Official Title of Study:

A Randomized, Placebo-Controlled, Double-blind, Multicenter Study to Assess the Efficacy and Safety of Multiple Doses of BMS-986165 in Subjects with Active Psoriatic Arthritis (PsA)

NCT Number: NCT03881059

Document Date (Date in which document was last revised): 05 June 2020

Statistical Analysis Plan

Sponsor:	Bristol-Myers Squibb
Protocol No:	IM011084
PRA Project Id:	BMS11084-011084 (IM011-084)
Version No.:	2.0/Amendment # 5

Title:	A Randomized, Placebo-Controlled, Double-blind, Multicenter Study to Assess the Efficacy and Safety of Multiple Doses of BMS-986165 in Subjects with Active Psoriatic Arthritis (PsA)
CRF Version No./Date:	3.1/05-Feb-2020
SAP No.	2.0

Approvals

Table of Contents

Abbreviations

Glossary of Abbreviations:	
ATC	Anatomic Therapeutic Classification
ACR	American College of Rheumatology
ADA	anti-drug antibody
AE	adverse event
AEI	adverse event of interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
AT	aminotransaminases
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BMI	body mass index
BMS	Bristol-Myers Squibb
BSA	Body surface area
CBC	Complete Blood Count
CI	Confidence Interval
CK	Creatine Kinase
CMH	Cochran-Mantel-Haenszel
CRF	Case Report Form, paper or electronic
Ctrough	Trough observed plasma concentration
DAPSA	Disease Activity Index for Psoriatic Arthritis Score
DAS	Disease Activity Score
DMARD	Disease-modifying antirheumatic drug
DILI	Drug-induced liver injury
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
eGFR	Estimated glomerular filtration rate
FAS	Full Analysis Set
GFR	Glomerular Filtration Rate
HAQ-DI	Health Assessment Questionnaire-Disability Index
HDL	High Density Lipoprotein
hsCRP	High-sensitivity C-reactive protein
ICH	International Conference on Harmonisation

Glossary of Abbreviations:	
IRT	Interactive Response Technology
LDH	Lactate dehydrogenase
LDL	Low Density Lipoprotein
LDI	Leeds Dactylitis Index
LEI	Leeds Enthesitis Index
LLOQ	Lower limit of quantification
LOCF	Last observation carried forward
LS	Least-squares
MDA	Minimal Disease Activity
MCS	Mental Component Summary
NRI	Non-responder imputation
PASI	Psoriasis Area and Seventy Index
PASDAS	Psoriatic Arthritis Disease Activity Score
PCS	Physical Component Summary
PDGD	Protocol deviation guidance document
PGA-F	Physician's Global Assessment-Fingernails
PK	Pharmacokinetics
PPS	Per Protocol Set
PROMIS	Patient Reported Outcome Measurement Information System
PsA	psoriatic arthritis
PsAID	Psoriatic Arthritis Impact of Disease
PsARC	Psoriatic Arthritis Response Criteria
PsO	Psoriasis
QD	Once daily
QTcB	QT interval corrected using Bazett's formula
QTcF	QT interval corrected using Fridericia's formula
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SF-36	Short Form Health Survey-36 Item
SPARCC	Spondyloarthritis Research Consortium of Canada
SQ	Subcutaneous
TEAE	Treatment-emergent adverse events
TNF	Tumor Necrosis Factor
TNFi	TNF-inhibitor

Glossary of Abbreviations:	
TYK2	Tyrosine Kinase 2
ULN	Upper limit of normal
VLDA	Very Low Disease Activity
WHO	World Health Organization
WLQ	work limitation questionnaire

1.0 Purpose

The statistical analysis plan (SAP) describes the statistical methods to be used during the reporting and analyses of data collected under BMS Protocol IM011084.

The SAP outlines the following:

- Study design
- Study objectives
- Endpoints and assessments
- Analysis sets
- Statistical methodology
- Conventions and definitions

The SAP should be read in conjunction with the study protocol and case report form (CRF). This version of the plan has been developed using the protocol v2.0 and CRF dated 05FEB2020. Any further changes to the protocol or CRF may necessitate updates to the SAP. Changes following approval of the first version of the SAP will be tracked in the SAP Change Log and a final version of the amended SAP will be approved prior to final database lock.

2.0 Study Description

2.1 Study Design

This is a Phase 2 randomized, double-blind, placebo-controlled, multicenter study to evaluate the safety and efficacy of two doses of BMS-986165 in subjects with active psoriatic arthritis (PsA). A total of approximately 180 subjects with active PsA will be randomized in Part A. This Phase 2 study will be conducted in a mixed population of subjects with active PsA who have either not received prior treatment with a biologic (biologic-naïve, approximately 70% targeted) or who have failed or been intolerant to 1 TNF-inhibitor (TNFi-experienced, approximately 30% targeted).

Part A

Subjects will undergo screening evaluations to determine eligibility within 28 days prior to administration of study medication. Following the screening process, if subjects are eligible for randomization then subjects will be randomized in a 1:1:1 ratio (approximately 60 subjects per arm) to 1 of the following 3 treatment groups:

- Placebo, oral
- BMS-986165 6 mg QD, oral
- BMS-986165 12 mg QD, oral

The primary endpoint is American College of Rheumatology (ACR) 20 response at Week 16 in Part A. Subjects who discontinue study treatment early for any reason other than consent withdrawal will be expected to complete the early termination visit followed by the safety follow-up period.

Part B

At Week 16, subjects will have the option of participating in the double-blind Part B of the study. All subjects who received placebo in Part A will receive ustekinumab subcutaneously (SQ) in Part B according to approved labeling for PsA; subjects who receive BMS-986165 in Part A and achieve Minimal Disease Activity (MDA) will continue treatment with BMS-986165 at the same dose in Part B, and subjects who receive BMS986165 in Part A and fail to achieve MDA will receive ustekinumab SQ in Part B. The treatment assignment in Part B will be double-blind double-dummy as follows:

- Placebo subjects in Part A: ustekinumab SQ injection plus oral placebo matching BMS-986165 QD in Part B
- BMS-986165 6 mg QD subjects in Part A with MDA at Week 16: BMS-986165 6 mg QD orally plus placebo matching ustekinumab SQ injection in Part B

- BMS-986165 6 mg QD subjects in Part A not achieving MDA at Week 16: ustekinumab SQ injection plus oral placebo matching BMS-986165 QD in Part B
- BMS-986165 12 mg QD subjects in Part A with MDA at Week 16: BMS-986165 12 mg QD orally plus placebo matching ustekinumab SQ injection in Part B
- BMS-986165 12 mg QD subjects in Part A not achieving MDA at Week 16: ustekinumab SQ injection plus oral placebo matching BMS-986165 QD in Part B

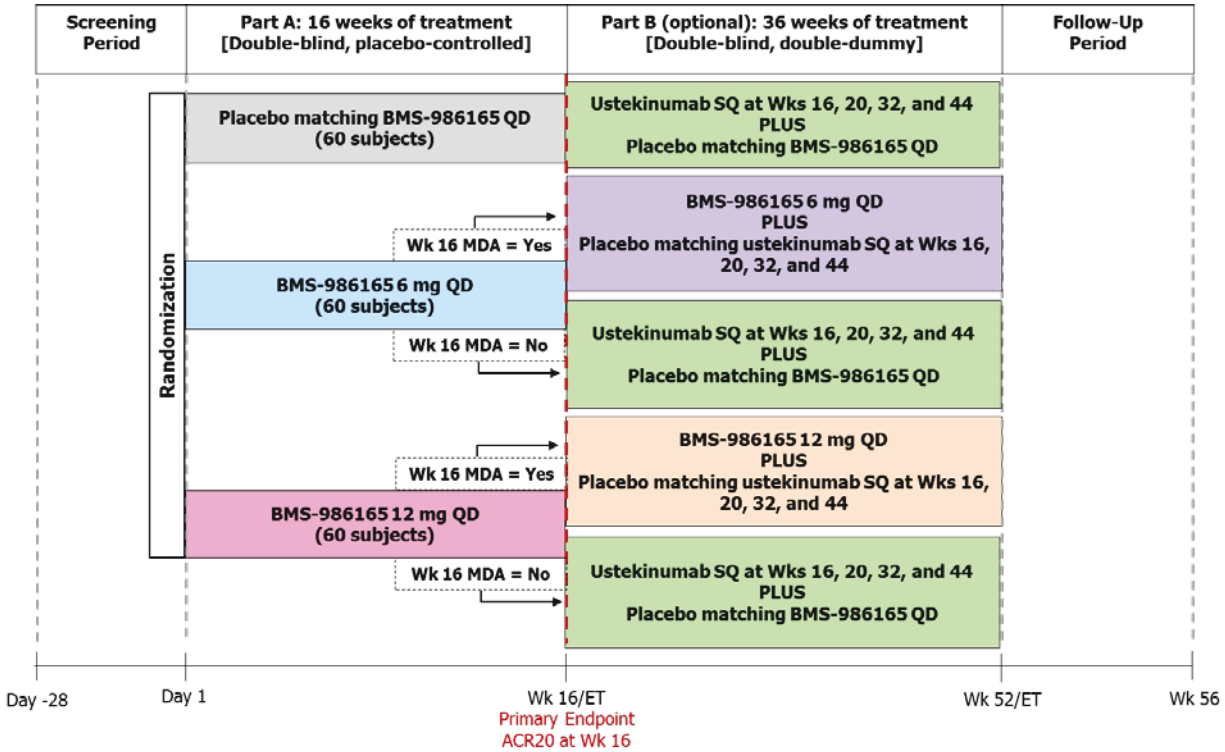
Study Design elements

The total duration of participation in Part A is approximately 24 weeks and will be divided into the following periods: screening (up to 4 weeks), double-blind placebo-controlled treatment for 16 weeks (Week 1 to Week 16, Part A), and follow-up (up to 4 weeks, if discontinued from study at Week 16).

For subjects completing both Part A and B of the study, the anticipated study duration is approximately 60 weeks: screening (up to 4 weeks), double-blind placebo-controlled treatment for 16 weeks (Week 1 to Week 16, Part A), double-blind double-dummy treatment for 36 weeks (Week 16 to Week 52, Part B), and follow-up (up to 4 weeks).

A schedule of assessments can be found in the protocol. A study design schematic is provided in Figure 1.

Figure 1: Study Design Schematic



Notes:
1) Subjects that complete Part A (16 weeks of treatment) will be offered the opportunity to enroll in Part B (36 weeks of treatment)
2) Completion of Week 16 of Part A and the start of Part B will occur on the same day
3) Subjects that complete Part A but decline Part B will proceed into the Follow-Up Period
4) Subjects who discontinue investigational product during Part A or Part B will have an Early Termination (ET) visit followed by the Follow-Up Period
5) Treatment assignment in Part B will depend on the subject's Part A randomized treatment as well as if the subject met MDA at Week 16 (Part A)
6) At Week 48, subjects will discuss with their Investigator/treating physician plans for treatment with ustekinumab following study completion. Subjects electing to take ustekinumab following study completion will receive either an ustekinumab SQ injection or placebo matching ustekinumab SQ injection at Week 52.

2.2 Sample Size and Power

Sample size considerations are based on providing exposure in a sufficient number of subjects for two dose levels of BMS-986165, 6 mg QD and 12 mg QD, to be able to observe a true dose-response trend that is significantly greater than placebo in the primary endpoint of ACR 20 after 16 weeks of treatment.

BMS-986165 is expected to provide similar PsA efficacy as observed in recent PsA trials with effective agents. With this assumption, the expected overall placebo response rate is assumed to be 30% and BMS-986165 response rates to be 50% and 55% in the 6 mg QD and 12 mg QD doses respectively.

A total sample size of 60 subjects per arm randomized in a blinded fashion at a 1:1:1 ratio to BMS-986165 6 mg QD arm, BMS-986165 12 mg QD arm, and placebo will provide 87.5% power in a mixed population to detect a significant dose-response trend compared to placebo for the primary endpoint of ACR 20 response at Week 16 using a trend test with α=0.05 (1-sided) or α=0.10 (2-sided).

2.3 Treatment Assignment and Randomization

At the time of the screening visit, before any study-related procedures are performed, the investigative site will access the enrollment option of the Interactive Response Technology (IRT) system for assignment of a subject number. This number is assigned sequentially by the system and will be unique across all sites. If a potential subject is rescreened, a new identification number will be assigned at the time of rescreening.

At Week 0 (Day 1) of Part A, subjects who meet all criteria for enrollment at Screening and Day 1 will be centrally randomized in a 1:1:1 ratio to BMS-986165 6 mg QD, BMS-986165 12 mg QD, or placebo matching BMS-986165, as determined by a computer-generated randomization schedule using the IRT. The randomization lists will be generated by the IRT using a permuted block design within each combination of stratum level. Randomization will be stratified by prior use (experienced/naïve) of TNF inhibitor (TNFi) and by body weight (< 90 kg and ≥ 90 kg). After all inclusion/exclusion criteria have been met for a subject, the investigative site will access the IRT on Day 1 for the purposes of randomizing the subject. A treatment group will be assigned by the IRT based on the above-described randomization schedule and each subject will be assigned a unique randomization number. In addition, a kit number will be assigned to the subject corresponding to the treatment assignment.

Subjects completing Part A will have the option to enroll into Part B of the study. Treatment assignment in Part B will be based on the subject's treatment assignment in Part A as well as the patient's MDA score at Week 16. An adequate supply of blister cards will be dispensed to cover the subject until their next scheduled visits

2.4 Unblinding Information

The Data Monitoring Committee (DMC) provides oversight of safety consideration throughout the study. A separate unblinded team, comprised of an unblinded Independent Reporting Statistician (IRS) and unblinded programmer(s), will produce output for the DMC using masked treatments. Treatment decodes may only be requested by DMC Chair and will be provided by the IRS. Data summaries and listings will be transmitted via a secure portal by IRS to only the DMC members. Additional details regarding the DMC process and unblinding are provided in the DMC charter.

Unblinded endpoint analyses will be performed after all subjects complete assessments through the Week 16 visit (Part A). An unblinding plan will be finalized prior to unblinding for endpoint analyses. The analysis will be performed by an independent statistical team who are not involved in the conduct of the study otherwise and reviewed by non-study team members who will use the data to support Phase 3 development decision making. Study team members who are site-facing will remain blinded until database lock of Part B.

Additionally, bioanalytical scientists involved in the processing of bioanalytical samples will be unblinded to randomized treatment assignments to minimize unnecessary sample bioanalysis of subjects who are on placebo.

2.5 Changes in Statistical Considerations from the Protocol

- The Full Analysis Set was changed from “all randomized subjects who are dispensed study drug” in the protocol to “all subjects who were randomized”.
- The Safety Set (All Treated) population was changed to the As-treated population to align with the other studies for this compound.
- The initial list of relevant protocol deviations provided in the protocol was updated for determination of the Per Protocol population.
- The RAPID3 endpoint was removed from the additional endpoints as this endpoint requires a different version of the Health Assessment Questionnaire than was used in IM011-084.
- Imputation methods were updated to remove the prohibited medication/ therapy criteria for binary and continuous endpoints.

3.0 **Objectives**

3.1 Primary Objective

The primary efficacy objective of this study is:

- Assess the dose-response relationship of BMS-986165 (6 or 12 mg once daily [QD]) at Week 16 in the treatment of subjects with active PsA (in terms of ACR 20 response)

3.2 Secondary Objectives

The secondary efficacy objectives of this study are:

- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in daily functional activities (in terms of HAQ-DI score)
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in reducing PsO severity (in terms of PASI 75 response)
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in Patient-reported Outcomes (in terms of SF-36, Physical Component Summary (PCS) score)

3.3 Additional Objectives

The Additional efficacy objectives of this study are:

- Assess the extent of the dose-response relationship of BMS-986165 treatment (6 or 12 mg QD) at Week 16 in active PsA (in terms of ACR 50 and ACR 70 responses)
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in Disease Activity Score (DAS) 28 CRP
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in dactylitis
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in enthesitis
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in Physician’s global assessment - fingernails (PGA-F) psoriasis (PsO)
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in minimal disease activity (MDA)
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in psoriatic arthritis activity (in terms of Psoriatic Arthritis Disease Activity Score, Disease Activity Index for Psoriatic Arthritis Score, and Psoriatic Arthritis Response Criteria)

- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in ankylosing spondylitis activity (in terms of Bath Ankylosing Spondylitis Disease Activity Index)
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in Patient-Reported Outcomes (in terms of SF-36 mental component summary (MCS) score, PROMIS-Fatigue) or FACIT-Fatigue score)
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in daily functional activities (in terms of HAQ-DI 0.35 response)
- Assess the dose-response relationship of BMS-986165 (6 or 12 mg QD) at Week 16 in reducing PsO severity (in terms of Psoriasis Activity and Severity Index 90 response)

3.4 Safety Objectives

The safety objectives of this study are to assess the safety and tolerability of BMS-986165 in subjects with active PsA.

4.0 Outcomes

The description of assessments for efficacy and safety can be found in Section 8 of the protocol. The calculation of key measures is provided in Section 8.2 of the SAP.

4.1 Efficacy

4.1.1 Primary Endpoint

The primary endpoint is the American College of Rheumatology (ACR) 20 response at Week 16. A subject is considered an ACR 20 responder if the following three conditions are met:

- ≥ 20% improvement from baseline in the number of tender joints (68 joint count)
- ≥ 20% improvement from baseline in the number of swollen joints (66 joint count)
- ≥ 20% improvement from baseline in at least 3 of the following 5 domains:
    - Subject Global Assessment of disease activity (see Protocol: APPENDIX 7)
    - Physician Global Assessment of psoriatic arthritis
    - Subject Global Assessment of pain (see Protocol: APPENDIX 8)

- Health Assessment Questionnaire-Disability Index (HAQ-DI)
- High-sensitivity C-reactive protein (hsCRP)

4.1.2 Secondary Endpoints

The secondary endpoints assessed at Week 16 include the following:

- Change from baseline in HAQ-DI score
- Psoriasis Area and Seventy Index (PASI) 75 response in subjects with at least 3% Body Surface Area (BSA) involvement at baseline
- Change from baseline in Short Form Health Survey-36 Item (SF-36) Physical Component Summary (PCS) score

4.1.3 Additional Endpoints

Additional endpoints at Week 16 include:

- ACR 50/70 response
- HAQ-DI 0.35 response, where HAQ-DI 0.35 responder is defined as a subject with an improvement from baseline in HAQ-DI score of at least 0.35
- Psoriasis Area and Severity Index (PASI) 90 response in subjects with at least 3% body surface area (BSA) involvement at baseline
- Dactylitis mean change from baseline (dactylitis count) in subjects with dactylitis at baseline
- Dactylitis change from baseline (LDI)
- Dactylitis resolution, where resolution is defined as a dactylitis count of 0 in subjects with dactylitis count ≥ 1 at baseline
- Enthesitis mean change from baseline using the Leeds Enthesitis Index (LEI)
- Enthesitis resolution, where resolution is defined as an LEI score of 0, in subjects with LEI ≥ 1 at baseline
- Change from baseline in Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis index
- Enthesitis resolution, where resolution is defined as a SPARCC enthesitis index score of 0, in subjects with SPARCC ≥ 1 at baseline
- Minimal Disease Activity (MDA) response, where an MDA responder is based on a subject fulfilling at least 5 of the below outcomes:
  - Tender joint count ≤ 1
  - Swollen joint count ≤ 1
  - PASI ≤ 1 or BSA ≤ 3%
  - Subject Global Assessment of pain ≤ 15 (see Protocol APPENDIX 8)
  - Subject Global Assessment of disease activity ≤ 20 (see Protocol APPENDIX 7)
  - HAQ-DI ≤ 0.5
  - Tender entheseal points ≤ 1
- Nail psoriasis assessment by PGA-F 0/1, assessed as a proportion of subjects with a PGA-F score of 0 or 1 amount subjects with a baseline PGA-F score ≥ 3
- Change from baseline in Psoriatic Arthritis Disease Activity Score (PASDAS)

- Change from baseline in Disease Activity Index for Psoriatic Arthritis Score (DAPSA)
- Psoriatic Arthritis Response Criteria (PsARC) response
- Disease Activity Score (DAS) 28 CRP response:
    - Low Disease Activity (LDA) defined as DAS 28 CRP score < 3.2
    - Proportions in remission and time to remission; remission is defined as a DAS 28 CRP score < 2.6
- Change from baseline in DAS 28 CRP score
- Change from baseline in Psoriatic Arthritis Impact of Disease (PsAID) score
- Change from baseline in Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), in subjects with baseline evidence of PsA spondylitis
- Change from baseline in SF-36 mental component summary (MCS) score
- Change from baseline in Patient Reported Outcome Measurement Information System-Fatigue (PROMIS-Fatigue) or FACIT-Fatigue score
- Change from baseline in Work Limitation Questionnaire (WLQ) At-Work Productivity Loss score

[redacted]

4.2 Safety

The safety outcomes include the following:

- Adverse events (AEs)
    - Treatment-emergent adverse events (TEAEs) – defined as:
        - Adverse events (AEs) which occur after the first dose of study treatment through 30 days after the final dose of the study treatment or subject's participation in the study if the last scheduled visit occurs at a later time.
    - Treatment-emergent adverse events of interest (AEIs) for the following events:
        - Skin-related AEs
        - Infection AEs, including influenza
        - Malignancy
    - SAEs
- Selected clinical laboratory parameters
    - Absolute and change from baseline values
    - Laboratory abnormalities (as determined by Common Terminology Criteria for Adverse Events [CTCAE v5.0] grading)
    - Shifts from baseline
    - Potential drug induced liver injury (DILI) is defined as a subject who meets the following criteria:

- 1) ALT or AST elevation >3 times ULN

AND

- 2) Total bilirubin >2 times ULN, without initial findings of cholestasis (elevated serum alkaline phosphatase)

AND

- 3) No other immediately apparent possible causes of aminotransaminase (AT) elevation and hyperbilirubinemia, including, but not limited to, viral hepatitis, pre-existing chronic or acute liver disease, or the administration of other drug(s) known to be hepatotoxic.

- Vital signs
  - Absolute and change from baseline values
  - Marked abnormalities defined by below categories:
    - Heart rate:
      - Value > 100 and change from baseline > 30
      - Value < 55 and change from baseline < -15
    - Systolic blood pressure:
      - Value > 140 and change from baseline > 20
      - Value < 90 and change from baseline < -20
    - Diastolic blood pressure:
      - Value > 90 and change from baseline > 10
      - Value < 55 and change from baseline < -10
- Electrocardiograms (ECGs)
  - Absolute and change from baseline values
  - Marked abnormalities defined by below categories:
    - QT interval corrected using Fridericia's formula (QTcF):
      - 450 -<480 msec
      - 480 -< 500 msec
      - ≥ 500 msec
      - 30 < change from baseline ≤ 60 msec
      - Change from baseline > 60 msec
      - Males: < 450 msec, ≥ 450 msec
      - Females: < 470 msec, ≥ 470 msec
    - PR interval ≥ 200 msec
    - QRS interval ≥ 200 msec

In cases where the QT interval is corrected using Bazett's (QTcB) formula is captured instead of QTcF, then QTcB will be converted to QTcF for analyses.

[redacted]

5.0 Populations for Analyses

The following analysis sets will be used in the summary and analysis of study data:

- **Enrolled population:** All subjects who sign informed consent.
- **Full Analysis Set (FAS):** All subjects who were randomized. Following the intent-to-treat (ITT) principle, subjects will be analyzed according to the treatment group assigned at randomization. The FAS will be the primary efficacy analysis population.
- **Per Protocol Set (PPS):** A subset of the FAS who are compliant with study treatment and who do not have any relevant protocol deviations (see Section 5.1 below) that may impact the primary efficacy endpoint assessments. The PPS will be analyzed according to the treatment assigned at randomization. The PPS will be a supportive efficacy analysis population.
- **As-treated Population:** All randomized subjects who receive at least one dose of double-blind study treatment. Subjects will be analyzed according to treatment received.
- [redacted]

[redacted]

5.1 Relevant Protocol Deviations

Relevant protocol deviations are deviations that can have an impact on the primary efficacy endpoints. The impact of relevant protocol deviations on the primary efficacy results will be assessed by excluding subjects from the FAS to define the PPS in supportive analyses of the primary efficacy endpoints. Relevant protocol deviations to be considered regarding exclusion of subjects from the FAS will include the following:

- Randomized but did not take any study treatment
- Failed to meet study inclusion criteria, 1a, 2a (Classification Criteria for Psoriatic Arthritis portion), 2b, 2d, 2e, 2f, 2g, 2i, 2k, and 3a, but were entered into the study

- Met study exclusion criteria, 4a, 4b, 4m, 4n, and 4s, but were entered into the study (only exclusion criteria expected to have an impact on the primary efficacy endpoints will be considered relevant)
- Failed to maintain compliance to study medication of >=75% for the overall treatment period to Week 16
- Failure to adhere to prohibited concomitant medication restrictions as described below:
    - Receiving a single intramuscular, intra-articular, intravenous, or oral course of high dose corticosteroid (prednisone > 10 mg/day or equivalent) within 28 days of the primary time point at Week 16
    - Receiving more than one instance of a single intramuscular, intra-articular, intravenous, or oral course of high dose corticosteroid (prednisone > 10 mg/day or equivalent) up to Week 16
    - Starting a new non-biologic disease-modifying antirheumatic drug (DMARD) at any time before the primary endpoint at Week 16
    - Starting a biologic DMARD at any time before the primary endpoint at Week 16
- Actual treatment received is different than randomized treatment

All subjects with relevant protocol deviations will be identified prior to unblinding of treatment assignment with the exception of subjects where actual treatment received is different than randomized treatment which will be determined after treatment unblinding. Relevant protocol deviations will be summarized by treatment group and deviation category for the FAS population.

6.0 Statistical Analyses

Descriptive summaries and analyses will be presented for data captured throughout the study using the following treatment groups.

During the first 16 weeks of treatment, data will be presented for the following treatments:

- BMS-986165 6 mg QD
- BMS-986165 12 mg QD
- Placebo

After Week 16, subjects continuing in Part B of the study will have their data presented for the following treatments:

- BMS-986165 6 mg QD
- BMS-986165 12 mg QD
- Placebo to ustekinumab
- BMS-986165 6 mg QD to ustekinumab
- BMS-986165 12 mg QD to ustekinumab

Data from Part B will be summarized descriptively only. There will not be any formal statistical analysis of the Part B data.

6.1 Efficacy Analyses

All efficacy analyses will be performed using the FAS, unless otherwise specified.

All dose-response trend tests will be tested using a 2-sided 0.10 level of significance. All other analyses will also be tested using a 2-sided 0.10 level of significance, unless otherwise specified.

6.1.1 Primary Endpoint(s)

6.1.1.1 Primary Analysis

Analysis Model

The primary efficacy analysis model for the binary endpoint, ACR 20 (responder/non-responder), will use a logistic regression model on FAS to assess whether there is a dose-response trend between ACR 20 response and dose level. The model will include dose level (0, 6, 12) as a continuous variable, and the following covariates: TNFi use (experienced/naïve) and body weight (≥ 90 kg and < 90 kg) from IRT in the model. The slope coefficient in the logistic regression of ACR 20 vs. dose will tested as H_0: Beta = 0 vs H_1: Beta ≠ 0 at α=0.10 (2-sided).Adjusted odds ratio and the corresponding 2-sided 95% confidence intervals (CI) will be provided from a logistic regression with responder as the dependent variable and treatment, TNFi use (experienced/naïve), and (≥ 90 kg and < 90 kg) from IRT as fixed factors.

Imputation Methodology

The estimand of interest is to assess the dose-response of multiple doses of BMS-986165 on subject response to ACR 20 criteria at Week 16 for all randomized subjects receiving at least one dose of study treatment. The primary imputation method of non-responder imputation (NRI) will be used for subjects who have the below intercurrent events:

- Subjects who discontinue the treatment or study early (i.e. prior to Week 16) and have no ACR 20 assessments at Week 16
- Subjects who are lost to follow-up
- Subjects who otherwise have missing endpoint data at Week 16 or not sufficient for a definitive determination of response status will be classified as non-responders at that time point

6.1.1.2 Sensitivity Analyses

The following strategies for addressing intercurrent events will be used in sensitivity analyses of the primary efficacy endpoint:

Last Observation Carried Forward (LOCF)

1. The last observation while the subject was on blinded treatment will be considered as their Week 16 value. This is essentially the last observation carried forward (LOCF) imputation methodology.

Mixed LOCF and NRI

2. Placebo subjects with an intercurrent event defined in the Section 6.1.1.1 will be imputed using the LOCF method and BMS-986165 subjects will be imputed using the NRI method in this analysis.

6.1.1.3 Supportive Analyses

Per Protocol Population Analysis

The primary efficacy endpoint will be analyzed using the PPS using the primary analysis methodology.

6.1.2 Secondary Endpoint(s)

6.1.2.1 Binary Endpoints

Analysis Model

The secondary endpoint of PASI 75 response at Week 16 will be assessed in a dose-response analysis. The binary endpoint (responder/non-responder) will use a logistic regression analysis similar to the primary analysis. The model will include dose level (0, 6, 12) as a continuous variable, and the following covariates: TNFi use (experienced/naïve) and body weight (≥ 90 kg and < 90 kg) from IRT in the model. The slope

coefficient in the logistic regression of PASI 75 vs. dose will tested as H_0: Beta =0 vs H_1: Beta ≠ 0 at α=0.10 (2-sided).

Adjusted odds ratio and the corresponding 2-sided 95% confidence intervals (CI) will be provided from a logistic regression with responder as the dependent variable and treatment, TNFi use (experienced/naïve), and (≥ 90 kg and < 90 kg) from IRT as fixed factors.

Imputation Methodology
Similar strategies for addressing intercurrent events used in the primary efficacy endpoint will be used for binary secondary efficacy endpoints as described in Section 6.1.1.1.

6.1.2.2 Continuous Endpoints

The change from baseline score will be calculated at Week 16 and analyzed using an analysis of covariance (ANCOVA). Dose level (0, 6, 12) will be assessed in the model as a continuous variable and the following covariates as fixed effects: TNFi use (experienced/naïve) and body weight (< 90 kg and ≥ 90 kg) from IRT. The baseline value will be added into the model as a covariate. A general contrast-based test will be used to evaluate the dose-response trend.

Imputation Methodology

For continuous secondary efficacy endpoints, a modified baseline observation carried forward (mBOCF) approach will be used for missing data. The baseline observation will be carried forward for subjects who discontinue study treatment due to:

- Lack of efficacy
- AEs

The last valid observation will be carried forward for all other subjects with missing data. Subjects who discontinue study treatment for other reasons will have the last valid observation carried forward (including the baseline value as applicable).

6.1.3 Adjustment for Multiplicity

Statistical analysis of secondary endpoints will be performed in a hierarchical fashion. In order to preserve the overall Type I error rate, a fixed-sequence testing method will be implemented for the below specified endpoints. Testing order of secondary endpoints will be tested in the below order and may only proceed to the next secondary endpoint if the null hypothesis is rejected at alpha = 0.10 (2-sided) for the prior endpoint showing a positive dose-response in that endpoint. If an endpoint fails at any step, then all subsequent p-values will be considered nominal. Secondary endpoints will be tested in the following order:

a) Change from baseline in HAQ-DI score at Week 16
b) PASI 75 response at Week 16
c) Change from baseline in SF-36 PCS score at Week 16

There will be no multiplicity adjustment for testing of additional or exploratory endpoints and analyses. Nominal p-values will be provided as descriptive statistics only.

6.1.4 Subgroup Analyses

Subgroup analyses will be performed on the primary efficacy endpoint, ACR20 at Week 16, using the FAS. The primary imputation method will be applied for these analyses. The Cochran-Mantel-Haenszel (CMH) test will be the analysis method used where the stratification factor is the specified subgroup. The following subgroups that may be evaluated will be considered:

- Geographic region (EU [Czech Republic, Germany, Hungary, Spain, Poland], Non-EU [Russian Federation], and US)
- Country

- Sex (male, female)
- Age group (<40 y, 40-<65 y, ≥65 y)
- Body weight categories (<90 kg; ≥90 kg) – from case report form
- Race
- Baseline TNFi use history (experienced/naïve) – from case report form
- Baseline non-biologic DMARD use history (yes/no)
- Baseline DMARD use (yes/no)
- Baseline glucocorticosteroid use (yes/no)
- Baseline methotrexate use (yes/no)
- Baseline disease severity (DAS28 <5.1 vs DAS28 ≥5.1)
- Baseline number of swollen joints (≤4 vs >4)
- Baseline presence of enthesitis (yes/no)
- Baseline presence of dactylitis (yes/no)
- Baseline CRP (<10 mg/L vs ≥10 mg/L)
- Duration of disease (< median y, ≥ median y)

6.1.5 Additional Endpoints

6.1.5.1 Binary Endpoints

Analysis Model

Additional endpoints defined in the Section 4.1.3 at Week 16 will be assessed in a dose-response analysis. The binary endpoint (responder/non-responder) will use a logistic regression analysis similar to the primary and secondary endpoint analyses. The model will include dose level (0, 6, 12) as a continuous variable, and the following covariates: TNFi use (experienced/naïve), and body weight (≥ 90 kg and < 90 kg) from IRT in the model. The slope coefficient in the logistic regression of endpoint vs. dose will tested as H_0: Beta =0 vs H_1: Beta ≠ 0 at α=0.10 (2-sided).

Adjusted odds ratio and the corresponding 2-sided 95% confidence intervals (CI) will be provided from a logistic regression with responder as the dependent variable and treatment, TNFi use (experienced/naïve), and (≥ 90 kg and < 90 kg) from IRT as fixed factors.

Imputation Methodology

Similar strategies used in the binary secondary efficacy endpoints will be used for the additional endpoints as specified in Section 6.1.1.1.

6.1.5.2 Continuous Endpoints

Analysis Model

The change from baseline score will be calculated at Week 16 and analyzed using an analysis of covariance (ANCOVA). Dose level (0, 6, 12) will be assessed in the model as a continuous variable and the following covariates as fixed effects: TNFi use (experienced/naïve) and body weight (< 90 kg and ≥ 90 kg) from IRT. The baseline value will be added into the model as a covariate. A general contrast-based test will be used to evaluate the dose-response trend.

Imputation Methodology

Similar strategies used in the continuous secondary efficacy endpoints will be used for the additional continuous endpoints as specified in Section 6.1.2.2.

[redacted]

6.2 Safety

Summaries of safety data will be presented by period (Part A and B) and treatment group, as applicable, for the As-treated population.

During the first 16 weeks of treatment, data will be presented for the following treatments:

- BMS-986165 6 mg QD
- BMS-986165 12 mg QD
- Placebo

After Week 16, subjects continuing in Part B of the study will have their data presented depending on MDA response in Part A, as follows:

- BMS-986165 6 mg QD → BMS-986165 6 mg QD
- BMS-986165 12 mg QD → BMS-986165 12 mg QD

- Placebo → ustekinumab injection
- BMS-986165 6 mg QD → ustekinumab injection
- BMS-986165 12 mg QD → ustekinumab injection

6.2.1 Adverse Events

Adverse events (AE) will be presented for the number and percentage of subjects and the number of events. All adverse events (treatment-emergent [TEAE] and non-treatment emergent) will be provided in listings. Summary tables will be reported in decreasing frequency based on the total column. Counting for frequency analysis will be by subject and not by AEs, and subjects are only counted once for recurring AEs within each system organ class (SOC) or preferred term (PT); however, the number of total AEs per subject, including multiple occurrences of individual AEs, will also be presented. For summaries of severity, AEs will be reported only at their maximum severity in each subject. Subjects with multiple AEs within a particular SOC or PT will be counted under the category of their most severe event within that SOC or PT.

AE (including deaths) dates will be imputed according to algorithms detailed in Section 8.0. The imputed date of AE onset will be used to assess whether AEs should be considered as treatment-emergent and included in the safety summaries. The original, partial dates will be included in data listings. No imputation will be performed on missing AE seriousness, severity, or relationship; they will be reported as missing.

AEs will be included in a period if the start date of the AE is after the first dispensation date within a period.

An overall summary for the following categories will be presented:

- Deaths
- SAEs
- Related SAEs
- AEs
- Related AEs
- Discontinued treatment due to AEs

The following summaries will also be provided for the following:

- TEAEs by SOC and PT
- Treatment-related TEAEs by SOC and PT
- TEAEs by PT reported in ≥ 5% of subjects
- Treatment-related TEAEs by PT reported in ≥ 5% of subjects
- TEAEs categorized by severity by SOC and PT
- Exposure-adjusted incidence rate (EAIR) for TEAEs by SOC and PT – EAIR is defined in Section 8.1 of the SAP. EAIR will be reported for only BMS-986165 group in Part B.

A summary of TEAEs leading to discontinuation of study treatment or study will be provided, grouped by SOC and PT for all TEAEs and treatment-related TEAEs.

6.2.1.1 Adverse Events of Interest (AEI) and Other Important AEs

Summaries for treatment-emergent AEIs events will be provided by SOC and PT for each AEI category:

- Skin-related events
- Infection events
- Malignancy

Creatine kinase (CK) elevation for CK elevation >2.5 upper limit of normal will be summarized as CTCAE grade 2 in the clinical laboratory summaries.

Additional information collected as part of the clinical safety program and adjudicated events will also be summarized.

6.2.1.2 Serious Adverse Events

Summaries for treatment-emergent SAEs will be provided for the following:

- Treatment-emergent SAEs by SOC and PT
- Treatment-related treatment-emergent SAEs by SOC and PT

6.2.1.3 Adverse Events Leading to Discontinuation of Study Treatment

Summaries for TEAEs leading to discontinuation of study treatment will be provided for the following:

- TEAEs by SOC and PT
- Treatment-related TEAEs by SOC and PT

6.2.2 Deaths

All adverse events with an outcome of death will be listed.

6.2.3 Clinical Laboratory Data

Laboratory parameters will be summarized using the International System (SI) of Units and US conventional units. Data will be summarized by time point, as applicable. The following summaries will be provided for each parameter:

- Absolute and change from baseline values for continuous parameters
- Number and percentage of subjects for the following:
    - Maximum postbaseline CTCAE grade for each applicable laboratory parameter
    - Shifts from baseline based on maximum postbaseline CTCAE grade
- Drug-induced Liver Injury (DILI) and Hy's Law summaries

All laboratory data specified in the summary tables will be present in listings.

6.2.4 Vital Signs and Physical Findings

Vital signs, including weight, will be summarized by time point, as applicable. The following summaries will be provided for each parameter:

- Absolute and change from baseline values
- Number and percentage of subjects for the following:
    - Marked abnormality for each abnormality category as defined in Section 4.2

Physical examination results and other safety data will be listed.

6.2.5 ECGs

ECG parameters will be summarized by time point, as applicable. The following summaries will be provided for each parameter:

- Absolute and change from baseline values
- Number and percentage of subjects for the following:
    - Marked abnormality for each abnormality category as defined in Section 4.2

6.3 General Methodology

The following methods/standards will be used:

- Statistical package(s) planned to be used
  - All analyses will use SAS version 9.4 or higher.
- Descriptive summaries will be tabulated by period treatment group and time point, unless otherwise specified in the TFL shells.
- Standard summary statistics for continuous and categorical variables:
  - Unless otherwise noted, categorical variables will be summarized using counts and percentages, with the number of subjects in each category as the denominator for percentages. Percentages will be rounded to one decimal place, except 100% will be displayed without any decimal places and percentages will not be displayed for zero counts.
  - Continuous variables will be summarized using the number of observations (n), mean, standard deviation (SD), median, minimum and maximum. The median, minimum and maximum values will be displayed to the same level of precision as the raw data, the mean to a further decimal place and the SD to two additional decimal places.
- P-values will be presented to 4 decimal places where presented in the TFL shells.

6.3.1 Subject Populations and Disposition

The number of subjects enrolled/screened and the number and percentage of subjects randomized, treated, and in each analysis population will be presented.

The number and percentage of subjects randomized in each region and center will be presented by Part A and B.

Additionally, the following summaries will be provided for the FAS by Period (Part A and B), treatment group and overall:

- The number and percentage of subjects who completed the treatment period, those who withdrew from study treatment prematurely, and a breakdown of the corresponding reasons for withdrawal from study treatment will be presented.
- The number and percentage of subjects who completed the study, those who withdrew from the study prematurely, and a breakdown of the corresponding reasons for withdrawal will be presented. Denominator will be the number of subjects in the FAS.

6.3.2 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized by treatment group for the FAS. Demographic characteristics include the following:

- Gender (male, female)
- Race
- Ethnicity
- Age (in years, at time of signing informed consent) and age category (<40, 40-<65, ≥65)
- Weight (in kg, at baseline) and weight category (≥90 kg, <90 kg)
- Body mass index (BMI in kg/m^2, at baseline)
- Geographic region (EU, Non-EU, US)
- Prior biologic use for psoriatic arthritis (yes, no)
- Reason for discontinuation of prior biologic use
- Stratification factors obtained from IRT: TNFi use (experienced/naïve), and body weight (<90 kg and ≥90 kg)
- Stratification factors obtained from database: TNFi use (experienced/naïve), and body weight (< 90 kg and ≥90 kg)
- Duration of disease (<median y, ≥ median y)

- PsA phenotype - oligoarticular (≤4 joints), polyarticular (>4 joints), arthritis with predominant associated spinal symptoms
- Age at disease onset (<18, 18-39, ≥40)

Additional demographics or baseline data may be added to summary tables.

General medical history and medical history related to PsA will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) (version 21.0 or an updated version at the time of database lock). General medical history data will be summarized for each SOC and PT by treatment group and overall for the FAS population. Separate tables will be provided for PsA medical history.

[redacted]

6.3.4 Exposure

6.3.4.1 Duration of Treatment

Duration by Group

Overall duration of each treatment, in days, will be calculated for each subject in each treatment group. Subjects are dispensed study treatment at each visit, starting with Day 1 (Week 0) and at each subsequent visit. The date of first dose of study treatment is the Week 0 PK dosing date and is recorded on the eCRF. If this date is missing, then the earliest drug dispensation date will be used. The last dose date is recorded on the end of treatment page in the eCRF. If this date is missing, then the latest date from the PK exposure page or drug accountability return date will be used.

Duration of treatment will be summarized descriptively for Part A and B, by treatment group.

BMS-986165:

Subjects randomized to BMS-986165 will have their duration of treatment derived as:

- (Date of last dose – date of first dose +1)

Placebo:

For subjects randomized to placebo, duration is defined as:

- Placebo = (date of last dose of placebo in Part A – date of first dose +1)

Ustekinumab:

For subjects who participate in Part B and are allocated to receive ustekinumab, duration is defined as:

- Ustekinumab = (date of last dose of ustekinumab in Part B – date of first dose in Part B +1)

Duration by Period

Overall duration (in days) of each treatment received within each study period will be calculated. Duration within each period is defined for each treatment group as:

$$(Last\ dose\ date - first\ dose\ date + 1)$$

The date of first dose of study treatment for Part A is as described above. The date of first dose of study treatment for Part B is defined as the earliest drug dispensation date for Part B. The last dose date is as described above.

6.3.4.2 Summary of Dosing

The number of doses taken for each subject for each period will be determined using the eCRF drug accountability data and is defined as:

$$Doses\ Taken\ (BMS-986165) = (number\ of\ tablets\ dispensed - number\ of\ tablets\ returned)$$

$$Doses\ (\text{ustekinumab}) = (number\ of\ injection\ dispensed - number\ of\ injection\ returned)$$

The number of doses taken will be summarized descriptively by treatment group within each study period and overall.

6.3.4.3 Compliance

Treatment compliance will be determined from data captured on the Drug Accountability eCRF.

BMS-98615

$$Number\ of\ expected\ doses: (date\ of\ next\ visit - date\ of\ current\ visit)x2$$

$$Number\ of\ doses\ taken: Number\ of\ doses\ dispensed - number\ of\ doses\ returned$$

Treatment compliance will be derived for each period and overall. Compliance is defined as:

$$\left(\frac{Number\ of\ doses\ taken}{Number\ of\ expected\ doses}\right)x100$$

Ustekinumab

$$Number\ of\ expected\ injection: (date\ of\ next\ visit - date\ of\ current\ visit)$$

Treatment compliance will be derived for each period and overall. Compliance is defined as:

$$\left(\frac{Number\ of\ injections}{Number\ of\ expected\ injections}\right)x100$$

Period compliance will be calculated by summing over all visits within the period and overall compliance will be calculated by summing each visit in the study and summarized using descriptive statistics by treatment group. The number and percentage of subjects with <75%, 75% to 100%, and >100% compliance will be provided by treatment group for each period and overall. If a subject does not return the container, then this dispensation event will be excluded in the calculation for total number of doses taken, total number of expected doses, and total compliance.

[redacted]

7.0 Sequence of Planned Analyses

7.1 Interim Analyses

No interim analysis is planned for this study.

7.2 Week 16 Endpoint Analyses

A database lock (16-week database lock) will occur once all randomized subjects have completed the Week 16 efficacy assessments or have discontinued prior to Week 16. Analyses of the collected efficacy and safety results during the 16-week Part A will be performed in order to aid in planning for subsequent clinical development. The study team responsible for managing the study and are site-facing, including medical monitors, will remain blinded to treatment assignment and the results of this analysis throughout the study.

7.3 Final Analyses and Reporting

All final, planned analyses identified in this statistical analysis plan will be performed only after the last subject has completed the study and the database has been locked. Investigative site staff and subject will remain blinded to treatment assignment until the database has been locked..
Any exploratory analyses completed to support study analyses in the Clinical Study Report (CSR), which were not identified in the statistical analysis plan, will be documented as such in the CSR.

8.0 **Conventions**

8.1 General Definitions

Term	**Definition**
Study Day	Study day is calculated as: assessment date – date of first dose + 1
Baseline	Unless otherwise stated, Baseline is defined as the last measurement prior to dosing on Day 1 (Week 0). If the measurement on Day 1 is missing or not available, then a prior measurement during the screening period may be used as baseline. Baseline assessments must be performed per protocol and standard of care assessments may not be used for baseline.
Change from Baseline	Change from baseline is defined as (value at post-baseline visit – value at baseline).
Change in the maximum post-baseline value	Change from baseline in the maximum post-baseline value is defined as highest observed value post-baseline. The change is calculated using this value as the post-baseline value.
Concomitant and Prior Medication	Prior medications are defined as medications with a start or stop date prior to the first dose of study treatment. Concomitant medications are defined as any medications ongoing at the start of study treatment or with a start date on or after the first dose date.
End of Study (EOS) Date	The EOS date is the date recorded on the eCRF that a randomized subject either discontinued or completed the study. If the subject is lost to follow-up, the EOS date will be the date of the last visit assessment obtained.
Exposure-adjusted incidence rate (EAIR)	EAIR = 100*The number of subjects with a specific event divided by the total exposure-time (in years) among the subjects in the treatment group, where the total exposure time in years is calculated by dividing the sum of exposure time in days over all subjects included in the treatment group*365.25.
First Dose Date – Study	The date a subject received their first dose on Day 1 as recorded in the eCRF Week 0 PK dosing date or the earliest drug dispensation date.
Last Dose Date – Study	The date of last recorded dose on the eCRF for a randomized subject.
First Dose Date – Period	The date a subject received their first dose in the defined study period as recorded in the eCRF Week 0 PK dosing date or the earliest drug dispensation date for Part A and earliest drug dispensation date for Part B.

Last Dose Date – Period	The last dose date is the date recorded on the end of treatment page in the eCRF. If this date is missing, then the latest date from the PK exposure page or drug accountability return date will be used.
Percent Change from Baseline	Percent change from baseline is defined as ([value at post-baseline visit – value at baseline]/value at baseline) x 100. If the baseline value is 0 and the post-baseline value is also 0, then the percent change from baseline is set to 0. If the baseline value is 0 and the post-baseline value is > 0, then the percent change from baseline will be missing.

8.2 Calculation of Key Measures

The following efficacy assessments will be used to assess subjects' disease activity and severity during the study. The planned analyses are defined in the Appendix 1. The specific analysis performed for each of assessment are defined in the Appendix 2. Outcomes are reported via an eCOA tool at various times throughout the study as described in the protocol Schedule of Activities. At study visits, assessments by the investigator or subjects and results/responses will be reported directly into the eCOA tool at the time of the visit. The tool will open assessments in a sequential manner, meaning that the full assessment is to be completed prior to moving forward to the next assessment. This limits the possibility of partially missing data. Also, as investigators/subjects are prompted to enter data for each assessment for the visit, the possibility of a full assessment being missing is also minimized.

Scoring of assessments where validated algorithms are not required will be derived in SAS datasets.

Scoring of assessments where validated scoring tools are required, licenses for these tools will be purchased and used for scoring prior to incorporating into the SAS datasets.

8.2.1 Investigator-Administered Assessments

Assessments will be performed by a qualified physician or rheumatologist or dermatologist or trained designee who is experienced in the assessment of psoriasis patients. To limit variability, every effort will be made so that the same individual conducts the assessment at all subsequent visits.

8.2.1.1 American College of Rheumatology (ACR) Improvement Criteria

The ACR 20, ACR 50 or ACR 70 definition of improvement is a 20%, 50% or 70% improvement, respectively, over baseline in tender and swollen joint counts and a 20%, 50% or 70% improvement, respectively, in 3 of the 5 remaining core data set measures:

≥ 20%, 50% or 70% improvement from baseline in the number of tender joints (68 joint count)

≥ 20%, 50% or 70% improvement from baseline in the number of swollen joints (66 joint count)

≥ 20%, 50% or 70% improvement from baseline in 3 of the following 5 domains:

- Subject Global Assessment of disease activity
- Physician Global Assessment of psoriatic arthritis
- Subject global assessment of pain (APPENDIX 8 of the protocol)
- Health Assessment Questionnaire-Disability Index (HAQ-DI, APPENDIX 6 of the protocol)
- hsCRP

8.2.1.2 Psoriasis Area and Severity Index (PASI)

The PASI is a measure of the average redness, thickness, and scaliness of psoriatic skin lesions (each graded on a 0–4 scale; 0 = none to 4 = very severe), weighted by the area of involvement (head, arms,

trunk to groin, and legs to top of buttocks). The PASI produces a numeric score that can range from 0 to 72, with higher PASI scores denoting more severe disease activity. The PASI includes multiple subscores and a final total score that will be provided by the eCOA system. Individual plaque characteristic rating scores are provided for each body region as well as the weighted score. Additionally, the degree of involvement of each body region is assessed and that score is multiplied by the weighted plaque characteristic score for a final score for each body region. The total PASI score is a sum of the 4 body regions: Head, Upper Extremities, Trunk and Lower Extremities.

The PASI Total score will be used to assess response to treatment. The percent change from baseline will be calculated at each visit. The PASI 75 endpoint is the proportion of subjects who experience at least a 75% improvement in PASI score as compared with the baseline value.

1 = $If\ \left(\frac{Baseline\ PASI - Visit\ PASI}{Baseline\ PASI}\right) x100 \geq 75\ then\ subject\ is\ a\ PASI\ 75\ responder$

0 = otherwise

The PASI 90 is defined similarly. The endpoint derivations will be performed in the analysis datasets.

8.2.1.3 Disease Activity Score (DAS) 28 CRP

A Disease Activity Score (DAS) can be used to assess patients' PsA disease activity, to determine whether it is under control, and if any treatment adjustments are required. It can also assist in establishing a target score to aim for, to help inform treatment decisions, and optimize disease management.

DAS 28 CRP is a composite outcome measure that assesses:

- How many joints in the hands (including metacarpophalangeal and proximal interphalangeal joints, but excluding distal interphalangeal joints), wrists, elbows, shoulders, and knees are swollen and/or tender over a total of 28 joints
- CRP in the blood to measure the degree of inflammation
- Subject Global Assessment of disease activity

The results are combined to produce the DAS 28 CRP score, which correlates with the extent of disease activity:

- < 2.6: Disease remission
- 2.6 – 3.2: Low disease activity (LDA)
- 3.2 – 5.1: Moderate disease activity
- >5.1: High disease activity

The following formula will be used to compute the DAS28-CRP score;

DAS28-CRP = 0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*ln(CRP+1) + 0.014*GH + 0.96.

8.2.1.4 Dactylitis

The number of digits in hands and feet with dactylitis will be counted by a blinded assessor. The LDI Basic is a quantitative measurement of dactylitis in the 20 digits using a dactylometer. The circumference of the affected and contralateral digits, and tenderness of the affected digits are measured to generate a total score. A higher score indicates worse dactylitis and is based on the current evaluation. Training will be provided so dactylitis will be evaluated in a consistent manner throughout the study.

LDI measures the ratio of the circumference of the affected digit to the circumference of the digit on the opposite hand or foot. If both sides are considered involved, a table of normative values is used to provide the comparison. The ratio of circumference is multiplied by a tenderness score (0=nontender, 1=tender). For each dactylitic digit, the final score is defined as:

[{(A/B) – 1}*100]*C,

Where A is circumference of involved digit, B is circumference of opposite, unaffected, digit or reference, and C is tenderness (0 or 1). The total score is determined by summing over all digits.

8.2.1.5 Enthesitis

The number of sites with enthesitis will be evaluated by a blinded assessor using the Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis index and the Leeds Enthesitis index (LEI).

The SPARCC Enthesitis Index has a 0 to 16 score that is derived from the evaluation of 8 locations: the greater trochanter (R/L), quadriceps tendon insertion into the patella (R/L), patellar ligament insertion into the patella and tibial tuberosity (R/L), Achilles tendon insertion (R/L), plantar fascia insertion (R/L), medial and lateral epicondyles (R/L), and the supraspinatus insertion (R/L). A higher count indicates a higher enthesitis burden based on the current evaluation. It has been validated in ankylosing spondylitis and correlates well with the LEI.

The LEI was developed specifically for PsA. An overall score of 0 to 6 is derived from the presence or absence of tenderness at six entheseal sites (right and left: lateral epicondyle, medial femoral condyle, and Achilles tendon insertion) at the time of evaluation. A higher count indicates a greater enthesitis burden.

8.2.1.6 Physician's Global Assessment-Fingernails (PGA-F)

In this assessment, fingernail psoriasis is evaluated. The PGA-F will be performed at baseline. If a subject shows evidence of psoriatic fingernail involvement, the assessment will be performed at each subsequent visit to assess severity and improvement over time. Only subjects with PGA-F at baseline will be assessed throughout the study. The overall condition of the fingernails is rated on a 5-point scale:

0 = clear, 1 = minimal, 2 = mild, 3 = moderate, and 4 = severe

The rating score will be collected in the eCOA system.

8.2.1.7 Minimal Disease Activity (MDA) and Very Low Disease Activity (VLDA)

Minimal Disease Activity (MDA) response is where an MDA responder is defined as a subject fulfilling 5 of 7 of the following outcomes:

- tender joint count ≤ 1
- swollen joint count ≤ 1
- PASI ≤ 1 or body surface area (BSA) ≤ 3%
- Subject Global Assessment of pain ≤ 15 (APPENDIX 8)
- Subject Global Assessment of disease activity ≤ 20 (APPENDIX 7)
- HAQ-DI ≤ 0.5
- Tender entheseal points ≤ 1

Very low disease activity (VLDA) response is where a VLDA responder is defined as a subject fulfilling all 7 of 7 of the MDA outcomes.

8.2.1.8 Psoriatic Arthritis Disease Activity Score (PASDAS)

A composite measure calculated from the Physician Global Assessment of psoriatic arthritis, the Subject Global Assessment of disease activity, the Short Form Health Survey-36 Item (SF-36) Physical Component Summary (PCS), the swollen joint count, the tender joint count, the LEI, the LDI (Basic), and the hsCRP. The PASDAS is derived using the following formula:

PASDAS = (((0.18*sqrt(Physician's Global Assessment of psoriatic arthritis) + (0.159*sqrt(Subject Global Assessment of disease activity) – (0.253*sqrt(SF36-PCS) + (0.101*LN(swollen joint count + 1)) + (0.048*LN(tender joint count + 1)) + (0.23*LN(Leeds enthesitis count + 1)) + (0.37*LN(tender dactylitis count + 1)) + (0.102*LN(CRP + 1)) + 2)*1.5.

8.2.1.9 PsA Response Criteria (PsARC)

The Psoriatic Arthritis Response Criteria (PsARC) consists of 4 measurements: tender/painful joint count, swollen joint count, Physician Global Assessment of psoriatic arthritis, and Subject Global Assessment of psoriatic arthritis.

In order to be classified as 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. Improvement in each of the measures is defined below:

- Decrease of ≥ 30% in tender joint counts
- Decrease of ≥ 30% in swollen joint counts
- Decrease of ≥ 20% in Physician's Global Assessment of Psoriatic Arthritis
- Decrease of ≥ 20% in Subject's Global Assessment of Psoriatic Arthritis

8.2.1.10 Disease Activity Index for Psoriatic Arthritis Score (DAPSA)

The Disease Activity Index for Psoriatic Arthritis Score (DAPSA) is a composite measure to assess peripheral joint involvement that is based upon numerical summation of 5 variables of disease activity: tender/painful joint count, swollen joint count, Subject Global Assessment of disease activity, Subject Global Assessment of pain, and CRP. The DAPSA is derived using the following formula:

DAPSA = tender joint count + swollen joint count + CRP + Subject Global Assessment of disease activity + Subject Global Assessment of pain.

8.2.2 Subject-Reported Assessments

8.2.2.1 Health Assessment Questionnaire-Disability Index (HAQ-DI)

The HAQ-DI is a patient-reported outcome measure that 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 consists of 2 to 3 items. For reach item in the questionnaire the level of activity is cored from 0 to 3 with 0 representing "no difficulty", 1 as "some difficulty", 2 as much "difficulty", and 3 has "unable to do". Any activity that requires assistance from another individual or requires the use of an assistive device adjusts to a minimum score of 2 to represent a more limited functional status.

Scoring for each functional category and the disability index will be defined as follows:

- Dressing and Grooming includes items 1 and 2
- Arising includes items 3 and 4
- Eating includes items 5, 6, and 7
- Walking includes items 8 and 9
- Hygiene includes items 10, 11, and 12
- Reach includes items 13 and 14
- Grip includes items 15, 16, and 17
- Activities includes items 18, 19, and 20

The score for each functional category will be the highest score within the individual item scores. If any aids, devices, or help from others is used, then an item score is adjusted up to a "2" if less than 2. The disability index will be the mean of the eight functional scores. If more than two of categories, or 25%, are missing, the index will not be scored. Otherwise, divide the sum of the categories by the number of answered categories.

8.2.2.2 Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

In subjects with baseline evidence of PsA spondylitis, symptoms will be evaluated using the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), which consists of a 0 to 100 scale measuring discomfort, pain, and fatigue in response to 6 questions pertaining to the 5 major symptoms of spinal involvement originally developed for ankylosing spondylitis:

- Fatigue (medical)
- Spinal pain
- Joint pain and swelling
- Areas of localized tenderness
- Morning stiffness duration
- Morning stiffness severity

A higher count indicates worse disease. The recall period is one week. Each individual question response is scaled to a 0-10 score by dividing by 10 and the BASDAI is derived using the following formula:

BASDAI = ((Q1 + Q2 + Q3 + Q4) + ((Q5 + Q6) / 2)) / 5

A scale of 0 – 50 is possible for the BASDAI score.

8.2.2.3 Short Form Health Survey-36 Item (SF-36)

The SF-36 v2.0 was designed as an indicator of health status in population surveys, health policy evaluations, and for use as an outcome measure in clinical practice and research. The instrument includes 36 items in a Likert-type format to measure the following 8 health dimensions: 1) limitations in physical activities, such as bathing or dressing; 2) limitations in social activities because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4) bodily pain; 5) general mental health (psychological distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions. Scores for each domain range from 0 to 100, with higher scores indicating a better status.

The 8 health dimensions assessed by SF-36 questionnaire are grouped into 2 main components, physical and mental. Each of the 8 dimensions contribute in different proportions to both the physical component summary (PCS) score and the mental component summary (MCS) score. Individual question responses are collected in the eCOA system. Scoring of the total score, PCS and MCS will be performed via access to validated software and scores will be provided in the analysis datasets for analysis purposes.

8.2.2.4 Patient Reported Outcome Measures Information System-Fatigue (PROMIS-FATIGUE))

The Patient-Reported Outcome Measures Information System-Fatigue (PROMIS-Fatigue) and Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) instruments evaluate a range of self-reported symptoms over the past week, from mild subjective feelings of tiredness to an overwhelming, debilitating, and sustained sense of exhaustion that likely decreases one's ability to execute daily activities and function normally in family or social roles. Fatigue is divided into the experience of fatigue (frequency, duration, and intensity) and the impact of fatigue on physical, mental, and social activities. The first 13 questions of the PROMIS-Fatigue questionnaire comprise the FACIT-Fatigue questionnaire.

The PROMIS-Fatigue questionnaire is not yet available in all languages. The PROMIS-Fatigue questionnaire will be administered at sites where it is available in the local language. Sites in countries where the PROMIS-Fatigue questionnaire is not available in the local language will start the protocol using the FACIT-Fatigue questionnaire. Once the PROMIS-Fatigue questionnaire is available in the local language and relevant approvals from ethics committees are received, any new enrolled subjects will complete the PROMIS-Fatigue questionnaire. Participants who were administered the FACIT-Fatigue at Day 1 will continue to be administered the FACIT-Fatigue throughout the study.

PROMIS-Fatigue score will be summarized by treatment and time point for subjects who were administered PROMIS Fatigue. FACIT-Fatigue scores will be similarly presented as a pooled analysis, which will include FACIT-Fatigue scores combined with scores from the first 13 questions of PROMIS-Fatigue.

8.2.2.5 Work Limitation Questionnaire (WLQ)

The WLQ is a 8-item self-report that measures the on-the-job impact of chronic health conditions and treatment with a focus on assessing limitations while performing specific job demands from the following 4 scales:

1) Time management: difficulty with handling time and scheduling demands (2 items)
2) Physical demands: ability to perform job tasks that involve bodily strength, movement, endurance, coordination, and flexibility (2 items)
3) Mental-interpersonal demands: cognitively demanding tasks and on-the-job social interactions (2 items)
4) Output demands: concerns reduced work productivity (2 items)

The questionnaire includes an option for 'Does not apply to my job". Any items with this response will be given a score of 'missing' for scoring purposes. Scoring will be performed in the analysis datasets. Individual scores to each of the 8 items will be provided by the eCOA system. Each item is scored from 1 to 5. A score is interpreted as: 1=0%, 2=25%, 3=50%, 4=75% and 5=100%. Scores range from 0 to 100%, where 0% = health makes the job demand difficult none of the time and 100% = health makes the job demand difficult all of the time.

Each scale is derived as a mean of the items within the scale. If half or more of the items within the scale are missing, the scale cannot be computed and will be considered missing.

$$Scale = \frac{item_1 + item_2 + \cdots + item_n}{n}, where\ n\ is\ the\ number\ of\ items\ within\ the\ scale$$

The scale scores of the WLQ are derived as a mean of the 25-item scores minus 1 and then multiplied by 25.

If all 4 scale scores are non-missing, then the WLQ Index is calculated using the following formula:

WLQ Index = (0.00048*WLQ Time Management Scale + 0.00036*WLQ Physical Tasks Scale + 0.00096*WLQ Mental-Interpersonal Tasks Scale + 0.00106*WLQ Output Tasks Scale).

The WLQ At-Work Productivity Loss score is calculated using the following formula:

WLQ At-Work Productivity Loss Score = (1-exp(-WLQ Index))*100.

8.2.2.6 Psoriatic Arthritis Impact of Disease (PsAID)

The Psoriatic Arthritis Impact of Disease (PsAID) is a 12-item self-report that measures psoriatic arthritis symptoms and impact of disease. Each item is scored on a 0 to 10 numeric rating scale with a one week recall period. Each item is scaled by a factor and the total score is calculated by summing over the scaled scores, with a higher value indicating worse health.

8.3 Missing, Unknown, or Partial Dates

Start Date		Stop Date: Complete: *yyyymmdd* <1st dose	Complete: *yyyymmdd* ≥1st dose	Partial: *yyyymm* <1st dose *yyyymm*	Partial: *yyyymm* ≥1st dose *yyyymm*	Partial: *yyyy* <1st dose *yyyy*	Partial: *yyyy* ≥1st dose *yyyy*	Missing/ Ongoing
Partial: *yyyymm*	= 1st dose *yyyymm*	2	1	n/a	1	n/a	1	1
	≠ 1st dose *yyyymm*		2	2	2	2	2	2

Partial: *yyyy*	= 1st dose *yyyy*	3	1	3	1	n/a	1	1
	≠ 1st dose *yyyy*		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1 = Impute as the date of first dose
2 = Impute as the first of the month
3 = Impute as January 1 of the year
4 = Impute as January 1 of the stop year
Note: If the start date imputation leads to a start date that is after the stop date, then there is a data error and the start date is not imputed.

Imputation rules for partial or missing stop dates:

1. Initial imputation
    a. For partial stop date "mmyyyy", impute the last of the month.
    b. For partial stop date "yyyy", impute December 31 of the year.
    c. For completely missing stop date, do not impute.
2. If the stop date imputation leads to a stop date that is after the death date, then impute the stop date as the death date.
3. If the stop date imputation leads to a stop date that is before the start date, then there is a data error and the stop date is not imputed.

Imputation rules for partial or missing death dates:

1. If death year and month are available but day is missing:
    a. If "mmyyyy" for last contact date = "mmyyyy" for death date, set death date to the day after the last contact date.
    b. If "mmyyyy" for last contact date < "mmyyyy" for death date, set death date to the first day of the death month.
    c. If "mmyyyy" for last contact date > "mmyyyy" for death date, data error and do not impute.

If both month and day are missing for death date or a death date is totally missing, set death date to the day after the last contact date.

8.4 Study Periods

Part A = (Week 0 to Week 16 visit)

Part B = Optional (Week 16 to Week 52 visit)

Follow-up = 4 week follow-up period

8.5 Day Ranges for Analysis Visits

Below are the day ranges for the analysis visit definitions. If more than one visit occurs within an analysis visit, then the visit that is closest to the target date should be used for analysis.

Period **Week**	**Target Day**	**Day Range**
Baseline		Screening, 1
Part A		
Week 1	8	2, 11
Week 2	15	12, 18
Week 4	29	19, 43
Week 8	57	44, 71
Week 12	85	72, 99
Week 16	113	100, 127 (or Week 16 drug dispense date)
Part B		

Period Week	Target Day	Day Range
Week 20	141	1st day after Week 16 drug dispense date, 155
Week 24	169	156, 183 (or Week 24 drug dispense date if earlier)
Week 28	197	1st day after Week 24 drug dispense date, 211
Week 32	225	212, 239
Week 36	253	240, 267
Week 40	281	268, 295
Week 44	309	296, 323
Week 48	337	324, 351
Week 52	365	352, last visit date prior to Safety Follow-up
Safety Follow-up	393	Safety Follow-up visit

9.0 [redacted]

10.0 Document History

Version Number	Version Date	Summary of Changes
1.0 - Original Document	09-Sep-2020	Not applicable
2.0 – Amendment 01	05-Jun-2020	Revisions provided below.

Revisions for Amendment 01: In addition to the revisions specified below, there were some minor typographical and formatting changes made. Additions are noted by bold text. Removals are noted by strikethrough.

SAP Section	Revised Text	[redacted]
1.0 Purpose	This version of the plan has been developed using the protocol v2.0 and CRF dated **05FEB2020**.	
2.4 Unblinding Information	**Additionally, bioanalytical scientists involved in the processing of bioanalytical samples will be unblinded to randomized treatment assignments to minimize unnecessary sample bioanalysis of subjects who are on placebo.**	
2.5 Changes in Statistical Considerations from the Protocol	~~No changes.~~ **The Full Analysis Set was changed from "all randomized subjects who are dispensed study drug" in the protocol to "all subjects who were randomized".** **The Safety Set (All Treated) population was changed to the As-treated population to align with the other studies for this compound.** **The initial list of relevant protocol deviations provided in the protocol was updated for determination of the Per Protocol population.** **The RAPID3 endpoint was removed from the additional endpoints as this endpoint requires a different version of the Health Assessment Questionnaire than was used in IM011-084.** **Imputation methods were updated to remove the prohibited medication/ therapy criteria for binary and continuous endpoints.**	

		[redacted]
4.1.3 Additional Endpoints	• Dactylitis mean change from baseline (dactylitis count) **in subjects with dactylitis at baseline** • Enthesitis mean change from baseline **using the Leeds Enthesitis Index (LEI)** • Enthesitis resolution, where resolution is defined as a**n** ~~Leeds Enthesitis Index (~~LEI~~)~~ score of 0, in subjects with LEI ≥ 1 at baseline • Disease Activity Score (DAS) 28 CRP response: o Low Disease Activity (LDA) defined as DAS 28 CRP score ≦ **<** 3.2 • Change from baseline in Work Limitation Questionnaire (WLQ) **At-Work Productivity Loss** score • ~~Change from baseline in routine assessment of patient index data 3 (RAPID3) score~~	[redacted]
4.2 Safety	• ~~New nonserious~~**Adverse events (AEs)** which ~~first~~ occur after the first dose of study treatment through 30 days after the final dose of the study treatment or subject's participation in the study if the last scheduled visit occurs at a later time~~; New serious adverse events (SAEs) which first occur after the first dose of study treatment through 30 days after the final dose of the study treatment or subject's participation in the study if the last scheduled visit occurs at a later time;~~ ▪ ~~SAEs reported prior to first dose of study treatment that increase in severity or frequency after first dose of study treatment through 30 days after the final dose of the study treatment or subject's~~	[redacted]

	~~participation in the study if the last scheduled visit occurs at a later time~~. ○ Treatment-emergent adverse events of interest (AEIs) ~~as determined through the Clinical Safety Program (CSP)~~ for the following events: ▪ Skin-related AEs ▪ Infection AEs, including influenza ~~▪ Creatine kinase (CK) elevation~~ ▪ Malignancy ~~▪ Cardiovascular~~ ○ Laboratory abnormalities (as determined by Common Terminology Criteria for Adverse Events [CTCAE **v5.0**] grading) ≥ 500 **msec** **Males: < 450 msec, ≥ 450 msec** **Females: < 470 msec, ≥ 470 msec** **In cases where the QT interval is corrected using Bazett's (QTcB) formula is captured instead of QTcF, then QTcB will be converted to QTcF for analyses.**	[redacted]
5.0 Populations for Analysis	All **subjects who were** randomized ~~subjects who are assigned study treatment~~. ~~**Safety Set (All Treated)**~~**As-treated Population:** All randomized subjects who receive at least one dose of double-blind study treatment. Subjects will be analyzed according to treatment received.	[redacted]
5.1 Relevant Protocol Deviations	Relevant protocol deviations are deviations that can have an impact on the primary efficacy endpoints. The impact of relevant protocol deviations on the primary efficacy results will be assessed by excluding subjects from the FAS to define the PPS in supportive analyses of the primary efficacy endpoints. Relevant protocol deviations to be considered regarding exclusion of subjects from the FAS will include the following:	[redacted]

	• Randomized but did not take any study treatment • Failed to meet study inclusion criteria**, 1a, 2a (Classification Criteria for Psoriatic Arthritis portion), 2b, 2d, 2e, 2f, 2g, 2i, 2k, and 3a,** but were entered into the study • Met study exclusion criteria**, 4a, 4b, 4m, 4n, and 4s,** but were entered into the study (only exclusion criteria expected to have an impact on the primary efficacy endpoints will be considered relevant) ~~• Poor compliance to study medication within the first 16 weeks of treatment defined as <80% compliant with study treatment~~**Failed to maintain compliance to study medication of >=75% for the overall treatment period to Week 16** • Failure to adhere to prohibited concomitant medication restrictions **as described below:** ○ **Receiving a single intramuscular, intra-arterial, intravenous, or oral course of high dose corticosteroid (prednisone > 10 mg/day or equivalent) within 28 days of the primary time point at Week 16** ○ **Receiving more than one instance of a single intramuscular, intra-arterial, intravenous, or oral course of high dose corticosteroid (prednisone > 10 mg/day or equivalent) up to Week 16** ○ **Starting a new non-biologic disease-modifying antirheumatic drug (DMARD) at any time before the primary endpoint at Week 16** ○ **Starting a biologic DMARD at any time before the primary endpoint at Week 16** ~~• Subject overdosed, misused or abused study treatment prior to Week 16~~ • Actual treatment received is different than randomized treatment All subjects with relevant protocol deviations will be identified prior to ~~database lock and~~unblinding of treatment assignment **with the exception of subjects where actual treatment received is different than randomized treatment which will be determined after treatment unblinding**. Relevant protocol deviations will be summarized by treatment group and deviation category for the FAS population. ~~Additionally, all important protocol deviations, which are deviations that may impact the efficacy and safety of subjects, will be identified prior to database lock and unblinding of treatment assignment. Important protocol deviations will be summarized by treatment group and deviation category for the FAS population.~~	
6.1.1 Primary Endpoint(s)	The primary efficacy analysis model for the binary endpoint, ACR 20 (responder/non-responder), will use a logistic regression model on FAS to assess whether there is a dose-response trend between ACR	[redacted]

6.1.1.1 Primary Analysis	20 response and dose level. The model will include dose level (0, 6, 12) as a continuous variable, and the following covariates: TNFi use (experienced/naïve) and body weight (≥ 90 kg and < 90 kg) **from IRT** in the model. ~~The odds ratio and the corresponding 2-sided 95% confidence interval (CI) will be provided. Linear contrast will be used to test the dose-response trend.~~ The slope coefficient in the logistic regression of ACR 20 vs. dose will tested as H_0: Beta = 0 vs H_1: Beta ≠ 0 at α=0.10 (2-sided). ~~The Pearson residuals and the deviance residuals will be plotted against the indices of the observations, which will be used to assess extreme cases and and/or systematic patterns in variation.~~ **Adjusted odds ratio and the corresponding 2-sided 95% confidence intervals (CI) will be provided from a logistic regression with responder as the dependent variable and treatment, TNFi use (experienced/naïve), and (≥ 90 kg and < 90 kg) from IRT as fixed factors.** <u>Imputation Methodology</u> The estimand of interest is to assess the dose-response of multiple doses of BMS-986165 on subject response to ACR 20 criteria at Week 16 for all randomized subjects receiving at least one dose of study treatment. The primary imputation method of non-responder imputation (NRI) will be used for subjects who have the below intercurrent events: • Subjects who discontinue the treatment or study early (i.e. prior to Week 16) and have no ACR 20 assessments at Week 16 • ~~Start a protocol prohibited medication/therapy~~ • Subjects who are lost to follow-up • Subjects who otherwise have missing endpoint data at Week 16 **or not sufficient for a definitive determination of response status will be classified as non-responders at that time point**	[redacted]
6.1.2 Secondary Endpoint(s) 6.1.2.1 Binary Endpoints	The secondary endpoint of PASI 75 response at Week 16 will be assessed in a dose-response analysis. The binary endpoint (responder/non-responder) will use a logistic regression analysis similar to the primary analysis. The model will include dose level (0, 6, 12) as a continuous variable, and the following covariates: TNFi use (experienced/naïve) and body weight (≥ 90 kg and < 90 kg) **from IRT** in the model. ~~The odds ratio and the corresponding 2-sided 95% confidence interval (CI) will be provided. Linear contrast will be used to test the dose-response trend.~~ The slope coefficient in the logistic regression of PASI 75 vs. dose will tested as H_0: Beta =0 vs H_1: Beta ≠ 0 at α=0.10 (2-sided). ~~Logistics regression diagnostics will be performed similar to those of the primary analysis model.~~ **Adjusted odds ratio and the corresponding 2-sided 95% confidence intervals (CI) will be provided from a logistic regression with responder as the dependent variable and**	[redacted]

	treatment, TNFi use (experienced/naïve), and (≥ 90 kg and < 90 kg) from IRT as fixed factors. Imputation Methodology Similar strategies for addressing intercurrent events used in the primary efficacy endpoint will be used for binary secondary efficacy endpoints **as described in Section 6.1.1.1**.	
6.1.2 Secondary Endpoint(s) 6.1.2.1 Continuous Endpoints	The change from baseline score will be calculated at Week 16 and analyzed using an analysis of covariance (ANCOVA). Dose level (0, 6, 12) will be assessed in the model as a continuous variable and the following covariates as fixed effects: TNFi use (experienced/naïve) and body weight (< 90 kg and ≥ 90 kg) **from IRT**. The baseline value will be added into the model as a covariate. A general contrast-based test will be used to evaluate the dose-response trend. ~~Treatment differences based on least-squares (LS) means and the corresponding 2-sided 95% CI will also be provided.~~ ~~Regression diagnostics will be performed by plotting residuals versus fitted values, which will be used to assess model fit and verify homogeneity of error variance. Normal Q-Q plot of residuals will be performed to check normality assumption~~ Imputation Methodology For continuous secondary efficacy endpoints, a modified baseline observation carried forward (mBOCF) approach will be used for missing data. The baseline observation will be carried forward for subjects who discontinue study treatment due to: • Lack of efficacy • AEs ~~For subjects who start a protocol prohibited medication/therapy will also have their endpoint value imputed as the baseline value.~~ The last valid observation will be carried forward for all other subjects with missing data. Subjects who discontinue study treatment for other reasons will have the last valid observation carried forward (including the baseline value as applicable).	[redacted]
6.1.4 Subgroup Analyses	• Geographic region **(EU [Czech Republic, Germany, Hungary, Spain, Poland], Non-EU [Russian Federation], and US)** • Country • Sex (male, female) • Age group (<~~65~~**40** y, **40-<65 y**, ≥65 y) • Body weight categories (<90 kg; ≥90 kg) **– from case report form** • Race • Baseline TNFi use history (experienced/naïve) **– from case report form** • Baseline non-biologic DMARD use history (yes/no) • Baseline DMARD use (yes/no) • Baseline glucocorticosteroid use (yes/no) • Baseline methotrexate use (yes/no) • Baseline disease severity (DAS28 <5.1 vs DAS28 ≥5.1)	[redacted]

	• Baseline number of swollen joints (≤4 vs >4) • Baseline presence of enthesitis (yes/no) • Baseline presence of dactylitis (yes/no) • Baseline CRP (<10 mg/L vs ≥10 mg/L) • Duration of disease (< median y, ≥ median y)	[redacted]
6.1.5 Additional Endpoints 6.1.5.1 Binary Endpoints	Analysis Model Additional endpoints defined in the Section 4.1.3 at Week 16 will be assessed in a dose-response analysis. The binary endpoint (responder/non-responder) will use a logistic regression analysis similar to the primary and secondary endpoint analyses. The model will include dose level (0, 6, 12) as a continuous variable, and the following covariates: TNFi use (experienced/naïve), and body weight (≥ 90 kg and < 90 kg) **from IRT** in the model. ~~The odds ratio and the corresponding 2-sided 95% confidence interval (CI) will be provided. Linear contrast will be used to test the dose-response trend.~~ The slope coefficient in the logistic regression of endpoint vs. dose will tested as H_0: Beta =0 vs H_1: Beta ≠ 0 at α=0.10 (2-sided). **Adjusted odds ratio and the corresponding 2-sided 95% confidence intervals (CI) will be provided from a logistic regression with responder as the dependent variable and treatment, TNFi use (experienced/naïve), and (≥ 90 kg and < 90 kg) from IRT as fixed factors.** Imputation Methodology Similar strategies used in the binary secondary efficacy endpoints will be used for the additional endpoints **as specified in Section 6.1.1.1**.	[redacted]
6.1.5 Additional Endpoints 6.1.5.2 Continuous Endpoints	The change from baseline score will be calculated at Week 16 and analyzed using an analysis of covariance (ANCOVA). Dose level (0, 6, 12) will be assessed in the model as a continuous variable and the following covariates as fixed effects: TNFi use (experienced/naïve) and body weight (< 90 kg and ≥ 90 kg) **from IRT**. The baseline value will be added into the model as a covariate. A general contrast-based test will be used to evaluate the dose-response trend. ~~Treatment differences based on least-squares (LS) means and the corresponding 2-sided 95% CI will also be provided.~~ Imputation Methodology Similar strategies used in the continuous secondary efficacy endpoints will be used for the additional continuous endpoints **as specified in Section 6.1.2.2**.	[redacted]

	[redacted]	[redacted]
6.2 Safety	Summaries of safety data will be presented by period (Part A and B) and treatment group, as applicable, for the ~~All Treated~~**As-treated** population.	[redacted]
6.2.1 Adverse Events	An overall summary for the following categories will be presented: • ~~Subjects with at least one TEAE~~ • ~~Subjects with at least one related TEAE~~ • ~~Subjects with at least one treatment-emergent SAE~~ • ~~Subjects with at least one related treatment-emergent SAE~~ • ~~Subjects discontinuing study treatment due to a TEAE~~ • ~~Subjects discontinuing from study due to a TEAE~~ • ~~Subjects who died due to an AE~~ • ~~Subjects who died due to a TEAE~~ • **Deaths** • **SAEs** • **Related SAEs** • **AEs** • **Discontinued treatment due to AEs** The following summaries will also be provided for the following: • TEAEs by SOC and PT • Treatment-related TEAEs by SOC and PT • TEAEs by PT reported in ≥ 5% of subjects • ~~TEAEs by PT reported in ≥ 1% of subjects~~	Modified [redacted]
6.2.1.1 Adverse Events of Interest (AEI) **and Other Important AEs**	Summaries for treatment-emergent AEIs events will be provided **by SOC and PT** for each AEI category~~for the following~~: • ~~TEAEs by SOC and PT~~ ~~Skin Events~~	[redacted]

	~~The number and percentage of subjects reporting each type of skin event and the corresponding location will be summarized.~~ ~~Infections~~ ~~The number and percentage of subjects reporting infections will be summarized.~~ ~~Creatine kinase (CK) elevation~~ ~~The number and percentage of subjects reporting CK elevations will be summarized.~~ ~~Malignancies~~ ~~The number and percentage of subjects reporting malignancies will be summarized.~~ ~~Cardiovascular~~ ~~The number and percentage of subjects reporting cardiovascular event will be summarized.~~ • **Skin-related events** • **Infection events** • **Malignancy events** **Creatine kinase (CK) elevation for CK elevation >2.5 upper limit of normal will be summarized as CTCAE grade 2 in the clinical laboratory summaries.** **Additional information collected for some events as part of the clinical safety program and adjudicated events will also be summarized.**	[redacted]
6.2.3 Clinical Laboratory Data	Laboratory parameters will be summarized using the International System (SI) of Units~~, unless otherwise specified~~ **and US conventional units**. Data will be summarized by time point, as applicable. The following summaries will be provided for each parameter: • Absolute and change from baseline values for continuous parameters • Number and percentage of subjects for the following: ~~o Categorical urinalysis parameter results~~ o Maximum postbaseline CTCAE grade for each applicable laboratory parameter o Shifts from baseline based on maximum postbaseline CTCAE grade • Drug-induced Liver Injury (DILI) and Hy's Law summaries All laboratory data specified in the summary tables will be present in listings.	[redacted]
6.2.4 Vital Signs and Physical Findings	Vital signs**, including weight,** will be summarized by time point, as applicable. The following summaries will be provided for each parameter:	[redacted]

		[redacted]
6.3 General Methodology	P-values will be presented to ~~3~~ **4** decimal places where presented in the TFL shells	[redacted]
6.3.2 Demographic and Baseline Characteristics	~~**The analysis set(s) to be used in the summaries should be specified.**~~ Demographic **and baseline** characteristics will be summarized by treatment group for **the FAS** ~~all of the analysis populations~~**.** Demographic and baseline characteristics include the following: • Gender (male, female) • Race • Ethnicity • Age (in years, at time of signing informed consent) and age category (<~~65~~**40, 40-<65,** ~~vs~~≥65) • Weight (in kg, at baseline) and weight category (≥90 kg, <90 kg) • Body mass index (BMI in kg/m^2, at baseline) • Geographic region (**EU, Non-EU, US**~~U.S., Japan, China, Rest of World~~) • Prior biologic use for psoriatic arthritis (yes, no) • Reason for discontinuation of prior biologic use • Stratification factors obtained from IRT: TNFi use (experienced/naïve), and body weight (<90 kg and ≥90 kg) • Stratification factors obtained from database: TNFi use (experienced/naïve), and body weight (< 90 kg and ≥90 kg) • Duration of disease (<median y, ≥ median y) • PsA phenotype - oligoarticular (≤4 joints), polyarticular (>4 joints), arthritis with predominant associated spinal symptoms • Age at disease onset (<18, 18-39, ≥40) • ~~Smoking status (Current daily smoker and whether heavy vs. light, Current occasional smoker, former smoker, never smoker, smoker current status unknown, unknown if ever smoked)~~ Additional demographics or baseline data may be added to summary tables. General medical history and medical history related to PsA will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) (version 21.0 or an updated version at the time of database lock). General medical history data will be summarized for each SOC and PT by treatment group and overall for the FAS ~~and Safety Set populations~~. Separate tables will be provided for PsA medical history.	

[redacted]	[redacted]	[redacted]
6.3.3.1 Psoriatic Arthritis and Other Inflammatory Disease Related Systemic Medications	Prior **and medications that started prior to first treatment and were ongoing after first treatment start date for** biologic and non-biologic medications will be summarized as described above for the number and percentage of subjects using each reported medication. Concomitant use of DMARDs like methotrexate will be summarized.	[redacted] .
[redacted]	[redacted]	[redacted]
6.3.4 Exposure 6.3.4.1 Duration of Treatment	<u>Duration by Group</u> Overall duration of each treatment, in days, will be calculated for each subject in each treatment group. Subjects are dispensed study treatment at each visit, starting with Day 1 (Week 0) and at each subsequent visit. The ~~Day 1 dispensed date will be considered as the~~ date of first dose of study treatment **is the Week 0 PK dosing date** and is recorded on the eCRF. **If this date is missing, then the earliest drug dispensation date will be used.** The last **dose** date ~~of dose~~ is ~~defined as the last day a subject received drug and is~~ recorded on the end of treatment page in the eCRF. **If this date is missing, then the latest date from the PK exposure page or drug accountability return date will be used.** ~~If the date of last dose of study treatment is missing, i.e.: subject lost to follow-up, the end of study date or last contact date (whichever comes first) will be used to calculate duration of treatment.~~ Duration of treatment will be summarized descriptively for Part A and B, by treatment group. For subjects randomized to placebo, ~~the Week 16 date will be used as the date of last dose of placebo. Formula for~~ duration is defined as: ➢ Placebo = **(date of last dose of placebo in Part A – date of first dose +1)**~~Week 16 date – date of first dose +1~~	[redacted] .

	~~If a placebo subject discontinues study treatment on or before the Week 16 visit, the date of last dose will be used to calculate the duration of placebo.~~ **For subjects who participated in Part B and were allocated to receive ustekinumab, duration is defined as:** Ustekinumab = (date of last dose of ustekinumab in Part B – date of first dose of in Part B +1)~~Week 52 date – date of first dose + 1)~~ Duration by Period Overall duration (in days) of each treatment received within each study period will be calculated. Duration within each period is defined for each treatment group as: $(Last\ dose\ date - first\ dose\ date + 1)$ ~~Subjects are dispensed study treatment at each visit, starting with Day 1 (Week 0) and at each subsequent visit. The dispensed drug date of Day 1 of the study period will be considered as t~~The date of first dose of study treatment for **Part A is as described above** ~~that period and is recorded on the eCRF~~**. The date of first dose of study treatment for Part B is defined as the earliest drug dispensation date for Part B.** The last **dose** date **is as described above**. ~~of dose within the period will be considered as the day prior to the next period start date.~~ ~~If the date of last dose of study treatment is missing, i.e. subject lost to follow-up, the end of study date or last contact date (whichever comes first) will be used to calculate duration of treatment~~.	
6.3.4.2 Summary of Dosing	The number of doses taken will be summarized descriptively by treatment group **within each study period and overall**.	[redacted]
6.3.4.3 Compliance	Treatment compliance will be determined from data captured on the Drug Accountability eCRF. ~~The number and percentage of subjects who have missed at least one dose will be provided.~~ ~~Additionally, descriptive statistics for the number of missed doses within each treatment period and overall will be provided by treatment group. The number of missed doses for each subject will be calculated for each period.~~ BMS-98615 $Number\ of\ expected\ doses: (date\ of\ next\ visit - date\ of\ current\ visit) x2$ $Number\ of\ \sout{missed}\ doses\ \boldsymbol{taken}: Number\ of\ \sout{expected}\ doses\ \boldsymbol{dispense} - number\ of\ doses\ \sout{taken}\ \boldsymbol{returned}$ Treatment compliance will be derived for each period and overall. Compliance is defined as: $\left(\frac{Number\ of\ doses\ taken}{Number\ of\ expected\ doses}\right) x100$	[redacted]

Ustekinumab

$$Number\ of\ expected\ injection: (date\ of\ next\ visit - date\ of\ current\ visit)$$

~~*Number of missed injection: Number of expected injections – number of injections*~~

Treatment compliance will be derived for each period and overall. Compliance is defined as:

$$\left(\frac{Number\ of\ injections}{Number\ of\ expected\ injections}\right) x100$$

Period compliance will be calculated by summing over all visits within the period and overall compliance will be calculated by summing each visit in the study and summarized using descriptive statistics by treatment group. The number and percentage of subjects with <~~80~~**75**%, ~~80~~**75**% to 1~~2~~**0**0%, and >1~~2~~**0**0% compliance will be provided by treatment group for each ~~visit,~~ period and overall. **If a subject does not return the container, then this dispensation event will be excluded in the calculation for total number of doses taken, total number of expected doses, and total compliance.**

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
7.3 Final Analyses and Reporting	All final, planned analyses identified in this statistical analysis plan will be performed only after the last subject has completed the study and the database has been locked. **Investigative site staff and subject will remain blinded to treatment assignment until the database has been locked**. ~~The randomization codes for all subjects will not be unblended until the database has been locked.~~	[redacted]
8.1 General Definitions	First Dose Date – Study: The date a subject received their first dose on Day 1 as recorded in the eCRF ~~as date study treatment was dispensed~~**Week 0 PK dosing date or the earliest drug dispensation date**. First Dose Date – Period: The date a subject received their first dose on Day 1 as recorded in the eCRF ~~as date study treatment was dispensed~~**Week 0 PK dosing date or the earliest drug dispensation date for Part A and earliest drug dispensation date for Part B**. Last Dose Date – Period: ~~The date of the last visit in the periods – 1. If a subject prematurely discontinues study treatment within a period, the date of last recorded dose on the eCRF will be used as the last dose date for the period~~**The last dose date is the date recorded on the end of treatment page in the eCRF. If this date is missing, then the latest date from the PK exposure page or drug accountability return date will be used**. **If the baseline value is 0 and the post-baseline is >0, then the percent change from baseline value will be missing.**	[redacted]
8.2.1.3 Disease Activity Score (DAS) 28 CRP	**The following formula will be used to compute the DAS28-CRP score;** **DAS28-CRP = 0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*ln(CRP+1) + 0.014*GH + 0.96.**	[redacted]

8.2.1.4 Dactylitis	**LDI measures the ratio of the circumference of the affected digit to the circumference of the digit on the opposite hand or foot. If both sides are considered involved, a table of normative values is used to provide the comparison. The ratio of circumference is multiplied by a tenderness score (0=nontender, 1=tender). For each dactylitic digit, the final score is defined as:** **[{(A/B) – 1}*100]*C,** **Where A is circumference of involved digit, B is circumference of opposite, unaffected, digit or reference, and C is tenderness (0 or 1). The total score is determined by summing over all digits.**	[redacted]
8.2.1.8 Psoriatic Arthritis Disease Activity Score (PASDAS)	**The PASDAS is derived using the following formula:** **PASDAS = (((0.18*sqrt(Physician's Global Assessment of psoriatic arthritis) + (0.159*sqrt(Subject Global Assessment of disease activity) – (0.253*sqrt(SF36-PCS) + (0.101*LN(swollen joint count + 1)) + (0.048*LN(tender joint count + 1)) + (0.23*LN(Leeds enthesitis count + 1)) + (0.37*LN(tender dactylitis count + 1)) + (0.102*LN(CRP + 1)) + 2)*1.5.**	[redacted]
8.2.1.9 PsA Response Criteria (PsARC)	The Psoriatic Arthritis Response Criteria (PsARC) consists of 4 measurements: tender/painful joint count, swollen joint count, Physician Global Assessment of psoriatic arthritis, and Subject Global Assessment of **psoriatic arthritis**~~pain ≤ 15~~. In order to be classified as 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. **Improvement in each of the measures is defined below:** • **Decrease of ≥ 30% in tender joint counts** • **Decrease of ≥ 30% in swollen joint counts** • **Decrease of ≥ 20% in Physician's Global Assessment of Psoriatic Arthritis** • **Decrease of ≥ 20% in Subject's Global Assessment of Psoriatic Arthritis**	[redacted]
8.2.1.10 Disease Activity Index for Psoriatic Arthritis Score (DAPSA)	**The DAPSA is derived using the following formula:** **DAPSA = tender joint count + swollen joint count + CRP + Subject Global Assessment of disease activity + Subject Global Assessment of pain.**	[redacted]
8.2.2.1 Health Assessment Questionnaire-Disability Index (HAQ-DI)	**Scoring for each functional category and the disability index will be defined as follows:** • **Dressing and Grooming includes items 1 and 2** • **Arising includes items 3 and 4** • **Eating includes items 5, 6, and 7** • **Walking includes items 8 and 9**	[redacted]

	• **Hygiene includes items 10, 11, and 12** • **Reach includes items 13 and 14** • **Grip includes items 15, 16, and 17** • **Activities includes items 18, 19, and 20** **The score for each functional category will be the highest score within the individual item scores. If any aids, devices, or help from others is used, then an item score is adjusted up to a "2" if less than 2. The disability index will be the mean of the eight functional scores. If more than two of categories, or 25%, are missing, the index will not be scored. Otherwise, divide the sum of the categories by the number of answered categories.**	
8.2.2.2 Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)	**Each individual question response is scaled to a 0-10 score by dividing by 10 and the BASDAI is derived using the following formula:** **BASDAI = ((Q1 + Q2 + Q3 + Q4) + ((Q5 + Q6) / 2)) / 5** **A scale of 0 – 50 is possible for the BASDAI score.**	[redacted]
8.2.2.5 Work Limitation Questionnaire (WLQ)	The WLQ is a ~~25~~**8**-item self-report that measures the on-the-job impact of chronic health conditions and treatment with a focus on assessing limitations while performing specific job demands from the following 4 ~~domains~~**scales**: 1) Time management: difficulty with handling time and scheduling demands (~~5~~ **2** items) 2) Physical demands: ability to perform job tasks that involve bodily strength, movement, endurance, coordination, and flexibility (~~5~~ **2** items) 3) Mental-interpersonal demands: cognitively demanding tasks and on-the-job social interactions (~~5~~ **2** items) 4) Output demands: concerns reduced work productivity (~~5~~ **2** items) The questionnaire includes an option for 'Does not apply to my job". Any items with this response will be given a score of 'missing' for scoring purposes. Scoring will be performed in the analysis datasets. Individual scores to each of the ~~25~~ **8** items will be provided by the eCOA system. Each item is scored from 1 to 5. A score is interpreted as: 1=0%, 2=25%, 3=50%, 4=75% and 5=100%. Scores range from 0 to 100%, where 0% = health makes the job demand difficult none of the time and 100% = health makes the job demand difficult all of the time. Each ~~domain~~**scale** is derived as a mean of the items within the ~~domain~~**scale**. If half or more of the items within the scale are missing, the scale cannot be computed and will be considered missing. $Scale = \frac{item_1 + item_2 + \cdots + item_n}{n}, where\ n\ is\ the\ number\ of\ items\ within\ t$ The ~~domain~~**scale** scores of the WLQ are derived as a mean of the 25-item scores minus 1 and then multiplied by 25.	[redacted]

	If all 4 scale scores are non-missing, then the WLQ Index is calculated using the following formula: **WLQ Index = (0.00048*WLQ Time Management Scale + 0.00036*WLQ Physical Tasks Scale + 0.00096*WLQ Mental-Interpersonal Tasks Scale + 0.00106*WLQ Output Tasks Scale).** **The WLQ At-Work Productivity Loss score is calculated using the following formula:** **WLQ At-Work Productivity Loss Score = (1-exp(-WLQ Index))*100.**	
8.2.2.6 Psoriatic Arthritis Impact of Disease (PsAID)	The Psoriatic Arthritis Impact of Disease (PsAID) is a 12-item self-report that measures psoriatic arthritis symptoms and impact of disease. Each item is scored on a 0 to 10 numeric rating scale with a one week recall period. **Each item is scaled by a factor and the total score is calculated by summing over the scaled scores**~~The PsAID has a total score~~, with a higher value indicating worse health.	[redacted]
8.2.2.7 Routine Assessment of Patient Index Data 3 (RAPID3)	~~Routine Assessment of Patient Index Data 3 (RAPID3)~~ ~~RAPID3 (routine assessment of patient index data 3) is a pooled index of the three patient-reported ACR Core Data Set measures: function, Subject Global Assessment of pain, and Subject Global Assessment of disease activity. Each of the 3 individual measures is scored 0 to 10, for a total of 30. Disease severity may be classified on the basis of RAPID3 scores: >12 = high; 6.1-12 = moderate; 3.1-6 = low; < or =3 = remission. RAPID3 scores are correlated with the disease activity score 28 (DAS 28) in clinical trials and clinical care, and are comparable to DAS 28 scores in the capacity to distinguish active from control treatments in clinical trials.~~	[redacted]
8.4 Study Periods	Part A = (Week 0 to Week 16 **visit**) Part B = Optional (Week 20 to Week 5~~6~~**2 visit**) Follow-up = 4 week follow-up period	[redacted]
8.5 Day Ranges for Analysis Visits	Week 16 100, 127 (or Week 16 drug dispense date ~~if earlier~~)	[redacted]

Appendix 1 Planned Analyses

List of Planned Analyses: Dose-response relationship of BMS-986165 (6 or 12 mg once daily [QD]) at Week 16 in the treatment			
Measure of Interest	**Population**	**Analysis at Week 16**	**Over Time Through Week 16**
ACR 20 – Primary	FAS	NRI+Logistic regression - primary	Binary GEE
ACR 20 - Primary	FAS	LOCF +Logistic regression – sensitivity LOCF/NRI +Logistic regression – sensitivity	Binary GEE
ACR 20 - Primary	PPS	NRI+Logistic regression - supportive	N/A
ACR 20 - Primary	FAS	NRI+CMH - subgroups	N/A
Change from baseline in HAQ-DI score - Secondary	FAS	mBOCF+ANCOVA	MMRM analysis
PASI 75 response in subjects with at least 3% BSA involvement at baseline - Secondary	FAS	NRI+Logistic regression	Binary GEE
Change from baseline in SF-36 PCS score - Secondary	FAS	mBOCF+ANCOVA	
ACR 50/70 response – Additional	FAS	NRI+Logistic regression	Binary GEE
HAQ-DI 0.35 response, where HAQ-DI 0.35 responder is defined as a subject with an improvement from baseline in	FAS	NRI+Logistic regression	Binary GEE

List of Planned Analyses: Dose-response relationship of BMS-986165 (6 or 12 mg once daily [QD]) at Week 16 in the treatment			
Measure of Interest	**Population**	**Analysis at Week 16**	**Over Time Through Week 16**
HAQ-DI score of at least 0.35 - Additional			
PASI 90 response in subjects with at least 3% body surface area (BSA) involvement at baseline - Additional	FAS	NRI+Logistic regression	Binary GEE
Dactylitis mean change from baseline (dactylitis count) - Additional	FAS	mBOCF+ANCOVA	MMRM analysis
Dactylitis change from baseline (LDI) - Additional	FAS	mBOCF+ANCOVA	MMRM analysis
Dactylitis resolution, where resolution is defined as a dactylitis count of 0 in subjects with dactylitis count ≥ 1 at baseline - Additional	FAS	NRI+Logistic regression	Binary GEE
Enthesitis mean change from baseline	FAS	mBOCF+ANCOVA	MMRM analysis
Leeds Enthesitis Index (LEI) score of 0, in subjects with LEI ≥ 1 at baseline	FAS	NRI+Logistic regression	Binary GEE

List of Planned Analyses: Dose-response relationship of BMS-986165 (6 or 12 mg once daily [QD]) at Week 16 in the treatment			
Measure of Interest	**Population**	**Analysis at Week 16**	**Over Time Through Week 16**
Change from baseline in Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis index	FAS	mBOCF+ANCOVA	MMRM analysis
Minimal Disease Activity (MDA) response	FAS	NRI+Logistic regression – Additional	Binary GEE
PGA-F 0/1, assessed as a proportion of subjects with a PGA-F score of 0 or 1 amount subjects with a baseline PGA-F score ≥ 3	FAS	NRI+Logistic regression – Additional	Binary GEE
Change from baseline in Psoriatic Arthritis Disease Activity Score (PASDAS)	FAS	mBOCF+ANCOVA - Additional	MMRM analysis
Change from baseline in Disease Activity Index for Psoriatic Arthritis Score (DAPSA)	FAS	mBOCF+ANCOVA - Additional	MMRM analysis
Psoriatic Arthritis Response Criteria (PsARC) response	FAS	NRI+Logistic regression – Additional	Binary GEE
Disease Activity Score (DAS) 28 CRP response:	FAS	NRI+Logistic regression – Additional	Binary GEE

List of Planned Analyses: Dose-response relationship of BMS-986165 (6 or 12 mg once daily [QD]) at Week 16 in the treatment			
Measure of Interest	**Population**	**Analysis at Week 16**	**Over Time Through Week 16**
Change from baseline in DAS 28 CRP score	FAS	mBOCF+ANCOVA - Additional	MMRM analysis
Change from baseline in Psoriatic Arthritis Impact of Disease (PsAID) score	FAS	mBOCF+ANCOVA - Additional	MMRM analysis
Disease Activity Index (BASDAI), in subjects with baseline evidence of PsA spondylitis	FAS	mBOCF+ANCOVA - Additional	MMRM analysis
Change from baseline in SF-36 mental component summary (MCS)score	FAS	mBOCF+ANCOVA - Additional	
Change from baseline in Patient Reported Outcome Measurement Information System (PROMIS-Fatigue) score	FAS	mBOCF+ANCOVA - Additional	
Change from baseline in Work Limitation Questionnaire (WLQ) score	FAS	mBOCF+ANCOVA - Additional	

Appendix 2 Summary of Efficacy Assessments

Assessment	Outcome Measure	Endpoint	Comparison	Analyses
ACR (Sec 8.2.1.1)	ACR 20	W16	Trend test	Logistic (Primary), sensitivity, supplemental analyses (Sec 6.1.1.1-3)
			[redacted]	CMH (Subgroup analyses; Sec 6.1.4; [redacted]
		[redacted]	[redacted]	[redacted]
	ACR50	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
	ACR 70	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			BMS vs. PBO	CMH (Exploratory; Sec 6.1.6.1)
		[redacted]	[redacted]	[redacted]
PASI (Sec 8.2.1.2)	PASI 75	W16	Trend test	Logistic (Secondary; Sec 6.1.2.1)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
	PASI 90	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
HAQ-DI (Sec 8.2.2.1)	CFB	W16	General contrast-based test	ANCOVA (Secondary; Sec 6.1.2.2)

Assessment	Outcome Measure	Endpoint	Comparison	Analyses
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
	HAQ-DI 0.35 response, where HAQ-DI 0.35 responder is defined as a subject with an improvement from baseline in HAQ-DI score of at least 0.35	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
DAS Sec 8.2.1.3	Low Disease Activity defined as DAS 28 CRP score < 3.2	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
	Remission is defined as a DAS 28 CRP score < 2.6	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			[redacted]	[redacted]
	CFB	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
SF-36 / MCS (Sec 8.2.2.3)	CFB	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)

Assessment	Outcome Measure	Endpoint	Comparison	Analyses
			[redacted]	[redacted]
		Baseline->W16 over time	BMS vs. PBO	
SF-36 / PCS (Sec 8.2.2.3)	CFB	W16	General contrast-based test	ANCOVA (Secondary; Sec 6.1.2.2)
			[redacted]	[redacted]
		Baseline->W16 over time	BMS vs. PBO	
SF-36 /total (Sec 8.2.2.3)	CFB	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
			[redacted]	[redacted]
		Baseline->W16 over time	BMS vs. PBO	
WLQ (Sec 8.2.2.10)	CFB	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
			[redacted]	[redacted]
		Baseline->W16 over time	BMS vs. PBO	
PGA-F Section 8.2.1.6	PGA-F 0/1 among subjects with a baseline PGA-F score ≥3	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]

Assessment	Outcome Measure	Endpoint	Comparison	Analyses
PsAID Sec 8.2.2.6	CFB	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
PROMIS-Fatigue (sec 8.2.2.4)	CFB	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
		Baseline->W16 over time	BMS vs. PBO	
PsARC (sec 8.2.1.9)	PsARC response	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
DAPSA Score (8.2.1.10)	CFB	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
PASDAS Score	CFB	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)

Assessment	Outcome Measure	Endpoint	Comparison	Analyses
(Sec 8.2.1.8)			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
(MDA) response (Sec 8.2.1.7)	(MDA) response	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
BASDAI Score (Sec 8.2.2.2)	CFB - DASDAI score in subjects with baseline evidence of PsA spondylitis	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
Enthesitis (Sec 8.2.1.5)	CFB - Enthesitis mean change from baseline	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
	Enthesitis resolution, where resolution is defined as a Leeds Enthesitis Index (LEI) score of 0, in subjects with LEI ≥ 1 at baseline	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
			[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]

Assessment	Outcome Measure	Endpoint	Comparison	Analyses
	CFB - in Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis index	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
	Enthesitis resolution, where resolution is defined as a SPARCC enthesitis index score of 0, in subjects with SPARCC ≥ 1 at baseline	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)
Dactylitis (Sec 8.2.1.4)	CFB -Dactylitis mean change from baseline (dactylitis count)	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
	CFB -Dactylitis change from baseline (LDI)	W16	General contrast-based test	ANCOVA (Additional; Sec 6.1.5.2)
	Dactylitis resolution, where resolution is defined as a dactylitis count of 0 in subjects with dactylitis count ≥ 1 at baseline	W16	Trend test	Logistic (Additional; Sec 6.1.5.1)

Assessment	Outcome Measure	Endpoint	Comparison	Analyses