Protocol B7931030

A PHASE 2B, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY OF PF-06700841 TO EVALUATE THE EFFICACY AT 16 WEEKS AND TO EVALUATE THE SAFETY AND EFFICACY UP TO 1 YEAR IN SUBJECTS WITH ACTIVE PSORIATIC ARTHRITIS

Statistical Analysis Plan
(SAP)

Version: **3**

Date: **2 Feb 2021**

[redacted]

[redacted]

TABLE OF CONTENTS

LIST OF TABLES 5
LIST OF FIGURES 5
APPENDICES 5
1. VERSION HISTORY 6
2. INTRODUCTION 7
2.1. Study Objectives, Endpoints, and Estimands 7
2.1.1. Primary Estimand 10
2.1.2. Secondary Estimand(s) 10
2.1.2.1. ACR50/90 response at Week 16 10
2.1.2.2. PASI75/90 response at Week 16 10
2.1.2.3. Change from baseline in HAQ-DI at Week 16 10
2.2. Study Design 10
3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS 12
3.1. Primary Endpoint 12
3.2. Secondary Endpoints 12
3.2.1. Secondary Endpoints – PsA Core Domain Set 12
3.2.2. Secondary Endpoints - Patient Reported Outcomes 13
3.2.3. Secondary Endpoints – Composite Outcomes 13
CCI
3.4. Baseline Variables 14
3.5. Safety Endpoints 14
3.5.1. Adverse Events 14
3.5.2. Laboratory Data 15
3.5.3. Vital Signs 15
3.5.4. Electrocardiograms 15
4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS) 15
5. GENERAL METHODOLOGY AND CONVENTIONS 16
5.1. Hypotheses and Decision Rules 16
5.2. General Methods 16
5.2.1. Analyses for Binary Endpoints 17

5.2.1.1. Normal Approximation 17
5.2.1.2. Cochran-Mantel-Haenszel Method 18
5.2.1.3. Logistic Regression Analysis 18
5.2.1.4. Longitudinal Analysis for Binary Endpoint 18
5.2.2. Analyses for Continuous Endpoints 19
5.2.2.1. Longitudinal Model for Repeated Measures 19
5.2.2.2. Jump to Reference Imputation Model 19
5.3. Methods to Manage Missing Data 20
5.3.1. Binary Endpoints 20
5.3.2. Continuous endpoints 20
CCI
5.3.4. Missing Dates 21
6. ANALYSES AND SUMMARIES 22
6.1. Primary Endpoint 22
6.1.1. The proportion of subjects achieving ACR20 response at Week 16 22
6.1.1.1. Main Analysis 22
6.1.1.2. Sensitivity/Supplementary Analyses 23
6.2. Secondary Endpoint(s) 23
6.2.1. The proportion of subjects achieving ACR20 response at Week 16 in the subgroup of subjects who are TNFα inhibitor naïve. 23
6.2.1.1. Main Analysis 23
6.2.1.2. Sensitivity/Supplementary Analyses 24
6.2.2. The proportion of subjects achieving an ACR20 response at all treatment timepoints (other than Week 16) 24
6.2.3. The proportion of subjects achieving ACR50 and ACR70 response at all treatment timepoints 25
6.2.4. Change from baseline in the ACR response criteria components 25
6.2.5. The proportion of subjects achieving PASI75/90/100 response 26
6.2.6. Change from baseline in the enthesitis score 26
6.2.7. Change from baseline in the Dactylitis Severity Score 27
6.2.8. Change from baseline in the Nail Psoriasis Severity Index 27
6.2.9. Change from baseline in the Patient's Global Joint and Skin Assessment – Visual Analog Scale (PGJS-VAS) 27

6.2.10. Change from baseline in the Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-Fatigue) 27
6.2.11. Change from baseline in the Short-Form-36 Health Survey (SF-36) Version 2, Acute 27
6.2.12. The proportion of subjects achieving Minimal Disease Activity (MDA) and Very Low Disease (VLDA) response 27
6.2.13. Change from baseline in the Disease Activity Index for Reactive Arthritis/PsA (DAREA/DAPSA)....27
6.2.14. The proportion of subjects achieving the Psoriatic Arthritis Response Criteria (PsARC) 28
6.2.15. Change from baseline in the Psoriatic Arthritis Disease Activity Score (PASDAS) 28
CCI
6.4. Subset Analyses....30
6.5. Baseline and Other Summaries and Analyses....30

6.5.1. Baseline Summaries....30
6.5.2. Study Conduct and Participant Disposition....32
6.5.3. Study Treatment Exposure....32
6.5.4. Concomitant Medications and Nondrug Treatments....32
6.6. Safety Summaries and Analyses....32
6.6.1. Adverse Events....32
6.6.2. Laboratory Data....33
6.6.3. Vital Signs....33
6.6.4. Electrocardiograms....33
7. INTERIM ANALYSES....33
7.1. Introduction....33
7.2. Interim Analyses and Summaries....34
8. REFERENCES....34
9. APPENDICES....36

LIST OF TABLES

Table 1. Summary of Changes....6
Table 2. Treatment Sequence....11
Table 3. Safety QTc Assessment....33

LIST OF FIGURES

Figure 1. Schematic of Study Design....11

APPENDICES

Appendix 1.1. Endpoint Derivations....36

1. VERSION HISTORY

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
v3 / 2 Feb 2021	Protocol Administrative Change Letter (PACL) 26 Oct 2020 Amendment 2 22 Oct 2019	Revisions and clarifications made in statistical analysis post BDR	• Section 3.4 – updated that the stratification factor (prior TNFi exposure) will be based on randomization file (not CRF). • Sections 6.2.4, 6.3.13 and 6.3.14 – removed summaries of absolute values • Section 6.4 – removed all subgroup analyses except ACR20 for Prior TNFi exposure (yes, no) – no longer needed. • Section 6.5.1 - added Concomitant Methotrexate Dose at Baseline (mg/wk), Baseline Rheumatoid Factor Positive, Baseline Cyclic Citrullinated Peptide Antibody Positive, Day 1 usage of DMARDS, MTX and NSAIDS to the summary of baseline disease characteristics; removed Baseline PGJA-VAS (incorrect endpoint but no longer needed). • Section 6.5.2 – removed entire section and moved Prior TNFi exposure to Baseline Disease Characteristics table • Sections 3.5.4 & 6.6.4 – changed how ECG data will be presented in the CSR (due to the change in ECG machines during the study). • Section 6.6.5 – removed listing of physical examination • Other minor changes.
v2 / 24 Apr 2020	Protocol Administrative Change Letter (PACL) 17 Apr 2020 Amendment 2 22 Oct 2019	Revisions and clarifications made in statistical analysis	• Section 2.1.1, 2.1.2.3, 6.1.1.1 and 6.2.1.1. Rescue medication taking is removed from the intercurrent event list due to its highly overlapping with the other intercurrent event of premature treatment discontinuation. • Section 5.3.1 and 5.3.2. Details regarding missing data due to COVID-19 are added.

			• Section 6.2.4. Definition of HAQ-DI responder is corrected to 'change from baseline in HAQ-DI<-0.3'. [redacted] • With a small overall proprotion of study participants having prior exposure to TNFi, this covariate was removed from model-based analysis to avoid convergence issues. • Other minor changes.
v1/ 10 Dec 2019	Amendment 2 22 Oct 2019	N/A	N/A

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study B7931030. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives, Endpoints, and Estimands

Primary Objective(s):	**Primary Endpoint(s):**
• To evaluate the efficacy of PF-06700841 compared to placebo in subjects with active psoriatic arthritis (PsA).	• The proportion of subjects achieving an American College of Rheumatology 20 (ACR20) response at Week 16.
Secondary Objective(s):	**Secondary Endpoint(s):**
• To evaluate the efficacy of PF-06700841 compared to placebo in subjects with active PsA who are TNFα inhibitor naïve.	• The proportion of subjects achieving an ACR20 response at Week 16 in the subgroup of subjects who are TNFα inhibitor naïve.
• To evaluate the improvement in signs and symptoms related PsA Core Domain Set in PF-06700841 treated subjects.	Assessed at all treatment timepoints: • The proportion of subjects achieving an ACR20 response at all treatment timepoints (other than Week 16), and proportion of subjects achieving an ACR50 and ACR70 response;

	• Change from baseline in the ACR response criteria components (Tender/painful joint count, Swollen joint count, Patient's Assessment of Arthritis Pain, Patient's Global Assessment of Arthritis, Physician's Global Assessment of Arthritis, Health Assessment Questionnaire [HAQ] disability index [DI], and hsCRP); • The proportion of subjects achieving a Psoriasis Area and Severity Index 75/90/100 (PASI75/90/100) response; • Change from baseline in the enthesitis score (using the Spondyloarthritis Research Consortium of Canada [SPARCC] Enthesitis Index and Leeds Enthesitis Index); • Change from baseline in the Dactylitis Severity Score; • Change from baseline in the Nail Psoriasis Severity Index (NAPSI) Score.
• To evaluate the improvement in patient reported outcome measures related PsA Core Domain Set in PF-06700841 treated subjects.	Assessed at all treatment timepoints: • Change from baseline in the Patient's Global Joint and Skin Assessment Visual Analog Scale (PGJSVAS); • Change from baseline in the Functional Assessment of Chronic Illness Therapy Fatigue (FACITFatigue); Assessed at all treatment timepoints except Week 2: • Change from baseline in the ShortForm36 Health Survey (SF36) Version 2, Acute.
• To evaluate the improvement in additional composite outcome measures in PF-06700841 treated subjects.	Assessed at all treatment timepoints except Week 2: • The proportion of subjects achieving Minimal Disease Activity (MDA) and Very Low Disease Activity (VLDA) response; • Change from baseline in the Disease Activity Index for Reactive Arthritis/PsA (DAREA/DAPSA); • The proportion of subjects achieving the Psoriatic Arthritis Response Criteria (PsARC); • Change from baseline in the Psoriatic Arthritis Disease Activity Score (PASDAS).
• To evaluate the safety and tolerability of PF-06700841.	• Incidence of adverse events (AEs), serious adverse events (SAEs) and serious infectious events (SIEs), withdrawals due to AEs and SAEs, and laboratory

	abnormalities, changes in vital signs, and electrocardiogram (ECG) findings throughout the study.
CCI	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
CCI	[redacted]
[redacted]	[redacted]

2.1.1. Primary Estimand

The primary estimand of this study is defined according to the primary objective and in alignment with the primary endpoint. It includes the following 4 attributes:

- Population: Subjects who have active psoriatic arthritis (PsA) and who are randomized;
- Variable: binary endpoint of American College of Rheumatology 20 (ACR20) response at Week 16.
- Intercurrent event (IE): premature treatment discontinuation.
- Population-level summary: Treatment differences in proportion of subjects who achieving ACR20 response at Week 16 between each dose group of PF-06700841 and placebo.

2.1.2. Secondary Estimand(s)

2.1.2.1. ACR50/90 response at Week 16

The estimands for ACR50/90 at Week 16 are defined in the same manner as the primary estimand for the ACR20 at Week 16.

2.1.2.2. PASI75/90 response at Week 16

The estimands for PASI75/90/100 at Week 16 are defined in the same manner as the primary estimand for the ACR20 at Week 16 with the exception that the population will consist of a subgroup of patients who have baseline BSA≥3% and PASI>0.

2.1.2.3. Change from baseline in HAQ-DI at Week 16

The estimand for this endpoint is defined as a hypothetical estimand as follows.

- Population: Subjects who have active psoriatic arthritis (PsA) and who are randomized;
- Variable: Change from baseline in HAQ-DI at Week 16.
- Intercurrent event (IE): premature treatment discontinuation.
- Population-level summary: Treatment differences in change from baseline in HAQ-DI at Week 16 between each dose group of PF-06700841 and placebo.

2.2. Study Design

This is a Phase 2b, randomized, double-blind, placebo-controlled, dose-ranging, parallel treatment group, efficacy and safety study designed to characterize the dose response of PF-06700841 in subjects with active PsA. A total of approximately 196 subjects will be

randomized in a 4:4:1:1:2:2 ratio to one of the six parallel treatment sequences as shown in Table 1.

Table 2. Treatment Sequence

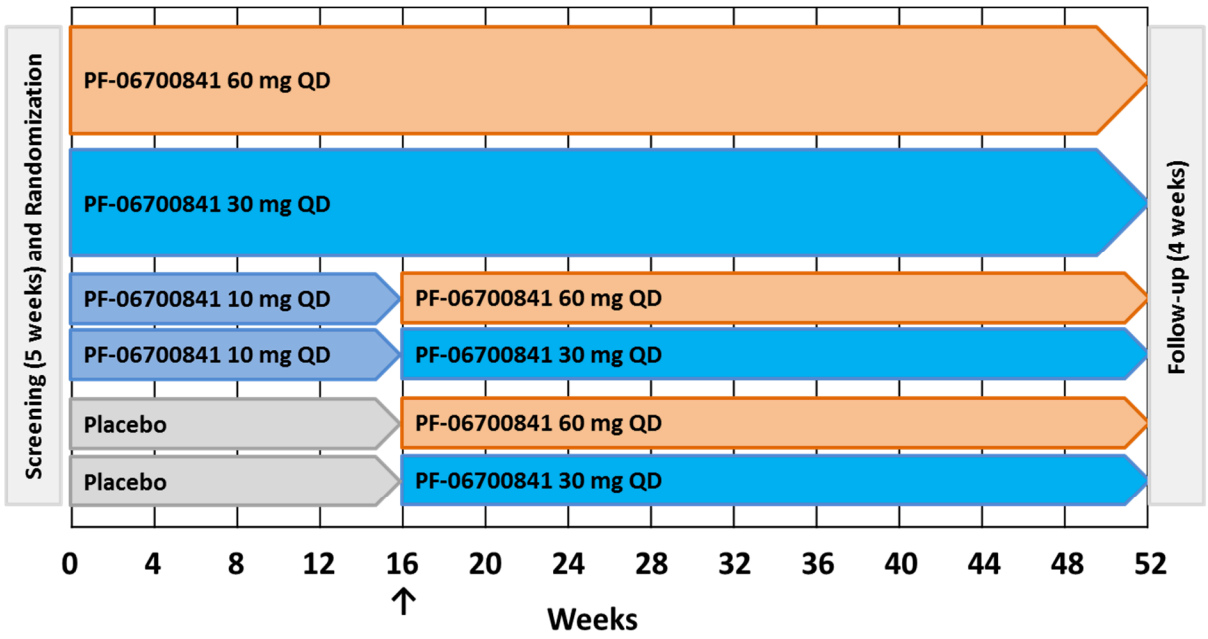
Treatment Sequence	Treatment	Subject Size
A	PF-06700841 60 mg QD	56
B	PF-06700841 30 mg QD	56
C[1]	PF-06700841 10 mg QD first & advance to 60 mg QD after the Week 16 visit	14
D[2]	PF-06700841 10 mg QD first & advance to 30 mg QD after the Week 16 visit	14
E[1]	Placebo QD first & advance to PF-06700841 60 mg QD after the Week 16 visit	28
F[2]	Placebo QD first & advance to PF-06700841 30 mg QD after the Week 16 visit	28

1. Starting after the Week 16 visit, subjects randomized to treatment sequences C or E will receive PF-06700841 60 mg QD in a blinded fashion for the remainder of the study.
2. Starting after the Week 16 visit, subjects randomized to treatment sequences D or F will receive PF-06700841 30 mg QD in a blinded fashion for the remainder of the study.

The primary efficacy endpoint is the ACR20 response rate at Week 16. For the treatment comparison of each of three active treatment regimens (60 mg QD, 30 mg QD and 10 mg QD) versus placebo in the primary endpoint, this results in a 2:2:1:2 allocation ratio (in the order of 60 mg QD, 30 mg QD, 10 mg QD and placebo).

A schematic of the study design is shown below:

Figure 1. Schematic of Study Design



↑ Primary study endpoints of ACR20 will be obtained at Week 16 visit. All subjects randomized to placebo will receive PF-06700841 in a blinded manner after Week 16 visit.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

Study endpoints are listed in this section. Details regarding the endpoint derivations are included in the Appendix 1.1. The baseline values of variables used to derive these endpoints will be measurements at the Baseline (Day 1) Visit. If the measurement at the Baseline Visit is not available but the variable is measured in the screening visit, then the measurement at the screening visit will be used as the baseline value.

Visual analog scale (VAS) data will need to be rescaled prior to any calculation and analysis. VAS is recorded 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 × 100mm or X/Y × 10cm depending on endpoints or the use in defining other endpoints (see Appendix 1.1 for details).

3.1. Primary Endpoint

- The proportion of subjects achieving an American College of Rheumatology 20 (ACR20) response at Week 16.

3.2. Secondary Endpoints

3.2.1. Secondary Endpoints – PsA Core Domain Set

- The proportion of subjects achieving an ACR20 response at Week 16 in the subgroup of subjects who are TNFα inhibitor naïve.
- The proportion of subjects achieving an ACR20 response at all treatment timepoints (other than Week 16)
- The proportion of subjects achieving an ACR50 and ACR70 response.
- Change from baseline in the ACR response criteria components, including
  - Tender/painful joint count,
  - Swollen joint count,
  - Patient's Assessment of Arthritis Pain (VAS),
  - Patient's Global Assessment of Arthritis (VAS),
  - Physician's Global Assessment of Arthritis (VAS),
  - Health Assessment Questionnaire (HAQ) disability index (DI) [HAQ-DI], and
  - hsCRP.
- The proportion of subjects achieving a Psoriasis Area and Severity Index 75/90/100 (PASI75/90/100) response.
- Change from baseline in the enthesitis score, using
  - Spondyloarthritis Research Consortium of Canada [SPARCC] Enthesitis Index, and
  - Leeds Enthesitis Index.
- Change from baseline in the Dactylitis Severity Score.
- Change from baseline in the Nail Psoriasis Severity Index (NAPSI) Score.

3.2.2. Secondary Endpoints - Patient Reported Outcomes

- Change from baseline in the Patient's Global Joint and Skin Assessment-Visual Analog Scale (PGJS-VAS) at all treatment timepoints;
- Change from baseline in the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) at all treatment timepoints;
- Change from baseline in the Short-Form-36 Health Survey (SF-36) Version 2, Acute, at all treatment timepoints except Week 2.

3.2.3. Secondary Endpoints – Composite Outcomes

- The proportion of subjects achieving Minimal Disease Activity (MDA) and Very Low Disease Activity (VLDA) response at all treatment timepoints except Week 2;
- Change from baseline in the Disease Activity Index for Reactive Arthritis/PsA (DAREA/DAPSA) at all treatment timepoints except Week 2;
- The proportion of subjects achieving the Psoriatic Arthritis Response Criteria (PsARC) at all treatment timepoints except Week 2;
- Change from baseline in the Psoriatic Arthritis Disease Activity Score (PASDAS) at all treatment timepoints except Week 2.

CCI

[redacted]

[redacted]

CCI [redacted]

3.4. Baseline Variables

The study procedure and clinical assessments are described in the protocol (Section 6.2.1). All patient reported outcomes (PROs) should be completed prior to any other study procedures. Some clinical assessments will be assessed in both Screening and Baseline visits for the determination of the eligibility for enrolment, but the measurements from the Baseline (Day 1) visit will be used in the summaries of baseline variables unless the Day 1 data are missing. Data from the screening period may be used if Day 1 data are missing. If multiple data points are available, we will use the last observation before Day 1 dosing except the ECG. The stratification factor (prior TNFi exposure) will be based on the randomization file.

3.5. Safety Endpoints

- Incidence of adverse events (AEs)
- Serious adverse events (SAEs)
- Serious infectious events (SIEs)
- Withdrawals due to AEs and SAEs
- Laboratory abnormalities
- Changes from baseline in vital signs
- Electrocardiogram (ECG) measurements

ECGs will be read and interpreted locally and centrally. Only locally read ECG data will be reported in the CSR. According to study protocol, the ECG at screening will be repeated two more times if QTcF exceeds 450 ms in the first reading and the average of the three QTcFs will be used to determine the subject eligibility. Triplicate ECGs will be obtained on Baseline (Day 1). In these cases, the average of the three QTcF values will be used in the summaries.

3.5.1. Adverse Events

In additional to standard safety displays, a 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers. Different analyses will be performed for different tiers (see Section 6.6.1).

- Tier 1 events: These are prespecified events of clinical importance and are maintained in a list in the product's Safety Review Plan.
- Tier 2 events: These are events that are not tier 1 but are "common." A Medical Dictionary for Regulatory Activities (MedDRA) preferred term (PT) is defined as a Tier 2 event if there are at least 3 subjects in any treatment group reported the event.
- Tier 3 events: These are events that are neither Tier 1 nor Tier 2 events.

The 3 tiers are mutually exclusive. Tier 3 events will be included in standard safety displays and not separately displayed in specified Tier 3 tables.

3.5.2. Laboratory Data

The laboratory tests will be performed at time points identified in the Schedule of Activities. Unscheduled clinical laboratory measurements may be obtained at any time to assess any perceived safety concerns. Data from central laboratories will be used for statistical analyses.

3.5.3. Vital Signs

Vital signs (blood pressure, pulse rate, and temperature) will be measured at times specified in the Schedule of Activities.

3.5.4. Electrocardiograms

Standard 12-lead ECGs should be collected at times specified in the Schedule of Activities.

The baseline ECG values (the last measurement prior to receive study treatment on Day 1) will serve as each subject's baseline values. If Day 1 ECG is missing, then screening ECG will serve as baseline value. Data from local ECG readings will be summarized in statistical analyses.

Baseline and change from baseline in ECG will be summarized descriptively. Categorical summaries of the ECG data will also be provided.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

Below is a description of the Analysis Sets (i.e. Population for Analysis) defined for the study. Data for all participants will be assessed to determine if participants meet the criteria for inclusion in each analysis population prior to unblinding and releasing the database and classifications will be documented per standard operating procedures.

Population	Description
Evaluable Population	All participants who were randomized to the study and received at least one dose of the randomized study treatment. The Evaluable Population is the primary efficacy analysis population. Participants will be analyzed according to the randomized treatment group.
Safety Analysis Population	All participants who received at least one dose of the randomized study treatment. Participants will be analyzed according to study treatment they actually received. A randomized but not treated participant will be excluded from the safety analyses.
CCI [redacted]	[redacted]

Population	Description
PD Analysis Population	All participants who received at least one dose of randomized study treatment and in whom at least one value of the PD parameter of interest is reported.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

The study is designed to evaluate the efficacy of PF-06700841 compared to placebo in subjects with active PsA. Each of the three dose groups of PF-06700841 will be formally tested in the primary endpoint (ACR20 at Week 16) and a selected list of secondary endpoints (PASI75, ACR50, PASI90, PASI100 and HAQ-DI at Week 16) for superiority against placebo.

For each dose group of PF-06700841 within each endpoint, the null hypothesis is that there is no difference between PF-06700841 and placebo, and the alternative hypothesis is that there is a difference between PF-06700841 and placebo.

In order to control for type-I error, the Dunnett's method will be used for the primary endpoint of ACR20 at Week 16 to adjust for multiplicity. The overall family-wise type-I error rate is set at 1-sided 5% (or equivalently 2-sided 10%). This test will serve as a gatekeeper for further formal testing of secondary endpoints. Specifically, if any PF-06700841 dose regimen is superior to placebo under the Dunnett's test, selected secondary endpoints at Week 16 for that dose regimen will be tested for superiority against placebo hierarchically in the order of PASI75 →ACR50 →PASI90 →PASI100 →HAQ-DI. Each endpoint will be tested at the significance level of 1-sided 5% (or equivalently 2-sided 10%).

5.2. General Methods

In general, the data for all continuous endpoints will be summarized by timepoint and treatment sequence group (or combined treatment sequence when appropriate) in tables containing descriptive statistics (N, mean, standard deviation, standard error of the mean, minimum, 1^{st}, 2^{nd} (median) and 3^{rd} quartiles and maximum) for baseline and change or percent change from baseline for those endpoints measured at baseline. The displays described above will only contain available data with no imputation. The data for all categorical endpoints will be summarized by contingency tables that show the counts/frequency in the various categories by treatment sequence group at each timepoint. For binary endpoint (e.g. ACR20), non-response imputation will be applied for the missing data in the summary tables.

For analyses through Week 16, the treatment sequence E and F will be combined into a single placebo group and the treatment sequence C and D will be combined into a single PF-06700841 10mg group. The following treatment comparisons with placebo will be made at each visit, where applicable,

- PF-06700841 60 mg QD (treatment sequence A) vs. placebo (combined treatment sequences E and F);
- PF-06700841 30 mg QD (treatment sequence B) vs. placebo (combined treatment sequences E and F);
- PF-06700841 10 mg QD (combined treatment sequences C and D) vs. placebo (combined treatment sequences E and F).

For analyses through the Follow-up Visit (Week 56), endpoints will be summarized by treatment sequence using descriptive statistics.

The primary efficacy analysis will be conducted on data collected during the first 16 weeks where comparisons to placebo will entail combining treatment sequences E and F to form one placebo group. The primary comparisons will be between each PF-06700841 dose group and placebo (combined) at Week 16. Formal hypothesis testings for treatment comparisons will be performed for ACR20, ACR50, PASI75, PASI90, PASI100 and HAQ-DI endpoints at Week 16 only in the sequence as described in the Section 5.1.

Treatment comparisons for these endpoints at post-baseline visits other than Week 16 or other endpoints at all post-baseline visits will be summarized using descriptive statistics only.

5.2.1. Analyses for Binary Endpoints

Data for all binary endpoints will be summarized by treatment group and visit in tabular and/or graphic format with descriptive statistics, including N (number of subjects evaluable for the endpoint in the corresponding treatment group/visit), n (number of responders), response rates (%), standard errors of the response rate, and 95% confidence intervals (CIs) based on normal approximation (Section 5.2.1.1). Treatment differences between each dose group of PF-06700841 and placebo will be summarized by point estimates and two-sided 90% CIs (Section 5.2.1.1).

5.2.1.1. Normal Approximation

The normal approximation for the 95% CIs of a binomial proportion will be calculated as

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

where $\hat{p}$ is the observed proportion for the binary endpoint and n is the sample size. If there is no (i.e. 0) or 100% response, 0.5 will be added to the number of responses and 1 will be added to the denominator for the calculation of standard errors and 95% CIs.

The normal approximation for the difference in binomial proportion will be used to test the superiority of each dose group of PF-06700841 against placebo in the binary endpoints, including ACR20/50 and PASI 75/90/100 under the testing procedure described in Section 5.1. Specifically, the test statistics for the difference in binomial random variables between each dose group of PF-06700841 and placebo are calculated as

$$Z_k = \frac{\hat{p}_k - \hat{p}_P}{\sqrt{\frac{\hat{p}_k(1-\hat{p}_k)}{n_k} + \frac{\hat{p}_P(1-\hat{p}_P)}{n_P}}}, k = 1,2,3,$$

where $\widehat{p_k}$ (k=1,2,3) and $\widehat{p_P}$ are observed proportion of responders in three PF-06700841 dose groups (60mg QD [k=1], 30mg QD [k=2] and 10mg QD [k=3]) and placebo group, respectively, and n_k (k=1,2,3) and n_P are sample sizes in the respective treatment groups. The normal approximation to each of the test statistics will be used for each dose group of PF-06700841 in comparison to the placebo group.

Treatment differences in binary response rates between each dose group of PF-06700841 and placebo will be summarized by point estimates and two-sided 90% CIs using normal approximation method. Two-sided 90% CIs will be calculated as

$$(\hat{p}_k - \hat{p}_P) \pm 1.645\sqrt{\frac{\hat{p}_k(1-\hat{p}_k)}{n_k} + \frac{\hat{p}_P(1-\hat{p}_P)}{n_P}}, k = 1,2,3$$

If there is no (i.e. 0) or 100% response in any of the treatment comparison with the placebo group, 0.5 will be added to the number of responses and 1 will be added to the denominator for each group within the treatment comparison in the calculation of treatment differences, standard errors, 90% CIs and 2-sided p-values (Agresti, 2013).

5.2.1.2. Cochran-Mantel-Haenszel Method

To adjust for the effect of prior TNFi exposure, treatment differences in the binary response rates can be summarized using the Cochran-Mantel-Haenszel method (Agresti, 2013). This analysis will be conducted as a sensitive analysis to the primary analysis using the normal approximation method without adjusting for the effect of prior TNFi exposure. This method will also be used to summarize treatment differences in other endpoints when appropriate.

5.2.1.3. Logistic Regression Analysis

A logistic regression model will be fit for the primary endpoint as a sensitivity analysis, with treatment as covariate. When appropriate, this analysis will be conducted for other binary endpoints.

5.2.1.4. Longitudinal Analysis for Binary Endpoint

As a supplementary analysis of longitudinal binary responses (ACR20, ACR50 and ACR70), scheduled to be collected multiple times during the first 16 weeks of the study, a generalized linear mixed model (GLMM) for repeated binary measures will be used. This model will have fixed effects for time (discrete), treatment, and time by treatment interaction; the dependent variable will be logit of the probability of "response". A common AR(1) variance-covariance matrix for all treatment groups will be used to model the variability among observations within a subject. If AR(1) matrix fails to converge, compound symmetry will be attempted. The parameters of the model will be estimated with pseudo-likelihood type methods using PROC GLIMMIX in SAS (SAS/STAT User's Guide). Simulations investigating generalized linear models (GLM) for correlated binary data fit with

PROC GLIMMIX indicate good performance as measured by type I error, CI coverage and bias even in the presence of data that is MAR (Liu and Zhan, 2011). From this model, one can obtain estimate of the probability of "response" for each treatment group at each timepoint (at Week 16 or prior) as well inferential comparisons (odds ratio, CI and p-value) between treatment groups.

If a response rate is 0 or 100% at a visit for a treatment group that may lead to convergence issue, data from that visit for all the treatment groups will be removed from the model fitting.

5.2.2. Analyses for Continuous Endpoints

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

5.2.2.1. Longitudinal Model for Repeated Measures

Data for longitudinally collected continuous endpoints at 2 or more post-baseline visits during the first 16 weeks may be analyzed as change or percent change from baseline with a marginal model for repeated measures (MMRM) that includes the fixed effect of treatment, visit (discrete variable, up to Week 16), treatment by visit interaction, and baseline value. An unstructured (UN) variance-covariance matrix will be fit whenever possible. If there is convergence issue with UN matrix, the model will be fit with heterogeneous compound symmetry (CSH) matrix. This model will generate maximum likelihood estimates under the assumption of MAR, but no imputations will be made for missing data. Subjects in treatment sequences C and D will be combined into a single group for PF-06700841 10 mg QD; subjects in treatment sequences E and F will be combined into a single group for placebo. Pairwise comparisons between each PF-06700841 dose group and placebo at each time points will be generated from this model with least square means for treatment differences, 90% CIs and 2-sided p-values.

5.2.2.2. Jump to Reference Imputation Model

As a supplementary analysis for MMRM, a multiple imputation (MI) approach will be used (Carpenter et al., 2013). The specific MI that will be implemented applies the "jump to reference" (JTR) imputation approach to the active dose groups but MAR imputation approach to the combined placebo group. The joint distribution of a subject's data across the timepoints up to and including Week 16 will be assumed to be multivariate normal (MVN). Each treatment group is assumed to have different sets of means. For dropouts randomized to an active group, applying JTR, the means of the MVN distribution across time for each dropout will be equal to those estimated means for that active group prior to dropping out, and equal to those estimated means for the placebo group post-dropout. For each imputation, the parameters of the MVN distributions will be sampled from the posterior distributions using Markov Chain Monte Carlo (MCMC) methods. The distribution of each subject's imputed values, conditional on his or her past values, will hence be MVN and a function of the subject's past values and the sampled parameters. The effect of past values on the

imputed values will depend on the strength of the correlation and the residuals (past values minus their expected values). For active group dropouts, the residuals will be computed using their respective active group means at the timepoints prior to dropout. While measures at timepoints prior to Week 16 are used in this MI approach, once one generates multiple complete datasets (100 to 1000 depending on feasibility) through imputation, the ANCOVA model described in 5.2.2.3 will be used to analyze just the data through Week 16 for treatment comparisons between each dose group of PF-06700841 and the combined placebo. Inference will be performed according to the Rubin's Rule (Rubin, 1987) for combining the multiple estimates and standard errors.

5.3. Methods to Manage Missing Data

In general, missing values in any endpoints will not be imputed when these endpoints are summarized using descriptive statistics. Missing values for safety endpoints will not be imputed.

5.3.1. Binary Endpoints

In the primary analysis of the primary endpoint, ACR20 response rate at Week 16, missing values due to a subject dropping from the study for any reason (e.g., lack of efficacy or adverse event), except for COVID-19, will be handled by non-responder imputation (NRI). In addition, if a subject discontinues treatment prematurely but not due to COVID-19, data collected at study visits after post treatment discontinuation will be censored, and subjects will be considered as non-responder for visits post treatment discontinuations. This approach will be used for all "response-type" endpoints (ACR20/50/70, PASI 75/90/100, MDA, VLDA and PsARC) at all timepoints under all the analysis methods specified in this document, with the exception of the GLMM method (Section 5.2.1.4). For the analysis of presence of dactylitis/enthesitis, missing data will not be imputed. Note that if a subject discontinues from the study at a visit of interest, say Week 16, but the value of the endpoint is not missing at Week 16, the value of the endpoint will be used for Week 16.

For a composite "response-type" endpoint (such as ACR20), if values 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. If one could not determine the response status in the presence of missing components at the timepoint, then the composite response-type endpoint status is considered "non-response" for that timepoint for all of the inferential analyses except the supplementary analysis for ACR endpoints up to Week 16 using a GLMM model (see Section 5.2.1.4), where MAR will be assumed for missing data.

If a subject misses a visit due to COVID-19, he/she will be excluded from the analysis for that visit; If a subject discontinues treatment or withdraws from the study due to COVID-19, he/she will be excluded from the analyses after the treatment discontinuation visit or study withdrawal visit, respectively.

5.3.2. Continuous endpoints

Continuous endpoints that are collected at multiple post-baseline timepoints during the first 16 weeks will be analyzed with a MMRM model (Section 5.2.2.1). This model will yield

unbiased estimates and valid inferences in the presence of data that is missing at random (MAR). If a subject misses a visit, discontinues treatment, or withdraws from the study at or prior to Week 16 due to COVID-19, the subject will NOT be excluded from the MMRM analysis. The missing data due to COVID-19 are assumed to be missing completely at random (MCAR) and the methodology in the MMRM model could apply.

As a supplementary analysis, a multiple imputation (MI) approach (Section 5.2.2.2) will also be used for the analysis of change from baseline in HAQ-DI at Week 16. If a subject misses a visit, discontinues treatment, or withdraws from the study at or before Week 16 due to COVID-19, the subject will be excluded from the MMRM analysis.

For descriptive summary of the continuous endpoints, if a subject misses a visit due to COVID-19, he/she will be excluded from the analysis for that visit; If a subject discontinues treatment or withdraws from the study due to COVID-19, he/she will be excluded from the analyses after the treatment discontinuation visit or study withdrawal visit, respectively.

For PRO instructions (SF-36, CCI [redacted] and FACIT-F), 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, and the methods to manage missing values as specified above will follow in the analyses of these endpoints.

5.3.4. Missing Dates

In compliance with Pfizer standards, if the day of the month is missing for any date used in a calculation, the 1st of the month will be used to replace the missing date unless the

calculation results in a negative time duration (eg, date of resolution cannot be prior to date of onset; if replacing resolution date with the 1st of the month results in a negative duration, the resolution date will be set to the onset date). Pfizer standards are similarly used if both month and day are missing (January 1 unless negative time duration).

If the start date is missing for an AE, the AE is considered to be treatment emergent unless the collection date is prior to the treatment start date.

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint

6.1.1. The proportion of subjects achieving ACR20 response at Week 16

6.1.1.1. Main Analysis

- Estimand strategy: Composite (Section 2.1.1).
- Analysis set: Evaluable Population (Section 4). All participants who were randomized to the study and received at least one dose of the randomized study treatment.
- Analysis methodology: Proportion of subjects achieving ACR20 response at Week 16 will be estimated by point estimates. Treatment comparison with placebo and two-sided 90% CIs will be tested using the normal approximation method, where the variance will be estimated by unpooled method (Section 5.2.1.1).
- Intercurrent events (post baseline but prior to Week 16 assessments) and missing data:
    - Treatment discontinuation: Per protocol Appendix 6, a participant will be withdrawn from treatment if experiences treatment compliance issues, certain AE, or abnormality findings in laboratory or ECG values. Any participants meeting discontinuation criteria must enter into the Follow-up Period with additional follow up visit occurring 1 week after their last dose of Early Termination Visit whenever possible. Data post treatment discontinuation will be censored and study participants who discontinue treatment prior to Week 16 will be considered as ‘non-responders’.
    - Handling of missing data due to COVID-19 is detailed in Section 5.3.1.
- The number and proportion of participants achieving ACR20 response at Week 16 will be presented for each treatment group.
- Treatment differences, 2-sided 90% CIs, raw p-values and adjusted p-values (Dunnett’s method) between each dose group of PF-06700841 and placebo in the ACR20 response rate at Week 16 will be presented, using normal approximation method (Section 5.2.1.1).
- Plot of the proportion of participants achieving ACR20 response at Week 16 by treatment group and plot of risk difference at Week 16 with 2-sided 90% CIs by dose group of PF-06700841 compared to placebo.

6.1.1.2. Sensitivity/Supplementary Analyses

Sensitivity Analyses

- Cochran-Mantel-Haenszel Analysis: Treatment differences, 2-sided 90% CIs and raw p-values between each dose group of PF-06700841 and placebo in the ACR20 response rate at Week 16 will be presented, using Cochran-Mantel-Haenszel method (Section 5.2.1.2).

- Logistic regression analysis of ACR20 response rate at Week 16 (Section 5.2.1.3). Missing data will be handled in the same way as the primary analysis. The odds ratio for the comparison of each dose group of PF-06700841 with the placebo group, the 2-sided 90% CIs, raw p-values and adjusted p-values (Dunnett's method) for odds ratio will be presented. The risk difference between each dose group of PF-06700841 and placebo (calculated using the observed placebo rate and estimated odds ratio from the logistic regression model) and 2-sided 90% confidence interval for risk difference will also be presented.

Supplementary Analyses

- Analysis of longitudinal ACR20 response rate (up to Week 16) using GLMM model (Section 5.2.1.4) will be performed. The odds ratio for the comparison of each dose group of PF-06700841 with the placebo group, the 2-sided 90% CIs will be presented by visit.

6.2. Secondary Endpoint(s)

6.2.1. The proportion of subjects achieving ACR20 response at Week 16 in the subgroup of subjects who are TNFα inhibitor naïve.

6.2.1.1. Main Analysis

- Estimand strategy: Composite

- Analysis set: the subgroup of subjects who are TNFα inhibitor naïve in the Evaluable Population (Section 4).

- Analysis methodology: Proportion of subjects achieving ACR20 response at Week 16 will be estimated by point estimates. Treatment comparison with placebo and two-sided 90% CIs will be tested using the normal approximation method, where the variance will be estimated by unpooled method (Section 5.2.1.1).

- Intercurrent events (post baseline but prior to Week 16 assessments) and missing data:
    - Same as in Section 6.1.1.1
- The number and proportion of participants achieving ACR20 response at Week 16 will be presented for each treatment group.

- Treatment differences, 2-sided 90% CIs and raw p-values between each dose group of PF-06700841 and placebo in the ACR20 response rate at Week 16 will be presented, using normal approximation method (Section 5.2.1.1).

- Plot of the proportion of participants achieving ACR20 response at Week 16 by treatment group and plot of risk difference at Week 16 with 2-sided 90% CIs by dose group of PF-06700841 compared to placebo.

6.2.1.2. Sensitivity/Supplementary Analyses

Sensitivity Analyses

- Logistic regression analysis of ACR20 response rate at Week 16 (Section 5.2.1.3). Missing data will be handled in the same way as the primary analysis. The odds ratio for the comparison of each dose group of PF-06700841 with the placebo group, the 2-sided 90% CIs, raw p-values and adjusted p-values (Dunnett's method) for odds ratio will be presented. The risk difference between each dose group of PF-06700841 and placebo (calculated using the observed placebo rate and estimated odds ratio from the logistic regression model) and 2-sided 90% confidence interval for risk difference will also be presented.

Supplementary Analysis – GLMM model

- Analysis of longitudinal ACR20 response rate (up to Week 16) using GLMM model (Section 5.2.1.4) will be performed.

- The ACR20 response rates by visit will be estimated from the fitted GLMM model.

- The odds ratio for the comparison of each dose group of PF-06700841 with the placebo group and the corresponding 2-sided 90% CIs will be presented by visit.

6.2.2. The proportion of subjects achieving an ACR20 response at all treatment timepoints (other than Week 16)

- Analysis set: Evaluable Population (Section 4).

- Analysis methodology: Proportion of subjects achieving ACR20 response at all visits will be estimated by point estimates. Treatment comparisons with placebo and two-sided 90% CIs will be made using the normal approximation method, where the variance will be estimated by unpooled method (Section 5.2.1.1).

- The number and proportion of participants achieving ACR20 response will be presented for each treatment group by timepoint.

- Treatment differences and 2-sided 90% CIs between each dose group of PF-06700841 and placebo in the ACR20 response rate at all visits (other than Week 16) will be presented, using normal approximation method (Section 5.2.1.1).

- Plot of proportion of participants achieving ACR20 response by treatment group and timepoints (including Week 16).

- Plot of treatment differences between each dose group of PF-06700841 and placebo with 2-sided 90% CIs by timepoints (including Week 16).

6.2.3. The proportion of subjects achieving ACR50 and ACR70 response at all treatment timepoints

- Analysis set: Evaluable Population (Section 4).

- Analysis methodology: Proportion of subjects achieving ACR50/ACR70 response at all visits will be estimated by point estimates. Treatment comparisons with placebo and two-sided 90% CIs will be made using the normal approximation method, where the variance will be estimated by unpooled method (Section 5.2.1.1).

- The number and proportion of participants achieving ACR50/ACR70 response will be presented for each treatment group.

- Treatment differences and 2-sided 90% CIs between each dose group of PF-06700841 and placebo in the ACR50/ACR70 response rate at all visits will be presented, using normal approximation method (Section 5.2.1.1).

- Plot of the proportion of participants achieving ACR50/ACR70 response by treatment group and timepoints.

- Plot of the treatment difference between each dose group of PF-06700841 and placebo with 2-sided 90% CIs by timepoints.

6.2.4. Change from baseline in the ACR response criteria components

- Analysis set: PD Analysis Population for hsCRP; Evaluable Population for all other criteria components (Section 4).

- ACR response criteria components include tender/painful joint count, swollen joint count, patient's assessment of arthritis pain, patient's global assessment of arthritis, physician's global assessment of arthritis, HAQ-DI and hsCRP.

- Each component endpoint will be summarized descriptively by the treatment group and visit.

    - The sample size, mean, standard deviation, median, minimum, and maximum at the baseline and postbaseline visits for baseline and change from baseline values will be presented by treatment group and visits (Section 5.2.2).

    - Treatment differences and 2-sided 90% CIs between each dose group of PF-06700841 and placebo at all postbaseline visits will be presented for each component endpoint.

- For each component endpoint, an MMRM model (Section 5.2.2.1) will be fit to the longitudinal data (up to Week 16). MAR is assumed for missing data in MMRM model analysis.
    - The least-squares (LS) means, 90% CIs for the LS means, treatment differences in the LS means between each dose group of PF-06700841 and placebo, and the corresponding 90% CIs will be presented for change from baseline in each component endpoint for all post-baseline visits.
- Analysis of HAQ-DI
    - Hypothesis testing of HAQ-DI at Week 16 will be based on the MMRM model. Treatment comparisons will be performed based on the fitted MMRM model. Handling of missing data due to COVID-19 is detailed in Section 5.3.2.As a supplementary analysis, the HAQ-DI will also be analyzed using a multiple imputation approach for missing data (Section 5.2.2.2).
    - A responder analysis of HAQ-DI will be performed, where a responder in HAQ-DI is defined as the change from baseline in HAQ-DI<-0.3. The analysis methodology will be the same as Section 6.2.3.

6.2.5. The proportion of subjects achieving PASI75/90/100 response

- Analysis Set: the subgroup of subjects with baseline BSA≥3% and PASI>0 in the Evaluable Population (Section 4).
- Analysis methodology: Same as Section 6.2.3.
- The number and proportion of participants achieving PASI75/90/100 response will be presented for each treatment group.
- Treatment differences and 2-sided 90% CIs between each dose group of PF-06700841 and placebo in the PASI75/90/100 response rate at all visits will be presented, using normal approximation method (Section 5.2.1.1).
- Plot of the proportion of participants achieving PASI75/90/100 response by treatment group and timepoints.
- Plot of treatment differences between each dose group of PF-06700841 and placebo with 2-sided 90% CIs by timepoints.

6.2.6. Change from baseline in the enthesitis score

- Enthesitis will be evaluated based on both SPARCC Enthesitis Index and Leeds Enthesitis Index. Analysis of enthesitis score will be performed for both indicies.
- Analysis Set: the subgroup of subjects with baseline enthesitis score >0 in the Evaluable Population (Section 4).

- Analysis strategy: Same as in Section 6.2.4.

6.2.7. Change from baseline in the Dactylitis Severity Score

- Analysis Set: the subgroup of subjects with baseline Dactylitis Severity Score (DSS)>0 in the Evaluable Population (Section 4).
- Analysis strategy: Same as in Section 6.2.4.

6.2.8. Change from baseline in the Nail Psoriasis Severity Index

- Analysis Set: the subgroup of subjects with baseline Nail Psoriasis Severity Index (NAPSI) >0 in the Evaluable Population (Section 4).
- Analysis strategy: Same as in Section 6.2.4.

6.2.9. Change from baseline in the Patient's Global Joint and Skin Assessment – Visual Analog Scale (PGJS-VAS)

- Analysis Set: Evaluable Population (Section 4).
- Analysis strategy: Same as in Section 6.2.4.

6.2.10. Change from baseline in the Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-Fatigue)

- Analysis Set: Evaluable Population (Section 4).
- Analysis strategy: Same as in Section 6.2.4.

6.2.11. Change from baseline in the Short-Form-36 Health Survey (SF-36) Version 2, Acute

- Analysis Set: Evaluable Population (Section 4).
- Analysis strategy: Same as in Section 6.2.4.

6.2.12. The proportion of subjects achieving Minimal Disease Activity (MDA) and Very Low Disease (VLDA) response

- Analysis Set: Evaluable Population (Section 4).
- Analysis strategy: Same as in Section 6.2.2.

6.2.13. Change from baseline in the Disease Activity Index for Reactive Arthritis/PsA (DAREA/DAPSA)

- Analysis Set: subjects in the Evaluable Population (Section 4) with baseline DAREA/DAPSA
- Analysis strategy: Same as in Section 6.2.4.

6.2.14. The proportion of subjects achieving the Psoriatic Arthritis Response Criteria (PsARC)

- Analysis Set: Evaluable Population (Section 4)
- Analysis strategy: Same as in Section 6.2.2.

6.2.15. Change from baseline in the Psoriatic Arthritis Disease Activity Score (PASDAS)

- Analysis Set: Evaluable Population (Section 4).
- Analysis strategy: Same as in Section 6.2.4.

CCI

[redacted]

CC
CI

[redacted]

[redacted]

[redacted]

[redacted]

CCI

[redacted]

6.4. Subset Analyses

Subgroup analysis will be performed for ACR20 for Prior TNFi exposure (yes, no).

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

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. In addition to displays by treatment groups, the summaries will also be provided for all the treatment groups combined.

Demographic characteristics:

- Baseline age (2 categorizations:18-44, 45-64, ≥65 years; and continuous in years);
- 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 in kg),

- Baseline height (cm, continuous);
- Baseline Body Mass Index (BMI: <18.5, 18.5 to <25, 25 to <30, 30 to <40, and ≥40 kg/m^2; and continuous kg/m^2);
- Screening smoking status (never smoked, ex-smoker, smoker);
- Alcohol use [yes, no; continuous (units/week) for subjects who consumed any alcohol, ie, responded Yes].

Baseline disease characteristics

- PsA duration (<2, ≥2 years; and continuous in years);
- Baseline PsA Subtype: (<5 joints, ≥5 joints);
- Baseline PASDAS: (≤3.2, >3.2 to <5.4, ≥5.4; and continuous);
- 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;
- [redacted]
- Baseline total psoriatic BSA (>=0 to <3, ≥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≥3% and PASI>0 at baseline,
- [redacted]
- Baseline SF-36 (physical functioning domain, physical component score [PCS] and mental component score [MCS]);
- Baseline FACIT-F (FACIT-F total score);
- Baseline SF-36 MCS (<42, >=42)
- [redacted]
- Baseline CRP (≤2.87, >2.87 mg/L; and continuous in mg/L);

- Concomitant Methotrexate Dose at Baseline (mg/wk)
- Baseline Rheumatoid Factor Positive
- Baseline Cyclic Citrullinated Peptide Antibody Positive
- Concomitant DMARDs at Day 1
- Concomitant MTX at Day 1
- Concomitant NSAIDs at Day 1
- Concomitant ORAL STEROIDS at Day 1
- Prior TNFi exposure (yes, no);

6.5.2. Study Conduct and Participant Disposition

Participant disposition will be summarized by treatment sequence.

6.5.3. Study Treatment Exposure

Study treatment exposure will be summarized by treatment sequence and combined when appropriate.

6.5.4. Concomitant Medications and Nondrug Treatments

All concomitant medication(s) and nondrug treatment will be provided in the listings.

6.6. Safety Summaries and Analyses

6.6.1. Adverse Events

In the 3-tier approach, the number and percentage of subjects with AEs over the first 16 weeks will be provided for each treatment group.

- Tier 1 events will not be analyzed in this study.
- Tier 2 events will be analyzed using asymptotic methods proposed by Miettinen and Nurminen (1985). Risk differences between each dose group of PF-06700841 and placebo and the corresponding 2-sided 95% CIs will be reported. Results will be presented in tabular displays and sorted by System Organ Class (SOC) and preferred terms (PT) within SOC, alphabetically, as well as by magnitude in risk difference. Graphic displays (e.g. forest plots) may be provided when appropriate. P-values will not be reported for Tier-2 events thus no multiplicity adjustment will be applicable.
- For Tier 2 events outputs, include footnotes to provide proper interpretation of p-values (Tier 1 events) and CIs, describe how comparison is conducted, e.g.: p-values and CIs are not adjusted for multiplicity and should be used for screening purpose only. The 95% CIs are provided to help gauge the precision of the estimates for risk difference. Risk difference is computed as PF-06700841 versus placebo. Footnotes should also state which methods were used to derive the p-values and confidence intervals.
- Tier-3 AEs will not be summarized separately but will be included as part of the overall AE summary. Tier-3 events will be reported with the observed proportions without comparative statistics.

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. The 3-tier approach facilitates this exploratory analysis.

6.6.2. Laboratory Data

Laboratory data will be listed and summarized by treatment in accordance with the sponsor reporting standards. Results from central labs will be used for statistical analyses. Baseline is defined in Section 3.4.

6.6.3. Vital Signs

Absolute values and changes from baseline in systolic and diastolic blood pressure, respiratory rate, pulse rate and temperature will be summarized by treatment and time post-dose, according to sponsor reporting standards. Baseline is defined in Section 3.4.

6.6.4. Electrocardiograms

Data from central readings will not be included in the CSR - all outputs will be produced using local ECG readings.

Baseline and change from baseline in ECG will be summarised descriptively up to Week 16 using 4 treatment groups (as described in Section 6.2). Baseline and change from baseline in ECG will be summarised at Week 52 using only those subjects using the same local read ECG machine at both Baseline and Week 52 visits using 6 treatment sequences (as described in Section 6.2).

Categorical summaries of the ECG data will be provided up to Week 16. For all visits from Week 20 to Week 52, only absolute categories will be reported.

Table 3. Safety QTc Assessment

Degree of Prolongation	Mild (ms)	Moderate (ms)	Severe (ms)
Absolute value	>450-480	>480-500	>500
Increase from baseline		30-60	>60

In addition, the number of participants with uncorrected QT values >500 ms will be summarized.

7. INTERIM ANALYSES

7.1. Introduction

An administrative interim analysis may be performed when approximately 60% of subjects have completed Week 16 visit or have the opportunity to complete the Week 16 visit. If performed, this interim analysis would be conducted by an unblinded team independent of this study team and the safety and efficacy results from the interim analysis may be reviewed by an Internal Review Committee (IRC). No members of the study team will be part of the

IRC. The interim analysis, if performed, will not be used to drive any decisions for the current study; however, it may be used for internal business decisions regarding future study planning.

During the interim analysis, some members of the study team may be unblinded and replaced with blinded colleagues. The subjects, investigators, and individuals from the sponsor (or designee) who interact with the investigators and monitor safety will continue to be blinded to individual study treatments throughout the follow up period of the study.

Before any interim analysis is conducted, the details of the objectives, decision criteria, dissemination plan, and method of maintaining the study blind as per Pfizer's standard operating procedures (SOPs) will be documented and approved in an internal review committee (IRC) charter.

7.2. Interim Analyses and Summaries

If performed, interim analyses will use the same analyses and summaries as described in Section 5 and 6, with the exception that hypothesis testings will not be formally performed because the sample size at IA would not provide adequate power. Due to the rapid study enrolment, the interim analysis was not conducted.

8. REFERENCES

1. Agresti A, Categorical Data Analysis, 2013, 3rd Edition, New York: John Wiley & Sons.

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 JF. The health assessment questionnaire (HAQ). Clinical and Experimental Rheumatology; 2005, 23 (Supp. 39), S14-S18.

4. Carpenter JR, Roger JH and Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. Journal of Pharmaceutical Statistics; 2013, **23**: 1352-1371.

5. 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.

6. 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.

7. 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.

8. Liu, GR and Zhan X Comparisons of methods for analysis of repeated binary responses with missing data. Journal of Biopharmaceutical Statistics, 2011, **21**: 371-392.

9. 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.

10. Mease PJ, Woolley JM, Bitman B, Wang BC, Globe DR, Singh A. Minimally important difference of Health Assessment Questionnaire in psoriatic arthritis: relating thresholds of improvement in functional ability to patient-rated importance and satisfaction. J Rheumatol 2011, 38: 2461-2465.

11. Miettinen O and Nurminen M. Comparative analysis of two rates. Statistics in Medicine, 1985, **4**: 213-226.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.

12. 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.

13. Rubin DB (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.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.

9. APPENDICES

Appendix 1.1. Endpoint Derivations

ACR assessments

The American College of Rheumatology's definition for calculating improvement in RA (ACR20) is calculated as a ≥20% improvement in tender and swollen joint counts and ≥20% improvement in 3 of the 5 remaining ACR-core set measures: patient and physician global assessments, pain, disability, and an acute-phase reactant. 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 in mm);
- Patient's Global Assessment of Arthritis (VAS in mm);
- Physician's Global Assessment of Arthritis (VAS in mm);
- 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. Explicitly,

1) If either tender or swollen joint count is missing or both are missing, ACR20/50/70 will be missing,
2) Of the other 5 remaining components,
   a. if >2 components are missing, ACR20/50/70 will be missing,
   b. if 1 of the components is missing and if < 3 components satisfy the improvement criteria, then ACR20/50/70 will be missing,
   c. if 2 of the components is missing and if < 3 components satisfy the improvement criteria, then ACR20/50/70 will be missing.

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. The unit for CRP will be mg/dL or mg/L depending on endpoint of interest. The normal range for the CRP is ≤2.87 mg/L.

The VAS data will need to be rescaled prior to any calculation (Section 3).

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 reach" 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 & Fries, 2005). A higher score represents a more limited physical functional status/ability.

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).

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).

CCI [redacted]

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) - TJC;

Swollen joint count (66) - SJC;

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) ≥20% improvement in Physician's Global Assessment of Arthritis (VAS); (2) ≥20% improvement in Patient's Global Assessment of Arthritis (VAS); (3) ≥30% improvement in tender joint count (68); and (4) ≥30% improvement in swollen joint count (66) (Gottlieb et al 2008).

If values in any of the components at a timepoint were missing, the component variables that were not missing were used to determine the response status. If one could not determine the response status in the presence of missing components at the timepoint, then the PsARC response will be considered as missing. Specifically, the following algorithm will be used for the derivation of the PsARC response.

PsARC Response	Algorithm
Responder	Meet improvement criteria for at least 2 components, at least one of which is TJC or SJC; AND
	No worsening component and no missing component.
Non-responder	Both TJC and SJC available but not meeting improvement criterion; OR

	(Number of components meeting the improvement criteria) + (Number of missing components) <= 1; OR
	At least one worsening component
Missing	Otherwise

The VAS data will need to be rescaled prior to any calculation (Section 3).

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

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

DAREA/DAPSA (Schoels et al 2010) 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 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 at every study visit during the treatment period except Week 2 visit 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 3).

Minimal Disease Activity (MAD) Score and Very Low Disease Activity (VLDA) Score

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

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

The VAS data will need to be rescaled prior to any calculation (Section 3).

Psoriatic Arthritis Disease Activity Score (PASDAS)

PASDAS (Helliwell et al 2013) is a composite PsA disease activity score that includes the following components: patient's global joint and skin assessment (VAS in mm), physician's 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-36 and CRP (mg/L) (Helliwell et al. 2013).

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

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 3). A higher PASDAS score indicates a higher disease activity.

Psoriasis Area and Severity Index (PASI)

The Psoriasis Area and Severity Index (PASI) 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 and neck, upper limbs, trunk (including axillae and groin), and lower limbs (including buttocks). 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.

$$PASI = 0.1Ah\,(Eh + Ih + Sh) + 0.2Au\,(Eu + Iu + Su) + 0.3At\,(Et + It + St) + 0.4Al\,(El + Il + Sl)$$

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:

$$PASI\_E = 0.1Ah\,(Eh) + 0.2Au\,(Eu) + 0.3At\,(Et) + 0.4Al\,(El)$$
$$PASI\_I = 0.1Ah\,(Ih) + 0.2Au\,(Iu) + 0.3At\,(It) + 0.4Al\,(Il)$$
$$PASI\_S = 0.1Ah\,(Sh) + 0.2Au\,(Su) + 0.3At\,(St) + 0.4Sl\,(Sl)$$

Any missing component will result in PASI as missing. In addition to the analysis of PASI scores, BSA (%) will be summarized descriptively.

CCI [redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

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: ΔFACIT-F total score, ΔFACIT-F experience domain score and ΔFACIT-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 (ie, 0 →4, 1→3, 3→1, 4→0) 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 (i.e., 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.

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.

The summary component scores are:

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

In addition, there is another subscale in the SF-36: 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, 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-36

VARIABLE	DERIVATION
SF-36 PF scale score	raw score = sum (items 3A, 3B, 3C, 3D, 3E, 3F, 3G, 3H, 3I, 3J) PF = (raw score – 10) * 5 PF_Z = (PF – 82.62455) / 24.43176 PF scale score = (PF_Z*10) + 50 When calculating the raw score, if 5 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 5 of the items are non-missing then PF scale score is missing. 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. A higher PF scale score indicates better physical functioning.
SF-36 RP scale score	raw score = sum (items 4A, 4B, 4C, and 4D) RP = [(raw score -4)/16] * 100 RP_Z = (RP – 82.65109) / 26.19282 RP scale score = (RP_Z * 10) + 50 When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then RP scale score is missing. 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.

SF-36 BP scale score	raw score = sum (reversed item 7 and reversed item 8) BP = (raw score -2) * 10 BP_Z = (BP – 73.86999) / 24.00884 BP scale score = (BP_Z * 10) + 50 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 Reverse direction of item 8 as follows: if =1 and original item 7 value =1, set to 6 if =1 and original item 7 value >=2, set to 5 if =2, set to 4 if =3, set to 3 if =4, set to 2 if =5, set to 1 If item 7 is answered and item 8 is missing, set 8 = reversed 7 as defined above. If item 8 is answered and item 7 is missing, set item 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 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. 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.

SF-36 GH scale score	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 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 Reverse direction of item 11B and 11D by subtracting score from 6. When calculating the raw score, if 3 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 3 of the items are non-missing then GH scale score is missing. 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.

SF-36 VT scale score	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 Reverse direction of Items 9a and 9e by subtracting score from 6. When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then VT scale score is missing. 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.
SF-36 SF scale score	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 Reverse direction of score for item 6 by subtracting score from 6. 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. 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.

SF-36 RE scale score	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 When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then RE scale score is missing. 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.
SF-36 MH scale score	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 Reverse direction of scores for 9D and 9H, by subtracting score from 6. If 3 or more of the items are non-missing, then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 3 of the items are non-missing then MH scale score is missing. 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.

SF-36 TR scale score	raw score = item 2 TR scale score = raw score 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.
SF-36 PCS score	PCS score includes the 8 scales for GH, PF, RP, RE, SF, MH, BP, and VT. 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; Raw Score = ((GH1*.24954)+(PF1* .42402)+(RP1*.35119) + (RE1*-.19206)+(SF1*-.00753)+(MH1*-.22069)+(BP1*.31754) + (VT1*.02877)) PCS Summary Scale Score = (raw score *10) + 50 Raw Score is missing if one of the component scale scores is missing.
SF-36 MCS score	MCS score includes the 8 scales for GH, PF, RP, RE, SF, MH, BP, and VT. 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; Raw Score =((GH1*-.01571)+(PF1*-.22999)+(RP1*-.12329) + (RE1*.43407)+(SF1*.26876)+(MH1*.48581)+(BP1*- 0.09731) + (VT1*.23534)) MCS Summary Scale Score = (raw score*10)+50 Raw score is missing if one of the component scale scores in missing.

CCI

CCI

List of Abbreviations

Abbreviation	Term
ACR	American College of Rheumatology
AE	Adverse Event
ANCOVA	Analysis of Covariance
AR	Autoregressive
BASDAI	Bath Ankylosing Spondylitis Disease Assessment Index
BLQ	Below the limit of quantification
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CRF	Case Report Form
CRP	C-Reactive Protein
DAREA/DAPSA	Disease Activity Index for Reactive Arthritis/Psoriatic Arthritis
DAS	Disease Activity Score
DLQI	Dermatology Life Quality Index
DMARDs	Disease-modifying Antirheumatic Drugs
ECG	Electrocardiogram
FACS	Fluorescence-Activated Cell Sorting
FAS	Full Analysis Set
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
GLIMMIX	Generalized linear mixed-effects model with repeated measures
HAQ-DI	Health Assessment Questionnaire – Disability Index
IE	Intercurrent Event
IRC	Internal Review Committee
CC	[redacted]
ITT	Intent-to-Treat
JTR	Jump to Reference
LLOQ	Lower Limit of Quantitation
LOCF	Last Observation Carried Forward
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
MCMC	Markov Chain Monte Carlo
MCS	Mental Component Summary
MDA	Minimal Disease Activity
MI	Multiple Imputation
MMRM	Marginal Model for Repeated Measures
MVN	Multivariate Normal
NAPSI	Nail Psoriasis Severity Index
NRI	Non-Responder Imputation
PASDAS	Psoriatic Arthritis Disease Activity Score
PASI	Psoriasis Area and Severity Index
PCS	Physical Component Summary

Abbreviation	Term
PD	Pharmacodynamic(s)
CCI [redacted]	[redacted]
PGJS	Patient's Global Joint and Skin Assessment
PK	Pharmacokinetic(s)
PsA	Psoriatic Arthritis
CCI [redacted]	[redacted]
CCI [redacted]	[redacted]
PsARC	Psoriatic Arthritis Response Criteria
PT	Preferred Terms
QTc	Corrected QT
QTcF	Corrected QT (Fridericia method)
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SF-36	Short-Form-36
SIE	Serious Infectious Events
SJC	Swollen Joint Count
SOC	System Organ Class
SOP	standard operating procedure
SPARCC	Spondyloarthritis Research Consortium of Canada
TJC	Tender Joint Count
UN	Unstructured
VAS	Visual Analog Scale
VLDA	Very Low Disease Activity