment Misallocations	12
5.6. Protocol Deviations	12
6. ENDPOINTS AND COVARIATES	12
6.1. Primary Endpoints	12
6.2. Secondary Endpoints	12
6.3. Other Endpoints	13
6.4. Covariates	15
7. HANDLING OF MISSING VALUES	15
8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES	15
8.1. Statistical Methods	15
8.1.1. Analyses for Continuous Data	15
8.1.2. Analyses for Categorical Data	15
8.2. Statistical Analyses	15
8.2.1. Analysis of Primary Endpoints	16
8.2.2. Analysis of Secondary and Other Efficacy/Health Outcome Endpoints	17

8.2.3. Analysis of Other Safety Data	19
8.3. Baseline and Other Summaries and Analyses	19
8.3.1. Baseline Summaries.....	19
8.3.2. Prior Drug Treatments for Psoriatic Arthritis	22
8.3.3. Concomitant Drug Treatments for Psoriatic Arthritis	23
9. REFERENCES	24
10. APPENDICES	25
1. INTRODUCTION	48
1.1. Study Design	48
1.2. Study Objectives	49
2. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING.....	49
3. HYPOTHESES AND DECISION RULES	49
3.1. Statistical Hypotheses	49
3.2. Statistical Decision Rules.....	49
4. ANALYSIS SETS	49
4.1. Full Analysis Set	49
4.2. ‘Per Protocol’ Analysis Set	50
4.3. Safety Analysis Set.....	50
4.4. Other Analysis Sets	50
4.4.1. Endpoint Specific Analysis Sets.....	50
4.5. Treatment Misallocations.....	50
5. ENDPOINTS AND COVARIATES	51
5.1. Primary Endpoints.....	51
5.2. Secondary Endpoints.....	51
5.3. Other Endpoints.....	52
5.4. Covariates.....	52
6. HANDLING OF MISSING VALUES	52
7. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES	52
7.1. Statistical Methods	53
7.1.1. Analyses for Continuous Data.....	53
7.1.2. Analyses for Categorical Data.....	53
7.2. Statistical Analyses	54

7.2.1. Analysis of Primary Endpoints	54
7.2.2. Analysis of Secondary and Other Efficacy/Health Outcome Endpoints	54
7.2.3. Analysis of Safety Data	55
7.3. Baseline and Other Summaries and Analyses	55
7.3.1. Baseline Summaries.....	55
7.3.2. Prior Drug Treatments for Psoriatic Arthritis.....	55
7.3.3. Concomitant Drug Treatments for Psoriatic Arthritis	56
7.4. Sub-study Definition and Use of Visit Windows in Reporting.....	56

LIST OF TABLES

Table 1 Endpoint Specific Analysis Sets	10
Table 2 AEs of Special Interest	16
Table 3 Endpoint Specific Analysis Sets	50

APPENDICES

Appendix 1. Definition and Use of Visit Windows in Reporting.....	25
Appendix 2. Further Definition of Endpoints	27
Appendix 2.1. ACR Assessments	27
Appendix 2.2. Health Assessment Questionnaire – Disability Index (HAQ-DI).....	28
Appendix 2.3. DAS 28-3 (CRP).....	29
Appendix 2.4. PsA Response Criteria (PsARC).....	29
Appendix 2.5. PsA Joint Activity Index (PsAJAI).....	30
Appendix 2.6. Disease Activity Index for Reactive Arthritis/Psoriatic Arthritis (DAREA/DAPSA).....	31
Appendix 2.7. Composite Psoriatic Disease Activity Index in Psoriatic Arthritis (CPDAI).....	31
Appendix 2.8. Minimal Disease Activity (MDA) Response	33
Appendix 2.9. Psoriatic Arthritis Disease Activity Score (PASDAS).....	34
Appendix 2.10. Psoriasis Area and Severity Index (PASI)	34
Appendix 2.11. Physician Global Assessment (PGA) of Psoriasis (PGA-PsO).....	35
Appendix 2.12. Dermatology Life Quality Index (DLQI).....	35
Appendix 2.13. Itch Severity Item (ISI)	36

Appendix 2.14. Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)	36
Appendix 2.15. EuroQol-5D Health Questionnaire 3-Level (EQ-5D-3L)	37
Appendix 2.16. Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F).....	37
Appendix 2.17. Short Form 36 (SF-36, version 2, acute).....	37
Appendix 3. Criteria for Clinical Diagnosis of the Metabolic Syndrome	44
Appendix 4. Definition of Diabetes Mellitus at Baseline	46
Appendix 5. Handling of Joint Count Assessment (Missing Assessment, “NOT DONE” and “NOT APPLICABLE”) and Injected Joint.....	47
Appendix 6. Methotrexate Withdrawal Sub-Study.....	48

1. AMENDMENTS FROM PREVIOUS VERSION(S)

Version 4 (April 12, 2018):

This SAP amendment is mainly updated to reflect the addition of a sub-study. It is based on Protocol Amendment 4 dated April 12, 2017.

Version 3 (April 14, 2016):

This SAP amendment occurred after reviews of the data by the external independent Data Monitoring Committee, but changes were not based on any recommendations by the DMC.

[Appendix 2.1](#): Updated the method of handling missing components in the evaluation of ACR endpoints to a general principle rather than listing out all possible cases of missingness.

[Appendix 2.1](#), [Appendix 2.4](#) & [Appendix 2.5](#): Clarified the evaluation of percent change from baseline for the components used in deriving ACR, PsARC, and PsAJAI endpoints when the baseline component value is equal to zero.

[Appendix 5](#): Added this section to clarify the handling of tender/painful joint (68) and swollen joint (66) assessment that is missing, “NOT DONE” or “NOT APPLICABLE,” as well as when a joint receives intra-articular injection at baseline or post-baseline.

Version 2.1 (February 1, 2016):

[Appendix 3](#) and [Appendix 4](#): Clarified drug treatment (other than study drug treatment) taken on Day 1 of the qualifying study will be considered as concomitant drug treatment, and taken on Day 0 will be considered as prior drug treatment to align with the Pfizer Data Standard for reporting drug treatments.

Version 2 (January 30, 2016):

Changes to the original SAP, version 10JUL2013, are summarized below. This SAP amendment occurred after reviews of the data by the external independent Data Monitoring Committee, but changes were not based on any recommendations by the DMC.

- [Section 5.4.1](#), Added the ‘Endpoint Specific Analysis Sets’ section to define analysis populations suitable for certain endpoints since disease may not be present at baseline as measured by certain endpoints.
- [Section 6](#), clarified that VAS will be re-scaled to 100 mm or 10 cm prior to any calculation and analysis.
- [Section 6.2](#), added additional details/clarifications for the secondary endpoints.

- [Section 6.3](#), added HAQ-DI response, components of ACR, PASI and its clinical sign components and BSA to the other endpoints, and other additional details/clarification for the other endpoints.
- [Section 8.1.2](#), added normal approximation for construction of confidence interval for responder rate.
- [Section 8.2](#), added clarification for analysis population.
- [Section 8.2.2](#), added analyses for components of ACR, PASI and its clinical sign components and BSA, and additional clarifications to analysis populations.
- [Section 8.3.1](#), added baseline summaries.
- [Section 8.3.2](#), added summaries of prior drug treatment for psoriatic arthritis.
- [Section 8.3.3](#), added summaries of concomitant drug treatment for psoriatic arthritis.
- [Section 9](#), updated the references.
- [Appendix 1](#), updated visit windows.
- [Appendix 2](#), added clarifications and additional details/update for definition/calculation of endpoints.

Version 1 (initial version, July 10, 2013):

2. INTRODUCTION

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

Psoriatic arthritis (PsA) is a chronic inflammatory autoimmune disease characterized by joint inflammation and destruction, psoriatic skin lesions, enthesitis, dactylitis, spondylitis, progressive disability and adverse effects on quality of life. Tofacitinib (CP-690,550) is a potent and selective inhibitor of the Janus Kinase (JAK) family of kinases. While tofacitinib shows nanomolar inhibitory potency against all JAK family kinases in enzymatic assays, it shows functional specificity for JAK1 and JAK1/3 over JAK2 in cell-based assays. The broad effects of JAK1/3 inhibition on multiple cytokine pathways provide rationale for developing tofacitinib as treatment for PsA.

This Phase 3 study is a long-term extension (LTE) study designed to evaluate the safety and tolerability of tofacitinib as a treatment for PsA.

An optional sub-study is included to assess the efficacy, safety and tolerability of tofacitinib 5 mg BID administered as monotherapy after methotrexate (MTX) withdrawal compared to tofacitinib 5 mg BID in combination with MTX. Subjects who have completed at least 24 months of participation in the long term extension study and are currently receiving tofacitinib 5 mg BID in combination with oral methotrexate are eligible to participate.

The analyses for the LTE study are covered in Sections 2 through 8, while analyses for the sub-study are covered in [Appendix 6](#).

2.1. Study Design

This is a Phase 3, long-term, open-label extension study designed to evaluate the safety, tolerability and efficacy of tofacitinib in subjects with active PsA. Subjects with active PsA will have previously participated in randomized studies of tofacitinib. For subjects who are completing participation in a randomized study of tofacitinib, the final visit of the qualifying study can be combined with screening and baseline visit for this study. Additional assessments and inclusion/exclusion criteria are required for subjects who enroll >14 days after completing treatment in their qualifying study. In this case, a separate screening visit to determine subject eligibility is required followed by a baseline visit.

All eligible subjects from qualifying studies A3921091 and A3921125 will receive open-label tofacitinib 5 mg BID upon entry into A3921092. Subjects from A3921091 will receive first dose of study medication ≥ 1 week after last injection of study medication in that qualifying study. Tofacitinib dose may be increased to 10 mg BID at study visits if, based upon investigator's discretion, subjects receiving tofacitinib 5 mg BID would benefit from a higher dose and are not experiencing any tofacitinib-related adverse events, including abnormalities in laboratory test results that are judged to be related to tofacitinib. Tofacitinib dose may be decreased (ie, 10 mg BID to 5 mg BID) for safety reasons at any time. Treatment duration for subjects participating in the main LTE study is approximately 3 years (36 standardized 4-week months). Subjects participating in the sub-study may have up to an additional 1 year (12 standardized 4-week months) of treatment for a maximum potential total of approximately 4 years (48 standardized 4-week months).

At various timepoints in this trial, safety measurements, including physical examination, clinical laboratory tests, adverse event monitoring, electrocardiograms (ECGs) and vital signs will be performed. All subjects will be monitored for clinical evidence of PsA response to treatment. Health Outcomes Measures (ie, Patient Reported Outcomes assessments for pain, quality of life, physical function, fatigue, work limitations, health care resource utilization and health status) will also be performed at various timepoints in this trial. In addition, subjects will be monitored for serious infections, lymphadenopathy and lymphoproliferative disorder (LPD).

2.2. Study Objectives

Primary Objectives

- To evaluate the long term safety and tolerability of treatment with tofacitinib (5 mg BID and 10 mg BID) in adult subjects with active Psoriatic Arthritis (PsA).

Secondary Objectives

- To evaluate the long term efficacy of treatment with tofacitinib (5 mg BID and 10 mg BID) in adult subjects with active Psoriatic Arthritis (PsA).

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

Interim analyses to review safety data along with some efficacy data may be performed. As this is an open-label study with no formal hypothesis testing, there are no issues of protecting the Type I error rate.

Data-cuts will be performed as needed to support tofacitinib registration. Additional data reviews may be performed as needed by the study team for safety evaluations or for administrative purposes.

This study will use an External Data Monitoring Committee (E-DMC).

The E-DMC will be responsible for ongoing monitoring of the efficacy and safety of subjects in the study according to the Charter.

Information about the E-DMC can be found in the E-DMC Charter, which outlines the operating procedures of the committee, including specific description of the scope of their responsibilities, including a plan where communication timelines are defined.

Two final analyses will occur when all subjects finish the study or terminate early, one for the main LTE study and the other one for the sub-study.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

There are no formal hypotheses being tested in this study.

4.2. Statistical Decision Rules

There are no statistical decision rules in this study.

5. ANALYSIS SETS

Below is a description of the Analysis Sets defined for this study.

5.1. Full Analysis Set

The full analysis set (FAS) is defined as all subjects enrolled in this study who were part of a prior qualifying study, and who received at least one dose of open-label study medication in A3921092.

5.2. ‘Per Protocol’ Analysis Set

There is no Per Protocol Analysis Set defined for this study.

5.3. Safety Analysis Set

The FAS and the Safety Analysis Set are the same.

5.4. Other Analysis Sets

5.4.1. Endpoint Specific Analysis Sets

Subjects will be excluded from FAS for analyzing a specific endpoint if the criterion for inclusion is not met for the endpoint as described in

Table 1.

Table 1 Endpoint Specific Analysis Sets

Instrument/Endpoint	Inclusion	Rationale
PASI75 or PASI	Include subjects with baseline BSA \geq 3% and baseline PASI $>$ 0	PASI is not assessed if baseline BSA $<$ 3% per the study protocol; PASI=0 means absence of psoriasis as measured by PASI.
Δ PGA-PsO	Include subjects with baseline PGA-PsO $>$ 0	PGA-PsO=0 means no PsO disease at baseline as measured by PGA.
Δ BSA	Include subjects with baseline BSA $>$ 0%	BSA=0% means no PsO disease at baseline as measured by BSA.
HAQ-DI responder (decrease Δ HAQ-DI \geq 0.30)	Include subjects with baseline HAQ-DI \geq 0.30	Restrict baseline HAQ-DI to \geq 0.30 to allow room for response
HAQ-DI responder (decrease Δ HAQ-DI \geq 0.35)	Include subjects with baseline HAQ-DI \geq 0.35	Restrict baseline HAQ-DI to \geq 0.35 to allow room for response
Dactylitis presence	Include subjects with baseline DSS $>$ 0	Baseline DSS=0 means absence of dactylitis. Dactylitis is not an enrollment criterion so it is expected that only a subset of the subjects will have dactylitis at baseline.
Δ DSS	Include subjects with baseline dactylitis severity score (DSS) $>$ 0	Baseline DSS=0 means absence of dactylitis. Dactylitis is not an enrollment criterion so it is expected that only a subset of the subjects will have dactylitis at baseline.
Δ Leeds Enthesitis Index (Δ LEI)	Include subjects with baseline LEI $>$ 0	Baseline LEI score=0 means absence of enthesitis as measured by LEI. Enthesitis is not an enrollment criterion so it is expected that only a subset of the subjects will have enthesitis at baseline.
Δ SPARCC Enthesitis Index	Include subjects with baseline SPARCC Enthesitis Index $>$ 0	Baseline SPARCC Enthesitis Index score=0 means absence of enthesitis as measured by SPARCC. Enthesitis is not an enrollment criterion so it is expected that only a subset of the subjects will have enthesitis at baseline.
Δ BASDAI	Include subjects with presence of spondylitis at screening and baseline BASDAI score $>$ 0 cm	Absence of spondylitis at screening or baseline BASDAI score=0 cm means absence of spondylitis. Spondylitis is not an enrollment criterion so it is expected that only a subset of the subjects will have spondylitis at baseline.
Δ BASDAI	Include subjects with presence of spondylitis at screening and baseline BASDAI score \geq 4 cm	Baseline BASDAI score \geq 4 cm is a more severe disease population in spondylitis.
Δ CPDAI	Include subjects with baseline BSA \geq 3%	Per the study protocol, PASI should only be performed if \geq 3% of subject's BSA is affected at baseline.
Δ ISI	Include subjects with baseline ISI score $>$ 0	Baseline ISI score=0 means absence of itching as measured by ISI. Itching is not an enrollment criterion so it is expected that only a subset of the subjects will have itching at baseline.
Δ NAPSI	Include subjects with baseline NAPSI score $>$ 0 on target nail	Baseline NAPSI score=0 means absence of nail psoriasis as measured by NAPSI on the target nail. Presence of nail psoriasis is not an enrollment criterion so it is expected that only a subset of the subjects will have nail psoriasis at baseline.

Abbreviations: Δ =change from baseline; baseline is the value obtained prior to the first dose of the qualifying study A3921091 or A3921125; BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; BSA = body surface area; CPDAI = Composite Psoriatic Disease Activity Index; DSS = Dactylitis Severity Score; HAQ-DI = Health Assessment Questionnaire - Disability Index; ISI = Itch Severity Item; LEI = Leeds Enthesitis Index; NAPSII = Nail Psoriasis Severity Index; PASI = Psoriasis Area and Severity Index; PGA-PsO = Physician's Global Assessment of Psoriasis; SPARCC = Spondyloarthritis Research Consortium of Canada.

5.5. Treatment Misallocations

If a subject was:

- Admitted to the study but not treated, then the subject will not be reported.

5.6. Protocol Deviations

No subjects will be excluded from analyses due to protocol deviations.

6. ENDPOINTS AND COVARIATES

Visual analog scale (VAS) data will need to be rescaled prior to any calculation and analysis. VAS is recorded on CRF in terms of length at mark (X in mm) and overall length of line (Y in mm). The rescaled VAS for use in analysis will be: $Z = X/Y \times 100$ mm or $X/Y \times 10$ cm, depending on endpoints or their use in defining other endpoints.

6.1. Primary Endpoints

- Incidence and severity of adverse events.
- Incidence of clinical abnormalities and change from baseline (in this and/or prior study) in clinical laboratory values during treatment.

6.2. Secondary Endpoints

Change from baseline will be denoted by Δ .

- *ACR20, ACR50 and ACR70 response rate at all time points* (see [Appendix 2.1](#)).
- *Δ HAQ-DI score at all time points* (see [Appendix 2.2](#)).
- Psoriatic Arthritis Response Criteria (PsARC) response at all time points (see [Appendix 2.4](#)).
- Physician's Global Assessment of Psoriasis (Δ PGA-PsO) response at all time points (see [Appendix 2.11](#)).
- Psoriasis Area and Severity Index 75 (PASI75) response ie, the proportion of subjects achieving at least a 75% reduction in PASI relative to baseline and PASI/PASI component scores ($\% \Delta$ PASI and its clinical signs component scores) at all time points (see [Appendix 2.10](#)).
- Dactylitis Severity Score (Δ DSS) at all time points.

This is the sum of the 20 digit disease severity ratings. A missing score or not assessed (ie, NOT DONE with a code of 9 in the CRF) for any digit will result in DSS as missing. A score of 8 (ie, not applicable) should be excluded from summing.

- Enthesitis score (using Δ in Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index and Leeds Enthesitis Index (LEI)) at all time points.

Each score is the number of sites in which enthesitis is present (total number of possible sites: 16 sites for SPARCC Enthesitis Index and 6 sites for LEI). A missing assessment (ie, no box checked) or not assessed (ie, NOT DONE box checked) for any digit will result in the enthesitis score as missing.

- *Bath Ankylosing Spondylitis Disease Activity Index (Δ BASDAI)* at all time points (see [Appendix 2.14](#)).
- Physical function/other patient reported outcomes to be assessed at Month 1, Month 6 (and every 6 months thereafter): Δ Short-Form 36 (version 2, Acute); Δ EQ-5D-3L; Functional Assessment of Chronic Illness Therapy-Fatigue (Δ FACIT-F).

The Short-Form 36 yields 10 endpoints (norm-based): 8 general health domains: physical functioning, role limitations due to physical health, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, and mental health. These domains can also be summarized as physical and mental component scores (see [Appendix 2.17](#)).

The EQ-5D-3L yields 6 endpoints: an assessment of one's health status that day (Δ EQ-VAS) and scores from each of five domains (Δ EQ-5D-3L) (see [Appendix 2.15](#)).

The FACIT-F will have 3 endpoints: Δ FACIT-F total score, Δ FACIT-F experience domain score and Δ FACIT-F impact domain score (see [Appendix 2.16](#)).

6.3. Other Endpoints

- Δ DAS28-3 (CRP) at all time points (see [Appendix 2.3](#)).
- HAQ-DI response, defined as an improvement (ie, a decrease) from baseline of ≥ 0.30 and ≥ 0.35 (2 endpoints) in HAQ-DI scores at all timepoints (see [Appendix 2.2](#)).
- ACR response criteria components (Δ in following: HAQ-DI, CRP, Patient's Assessment of Arthritis Pain, Patient's Global Assessment of Arthritis, Physician's Global Assessment of Arthritis, swollen joint count, tender/painful joint count) at all time points (see [Appendix 2.1](#)). Note that Δ HAQ-DI is covered under [Section 6.2](#) as a secondary endpoint already.
- *Psoriatic Arthritis Joint Activity Index (raw PsAJAI)* (see [Appendix 2.5](#)), *Disease Activity Index for Reactive Arthritis/PsA (Δ DAREA/DAPSA)* (see [Appendix 2.6](#)), *Composite Psoriasis Disease Activity Index (Δ CPDAI)* (see [Appendix 2.7](#)), *Psoriatic Arthritis Disease Activity Score (Δ PASDAS)* (see [Appendix 2.9](#)), *Minimal Disease*

Activity (MDA, yes/no) and their components (yes/no response for each component) (see [Appendix 2.8](#)) at all time points.

- BSA (% Δ BSA) at all time points (see [Appendix 2.10](#)).
- Presence of dactylitis at all time points.
- Nail Psoriasis Severity Index (Δ NAPSI) Score at Month 1, Month 6 (and every 6 months thereafter).

This is a score ranging from 0-8 based on 1 targeted nail.

- Itch Severity Item (Δ ISI) at Month 1, Month 6 (and every 6 months thereafter) (see [Appendix 2.13](#)).
- Patient's Global Joint and Skin Assessment (Δ PGJS-VAS, 3 endpoints) at all time points.
- Ankylosing Spondylitis Quality of Life questionnaire (Δ ASQOL) at all time points.

The ASQOL has a total score which is calculated by summing the 18 items.

- Physical function/other patient reported outcomes to be assessed at Month 1, Month 6 (and every 6 months thereafter): Dermatology Life Quality Index (Δ DLQI); PsA Healthcare Resource Utilization Questionnaire (raw PsA HCRU and Δ PsA HCRU-Self Rating of Job Performance at Work); Work Limitations Questionnaire (Δ WLQ).

A total score for DLQI is computed from 10 questions.

The WLQ is a twenty-five item scale that evaluates the degree to which health problems interfere with an ability to perform job roles along four dimensions. The Time Management scale (Question 1) contains five items. The Physical Demands scale (Question 2) contains 6 items. The Mental/Interpersonal Demands Scale (Questions 3 and 4) has nine items. The Output Demands scale (Question 5) contains five items. Scores for each item are: 0 (none of the time), 1, 2, 3 or 4 (all of the time). "Does not apply to my job" is treated as missing. Each of the scale scores are computed by averaging the items that comprise the scale, and then multiplying by 25.

- Incidence of investigator-reported clinically significant changes in physical examination from baseline (in this and/or prior study) during treatment.
- Incidence of electrocardiogram (ECG) abnormalities and change from baseline (in this and/or prior study) in ECG measurements during treatment.
- Incidence of vital sign (blood pressure and pulse rate) abnormalities and changes from baseline (in this and/or prior study) in vital sign measurements during treatment.

6.4. Covariates

Reporting of the study data will be done with descriptive statistics. There will be no statistical models, and hence no covariates.

7. HANDLING OF MISSING VALUES

For the SF-36v2, ASQOL, DLQL, EQ-5D-3L, BASDAI, WLQ, PsA-HCRU, and FACIT-F instruments, rules suggested by the developers of these will be followed in calculating scores when individual question/items may be missing. If these rules are not enough for calculating a score, then the endpoint will be considered to have a missing value. Missing values in any of the endpoints will not be imputed when summarizing these endpoints using descriptive statistics. See [Appendix 2](#) for additional details for the calculation of endpoints, including handling of missing item values in the calculation.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

There will be 2 sets of analyses: one for the A3921092 LTE and the other for the sub-study. For the A3921092 LTE, the safety analysis will include cumulative data from all the subjects and all the data (including the data collected in the sub-study). The efficacy analysis for the LTE will only include the data collected in the A3921092 LTE (the efficacy data collected in the sub-study will be excluded from the LTE analysis).

Statistical methods for the sub-study are found in [Appendix 6](#).

8.1. Statistical Methods

8.1.1. Analyses for Continuous Data

In general, the data for all continuous endpoints will be summarized by time point in tables containing descriptive statistics (N, mean, standard deviation, standard error of the mean, minimum, 1st, 2nd (median) and 3rd quartiles and maximum) for baseline and change from baseline for those endpoints measured at baseline.

8.1.2. Analyses for Categorical Data

The data for all categorical endpoints (including binary) will be summarized by contingency tables that show the counts/frequency in the various categories at each time point. For some categorical endpoints, such as the presence or severity of adverse events, contingency or frequency tables will consider the data over the entire duration of the study and will not be constructed at each time point.

Normal approximation will be used for the construction of confidence interval (CI) for the responder rate over time. Let N be number of subjects assessed at each visit and n be number of responders, and $p_{\text{obs}} = n/N$ be the observed responder rate, then the two-sided 95% CI will be formed by $p_{\text{obs}} \pm 1.96 \times \text{squared root of } [p_{\text{obs}} \times (1 - p_{\text{obs}}) / N]$.

8.2. Statistical Analyses

Tofacitinib dose may be increased to 10 mg BID at study visits based upon investigator's discretion. Also, tofacitinib dose may be decreased (ie, 10 mg BID to 5 mg BID) for safety

reasons at any time. Hence changes in the dose are treatment related. Consequently, descriptive statistics will be displayed for just one treatment group: the entire cohort of subjects in which tofacitinib is dosed flexibly between 5 mg and 10 mg BID.

The baseline values for the safety endpoints come from the baseline of the qualifying study for patients who enroll within 14 days of the last visit of the qualifying study. For the patients whose enrollment are out of 14-day window, their baseline values of the safety endpoints will be from the baseline visit of this study, A3921092. The baseline values for all the efficacy or health-outcome endpoints come from the baseline of the qualifying study. In addition, as a sensitivity analysis for the ACR response rates and Δ HAQ-DI, the baseline value will be the baseline visit for A3921092 for patients whose enrollment is out of 14-day window.

All analyses will use FAS (Section 5.1) or FAS with some subjects excluded from analyses for certain endpoints (Section 5.4.1) except where indicated otherwise.

8.2.1. Analysis of Primary Endpoints

Safety analysis will include cumulative data from both the main LTE study and sub-study.

- Adverse events will be summarized according to Pfizer Data Standards (PDS).
- Incidence rate (IR) defined as the number of subjects with an event in the numerator divided by the study drug exposure in the denominator (expressed as the number of subjects with an event per 100 patient-years) will be calculated for deaths, SAEs, AEs leading to discontinuation and AEs of special interest. AEs of special interest include but may not be limited in the events listed in **Table 2** AEs of Special Interest. The event will be counted from subject's first dose to subject's last dose of study drug + 28 days. The denominator is the sum of the time to event or to subject's last dose of study drug + 28 days, whichever is earlier.

Table 2 AEs of Special Interest

Serious infections (SIs)
Herpes zoster (HZ)
Adjudicated opportunistic infections (OIs) excluding tuberculosis (TB) ^a
Hematologic events (anaemia, neutropenia, lymphopenia, and thrombocytopenia)
Adjudicated malignancies ^a
Adjudicated major adverse cardiovascular events (MACE) ^{a,b}
Adjudicated drug-induced liver injury (DILI) ^a and other hepatic events ^c
Renal events ^d
Adjudicated gastrointestinal (GI) perforations ^a
Adjudicated interstitial lung disease (ILD) ^e

Abbreviations: DILI = drug-induced liver injury; GI = gastrointestinal; HZ = herpes zoster; ILD = interstitial lung disease; MACE = major adverse cardiovascular events; MedDRA = Medical Dictionary for Regulatory Activities; OI = opportunistic infection; SI = serious infection; SMQ = Standardised MedDRA Queries; TB = tuberculosis.

a. Adjudicated by Adjudication Committees independent and external to Pfizer.

b. Components: cardiovascular death, non-fatal myocardial infarction, and non-fatal stroke.

c. Including SMQs: Hepatic failure, fibrosis, and cirrhosis, and other liver damage-related conditions; and Hepatitis, non-infectious.

d. Including Acute renal failure SMQ.

e. Adjudicated by an internal review committee.

- Safety laboratory tests will be summarized according to the PDS.
- The summaries for laboratory tests include contingency tables (eg, normal, high, low), and descriptive statistics for change from baseline by treatment group and visit. For the FACS, the endpoints to be summarized, which are subsets of lymphocytes and are measured at baseline, and every 3 months beginning at Month 3 visit, are as follows: $CD3^+$ (% , *abs*), $CD3^+CD4^+$ (% , *abs*), $CD3^+CD8^+$ (% , *abs*), $CD19^+$ (% , *abs*), $CD16^+/CD56^+$ (% , *abs*). Raw values as well as percent change from baseline will be summarized with descriptive statistics by visit.

8.2.2. Analysis of Secondary and Other Efficacy/Health Outcome Endpoints

Efficacy analysis will only include the data collected in the main LTE study.

Contingency/frequency tables as described in [Section 8.1.2](#) will summarize:

- ACR20, ACR50 and ACR70 response rate at all time points.

In addition, the ACR response rates will be displayed at all time points by the qualifying study. As a sensitivity analysis to using the baseline values from the qualifying study for all subjects, one will also provide tables of ACR response rates in which baseline is the baseline visit for A3921092 for patients whose enrollment are out of 14-day window.

- Psoriatic Arthritis Response Criteria (PsARC) response at all timepoints.
- Psoriasis Area and Severity Index 75 (PASI75) response, ie, the proportion of subjects achieving at least a 75% reduction in PASI relative to baseline for subjects with baseline BSA \geq 3% and baseline PASI $>$ 0 (see [Section 5.4.1](#)).
- Presence of dactylitis at all timepoints for subjects with baseline DSS $>$ 0.
- HAQ-DI response rates (2 endpoints: (decrease of \geq 0.30 and decrease of \geq 0.35) at all timepoints for subjects with baseline HAQ-DI \geq 0.30 and HAQ-DI \geq 0.35, respectively) (see [Section 5.4.1](#)).
- Minimal Disease Activity (MDA) at all timepoints.

Descriptive statistics as described in [Section 8.1.1](#) will be tabulated for:

- ACR response criteria components (Δ in HAQ-DI, CRP, Patient's Assessment of Arthritis Pain, Patient's Global Assessment of Arthritis, Physician's Global Assessment of Arthritis, swollen joint count, tender/painful joint count) at all timepoints.

In addition, the descriptive statistics for Δ HAQ-DI scores will be displayed at all timepoints by the qualifying study. As a sensitivity analysis to using the baseline values from the qualifying study for all subjects, one will also provide tables of

descriptive statistics in which baseline is the baseline visit for A3921092 for patients whose enrollment are out of 14-day window.

- *Physician's Global Assessment of Psoriasis (Δ PGA-PsO) response at all timepoints* for subjects with baseline PGA-PsO>0 (see [Section 5.4.1](#)).
- *PASI and PASI component scores (% Δ PASI and its clinical signs component scores)* at all timepoints for subjects with BSA \geq 3% and baseline PASI>0 (see [Section 5.4.1](#)).
- *BSA (% Δ BSA) at all timepoints* for subjects with baseline BSA>0 (see [Section 5.4.1](#)).
- *Dactylitis severity score (Δ DSS) at all timepoints* for subjects with baseline DSS>0 (see [Section 5.4.1](#)).
- *Enthesitis score (using Δ in Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index and Leeds Enthesitis Index) at all timepoints* for subjects with baseline SPARCC Enthesitis Index>0 and LEI>0, respectively (see [Section 5.4.1](#)).
- *Bath Ankylosing Spondylitis Disease Activity Index (Δ BASDAI) at all timepoints.* Two analysis populations will be used: one for subjects with presence of spondylitis at screening and baseline BASDAI score>0 cm, and the other for subjects with presence of spondylitis at screening and baseline BASDAI score \geq 4 cm (see [Section 5.4.1](#)).
- Δ FACIT-F (3 endpoints: total score, experience score and impact score) at Month 1, Month 6 (and every 6 months thereafter).
- Δ in components and domains of the Short-Form-36 Health Survey (SF-36) Version 2, Acute at Month 1, Month 6 (and every 6 months thereafter); for each component and domain, only the scaled score should be analyzed (see [Appendix 2.17](#) for derivations).
- Health status and domains from the EuroQol 5-Dimension Health State Profile (Δ EQ-5D-3L and Δ EQ-VAS) at Month 1, Month 6 (and every 6 months thereafter).
- Δ in DAS28-3 (CRP) at all timepoints.
- Psoriatic Arthritis Joint Activity Index (raw PsAJAI), Disease Activity Index for Reactive Arthritis/PsA (Δ DAREA/DAPSA), Composite Psoriasis Disease Activity Index (Δ CPDAI), Psoriatic Arthritis Disease Activity Score (Δ PASDAS). Δ CPDAI will be summarized for subjects with baseline BSA \geq 3% (see [Section 5.4.1](#)).
- *Nail Psoriasis Severity Index (Δ NAPSI) Score at Month 1, Month 6 (and every 6 months thereafter)* for subjects with baseline NAPSI>0 (see [Section 5.4.1](#)).

- *Itch Severity Item (ΔISI) at Month 1, Month 6 (and every 6 months thereafter)* for subjects with baseline $ISI > 0$ (see [Section 5.4.1](#)).
- Patient's Global Joint and Skin Assessment ($\Delta PGJS$ -VAS) at all timepoints.
- Total score on the Ankylosing Spondylitis Quality of Life questionnaire ($\Delta ASQOL$) at all timepoints.
- Total score for the $\Delta DLQI$ and the four scale scores for ΔWLQ at Month 1, Month 6 (and every 6 months thereafter).
- PsA-HCRU (each of the scales) and Δ PsA-HCRU self-rating of job performance at work.

8.2.3. Analysis of Other Safety Data

Safety analysis will include cumulative data from both the main LTE study and sub-study.

- Vital signs and 12-lead ECG parameters will be summarized according to the PDS. This will include summaries with contingency tables (eg, normal, high, low), and descriptive statistics for change from baseline by visit.
- Demographic parameters will be summarized (as number and percent or mean, standard deviation, and range) by gender following the PDS.
- Clinical findings on any physical examination during the study, including investigator-reported clinically significant changes in physical examination from baseline (in this and/or prior study) during treatment, will be tabulated by following the PDS.
- Previous and concomitant medication usage by medication type will be tabulated using the WHO-Drug dictionary following the PDS.

To help better understand and interpret the observed safety results, registry databases will be utilized to construct an appropriate contemporaneous and/or historical control group. However, such an analysis is outside the scope of this SAP and will include data not only from this study but the data from the qualifying studies as well.

8.3. Baseline and Other Summaries and Analyses

8.3.1. Baseline Summaries

Baseline/screening will be the baseline/screening of the qualifying study. Baseline characteristics will include but may not be limited to the ones listed below and will be summarized descriptively. For continuous variables, the summary will include N, mean, SD and range; for binary and categorical variables, the summary will include frequencies and percentages. A missing category will be included for those subjects with missing value.

- Demographic characteristics:
 - Baseline age (2 categorizations: <18, 18-44, 45-64, ≥65 years; <18, 18-44, 45-64, 65-74, 75-84, ≥85 years; and continuous);
 - Sex (female, male);
 - Race (white, black, Asian, other);
 - Ethnicity (Hispanic/Latino, non-Hispanic/Latino);
 - Baseline body weight (<60, ≥60 to ≤100, >100 kg; and continuous),
 - Baseline height (cm, continuous);
 - Baseline Body Mass Index (BMI: <18.5, 18.5 to <25, 25 to <30, 30 to <40, and ≥40 kg/m²; and continuous);
 - Geographic region (United States and Canada, Australia and Western Europe, Russia and Eastern Europe, Rest of World);
 - Screening smoking status (never smoked, ex-smoker, smoker);
 - Alcohol use [yes, no; continuous (units/week) for subjects who consumed any alcohol (responded Yes)].
- Baseline disease characteristics
 - PsA duration (<2, ≥2 years; and continuous);
 - Baseline PsA Subtype: (<5 joints, ≥5 joints) formed based on baseline swollen joint count (66) or baseline tender/painful joint count (68), ie, count any joint that is either swollen or tender/painful;
 - Baseline PASDAS: (≤3.2, >3.2 to <5.4, ≥5.4; and continuous);
 - Baseline CPDAI (≤4, >4 to <8, ≥8; and continuous) for subjects with baseline BSA≥3%;
 - Baseline swollen joint count (66) (continuous);
 - Baseline tender/painful joint count (68) (continuous);
 - Baseline HAQ-DI (continuous);
 - Screening presence of distal interphalangeal joints involvement (yes, no);
 - Screening presence of arthritis mutilans (yes, no);

- Baseline presence of enthesitis measured by SPARCC enthesitis index or LEI (yes, no). Yes is defined for those subjects with baseline SPARCC enthesitis index >0 or LEI >0 ;
- Baseline presence of enthesitis measured by SPARCC enthesitis index (yes, no). Yes is defined for those subjects with baseline SPARCC enthesitis index >0 ;
- Baseline presence of enthesitis measured by LEI (yes, no). Yes is defined for those subjects with baseline LEI >0 ;
- Baseline enthesitis index measured by SPARCC enthesitis index (continuous) for those subjects with SPARCC enthesitis index >0 at baseline;
- Baseline enthesitis index measured by LEI (continuous) for those with LEI >0 at baseline;
- Baseline presence of dactylitis (yes, no). Yes is defined for those subjects with baseline DSS >0 ;
- Baseline DSS (continuous) for those subjects with DSS >0 at baseline;
- Baseline presence of spondylitis (yes, no). Yes is defined for those subjects with presence of spondylitis at screening and baseline BASDAI >0 ;
- Baseline BASDAI (0, >0 to <4 , ≥ 4) for those subjects with presence of spondylitis at screening. BASDAI score ≥ 4 indicates active disease/suboptimal treatment;
- Baseline BASDAI (continuous) for those subjects with presence of spondylitis at screening and BASDAI >0 at baseline;
- Baseline total psoriatic BSA (0, >0 to <3 , $\geq 3\%$);
- Baseline total psoriatic BSA (continuous) for those with BSA $>0\%$ at baseline;
- Baseline PASI (0, >0 to ≤ 20 , >20);
- Baseline PASI (continuous) for those subjects with BSA $\geq 3\%$ and PASI >0 at baseline;
- Baseline PGA-PsO (0, 1, 2, 3, 4);
- Baseline PGA-PsO (continuous) for those subjects with PGA-PsO >0 at baseline;
- Number (%) of subjects with baseline PGA-PsO ≥ 3 , PASI ≥ 12 and BSA $\geq 10\%$;
- Baseline SF-36v2 (8 domains, physical component score [PCS] and mental component score [MCS]);

- Baseline FACIT-F (FACIT-F total score, FACIT-F experience score and FACIT-F impact score);
- Baseline DLQI (<5 , ≥ 5 ; and continuous);
- Baseline CRP (≤ 2.87 , > 2.87 mg/L; and continuous);
- Baseline diabetes mellitus (yes, no) (see [Appendix 4](#) for definition);
- Baseline metabolic syndrome (yes, no) (see [Appendix 3](#) for definition);
- Baseline cardiovascular risks per PDS and detailed in the statistical programming plan;
- Baseline rheumatoid factor positive (yes, no);
- Baseline cyclic citrullinated peptide antibody positive (yes, no).

8.3.2. Prior Drug Treatments for Psoriatic Arthritis

Prior drug treatment will include any treatment for PsA prior to the first dose of study medication in Study A3921092. It will include the drug treatment prior to the qualifying study, the treatment used during the qualifying study (which includes adalimumab but not tofacitinib) and any treatment used during the enrollment gap between the qualifying study and the Study A3921092.

- Prior DMARDs (disease-modifying antirheumatic drugs) drug treatment (yes, no), yes is defined for a subject who used any DMARDs;
- Prior bDMARDs (biologic DMARD) drug treatment (yes, no), yes is defined for a subject who used any bDMARDs;
- Prior TNFi (Tumor necrosis factors inhibitor) bDMARDs drug treatment (yes, no), yes is defined for a subject who used any TNFi bDMARDs;
- Prior non-TNFi bDMARDs drug treatment (yes, no), yes is defined for a subject who used any non-TNFi bDMARDs;
- Prior csDMARDs drug treatment (yes, no), yes is defined for a subject who used any csDMARDs;
- Prior non-DMARDs drug treatment (yes, no), yes is defined for a subject who used any non-DMARDs;
- Prior oral corticosteroids use (yes, no);
- Prior NSAIDs use (yes, no).

8.3.3. Concomitant Drug Treatments for Psoriatic Arthritis

Concomitant drug treatment will include any treatment for PsA during Study A3921092.

- DMARDs drug treatment (yes, no);
- Non-DMARDs drug treatment (yes, no);
- Oral corticosteroids use (yes, no);
- NSAIDs use (yes, no);
- Joint injection drug use (yes, no).

9. REFERENCES

1. Agresti A. *Categorical Data Analysis* (2nd Edn). Wiley: Hoboken, NJ, 2002.
2. Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation*, 2009, 120: 1640-1645.
3. Bruce, B., Fries, J.F. The health assessment questionnaire (HAQ). *Clinical and Experimental Rheumatology*; 2005, 23 (Supp. 39), S14-S18.
4. Gladman DD, Tom BDMM, Mease PJ and Farewell VT. Informing Response Criteria for Psoriatic Arthritis (PsA). II: Further Considerations and a Proposal - The PsA Joint Activity Index. *J Rheumatol* 2010, 37: 2559–2565.
5. Gottlieb A, Korman NJ, Gordon KB, et al. Guidelines of care for the management of psoriasis and psoriatic arthritis: Section 2. Psoriatic arthritis: overview and guidelines of care for treatment with an emphasis on the biologics. *J Am Acad Dermatol* 2008, 58: 851-64.
6. Helliwell PS, Fitzgerald O, Fransen J, et al. The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project). *Ann Rheum Dis*, 2013, 72: 986–991.
7. Madsen OR. Is DAS28-CRP with three and four variables interchangeable in individual patients selected for biological treatment in daily clinical practice? *Clin Rheumatol*, 2011, 30: 1577–1582.
8. Mumtaz A, Gallagher P, Kirby B, et al. Development of a preliminary composite disease activity index in psoriatic arthritis. *Ann Rheum Dis*, 2011, 70: 272-277.
9. Ramiro S, van Tubergen A, van der Heijde D, van den Bosch F, Dougados M, Landewé R. How to deal with missing items in BASDAI and BASFI. *Rheumatology*, 2014, 53: 374-376.
10. Schoels M, Aletaha D, Funovits J, Kavanaugh A, Baker D, Smolen JS. Application of the DAREA/DAPSA score for assessment of disease activity in psoriatic arthritis. *Ann Rheum Dis*, 2010, 69: 1441-1447.

10. APPENDICES

Appendix 1. Definition and Use of Visit Windows in Reporting

For reporting purposes, the following visit windows will be used for efficacy/PRO variables and for any safety displays that display by week/month. If two or more visits/observations fall into the same window, the visit/observation closest to the target day should be used in the analyses. If there is a tie, the later visit should be used. However, if two or more visits/observations fall in the same window, and the latest visit is the last visit of the study due to early termination, then this will be the visit used in the analysis for that visit window. Per the master visit schedule provided to the investigator sites, a month is defined as 4 weeks or 28 days.

Study day will be calculated as: date of assessment/collection – date of the first dose + 1.

Visit Label	Target Day	Definition [Day Window] - Lower	Definition [Day Window] - Upper
Screening		Prior to baseline visit	
Baseline		<p>The baseline values for the safety endpoints come from the baseline of the qualifying study for patients who enroll within 14 days of the last visit of the qualifying study. For the patients whose enrollment are out of 14-day window, their baseline values of the safety endpoints will be from the baseline visit of this study, A3921092.</p> <p>The baseline values for all the efficacy or health-outcome endpoints come from the baseline of the qualifying study.</p> <p>In addition, in the sensitivity analysis for the ACR response rates and ΔHAQ-DI, the baseline value will be the baseline visit for A3921092 for subjects whose enrollment is out of 14-day window.</p> <p>Day 1 = date of first dose of study treatment taken of study A3921092.</p>	
Month 1	29	1*	57
Month 3	85	58	127
Month 6	169	128	211
Month 9	253	212	295
Month 12	337	296	379
Month 15	421	380	463
Month 18	505	464	547
Month 21	589	548	631
Month 24	673	632	715
Month 27	757	716	799
Month 30	841	800	883
Month 33	925	884	967
Month 36	1009	968	1051 (for safety) ≥969 (for efficacy and PRO)
Month 39**	1093	1052	1135
Month 42**	1177	1136	1219
Month 45**	1261	1220	1303
Month 48**	1345	≥1304	

*The lower limit of the Month 1 visit window will be Day 2 for safety analysis for subjects who enroll out of 14-day window of the last visit of the qualifying study.

**For safety endpoints summarized by visit such as laboratory value and vital signs data at a windowed visit only.

Note some of the endpoints may not be measured at every visit. For example, SF-36v2, FACIT-F, EQ-5D-3L, PsA-HCRU, WLQ, DLQI, NAPSI and ISI will not be assessed at Months 3 per the study protocol and the visit window at Month 3 will be undefined. In the event that an observation falls into one of the undefined visit windows, that observation will not be used in analysis.

The following categories will be used in the treatment exposure and duration of treatment summaries. The duration will be calculated as: date of the last dose – date of the first dose +

1. Any missed doses/drug holidays will be ignored in the calculation.

Category (days)
1
2-14 (2 weeks)
15-28 (2 weeks-1 month)
29-84 (1-3 months)
85-168 (3-6 months)
169-252 (6-9 months)
253-336 (9-12 months)
337-504 (12-18 months)
505-672 (18-24 months)
673-840 (24-30 months)
841-1008 (30-36 months)
1009-1176 (36-42 months)
≥1177 (42-48 months)

Appendix 2. Further Definition of Endpoints

Appendix 2.1. ACR Assessments

The American College of Rheumatology's definition for calculating improvement in RA (ACR20) is calculated as a $\geq 20\%$ improvement in tender and swollen joint counts and $\geq 20\%$ improvement in 3 of the 5 remaining ACR-core set measures: patient and physician global assessments, pain, disability, and an acute-phase reactant. Similarly, ACR50 and 70 are calculated with the respective percent improvement. This efficacy measurement will be made at every study visit. The specific components of the ACR Assessments that will be used in this study are:

Tender/Painful Joint count (68);

Swollen Joint Count (66);

Patient's Assessment of Arthritis Pain (VAS);

Patient's Global Assessment of Arthritis (VAS);

Physician's Global Assessment of Arthritis (VAS);

C-Reactive Protein (CRP in mg/L);

Health Assessment Questionnaire – Disability Index (HAQ-DI).

If the value in any of the components at a timepoint is missing, the component variables that are not missing will be used to determine the response status. As a general principle, if there are sufficient non-missing components to determine whether the ACR endpoint is a response or non-response, then ACR endpoint is not missing, else if the available non-missing components are not sufficient to determine the response status of ACR endpoint then it is considered missing.

In order to avoid numerical difficulty, if the baseline value of any component is equal to 0, the following algorithm will be used in evaluating the percent change from baseline:

1. If change from baseline is also equal to 0, then percent change from baseline is set to be 0%;
2. If change from baseline is > 0 , then percent change from baseline is set to be 999999%.

These percentages will be used to derive the ACR endpoints. Change from baseline cannot be < 0 since none of the components should have negative value.

Lower limit of quantification (LLOQ) for CRP is 0.020 mg/dL (or 0.200 mg/L). Any CRP value below the LLOQ will be reported as " <0.020 " in database and will be set to 0.019 mg/dL (0.190 mg/L) and used in analyses throughout. The unit for CRP will be mg/dL or mg/L depending on endpoint of interest. The upper limit of the normal range is 2.87 mg/L.

The VAS data will need to be rescaled prior to any calculation ([Section 6](#)).

Appendix 2.2. Health Assessment Questionnaire – Disability Index (HAQ-DI)

The HAQ-DI assesses the degree of difficulty a subject has experienced during the past week in 8 domains of daily living activities: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and other activities. Each activity category or domain consists of 2-3 items. For each question in the questionnaire, the level of difficulty is scored from 0 to 3 with 0 representing “no difficulty,” 1 as “some difficulty,” 2 as “much difficulty,” and 3 as “unable to do”. The domain score for each domain is the maximum (ie, worst) of the scores from the items/questions within the domain. If this domain score is ≥ 2 , no further adjustment is needed. If it is < 2 (ie, 0 or 1) but no aids, devices or help from another person is indicated, then also no adjustment is needed. However, if it is < 2 and any aid, device or help from another person is indicated, then it is further adjusted upward to 2 as described below.

Any activity that requires assistance from another individual or requires the use of an aid or an assistive device adjusts to a minimum score of 2 to represent a more limited functional status.

Domain	If domain score is either 0 or 1, adjust to 2 when the following is satisfied.
Dressing and grooming	“Devices used for dressing (button hook, zipper pull, long-handled shoe horn, etc)” or help from others on “Dressing and Grooming” is checked
Arising	“Special or built up chair” or help from others on “Arising” is checked.
Eating	“Built up or special utensils” or help from others on “Eating” is checked.
Walking	“Cane”, “Walker”, “Crutches”, “Wheelchair”, or help form others on “Walking” is checked
Hygiene	“Raised toilet seat”, “Bathtub bar”, “Long-handled appliances in bathroom”, “Bathtub seat” or help from others on “Hygiene” is checked.
Reach	“Long-handled appliances for each” or help from others on “Reach” is checked
Grip	“Jar opener (for jars previously opened)” or help from others on “Gripping or opening things” is checked.
Activities	Help from others on “Errands and chores” is checked.
Note: For “Other, (specify)”, whether checked or unchecked or specifying the other “aids or devices” in this category, is not to be used in the adjustment of the domain score.	

For each domain, the domain score will be determined by non-missing scores from the questions. A domain score will only be missing if all scores within the domain are missing. The HAQ-DI score is the average of all the 8 domain scores. If > 2 domain scores are not complete or missing, HAQ-DI score is considered missing, else the HAQ-DI score is computed as the average of the non-missing domain scores (Bruce and Fries 2005).³

Change from baseline in HAQ-DI (Δ HAQ-DI) will be calculated.

HAQ-DI response will be defined as an improvement (ie, a decrease) from baseline of ≥ 0.30 and ≥ 0.35 (2 endpoints) in HAQ-DI scores.

Appendix 2.3. DAS 28-3 (CRP)

The Disease Activity Score (DAS) is a derived measurement with differential weighting given to each component. DAS 28-3 (CRP) will be calculated from measurements made at all visits (Madsen, 2011).⁷

The components of the DAS 28-3 arthritis assessment are:

Tender/Painful Joint Count (28);

Swollen Joint Count (28);

C-Reactive Protein (CRP in mg/L).

The 28 joint counts include the following joints: shoulders, elbows, wrists, metacarpophalangeal joints (MCP), proximal interphalangeal joints (PIP), and knees.

$$\text{DAS28-3(CRP)} = [0.56 * \text{sqrt}(\text{TJC28}) + 0.28 * \text{sqrt}(\text{SJC28}) + 0.36 * \ln(\text{CRP}+1)] * 1.10 + 1.15$$

where sqrt() refers to the square root, and ln() refers to the natural logarithm.

Any missing component will result in DAS28-3 (CRP) as missing. A higher score of DAS28-3 (CRP) represents a more severe disease activity.

Appendix 2.4. PsA Response Criteria (PsARC)

The PsARC will be collected/derived at all visits in addition to the ACR response criteria. The PsARC consists of 4 measurements:

Tender joint count (68);

Swollen joint count (66);

Physician's Global Assessment of Arthritis (VAS);

Patient's Global Assessment of Arthritis (VAS).

The same tender/painful joint count and swollen joint count used for ACR response criteria will be applied to the PsA Response Criteria. In order to be a 'PsARC responder', subjects must achieve improvement in 2 of 4 measures, one of which must be joint pain or swelling, without worsening in any measure.

Specifically, the PsARC response is defined as improvement in two of the following 4 criteria, one of which must be joint pain or swelling, without worsening in any measure: (1) $\geq 20\%$ improvement in Physician's Global Assessment of Arthritis (VAS); (2) $\geq 20\%$ improvement in Patient's Global Assessment of Arthritis (VAS); (3) $\geq 30\%$ improvement in tender joint count (68); and (4) $\geq 30\%$ improvement in swollen joint count (66) (Gottlieb et al 2008).⁵

Methods in calculating percent change from baseline when baseline value of any component is equal to 0 can be found in [Appendix 2.1](#).

If values in any of the components at a timepoint are missing, the components that are not missing are used to determine the response status. If one could not determine the response status in the presence of missing components at the timepoint, then PsARC response will be considered as missing. The VAS data will need to be rescaled prior to any calculation ([Section 6](#)).

Appendix 2.5. PsA Joint Activity Index (PsAJAI)

The PsA Joint Activity Index (PsAJAI) is a weighted sum of $\geq 30\%$ improvement in the 7 core measures of the ACR with weights of 2 given to the joint count measures (JNT, defined below), CRP and the physician's global assessment of disease activity (MDGDA). Weights of 1 are given to the remaining $\geq 30\%$ improvement measures including patient's assessment of arthritis pain (PAIN), patient's global assessment of Arthritis (PtGDA) and HAQ-DI (Gladman et al 2010).⁴ The values are summed to get a score out of 9 (range 0-9). The score is calculated as follows:

$$\text{PsAJAI} = 2 \times 30\% \downarrow \text{JNT} + 2 \times 30\% \downarrow \text{CRP} + 2 \times 30\% \downarrow \text{MDGDA} + 30\% \downarrow \text{PtGDA} + 30\% \downarrow \text{PAIN} + 30\% \downarrow \text{HAQ-DI}.$$

The formula contains indicator functions. For example, if a subject has a $\geq 30\%$ improvement in CRP, the term “ $30\% \downarrow \text{CRP}$ ” will be set to 1 in the above formula; otherwise it is set to 0. For JNT, “ $30\% \downarrow \text{JNT}$ ” will be set to 1 only if BOTH $\geq 30\%$ improvement/reduction in tender/painful joints (68) AND $\geq 30\%$ improvement/reduction in swollen joints (66), otherwise “ $30\% \downarrow \text{JNT}$ ” will be set to 0.

Methods in calculating percent change from baseline when baseline value of any component is equal to 0 can be found in [Appendix 2.1](#).

PsAJAI is set to missing if any of the 7 components (JNT is comprised of tender and swollen joint counts) is missing.

The VAS data will need to be rescaled prior to any calculation ([Section 6](#)).

Appendix 2.6. Disease Activity Index for Reactive Arthritis/Psoriatic Arthritis (DAREA/DAPSA)

DAREA/DAPSA (Schoels et al 2010)¹⁰ is a composite instrument to assess peripheral joint involvement that is based upon numerical summation of 5 variables of disease activity: tender/painful joint count + swollen joint count (using SJC66/ TJC68 assessments), patient's global assessment of Arthritis (PtGA in cm), patient's assessment of Arthritis Pain (PAIN in cm) and CRP (in mg/dL). Since DAREA/DAPSA reflects domains found important in PsA, it has been proposed to serve as a Disease Activity Index for Psoriatic Arthritis (DAPSA). DAREA/DAPSA is calculated as follows:

$$DAREA/DAPSA = SJC66 + TJC68 + PtGA + PAIN + CRP.$$

Any missing component will result in DAREA/DAPSA as missing.

The VAS data will need to be rescaled prior to any calculation (Section 6).

Appendix 2.7. Composite Psoriatic Disease Activity Index in Psoriatic Arthritis (CPDAI)

CPDAI (Mumtaz et al 2011)⁸ is a composite psoriatic disease activity index incorporating 5 domains including joint disease, skin involvement, enthesitis, dactylitis and spinal disease. For dactylitis, a simple dactylitic digit count was applied. For enthesitis, the Leeds enthesitis index (LEI) was used. For spinal involvement, the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and the Ankylosing Spondylitis Quality of Life (ASQOL) were used. In addition, the Dermatology Quality of Life Index (DLQI) was used as a measure of the impact of skin disease. Using these instruments, and ranges of values based upon literature review, disease activity under each domain was graded as none, mild, moderate and severe giving a range of attainable CPDAI scores of between 0 and 15. A higher CPDAI score indicates a higher disease activity.

Composite Psoriatic Disease Activity Index (CPDAI) score total 0-15

	<i>Not involved (0)</i>	<i>Mild (1)</i>	<i>Moderate (2)</i>	<i>Severe (3)</i>
<i>Peripheral arthritis</i>		<i>≤4 Joints (count a joint that is swollen and/or tender); normal function (HAQ-DI ≤0.5)*</i>	<i>≤4 Joints but function impaired; or >4 joints, normal function</i>	<i>> 4 Joints and function impaired</i>
<i>Skin disease</i>		<i>PASI ≤10 and DLQI ≤10</i>	<i>PASI ≤10 but DLQI >10; or PASI >10 but DLQI ≤10</i>	<i>PASI >10 and DLQI >10</i>
<i>Enthesitis</i>		<i>≤3 Sites; normal function (HAQ-DI ≤0.5)*</i>	<i>≤3 Sites but function impaired; or >3 sites but normal function</i>	<i>>3 Sites and function impaired</i>
<i>Dactylitis</i>		<i>≤3 Digits; normal function (HAQ-DI ≤0.5)*</i>	<i>≤3 Digits but function impaired; or >3 digits but normal function</i>	<i>>3 Digits and has function impaired</i>
<i>Spinal disease</i>		<i>BASDAI ≤4; normal function (ASQoL ≤6)</i>	<i>BASDAI >4 but normal function; BASDAI ≤4 but function impaired</i>	<i>BASDAI >4 and function impaired</i>

**Health assessment questionnaire-disability index (HAQ-DI) only counted if clinical involvement of domain (joint/ethesis/dactylitis) present. DLQI or ASQoL only counted if clinical involvement of domain (ie, PASI>0 or BASDAI>0) present.*

Explicitly, each disease dimension will be scored as follows (0=not involved, 1=mild, 2=moderate, 3=severe). For evaluating peripheral arthritis score (D1 below), if a joint is swollen and/or tender, it is only counted once.

D1. Peripheral arthritis scoring

	HAQ-DI	
Joints (swollen or tender)	≤0.5	>0.5
0	0	0
1 to 4	1	2
>4	2	3

D2. Skin disease scoring

	DLQI	
PASI	≤10	>10
0	0	0
0< to ≤10	1	2
>10	2	3

D3. Leeds enthesitis scoring

	HAQ-DI	
Sites	≤0.5	>0.5
0	0	0
1 to 3	1	2
>3	2	3

D4. Dactylitis scoring

	HAQ-DI	
Digits	≤0.5	>0.5
0	0	0
1 to 3	1	2
>3	2	3

D5. Spinal disease scoring

	ASQol	
BASDAI	≤6	>6
0	0	0
0< to ≤4	1	2
>4	2	3

$$CPDAI = D1 + D2 + D3 + D4 + D5.$$

CPDAI will be set as missing if any of the component score is missing. Per the study protocol, PASI should only be performed if ≥3% of subject's BSA is affected at baseline, thus CPDAI will be set as missing and excluded from any analysis for any subject (the entire subject) whose baseline BSA is <3%.

Appendix 2.8. Minimal Disease Activity (MDA) Response

A psoriatic arthritis patient is defined as having Minimal Disease Activity (MDA, yes/no) when the patient meets ≥5 of the 7 following criteria: 1) tender/painful joint count ≤1; 2) swollen joint count ≤1; 3) PASI score ≤1 or BSA ≤3%; 4) patient assessment of Arthritis Pain (VAS) ≤15 mm; 5) patient's global arthritis assessment (VAS) ≤20 mm; 6) HAQ-DI score ≤0.5; 7) tender enthesal points (using Leeds Enthesitis Index) ≤1.

If < 5 non-missing component criteria are satisfied and at least one of the remaining component criteria is missing, then MDA response is set to be as missing.

The VAS data will need to be rescaled prior to any calculation ([Section 6](#)).

Appendix 2.9. Psoriatic Arthritis Disease Activity Score (PASDAS)

PASDAS (Helliwell et al 2013)⁶ is a composite PsA disease activity score that includes the following components: patient global joint and skin assessment (VAS in mm), physician global psoriatic arthritis assessment (VAS in mm), swollen (66 joints) and tender joint counts (68 joints), Leeds Enthesitis Index score, tender dactylitic digit score, physical component summary score (PCS) of SF-36v2 and CRP (mg/L).

$$\begin{aligned} \text{PASDAS} = & (((0.18 \times \sqrt{\text{Physician global VAS}}) + (0.159 \times \sqrt{\text{Patient global VAS}}) \\ & - (0.253 \times \sqrt{\text{SF36-PCS}}) + (0.101 \times \text{LN}(\text{Swollen joint count} + 1)) \\ & + (0.048 \times \text{LN}(\text{Tender joint count} + 1)) \\ & + (0.23 \times \text{LN}(\text{Leeds enthesitis index score} + 1)) \\ & + (0.37 \times \text{LN}(\text{Tender dactylitis count} + 1)) \\ & + (0.102 \times \text{LN}(\text{CRP} + 1)) + 2) \times 1.5 \end{aligned}$$

where LN() is the natural logarithm. For patient’s global joint and skin assessment (VAS), it is obtained only from the first question in the CRF for assessing subject’s global joint and skin, “In all the ways in which your PSORIASIS and ARTHRITIS, as a whole, affects you, how would you rate the way you felt over the past week?” The physician’s global psoriatic arthritis assessment (VAS) is obtained from the CRF at the prompt: “THE SUBJECT’S OVERALL PSORIATIC ARTHRITIS AT THIS TIME IS...”. The tender dactylitis digit score is the number of digits out of 20 (fingers and toes) where rating of either 2 (digit is tender and swollen (active)) or 3 (digit is very swollen and tender or very tender and swollen (highly active)) is registered. Any missing component will result in PASDAS as missing. The VAS data will need to be rescaled prior to any calculation (Section 6). A higher PASDAS score indicates a higher disease activity.

Appendix 2.10. Psoriasis Area and Severity Index (PASI)

The Psoriasis Area and Severity Index quantifies the severity of a subject’s psoriasis based on both *lesion severity* and the percentage of *body surface area* (BSA) affected.

- **Lesion severity:** The basic characteristics of psoriatic lesions – erythema, induration and scaling – provide a means for assessing the severity of lesions. Assessment of these three main signs is performed separately for four areas of the body: head, upper limbs, trunk, and lower limbs. Average erythema, induration and scaling are rated for each body area according to a 5-point scale: 0, no involvement; 1, slight; 2, moderate; 3, marked; 4, very marked.
- **Body surface area involvement (%BSA):** The extent (%) to which each of the four areas of the body is affected by psoriasis is assigned a numerical score according to the following area scoring criteria: 0, no involvement; 1, >0 to 9%; 2, 10 to 29%; 3, 30 to 49%; 4, 50 to 69%; 5, 70 to 89%; 6, 90 to 100%.

In each area, the sum of the severity rating scores for erythema, induration and scaling is multiplied by the score representing the percentage of this area involved by psoriasis,

multiplied by a weighting factor (head 0.1; upper limbs 0.2; trunk 0.3; lower limbs 0.4). The sum of the numbers obtained for each of the four body areas is the PASI.

$$\text{PASI} = 0.1A_h (E_h + I_h + S_h) + 0.2A_u (E_u + I_u + S_u) + 0.3A_t (E_t + I_t + S_t) + 0.4A_l (E_l + I_l + S_l)$$

Where A = area of involvement score; E = erythema; I = induration; S = scaling; h = head; u = upper limbs; t = trunk; l = lower limbs.

The PASI score can vary in increments of 0.1 units from 0.0 to 72.0, with higher scores representing increasing severity of psoriasis.

Each PASI clinical signs component score can be computed as:

$$\text{PASI}_E = 0.1A_h (E_h) + 0.2A_u (E_u) + 0.3A_t (E_t) + 0.4A_l (E_l)$$

$$\text{PASI}_I = 0.1A_h (I_h) + 0.2A_u (I_u) + 0.3A_t (I_t) + 0.4A_l (I_l)$$

$$\text{PASI}_S = 0.1A_h (S_h) + 0.2A_u (S_u) + 0.3A_t (S_t) + 0.4A_l (S_l)$$

Any missing component will result in PASI as missing.

BSA (number of handprints; %) will be an endpoint of interest.

Appendix 2.11. Physician Global Assessment (PGA) of Psoriasis (PGA-PsO)

The Physician Global Assessment of Psoriasis is scored on a 5-point scale, reflecting a global consideration of the erythema, induration and scaling across all psoriatic lesions. Average erythema, induration and scaling are rated separately over the whole body according to a 5-point severity scale, scored as 0 (none), 1, 2, 3, or 4 (most severe). The severity rating scores are summed and the average taken – the total average is rounded to the nearest whole number score to determine the PGA.

Appendix 2.12. Dermatology Life Quality Index (DLQI)

The DLQI is a general dermatology questionnaire that consists of 10 items that assess patient health-related quality of life (daily activities, personal relationships, symptoms and feelings, leisure, work and school, and treatment). The DLQI is scored such that a higher score indicates greater impairment.

Total raw score = sum (items 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)

The scoring of each question other than Question 7 is:

Very much	is scored 3;
A lot	is scored 2;
A little	is scored 1;
Not at all	is scored 0;

Not relevant is scored 0.

For Question 7:

if yes, then scored 3;

if no or not relevant, then scored 0;

if it is answered “no” or “not relevant” but then either “a lot” or “a little” is checked this is then scored 2 or 1, respectively.

If a question is unanswered it is scored 0 and the scores are summed and expressed as usual out of a maximum of 30. If two or more questions are left unanswered the DLQI is not scored.

Appendix 2.13. Itch Severity Item (ISI)

The severity of itching due to psoriasis will be assessed using a single-item, 0 to 10 horizontal numeric rating scale (NRS). Subjects will be asked to assess their itch over the past 24 hours on the NRS anchored by the terms “no itching” (0) and “worst possible itching” (10) at the ends.

Appendix 2.14. Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

The BASDAI score is obtained by computing the mean score for the 2 questions related to morning stiffness (questions 5 and 6) and then adding that value to the sum of the scores for the first 4 questions and then dividing the total by 5. This can be written as:

$$\text{BASDAI} = (Q1 + Q2 + Q3 + Q4 + (Q5 + Q6)/2) / 5$$

The total score will range from 0.0 to 10.0 cm.

If the answer to either any 1 question is missing, then the BASDAI score will be calculated using the remaining non-missing answers. If >1 answers are missing, then the BASDAI is considered as missing (Ramiro, 2014).⁹

Explicitly, the calculation of BASDAI score for any 1 missing answer is detailed in the table below:

Missing Q5 or Q6	Missing any one of Q1-Q4	Formula
No	Yes	$\text{BASDAI} = [\text{sum of the 3 non-missing of } (Q1, Q2, Q3, Q4) + (Q5 + Q6)/2] / 4$
Yes	No	$\text{BASDAI} = [Q1 + Q2 + Q3 + Q4 + (\text{non-missing of } Q5 \text{ or } Q6)] / 5$

The raw VAS data will need to be rescaled to a 10 cm scale prior to any calculation ([Section 6](#)).

Appendix 2.15. EuroQol-5D Health Questionnaire 3-Level (EQ-5D-3L)

On the EQ-5D (subject version, 3 categories of response per question), 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) are assessed. The status of each dimension has three possible responses (no problem in the relevant health dimension to severe problems).

The EQ VAS records the patient's self-rated health on a vertical visual analogue scale where the endpoint is labelled 'Best imaginable health state' and 'Worst imaginable health state'. Based on the patient's mark on the VAS form a score ranging from 0 to 100 mm is recorded. The VAS data will need to be rescaled prior to any calculation ([Section 6](#)).

Appendix 2.16. Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)

FACIT-F is a 13 item questionnaire (each item score ranges from 0 to 4). There will be 3 endpoints derived: Δ FACIT-F total score, Δ FACIT-F experience domain score and Δ FACIT-F impact domain score. FACIT-F total score (range 0-52) is calculated by summing the 13 items. FACIT-F experience domain score (range 0-20) is calculated by summing 5 items of Q1 I feel fatigued, Q2 I feel weak all over, Q3 I feel listless ("washed out"), Q4 I feel tired and Q7 I have energy, while FACIT-F impact domain score (range 0-32) is calculated by summing the remaining 8 items. Note that all the scores except Q7 & Q8 should be reversed prior to summing for both the total score and the domain scores such that higher scores represent better functioning.

All responses are added with equal weight to obtain the total score. In cases where some answers are missing, a total score is prorated from the score of the answered items, so long as more than 50% of the items (ie, at least 7 of 13 for FACIT-F total score, at least 3 of 5 for FACIT-F experience domain score, and at least 5 of 8 for FACIT-F impact domain score) are answered.

Appendix 2.17. Short Form 36 (SF-36, version 2, acute)

The SF-36 v.2 (Acute version) is a 36-item generic health status measure. It measures 8 general health concepts or domains: physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional and mental health. These domains can also be summarized as physical and mental component scores.

These 8 domains are as follows:

- a. Physical Functioning (PF). This score is based on the responses to the 10 items that compose Question 3 and reflects the degree to which various physical activities have been limited in the previous week by the subject's health.

- b. Role-Physical (RP). This score is based on the responses to the four items that compose Question 4 and reflects the relative amount of time that the subject has had problems with work or other regular daily activities as a result of their physical health during the previous week.
- c. Bodily Pain (BP). This score is based on the responses to Questions 7 and 8 and reflects bodily pain and its effects on normal work during the previous week.
- d. General Health (GH). This score is based on responses to Question 1 and the four items in Question 11 and reflects the subject's perception of their general health during the previous week.
- e. Vitality (VT). This score is based on responses to Question 9 items a, e and g, and reflects the subject's physical energy level relative to time during the previous week.
- f. Social Functioning (SF). This score is based on responses to Questions 6 and reflects how physical health or emotional problems have interfered with social activities during the previous week.
- g. Role-Emotional (RE). This score is based on responses to the three items in Question 5 and reflects the amount of time during the previous week that emotional problems have interfered with work or regular daily activities.
- h. Mental Health (MH). This score is based on responses to Question 9 items b, c, d, f, and h and reflects various mental/emotional states relative to time during the previous week.
- i. Health Transition (TR). This score is based on the response to Question 2 and is a rating of current general health compared to one week previous.

The summary component scores are:

- Physical Component Summary (PCS).
- Mental Component Summary (MCS).

The summary component scores, PCS and MCS, are based on a normalized sum of the 8 scale scores PF, RP, BP, GH, VT, SF, RE, and MH. All domains and summary components are scored such that a higher score indicates a higher functioning or health level.

Data Derivation Details to Obtain Scale Scores for SF-36v2

VARIABLE	DERIVATION
SF-36v2 PF scale score	<p>raw score = sum (items 3A, 3B, 3C, 3D, 3E, 3F, 3G, 3H, 3I, 3J) $PF = (\text{raw score} - 10) * 5$ $PF_Z = (PF - 82.62455) / 24.43176$ PF scale score = (PF_Z*10) + 50</p> <p>When calculating the raw score, if 5 or more of the items are non-missing then replace any missing values as follows:</p> <p>Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores.</p> <p>Otherwise, if less than 5 of the items are non-missing then PF scale score is missing.</p> <p>The response scale for each activity ranges from 1 to 3 where 1=limited a lot, 2=limited a little, and 3=not limited at all.</p> <p>A higher PF scale score indicates better physical functioning.</p>
SF-36v2 RP scale score	<p>raw score = sum (items 4A, 4B, 4C, and 4D) $RP = [(\text{raw score} - 4) / 16] * 100$ $RP_Z = (RP - 82.65109) / 26.19282$ RP scale score = (RP_Z * 10) + 50</p> <p>When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows:</p> <p>Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores.</p> <p>Otherwise, if less than 2 of the items are non-missing then RP scale score is missing.</p> <p>The response scale for each item ranges from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher RP scale score indicates better role-physical functioning.</p>

SF-36v2 BP scale score	<p>raw score = sum (reversed item 7 and reversed item 8) $BP = (\text{raw score} - 2) * 10$ $BP_Z = (BP - 73.86999) / 24.00884$ BP scale score = (BP_Z * 10) + 50</p> <p>Reverse direction of Item 7 as follows: if=1, set to 6; if=2, set to 5.4; if=3, set to 4.2; if=4, set to 3.1; if=5, set to 2.2; if=6, set to 1.</p> <p>Reverse direction of item 8 as follows: if=1 and original value of item 7=1, set to 6; if=1 and original value of item 7>=2, set to 5; if=2, set to 4; if=3, set to 3; if=4, set to 2; if=5, set to 1.</p> <p>If item 7 is answered and item 8 is missing, set reversed 8 = reversed 7 as defined above. If 8 is answered and 7 is missing, set reversed 7 as reverse item 8 as follows: if=1, set to 6; if=2, set to 4.75; if=3, set to 3.5; if=4, set to 2.25; if=5, set to 1.</p> <p>If 1 or more questions were answered, calculate BP scale score as defined above. If neither question was answered then BP scale score is missing.</p> <p>The scale for Question 7, amount of bodily pain, ranges from 1 to 6 where 1=None, 2=Very mild, 3=mild, 4=Moderate, 5=Severe, and 6=Very severe. The scale for Question 8, the degree to which pain interfered with normal work, ranges from 1 to 5 where 1=Not at all, 2=A little bit, 3=Moderately, 4=Quite a bit, and 5=Extremely. A higher BP scale score indicates lack of bodily pain.</p>
------------------------	---

SF-36v2 GH scale score	<p>raw score = sum (reversed item 1, item 11A, reversed 11B, 11C and reversed 11D) GH = (raw score -5) * 5 $GH_Z = (GH - 70.78372) / 21.28902$ GH scale score = (GH_Z * 10) + 50</p> <p>Reverse direction of Item 1 as follows: if=1, set to 5; if=2, set to 4.4; if=3, set to 3.4; if=4, set to 2; if=5, set to 1.</p> <p>Reverse direction of item 11B and 11D by subtracting score from 6.</p> <p>When calculating the raw score, if 3 or more of the items are non-missing then replace any missing values as follows:</p> <p>Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores.</p> <p>Otherwise, if less than 3 of the items are non-missing then GH scale score is missing.</p> <p>Responses for Question 1, an assessment of self-perceived health status, range from 1 to 5 where 1=Excellent, 2=Very good, 3=Good, 4=Fair, and 5=Poor. Responses for the items in Question 11 range from 1 to 5 where 1=Definitely true, 2=Mostly true, 3=Don't know, 4=Mostly false, and 5=Definitely false and reflect the subject's perception of their relative health and expectations of their future health status. A higher GH scale score indicates better general health perceptions.</p>
SF-36v2 VT scale score	<p>raw score = sum (reversed item 9a, reversed 9e, 9g and 9i) $VT = [(raw\ score - 4) / 16] * 100$ $VT_Z = (VT - 58.41968) / 20.87823$ VT scale score = (VT_Z * 10) + 50</p> <p>Reverse direction of Items 9a and 9e by subtracting score from 6.</p> <p>When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows:</p> <p>Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores.</p> <p>Otherwise, if less than 2 of the items are non-missing then VT scale score is missing.</p> <p>The scale for these items ranges from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher VT scale score indicates more vitality.</p>

<p>SF-36v2 SF scale score</p>	<p>raw score = sum (reversed 6 and 10) $SF = [(raw\ score - 2) / 8] * 100$ $SF_Z = (SF - 85.11568) / 23.24464$ SF scale score = (SF_Z * 10) + 50</p> <p>Reverse direction of score for item 6 by subtracting score from 6.</p> <p>When calculating the raw score, if 1 of the items is missing then substitute the missing score with the score on the non- missing item. If both items are missing then SF scale score is missing.</p> <p>Responses to Question 6, an assessment of the extent to which health/emotional problems interfered with social activities, range from 1 to 5 where 1=Not at all, 2=Slightly, 3=Moderately, 4=Quite a bit, and 5=Extremely. Responses to Question 10 reflect the amount of time that health/emotional problems interfered with social activities and range from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher SF scale score indicates better social functioning.</p>
<p>SF-36v2 RE scale score</p>	<p>raw score = sum (items 5A, 5B, and 5C) $RE = [(raw\ score - 3) / 12] * 100$ $RE_Z = (RE - 87.50009) / 22.01216$ RE scale score = (RE_Z * 10) + 50</p> <p>When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows:</p> <p>Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores.</p> <p>Otherwise, if less than 2 of the items are non-missing then RE scale score is missing.</p> <p>Responses to the items in Question 5 range from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher RE scale score indicates better role-emotional functioning.</p>
<p>SF-36v2 MH <u>scale</u> score</p>	<p>raw score = sum (items 9B, 9C, reversed 9D, 9F and reversed 9H) $MH = (raw\ score - 5) * 5$ $MH_Z = (MH - 75.76034) / 18.04746$ MH scale score = (MH_Z * 10) + 50</p> <p>Reverse direction of scores for 9D and 9H, by subtracting score from 6.</p> <p>If 3 or more of the items are non-missing then replace any missing values as follows:</p> <p>Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores.</p> <p>Otherwise, if less than 3 of the items are non-missing then MH scale score is missing.</p> <p>The scale for these items ranges from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher MH scale score indicates better mental health.</p>

SF-36v2 TR scale score	<p>raw score = item 2 TR scale score = raw score</p> <p>The scale for this item ranges from 1 to 5 where 1=Much better now than one week ago, 2=Somewhat better now than one week ago, 3=About the same as one week ago, 4=Somewhat worse now than one week ago, and 5=Much worse now than one week ago. A higher TR scale score indicates worse general health currently relative to one week previous.</p>
SF-36v2 PCS score	<p>PCS score includes the 8 scales for GH, PF, RP, RE, SF, MH, BP, and VT.</p> <p>PF1= (PF-82.62455)/24.43176; RP1=(RP-82.65109)/26.19282; BP1=(BP-73.86999)/24.00884; GH1 = (GH-70.78372)/21.28902; VT1= (VT-58.41968)/20.87823; SF1=(SF-85.11568)/23.24464; RE1= (RE-87.50009)/22.01216; MH1=(MH-75.76034)/18.04746;</p> <p>Raw Score = ((GH1*.24954)+(PF1* .42402)+(RP1*.35119)+ (RE1*-.19206)+(SF1*-.00753)+(MH1*-.22069)+(BP1*.31754)+ (VT1*.02877))</p> <p>PCS Summary Scale Score = (raw score *10) + 50</p> <p>Raw Score is missing if one of the component scale scores is missing.</p>
SF-36v2 MCS score	<p>MCS score includes the 8 scales for GH, PF, RP, RE, SF, MH, BP, and VT.</p> <p>PF1= (PF-82.62455)/24.43176; RP1=(RP-82.65109)/26.19282; BP1=(BP-73.86999)/24.00884; GH1 = (GH-70.78372)/21.28902; VT1= (VT-58.41968)/20.87823; SF1=(SF-85.11568)/23.24464; RE1= (RE-87.50009)/22.01216; MH1=(MH-75.76034)/18.04746;</p> <p>Raw Score =((GH1*-.01571)+(PF1*-.22999)+(RP1*-.12329)+ (RE1*.43407)+(SF1*.26876)+(MH1*.48581)+(BP1*-.09731)+ (VT1*.23534))</p> <p>MCS Summary Concept Score = (raw score *10) + 50</p> <p>Raw Score is missing if one of the component scale scores is missing.</p>

Appendix 3. Criteria for Clinical Diagnosis of the Metabolic Syndrome

Criteria for clinical diagnosis of the metabolic syndrome will be based on the 2009 Joint Statement Harmonizing the Metabolic Syndrome (Alberti et al, 2009).¹ There are 5 measures or risk factors as detailed in Table 1 of the 2009 Joint statement and excerpted below:

Measure	Categorical Cut Points
Elevated waist circumference*	Population- and country-specific definitions
Elevated triglycerides (drug treatment for elevated triglycerides is an alternate indicator†)	≥150 mg/dL (1.7 mmol/L)
Reduced HDL-C (drug treatment for reduced HDL-C is an alternate indicator†)	<40 mg/dL (1.0 mmol/L) in males; <50 mg/dL (1.3 mmol/L) in females
Elevated blood pressure (antihypertensive drug treatment in a patient with a history of hypertension is an alternate indicator)	Systolic ≥130 and/or diastolic ≥85 mm Hg
Elevated fasting glucose‡ (drug treatment of elevated glucose is an alternate indicator)	≥100 mg/dL

HDL-C indicates high-density lipoprotein cholesterol.

*It is recommended that the IDF cut points be used for non-Europeans and either the IDF or AHA/NHLBI cut points used for people of European origin until more data are available.

†The most commonly used drugs for elevated triglycerides and reduced HDL-C are fibrates and nicotinic acid. A patient taking 1 of these drugs can be presumed to have high triglycerides and low HDL-C. High-dose ω-3 fatty acids presumes high triglycerides.

‡Most patients with type 2 diabetes mellitus will have the metabolic syndrome by the proposed criteria.

Any 3 abnormal findings out of the 5 measures will qualify a subject for the metabolic syndrome.

The baseline data of qualifying study will be used to derive the metabolic syndrome.

- Any drug treatment taken on Day 0 recorded on the “CONCOMITANT DRUG TREATMENT - ANTI-DIABETIC” CRF will be considered having the drug treatment of elevated glucose criterion satisfied.
- Any drug treatment taken on Day 0 recorded on the “CONCOMITANT DRUG TREATMENT - ANTI-HYPERTENSIVE MEDICATIONS” CRF will be considered having the drug treatment of elevated blood pressure criterion satisfied.

- Any drug treatment taken on Day 0 recorded on the “CONCOMITANT DRUG TREATMENT – LIPID LOWERING AGENTS” CRF will be considered having the drug treatment of elevated triglycerides criterion satisfied.
- For the HDL-C criterion, only the lab categorical cut points will be considered and the drug treatment will not be considered.
- The waist circumference thresholds will be population and country specific per Table 2 of the 2009 Joint Statement excerpted below. Explicitly, the US specific threshold will be used for the subjects enrolled from US, the Canada threshold for the subjects enrolled from Canada, the European threshold for the subjects enrolled from Western Europe, Eastern Europe and Russia, the Asian (including Japanese) threshold for the subjects enrolled from Asia, and Ethnic Central and South American threshold for the subjects enrolled from Central and South America.

Population	Organization (Reference)	Recommended Waist Circumference Threshold for Abdominal Obesity	
		Men	Women
Europid	IDF (4)	≥94 cm	≥80 cm
Caucasian	WHO (7)	≥94 cm (increased risk) ≥102 cm (still higher risk)	≥80 cm (increased risk) ≥88 cm (still higher risk)
United States	AHA/NHLBI (ATP III)* (5)	≥102 cm	≥88 cm
Canada	Health Canada (8,9)	≥102 cm	≥88 cm
European	European Cardiovascular Societies (10)	≥102 cm	≥88 cm
Asian (including Japanese)	IDF (4)	≥90 cm	≥80 cm
Asian	WHO (11)	≥90 cm	≥80 cm
Japanese	Japanese Obesity Society (12)	≥85 cm	≥90 cm
China	Cooperative Task Force (13)	≥85 cm	≥80 cm
Middle East, Mediterranean	IDF (4)	≥94 cm	≥80 cm
Sub-Saharan African	IDF (4)	≥94 cm	≥80 cm
Ethnic Central and South American	IDF (4)	≥90 cm	≥80 cm

Appendix 4. Definition of Diabetes Mellitus at Baseline

A subject is considered having diabetes mellitus at baseline of the qualifying study, if any of the conditions below is met:

- Subject had a diagnosis of diabetes mellitus recorded on the “CARDIOVASCULAR RISK FACTORS” CRF or on the “SIGNIFICANT MEDICAL HISTORY AND NON-DRUG ALLERGIES” CRF at screening;
- Subject received any drug treatment taken on Day 0 recorded on the “CONCOMITANT DRUG TREATMENT - ANTI-DIABETIC” CRF;
- Subject’s HbA1c $\geq 6.5\%$ at baseline or, if HbA1c is not available, baseline fasting plasma glucose ≥ 126 mg/dL.

Appendix 5. Handling of Joint Count Assessment (Missing Assessment, “NOT DONE” and “NOT APPLICABLE”) and Injected Joint

This Section applies to the tender/painful joint (68) and swollen joint (66).

For missing or “NOT DONE” assessment:

- A missing painful / tender assessment or one “NOT DONE” at baseline is set to “not painful / tender”;
- A missing swollen assessment or one “NOT DONE” at baseline is set to “not swollen”;
- A missing painful / tender assessment or one “NOT DONE” post-baseline is set to “painful / tender”;
- A missing swollen assessment or one “NOT DONE” at post-baseline is set to “swollen”.

For “NOT APPLICABLE” assessment:

- A joint marked “NOT APPLICABLE” is not to be counted in the summation of swollen and painful / tender joints at baseline and post baseline;
- Any new “NOT APPLICABLE” for a joint post baseline is set to “painful / tender” and “swollen”. That is, if a joint is assessed, “Present”, “Absent”, “NOT DONE”, or with missing assessment at baseline, the same joint marked as “NOT APPLICABLE” post-baseline will be counted as “painful / tender” and “swollen”.

If a joint received an intra-articular injection (baseline or post-baseline), then set the joint status to “painful/tender” and “swollen” on and after the date of injection.

Appendix 6. Methotrexate Withdrawal Sub-Study

1. INTRODUCTION

The purpose of this sub-study is to assess the efficacy, safety and tolerability of tofacitinib 5 mg BID administered as monotherapy after methotrexate (MTX) withdrawal compared to tofacitinib 5 mg BID continued in combination with MTX in subjects with active PsA. Subjects who have completed at least 24 months of participation in the long term extension study A3921092 and are currently receiving methotrexate may be included.

Subjects will be randomized to the 2 treatment arms in a 1:1 ratio, one in which they will receive open label tofacitinib 5 mg tablets BID and blinded methotrexate capsules (oral dose range from 7.5 mg to 20 mg per week) and the second in which they will receive open label tofacitinib 5 mg tablets BID and blinded MTX placebo capsules. The sub-study will be approximately 12 (standardized 4-week) months in duration and visits will occur at Baseline, Months 1, 3, 6, 9 and 12. The sub-study baseline will occur at the same time as a scheduled study visit at or after Month 24 of A3921092. This visit will be referred to as Sub-study Switch Visit or Baseline visit (Sub-study Switch/Baseline visit).

1.1. Study Design

This sub-study within Study A3921092 is a Phase 3, randomized, double-blind, placebo-controlled, parallel group estimation study designed to assess the efficacy, safety and tolerability of tofacitinib 5 mg BID versus tofacitinib 5 mg BID in combination with concomitant MTX in subjects with active PsA. Subjects with active PsA will have previously completed at least 24 months of treatment with tofacitinib in Study A3921092 and will be on a stable dose of tofacitinib 5 mg BID for at least 3 months prior to the first dose of sub-study drug and be on a stable dose of oral MTX at least 4 weeks prior to the first dose of sub-study drug. For subjects who are qualified to enter into the sub-study, the Combined A3921092 Sub-study Switch/Baseline visit will occur at a scheduled study visit at or after Month 24 of A3921092. There will be no separate Baseline visit for the sub-study. The visit window will remain as in the LTE in which a month is considered 28 days. Once a subject enters the sub-study, they will complete their participation in the sub-study portion and be unable to re-enter the main study. Subjects may be discontinued from the sub-study at any time during the sub-study in the event of deterioration of response as determined by the investigator.

Eligible subjects from study A3921092, will be randomized into A3921092 sub-study in a 1:1 ratio to tofacitinib 5 mg BID + MTX placebo or tofacitinib 5 mg BID + MTX. MTX will be blinded to the subjects, investigator and the sponsor. The subjects in the tofacitinib 5 mg BID + MTX group will receive blinded MTX capsules (range dose of 7.5-20 mg a week based upon the pre-existing dose) and the subjects in the tofacitinib 5 mg BID group will receive the matching MTX placebo capsules. For the duration of the sub-study tofacitinib dose must remain stable at 5 mg BID. The MTX dose and oral route of administration must remain stable for the 12 month duration of the sub-study. All subjects will receive open-label tofacitinib 5 mg BID throughout the sub-study without dose increase. The total number of subjects randomized will be approximately 180 with 90 in each treatment group. Treatment duration for subjects participating in this sub-study is approximately 12 months.

At various time points in this trial, safety measurements, including physical examination, clinical laboratory tests, adverse event monitoring, electrocardiograms (ECGs) and vital signs will be performed. All subjects will be monitored for clinical evidence of PsA response to treatment. Health Outcomes Measures (ie, Patient Reported Outcomes assessments for pain, quality of life, physical function, fatigue and health status) will also be performed at various time points in this trial. In addition, subjects will be monitored for serious infections, lymphadenopathy and lymphoproliferative disorder (LPD).

1.2. Study Objectives

Primary Objective

- *To assess the efficacy of tofacitinib 5 mg BID monotherapy as compared to tofacitinib 5 mg BID with background MTX in subjects who have received prior treatment of tofacitinib in combination with MTX.*

Secondary Objective

- *To assess the safety and tolerability of tofacitinib 5 mg BID monotherapy as compared to tofacitinib 5 mg BID with background MTX in subjects who have received prior treatment of tofacitinib in combination with MTX.*

2. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

No formal interim analysis is planned for the sub-study. However, data-cuts may be performed as needed to support tofacitinib regulatory submission.

This sub-study will use the same External Data Monitoring Committee (E-DMC) as for the main LTE study.

A final analysis will occur when all subjects finish the sub-study or terminate early.

3. HYPOTHESES AND DECISION RULES

3.1. Statistical Hypotheses

There are no formal hypotheses being tested in this sub-study.

3.2. Statistical Decision Rules

There are no statistical decision rules in this sub-study.

4. ANALYSIS SETS

Below is a description of the Analysis Sets defined for this sub-study.

4.1. Full Analysis Set

The full analysis set (FAS) is defined as all subjects who are randomized to the sub-study and receive at least one dose of the sub-study drug (tofacitinib, MTX or placebo). Subjects will be analyzed based on treatment randomized.

4.2. ‘Per Protocol’ Analysis Set

There is no Per Protocol Analysis Set defined for this sub-study.

4.3. Safety Analysis Set

The safety analysis set (SAS) is defined as all subjects who received at least one dose of the sub-study drug (tofacitinib, MTX or placebo). Subjects will be analyzed based on treatment received.

4.4. Other Analysis Sets

4.4.1. Endpoint Specific Analysis Sets

Subjects will be excluded from FAS for analyzing a specific endpoint if the criterion for inclusion is not met for the endpoint as described in **Table 3 Endpoint Specific Analysis Sets**.

Table 3 Endpoint Specific Analysis Sets

Instrument/Endpoint	Inclusion	Rationale
Δ PGA-PsO	Include subjects with baseline PGA-PsO>0	PGA-PsO=0 means no PsO disease at baseline as measured by PGA.
Δ BSA	Include subjects with baseline BSA>0	BSA=0 means no PsO disease at baseline as measured by BSA.
Δ DSS	Include subjects with baseline dactylitis severity score (DSS)>0	Baseline DSS=0 means absence of dactylitis. Dactylitis is not an enrollment criterion so it is expected that only a subset of the subjects will have dactylitis at baseline.
Dactylitis presence	Include subjects with baseline DSS>0	Baseline DSS=0 means absence of dactylitis. Dactylitis is not an enrollment criterion so it is expected that only a subset of the subjects will have dactylitis at baseline.
Δ Leeds Enthesitis Index (Δ LEI)	Include subjects with baseline LEI>0	Baseline LEI score=0 means absence of enthesitis as measured by LEI. Enthesitis is not an enrollment criterion so it is expected that only a subset of the subjects will have enthesitis at baseline.
Enthesitis presence	Include subjects with baseline LEI>0	Baseline LEI score=0 means absence of enthesitis as measured by LEI. Enthesitis is not an enrollment criterion so it is expected that only a subset of the subjects will have enthesitis at baseline.

Abbreviations: Δ =Change from baseline; baseline is the value obtained prior to the first dose of the sub-study; BSA = body surface area; DSS = Dactylitis Severity Score; PGA-PsO = Physician’s Global Assessment of Psoriasis.

4.5. Treatment Misallocations

If a subject was:

- Randomized but not treated, then the subject will appear on the subject evaluation table as randomized but not treated; this is the extent of how much data for this subject will be reported.
- Treated but not randomized, then by definition the subject will be excluded from the FAS and SAFETY analysis sets. A narrative will be provided for this subject in the clinical study report (CSR).

- Randomized but took the incorrect treatment exclusively through the entire duration of the study, then the subject will be reported under the randomized treatment group for the efficacy analysis but will be reported under the treatment they actually received for all safety analyses.
- Randomized but took at least one dose of randomized treatment and one dose of incorrect treatment, then the subject will be reported under the randomized treatment group for the efficacy and safety analyses.

Patient-reported outcomes (PRO) will be considered the same as for efficacy in this regard.

5. ENDPOINTS AND COVARIATES

Visual analog scale (VAS) data will be similarly rescaled prior to any calculation and analysis as in the main LTE study. Refer to main LTE study [Section 6](#) for the details. *The baseline will be the latest value obtained prior to the first dose of the sub-study.*

5.1. Primary Endpoints

- Δ HAQ-DI at 6 months.
- Δ PASDAS at 6 months.

5.2. Secondary Endpoints

- Δ HAQ-DI score at all time points except Month 6 (see [Appendix 2.2](#)).
- Δ PASDAS at all time points except Month 6 (see [Appendix 2.9](#)).
- PsARC response at all time points (see [Appendix 2.4](#)).
- Δ PGA-PsO at all time points (see [Appendix 2.11](#)).
- % Δ BSA at all time points (see [Appendix 2.10](#)).
- Δ DSS at all time points.
- Dactylitis absence at all time points.
- Δ in Enthesitis score (using LEI) at all time points.
- Enthesitis absence (using LEI) at all time points.
- MDA response at all time points (see [Appendix 2.8](#)).
- Δ in ACR components (tender/painful joint count (68), swollen joint count (66), Physician's Global Assessment of Arthritis, Patient's Global Assessment of Arthritis, Patient's Assessment of Arthritis Pain, CRP) at all time points (see [Appendix 2.1](#)).

- *Physical function/other patient reported outcomes to be assessed at all time points: Δ Short-Form 36 (version 2, Acute) (see [Appendix 2.17](#)); Δ FACIT-F (see [Appendix 2.16](#)); Δ EQ-5D-3L (see [Appendix 2.15](#)); EQ-VAS.*
- *Incidence and severity of adverse events.*
- *Incidence of clinical abnormalities and change from baseline (see [Appendix 6](#) Section 5 for baseline definition) in clinical laboratory values during treatment.*

5.3. Other Endpoints

- *Δ DAS28-3 (CRP) at all time points (see [Appendix 2.3](#)).*
- *MDA component response (yes/no) at all time points.*
- *Δ PGJS-VAS at all time points.*
- *Incidence of investigator-reported clinically significant changes in physical examination from baseline during treatment.*
- *Incidence of electrocardiogram (ECG) abnormalities and change from baseline in ECG measurements during treatment.*
- *Incidence of vital sign (blood pressure and pulse rate) abnormalities and changes from baseline in vital sign measurements during treatment.*

5.4. Covariates

Mixed Model for Repeated Measures (MMRM) will be used for analyzing continuous efficacy endpoints collected over multiple visits (eg, Δ HAQ-DI or Δ PASDAS). The baseline value will be included in the model as a covariate.

6. HANDLING OF MISSING VALUES

For the SF-36v2, EQ-5D-3L and FACIT-F instruments, the same approaches as for the main LTE study will be followed. Refer to main LTE study [Section 7](#) for the details. In general, missing values will not be imputed. For the binary efficacy endpoints (eg, PsARC Response), missing responses will be treated as non-response (MR=NR).

7. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

Both the efficacy and safety analysis will base on data collected in the sub-study only unless noted otherwise.

There are two treatment groups in the sub-study: Tofacitinib 5 mg BID + MTX placebo and Tofacitinib 5 mg BID + MTX. Subjects will be analyzed and reported by treatment groups.

Efficacy will also be assessed for a subset of the subjects who had more severe PsA as defined by ≥ 5 swollen or tender joints at the baseline of the sub-study, if sample size permits.

7.1. Statistical Methods

7.1.1. Analyses for Continuous Data

For the continuous endpoints, by-visit summary statistics (N, mean, standard deviation, standard error of the mean, minimum, 1st, 2nd (median) and 3rd quartiles and maximum) will be provided by treatment group for both the original and the change (or for some measures, percent change, eg, %ΔBSA) from baseline values over 12 months.

In addition, *continuous efficacy endpoints collected over multiple visits (eg, ΔHAQ-DI or ΔPASDAS) will be analyzed using a Mixed Model for Repeated Measures (MMRM) without imputation for missing values on the FAS.* The MMRM model will include fixed effects of treatment (Tofacitinib + Placebo and Tofacitinib + MTX), visit, treatment by visit interaction, and baseline value. The model will fit a common unstructured variance-covariance matrix for both treatment groups. If the unstructured matrix fails to converge, heterogeneous compound symmetry (CSH) will be attempted. *Treatment comparison between the 2 groups at each post-baseline time point will be generated from this model providing both 2-sided p-value and 95% CI.*

7.1.2. Analyses for Categorical Data

The data for all categorical endpoints (including binary) will be summarized by contingency tables that show the counts/frequency of the response at each visit by treatment group. For some categorical endpoints, such as the presence or severity of adverse events, contingency or frequency tables will consider the data over the entire duration of the study and will not be constructed at each time point.

Two-sided p-value and 95% CI will be generated by normal approximation to the binomial distribution approach for the treatment comparison between the 2 groups. Let N be number of subjects assessed at each visit, n be number of responders, and $\hat{p} = n/N$ be the observed responder rate. Use subscript t to refer to the treatment group (Tofacitinib + Placebo) and c to refer to the treatment group (Tofacitinib + MTX). Then the normal approximation to the 2-sided p-value will be calculated as

$$2Pr\left\{Z \geq \left| \frac{\hat{p}_t - \hat{p}_c}{\sqrt{\frac{\hat{p}_t(1-\hat{p}_t)}{N_t} + \frac{\hat{p}_c(1-\hat{p}_c)}{N_c}}} \right| \right\},$$

where Z is Normal (0, 1) distributed.

Two-sided 95% CI will be formed by

$$(\hat{p}_t - \hat{p}_c) \pm 1.96 \sqrt{\frac{\hat{p}_t(1-\hat{p}_t)}{N_t} + \frac{\hat{p}_c(1-\hat{p}_c)}{N_c}}.$$

If there is no (ie, 0) response or 100% response in any of treatments, 0.5 will be added to the number of responses and 1 will be added to the denominator in each treatment corresponding to the pair of comparison for calculating the treatment difference, standard error, 95% CI and 2-sided p-value (Agresti, 2002).

When response rate of 0% or 100% is observed in both treatments in comparison, no formal comparison will be performed. Estimated response rate of 0% or 100% will be reported as observed.

The final results will be expressed in percentages, ie, (proportions x 100)%.

7.2. Statistical Analyses

The primary efficacy analysis will use FAS (Appendix 6 [Section 4.1](#)). Continuous efficacy endpoints, for which change or percent change from baseline is the measure to be analyzed, would require that a subject have both a baseline value and at least one post-baseline value to be included in the FAS. The safety analysis will use SAS (Appendix 6 [Section 4.3](#)).

7.2.1. Analysis of Primary Endpoints

For each of the two primary endpoints in the sub-study: Δ HAQ-DI at Month 6 and Δ PASDAS at Month 6, a MMRM described in Appendix 6 [Section 7.1.1](#) on the FAS without imputation for missing values will be used to estimate the change from baseline at Month 6 for each treatment group. The treatment comparison between the 2 groups at Month 6 will also be generated from this model providing both 2-sided p-value and 95% CI. The analysis will include all the data through Month 12. Treatment comparisons between the 2 treatment groups at other visits will be made as well from this model.

7.2.2. Analysis of Secondary and Other Efficacy/Health Outcome Endpoints

The MMRM described in Appendix 6 [Section 7.1.1](#) will be used on FAS or endpoint specific FAS for:

- *Δ HAQ-DI score at all time points except Month 6* (see [Appendix 2.2](#)). Note that this will be accomplished by the same MMRM analysis in Appendix 6 [Section 7.2.1](#).
- *Δ PASDAS at all time points except Month 6*. Note that this will be accomplished by the same MMRM analysis in Appendix 6 [Section 7.2.1](#).
- *Δ PGA-PsO at all time points* (see [Appendix 2.11](#)).
- *% Δ BSA at all time points* (see [Appendix 2.10](#)).
- *Δ DSS at all time points*.
- *Δ in Enthesitis score (using LEI) at all time points*.
- *Δ in ACR components (tender/painful joint count (68), swollen joint count (66), Physician's Global Assessment of Arthritis, Patient's Global Assessment of Arthritis, Patient's Assessment of Arthritis Pain, CRP) at all time points* (see [Appendix 2.1](#)).
- *Physical function/other patient reported outcomes to be assessed at all time points: Δ Short-Form 36 (version 2, Acute) (see [Appendix 2.17](#)); Δ FACIT-F (see [Appendix 2.16](#)); Δ EQ5D (see [Appendix 2.15](#)); EQ-VAS.*

- Δ DAS28-3 (CRP) at all time points (see [Appendix 2.3](#)).
- Δ PGJS-VAS at all time points.

The analysis methods described in Appendix 6 [Section 7.1.2](#) will be used on the FAS or endpoint specific FAS for:

- PsARC response at all time points (see [Appendix 2.4](#)).
- Dactylitis absence at all time points.
- Enthesitis absence (using LEI) at all time points.
- MDA response at all time points.
- MDA component (yes/no) response at all time points.

7.2.3. Analysis of Safety Data

- Safety analysis by treatment groups will follow the similar approaches as for the main LTE study, except the incidence rate analysis will not be performed in the sub-study. Refer to main LTE study [Section 8.2.1](#) Analysis of Primary Endpoints and [Section 8.2.3](#) Analysis of Safety Data for the details.

7.3. Baseline and Other Summaries and Analyses

7.3.1. Baseline Summaries

Baseline will be from the combined A3921092 Sub-study Switch/Baseline visit unless noted otherwise. Baseline characteristics will include but may not be limited to the ones listed below and will be summarized similarly as in the main LTE study. Refer to main LTE study [Section 8.3.1](#) for more details.

- Demographic characteristics include but may not be limited to: Age, Sex, Race (from main LTE baseline), Ethnicity (from main LTE baseline), Weight, Height, BMI, Geographic Region (from main LTE baseline), Smoking Status, Alcohol Use.
- Baseline disease characteristics include but may not be limited to: PsA Duration, PsA Subtype, PASDAS, Swollen Joint Count, Tender/Painful Joint Count, HAQ-DI, Screening Presence of Distal Interphalangeal Joints Involvement, Screening Presence of Arthritis Mutilan, Presence of Enthesitis (LEI), Enthesitis Index (LEI), Presence of dactylitis, DSS, BSA, PGA-PsO, SF-36v2 (8 domains and PCS and MCS), FACIT-F (total score and 2 domains), CRP, Diabetes Mellitus, Metabolic Syndrome, CV Risks per PDS, Rheumatoid Factor Positive, Cyclic Citrullinated Peptide Antibody Positive.

7.3.2. Prior Drug Treatments for Psoriatic Arthritis

Prior drug treatment will include any treatment except tofacitinib used during the qualifying study, during the enrollment gap between the qualifying study and the main LTE study, and the main LTE study for PsA prior to the first dose of study medication in the sub-study. Refer to main LTE study [Section 8.3.2](#) for detailed category list of the prior drug treatments.

7.3.3. Concomitant Drug Treatments for Psoriatic Arthritis

Concomitant drug treatment will include any treatment for PsA during the sub-study. The summary by treatment groups will be similarly as for the main LTE study. Refer to main LTE study [Section 8.3.3](#) for the details.

MTX dose will be summarized by visit.

7.4. Sub-study Definition and Use of Visit Windows in Reporting

The visit windows are similarly defined as for the main LTE study in below table. The study day will be calculated as: date of assessment/collection – date of the first dose in the sub-study + 1.

If two or more visits/observations fall into the same window, the visit/observation closest to the target day should be used in the analyses. If there is a tie, the later visit should be used. However, if two or more visits/observations fall in the same window, and the latest visit is the last visit of the study due to early termination, then this will be the visit used in the analysis for that visit window. Per the master visit schedule provided to the investigator sites, a month is defined as 4 weeks or 28 days.

Visit Label	Target Day	Definition [Day Window] – Lower	Definition [Day Window] - Upper
Switch/Baseline	1	Day 1 = date of first dose of study treatment taken of sub-study Prior to first dose of study treatment taken of sub-study	
Month 1	29	2	57
Month 3	85	58	127
Month 6	169	128	211
Month 9	253	212	295
Month 12	337	≥296	

The following categories will be used in the treatment exposure and duration of treatment summaries for tofacitinib. The duration of treatment will be calculated as: date of the last dose in the sub-study – date of the first dose in the sub-study + 1. Any missed doses/drug holidays will be ignored in the duration calculation. Treatment exposure will be defined as the number of days that subjects took the study drug (gaps not counted).

Category (days)
1
2-28 (1 month)
29-84 (1-3 months)
85-168 (3-6 months)
169-252 (6-9 months)
≥253 (9-12 months)