Official Title of Study:

PHASE 2, RANDOMIZED, MULTI-CENTER, DOUBLE-BLIND, DOSE-RANGING, PLACEBO-CONTROLLED, ADAPTIVE DESIGN STUDY TO EVALUATE THE EFFICACY AND SAFETY/PHARMACOKINETICS OF BMS-986142 IN SUBJECTS WITH MODERATE TO SEVERE RHEUMATOID ARTHRITIS WITH AN INADEQUATE RESPONSE TO METHOTREXATE WITH OR WITHOUT TNF INHIBITORS

NCT Number: NCT02638948

Document Date (Date in which document was last revised): December 4, 2015

STATISTICAL ANALYSIS PLAN

FOR DOUBLE-BLIND PERIOD ANALYSES

PHASE 2, RANDOMIZED, MULTI-CENTER, DOUBLE-BLIND, DOSE-RANGING, PLACEBO-CONTROLLED, ADAPTIVE DESIGN STUDY TO EVALUATE THE EFFICACY AND SAFETY/PHARMACOKINETICS OF BMS-986142 IN SUBJECTS WITH MODERATE TO SEVERE RHEUMATOID ARTHRITIS WITH AN INADEQUATE RESPONSE TO METHOTREXATE WITH OR WITHOUT TNF INHIBITORS

PROTOCOL IM006016

VERSION 1.0

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Research Hypothesis:

At least one of the dose levels of BMS-986142 administered daily is more effective than placebo on a background of MTX in achieving ACR20 and ACR70 response after 12 weeks of treatment in subjects with moderately-to severely active RA with an inadequate response to MTX with or without TNF inhibitors.

Schedule of Analyses:

The following analyses will be performed:

- An interim analysis will be performed for futility and dose adaptation after about 20 patients per treatment arm have completed at least 4 weeks of treatment period or discontinued the treatment;
- The final analysis will be performed after the last treated subject in the study reaches Week 12 in the double-blind (DB) period.

2 STUDY DESCRIPTION

2.1 Study Design

This is a 12 week randomized, double-blind, placebo-controlled, dose-ranging study with adaptive design features based on an interim analysis. The study will initially have a Screening Period (with 2 screening visits for subjects that require DMARD washout) and be followed by up to 12 weeks of DB treatment and a 30 day follow-up period.

Four treatments will be administered: placebo, 100, 200, and 350 mg of BMS-986142 daily for 12 weeks. Based on the Week 4 interim analysis, one or more BMS-986142 doses and/or sample size may be subsequently modified.

Screening Period:

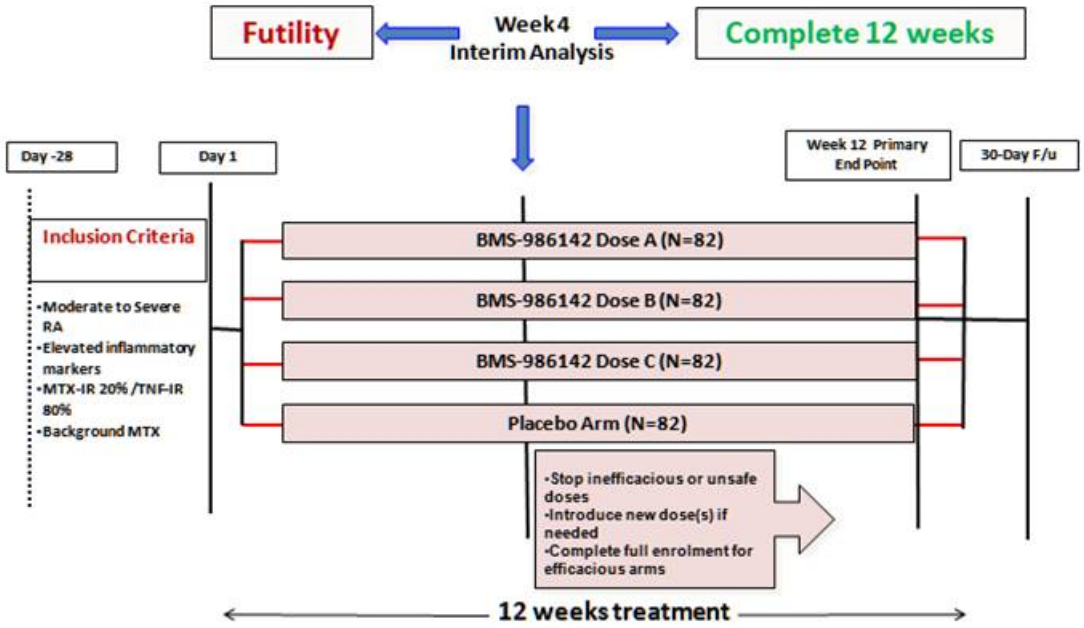
The standard duration of the Screening Period is up to 30 days, with two Screening Visits allowed for subjects who require washout of hydroxychloroquine/chloroquine and/or sulfasalazine. Subjects who are on hydroxychloroquine/chloroquine and/or sulfasalazine in combination with MTX will have hydroxychloroquine/chloroquine and/or sulfasalazine discontinued at the beginning of the Screening Period as they need to be off those medications for at least 30 days and need to be rescreened prior to randomization (Day 1). Should more time be needed, the duration of the Screening Period may be extended up to another week (total of five weeks) depending on current DMARDs washout, drug stabilization, MRI technical issues, and subject scheduling.

Subjects with moderate to severely active RA who have had an inadequate response to MTX or up to 2 TNF inhibitors will receive BMS-986142 or placebo daily for up to 12 weeks, followed by a 30 day follow-up period.

Double-Blind Period:

Upon meeting the inclusion criteria and none of the exclusion criteria and completing the screening period, approximately 328 subjects will be randomized to 1 of the 4 treatment arms in an equal ratio as shown in the study schematic below.

Figure 1: Study Design Schematic



During this period, the dose of methotrexate, NSAIDs, and oral prednisone (or its equivalent) should remain stable. Intra-articular corticosteroid injections and intra-muscular injections are not permitted. Analgesics are permitted with certain restrictions (see Section 3.2).

At the end of the DB period, alternate therapies for RA should be discussed with subjects.

Interim Analyses:

Interim Analyses (IA) of clinical and PK data will be performed after approximately 20 subjects per treatment arm reach Week 4 (complete 4 weeks of treatment or discontinue) and complete the specified assessments for efficacy (e.g., DAS28-CRP), safety, and PK. Details of interim analyses are described in Section 7.9 (Interim analysis).

2.2 Treatment Assignment

After completion of all screening evaluations, on Day 1, all eligible subjects will be randomly assigned to 1 of 4 treatment arms (BMS-986142, 100 mg; BMS-986142, 200 mg; BMS-986142, 350 mg; and Placebo) in an equal ratio. To randomize a subject, a phone call will be placed into the randomization option of the interactive voice response system (IVRS) in order to obtain a subject's randomized treatment assignment. Randomization will be assigned in the order in which subjects qualify for treatment, not in the order of study enrollment. The IVRS will be available 24-hours a day, 7 days a week, via a toll-free number. Randomization will be stratified by prior treatment status (MTX-IR vs TNF-IR) and geographic region.

Specific instructions (including an enrollment/randomization worksheet) for the central enrollment and randomization procedure using an IVRS will be provided to the site.

Randomized schedules will be generated and kept by the Randomization Group within Drug Supply Management of Bristol-Myers Squibb. The randomization ratio may be adjusted if any dose arm will be discontinued and/or new dose arm will be added based on Week 4 interim analysis.

2.3 Blinding and Unblinding

Blinding of treatment assignment is critical to the integrity of this clinical study. However, in the event of a medical emergency or pregnancy in an individual subject in which knowledge of the investigational product is critical to the subject's management, the blind for that subject may be broken by the investigator. The subject's safety takes priority over any other considerations in determining if a treatment assignment should be unblinded.

Before breaking the blind of an individual subject's treatment, the investigator should determine that the unblinded information is necessary, ie, that it will alter the subject's immediate management. In many cases, particularly when the emergency is clearly not related to the investigational product, the problem may be properly managed by assuming that the subject is receiving active product. It is highly desirable that the decision to unblind treatment assignment be discussed with the Medical Monitor, but the investigator always has ultimate authority for the decision to unblind. The Principal Investigator should only call in for emergency unblinding AFTER the decision to discontinue the subject has been made.

For this study, the method of unblinding for emergency purposes is through the IVRS.

For information on how to unblind in case of an emergency, consult the IVRS manual

In cases of accidental unblinding, contact the Medical Monitor and ensure every attempt is made to preserve the blind.

Any request to unblind a subject for non-emergency purposes should be discussed with the Medical Monitor.

The Bioanalytical Sciences department or its designate will be unblinded to the randomized treatment assignments in order to minimize unnecessary analysis of samples from control group subjects.

To implement the adaptive design that is planned in advance, an interim analysis will be conducted prior to the formal locking of the study database. In order to allow for a high quality and/or validated data analysis and a timely analysis of data, the unblinded team may consist of up to two people from each relevant function role within BMS.

- Global Biometric Sciences - Up to 2 statisticians and 2 programmers
- Clinical Pharmacology and Pharmacometrics - Up to 2 pharmacometric and pharmacokinetic scientists and 2 pharmacometric programmers
- Clinical Biomarkers - Up to 2 clinical biomarker scientists

- Bioinformatics -Up to 2 bioinformaticians

This configuration ensures that the unblinded team has the resources to independently conduct its own analyses. The study team does not have access to any unblinded data and is not involved in the review or discussion of unblinded data.

2.4 Protocol Amendments

Not applicable.

3 OBJECTIVES

3.1 Primary Objective

To compare the efficacy of BMS-986142 versus placebo (PBO) on a background of MTX as assessed by ACR20 and ACR70 response rates at Week 12.

3.2 Secondary Objectives

- Assess additional efficacy outcomes of BMS-986142 at Week 12 and over 12 weeks of treatment as measured by ACR20, ACR50 and ACR70 response rates, DAS28-CRP change from baseline, DAS28-ESR change from baseline, Clinical Disease Activity Index (CDAI), Simplified Disease Index (SDAI), and Boolean remission.
- Assess the safety and tolerability of BMS-986142.
- Evaluate Ctrough of BMS-986142.
- Assess the efficacy of BMS-986142 + MTX to MTX alone in reducing synovitis, osteitis, bone erosion and cartilage loss in hands/wrists by MRI at Weeks 4, and 12, from baseline.

[redacted]

4 ENDPOINTS

4.1 Primary Efficacy Endpoint(s)

Co-Primary endpoints:

- Proportion of subjects who achieve ACR20 response rate at Week 12
- Proportion of subjects who achieve ACR70 response rate at Week 12

4.2 Secondary Efficacy Endpoint(s)

- Proportion of subjects who achieve ACR20 over time from baseline to Week 12
- Proportion of subjects who achieve ACR50 over time from baseline to Week 12
- Proportion of subjects who achieve ACR70 over time from baseline to Week 12
- Proportion of subjects who achieve DAS28-CRP < 2.6 over time from baseline to Week 12
- Proportion of subjects who achieve DAS28-ESR < 2.6 over time from baseline to Week 12
- Proportion of subjects who achieve CDAI ≤ 2.8 over time from baseline to Week 12
- Proportion of subjects who achieve SDAI ≤ 3.3 over time from baseline to Week 12
- Proportion of subjects who achieve Boolean Remission over time from baseline to Week 12
- Change from baseline in DAS28-CRP score over time up to Week 12
- Change from baseline in DAS28-ESR score over time up to Week 12
- Change from baseline in CDAI score over time up to Week 12
- Change from baseline in SDAI score over time up to Week 12
- Change from baseline in RAMRIS scores of synovitis, osteitis (bone marrow edema), bone erosion and cartilage loss (joint-space narrowing) (MRI) to Week 4, and Week 12

4.3 Safety Endpoints

- Incidence and severity of all Adverse Events (AEs), Serious AEs (SAE), and Events of Special Interest (EOSI)
- Incidence and severity of clinically significant changes in vital signs
- Incidence and severity of clinically significant electrocardiogram (ECG) abnormalities
- Incidence and severity of clinically significant abnormalities in general laboratory tests

4.4 Pharmacokinetic Endpoints

Ctrough: Trough level plasma concentration of BMS-986142 at time points specified in Section 5.5 of the protocol.

[redacted]

5 SAMPLE SIZE AND POWER

The sample size calculation is driven by the power to compare the proportion of subjects who achieve ACR70 response at Week 12 between each BMS-986142 dose arm and placebo arm without multiplicity adjustment. The target population consists of mixed subjects who are TNF-IR (≥ 80%) and MTX-IR (≤ 20%). For co-primary endpoint of ACR70 at Week 12, 82 subjects per arm will provide ~80% power to detect a treatment difference of 14% at the type I error rate of $\alpha = 0.05$ (two-sided), assuming the placebo response rate of 2.5%. The assumed treatment difference of 14% is double of 7% that was median difference observed from historical data on approved drugs across different mechanism of actions (MOAs).

If the placebo response rate in ACR70 at Week 12 is higher than planned 2.5% (eg, 5%), 82 subjects per arm can still provide > 70% power to detect a difference of 14%.

For the population of TNF-IR (≥ 80%), 65 subjects per arm will provide 68% power to detect a treatment difference of 14% in ACR70 at the Type 1 error rate of $\alpha = 0.05$ (two-sided), assuming the placebo response rate is 2.5%.

For the other co-primary endpoint of ACR20 at Week 12, 82 subjects per arm will provide 98% power to detect a difference of 31% over placebo at the type I error rate of $\alpha = 0.05$ (two-sided), assuming placebo response rate of 22%. The assumed treatment difference of 31% and placebo response rate of 22% were based on the observed data from approved drugs across different MOAs.

For the population of TNF-IR (≥ 80%), 65 subjects per arm will provide 94% power to detect a treatment difference of 30% in ACR20 at the Type 1 error rate of $\alpha = 0.05$ (two-sided), assuming the placebo response rate is 18%.

An interim analysis (IA) will be conducted when about 20 subjects reach week 4 (complete 4 weeks of treatment or discontinue). Based on the totality of efficacy, safety, and PK at Week 4 IA, the following adaptation(s) may be recommended:

- Stop inefficacious or unsafe dose(s)
- Introduce new dose(s) arms if needed and enroll (up to about 82 subjects per arm)
- Complete full enrolment for all efficacious arms (up to about 82 subjects per arm)

- Stop the study enrollment if all dose arms are stopped for safety or futility;

The total sample size varies depending on the interim analysis results (range from 328 to 408 subjects). The details of interim analysis and the criteria to adapt dose arm(s) and stop for futility are specified in the section 7.9.

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

The following periods are defined in this SAP for the purpose of analyses and reporting.

Pre-treatment period: It covers the time period which starts from the day of enrollment and lasts until the initiation of randomized double-blind treatment

Double-blind (DB) period: It starts at the time of the first dose of blinded treatment (BMS-986142 or placebo) and lasts until the date of the last dose of double-blinded treatment + 30 days or the start of first dose in long-term extension period.

6.2 Treatment Regimens

The eligible subjects will be randomized to receive BMS-986142 (100mg, 200mg and 350mg) or Placebo daily (double-dummy) at equal ratio. There may be an instance of new treatment group(s) added and one or more treatment arms (other than Placebo) may be dropped based on interim analysis. For more details, see study design in the Protocol.

6.3 Populations for Analyses

- All Enrolled Subjects: All subjects who sign an informed consent.
- All Randomized Subjects: All subjects who are randomized to a treatment.
- Modified Intent-to-Treat (MITT) Analysis Population: All randomized subject who have received at least one dose of the study medication. Subjects will be grouped according to the treatment to which they were randomized by IVRS at the start of the study. The MITT analysis population will also be referred to as “All Randomized and Treated Subjects”.
- As-Treated Analysis Population: All Subjects who have received at least one dose of study medication. Subjects will be grouped according to the treatment that they actually received as opposed to the treatment to which they were randomized. Subjects will be grouped on an as randomized basis unless the subject received the incorrect medication for the entire period of treatment. In that case, the subject will be analyzed in the treatment group associated with the incorrect medication he/she received.
- Efficacy Analysis Population: This is a subset of the MITT Analysis Population where the subjects who are randomized to a treatment arm and discontinued based on the interim analysis are excluded. All efficacy results involve statistical comparison or modeling will be based on this analysis population unless otherwise specified.
- [redacted]
- Pharmacokinetic Population: All subjects who have received BMS-986142 and have any available concentration-time data.

7 STATISTICAL ANALYSES

7.1 General Methods

The table below provides an overview of the efficacy analyses to be performed.

Table 1: Planned Efficacy Analyses

Measure of Interest (Primary and Secondary Efficacy Endpoints):	Analysis Method
Co-primary endpoints: ACR20 and 70 response rate at Day 85	Chi-Square test, two sided p-value will be provided. Point estimates and 95% CI of treatment difference.
Binary endpoints : ACR20, ACR50 and ACR70 response rate over time from baseline to Day 85; DAS28-CRP<2.6, DAS28-ESR<2.6, CDAI<=2.8, SDAI<=3.3 and Boolean remission over time from baseline to Day 85	Point estimate and 95% CI of response rate, point estimates and 95% CI of treatment difference.
Continuous endpoints: DAS28-CRP, DAS28-ESR, CDAI, SDAI, MRI scores (synovitis, osteitis,bone erosion and cartilage loss)	Mixed model analysis, point estimates and 95% CI for mean change from baseline within treatment groups, point estimates and 95% CI for mean change from baseline between treatment groups.

The Chi-square tests will be used to compare the ACR20 and ACR70 response rates at Day 85 between each of the active treatment arms and the placebo arm. The Chi-square p-value will be provided for each comparison between the active treatment and the placebo arm. Within each treatment group, the statistical testing will be performed for ACR20 at level of 0.05 first. If it is statistically significant, then the statistical testing will be performed for ACR70. Otherwise, testing will be stopped for this treatment group. All treatment groups will follow the same testing procedure. In addition, the difference of the response rates between each active treatment and the placebo will be estimated and their corresponding 95% confidence interval will be calculated. All subjects who discontinue prematurely during the treatment period or receive prohibited medications (refer to Section 8.3) will be counted as non-responder at subsequent visits.

For each of the binary endpoints, the estimate and its corresponding 95% confidence interval will be calculated for the difference of the proportions between each active treatment arm and the placebo arm at the specified visit, similar to the primary analysis except no p-values will be provided.

The construction of confidence intervals for binary variables will be based on normal approximation, if the number of the events in each individual treatment arm is at least 5. Otherwise, confidence interval using an exact method will be provided. The exact confidence intervals for binary variable within treatment group and between treatment groups will be constructed.

For each of the continuous endpoints, the mixed effect model will be fit with treatment and visit as the fixed effects and measurements within each subject as the repeated measurements. The baseline value will be added into the model as a covariate if necessary. Based on the mixed effect model, the estimate and 95% confidence interval will be calculated for the difference between each active treatment and the placebo at each specified visit.

The results will be presented by treatment groups unless otherwise specified.

7.2 Study Conduct

Relevant protocol deviations, which could have an impact on the primary/key efficacy endpoint, will be identified for all subjects who are randomized and receive study medication during the double-blind period. Details of relevant protocol deviations are provided in Appendix 1.

All subjects with relevant protocol deviations that could affect the primary efficacy will be identified prior to database lock and unblinding of treatment assignment. All relevant protocol deviations will be listed and summarized by treatment group. The subjects identified as relevant protocol deviation will not be excluded from any analysis population described in Section 6.3.

7.3 Study Population

Frequency distributions or summary statistics of data pertaining to subject disposition, demographic characteristics, baseline disease characteristics, and medical history will be tabulated and displayed by treatment group. No statistical test will be carried out for comparison of any baseline measures among treatment groups.

These presentations will be provided by randomized treatment group for the MITT analysis population unless specified otherwise.

7.3.1 Subject Disposition

The number of subjects enrolled into the study at screening, the number of subjects randomized, the number of subjects entering the double-blind period (DB) and the number of subjects who complete the double-blind period will be summarized by treatment group. The number of subjects discontinuing from the DB period (subject disposition) will also be summarized, together with the reasons for discontinuation, taken from the CRF status pages. The subject disposition for the DB period will be based on the MITT analysis population.

7.3.2 Demography and Baseline Characteristics

Demographic and baseline disease characteristics will be summarized by randomized treatment group and overall based on the MITT analysis populations. For continuous variables, they will be summarized using means, standard deviations and ranges. The distribution of categorical variables such as gender, race etc. will be summarized by treatment group using frequency and percentage.

Demographic characteristics include the following:

- Age in years
- Weight in kgs
- Gender: Male, Female

- Race
- Ethnicity (Hispanic and Non-Hispanic)
- Region grouping scheme A (Latin America+East Europe and ROW) [details in Appendix 2]
- Region grouping scheme B (North America, Latin America, Europe and ROW) [details in Appendix 2]
- Stratification (MTX-IR vs. TNF-IR)

Baseline clinical characteristics include the following:

- Duration of disease in years
- Duration of disease (<=2 years, > 2 to 5 years, > 5 years to 10 years, > 10 years)
- ACR functional status classification
- RA classification (2010 ACR/EULAR criteria)
- Individual scores of each ACR components
- DAS28-CRP
- DAS28-ESR
- RF Status: Negative versus Positive
- ACPA Status: Negative versus Positive
- hsCRP (mg/L)
- ESR (mm/hr)
- MTX weekly dose
- Oral daily steroid dose (prednisone equivalents) in mg at Day 1
- Prior non-biologics DMARD therapy: Categorized number of non-biologic DMARDs use prior to randomization (0, 1, 2, >=3)
- Number of subjects with inadequate and/or intolerance to TNFi (overall Yes/No, then categorized by intolerant, no response, insufficient response, and loss of response. Note: subject can be included in more than one category)
- Number of subjects with a primary failure (i.e., insufficient response to the first course of Therapy) (Yes, No)
- Number of subjects with TNFi failures: 0, 1, 2, or >2

7.3.3 Medical History and Prior Medication

Medical history and prior medication information before the first study medication will be summarized by treatment group using relative frequencies. Listings of medical history and prior medication will be presented. The results will be based on the MITT analysis population.

7.4 Extent of Exposure

7.4.1 Study Therapy

In general, extent of exposure will be summarized for the As-treated analysis population in two ways:

- by the number of study drug doses. Frequency distribution of number of doses will be summarized
- by the number of days the subject is known to be on study drug, ignoring any dosing interruptions

In general, exposure to study drug during the double-blind (DB) period in days is calculated as:

For subjects who discontinue study medication,

- Exposure = (date of last DB dose - date of first DB dose + 1) + 30.

The offset of 30 days represents 30 days of follow-up after last dose and is a lot longer than 5 times of half-life of the study medication. The frequency will be presented according to the duration ranges (in days) : <28, 29-56, 57-84 and so on.

7.4.2 Discontinuations from Study Therapy

Discontinuation from study therapy is defined as subject's termination of the study medication without resumption prior to study completion. Discontinuation from study therapy during the treatment and DB period will be summarized by reason for the premature termination, taken from the study status case report form (CRF) pages. These summaries will be provided based on the MITT analysis population by randomized treatment group and overall.

7.4.3 Treatment Compliance

The CRF for this study will capture information on doses of the study medication. All subjects who skip any study drug dose will be listed. The number of subjects with missed doses (excluding missed doses due to premature discontinuation from the study) by number of missed doses will be summarized based on the As-treated analysis population for the DB period by treatment groups.

7.4.4 Anti-rheumatic Medication

For the purpose of this analysis plan, the term DMARDs will refer to non-biologic DMARDs. Anti-rheumatic medication summaries will be provided to present the medication use of corticosteroids, NSAIDs, and DMARDs. Number and percent of subjects receiving corticosteroids, NSAIDs, and DMARDs will be summarized by treatment group for MITT analysis population during the DB period.

In addition, the mean weekly MTX dose at scheduled visit and mean daily dose of corticosteroids during the DB period will be summarized.

A comprehensive listing of all concomitant medications will be also provided for the MITT analysis population.

7.5 Efficacy

The efficacy analyses will be performed using the efficacy analysis population. The results will be grouped by "as randomized" treatment groups, unless otherwise specified. No formal statistical testing was pre-specified for other efficacy endpoints except the primary endpoint of ACR20 and ACR70 response at Day 85. The analyses of binary endpoints and continuous endpoints will be provided over time including the milestone visits (e.g. Day 85).

Note: If the enrollment is terminated for all treatment arms based on Week 4 interim analysis, the sensitivity analyses, per-protocol analyses, sub-group analyses and additional analyses as defined in the following sections may not be performed based on smaller sample size.

7.5.1 Primary Efficacy Endpoints in Double-blind Period

The co-primary efficacy endpoints are ACR20 and ACR70 response rate at Day 85.

The Chi-square tests will be used to compare the ACR20/70 response rates at Day 85 between each of the active treatment arms and the placebo arm. The Chi-square p-value will be provided for each comparison between the active treatment and the placebo arm. Within each treatment group, the statistical testing will be performed for ACR20 at level of 0.05 first. If it is statistically significant, then the statistical testing will be performed for ACR70. Otherwise, testing will be stopped for this treatment group. All treatment groups will follow the same testing procedure.

In addition, the difference of the response rates between each active treatment and the placebo will be estimated and their corresponding 95% confidence interval will be calculated for Day 85 visit and other applicable visits.

The point estimate for ACR response rate along with 95% confidence interval will be calculated for each treatment arm for each applicable visit.

All subjects who prematurely discontinue the trial after receiving study medication will have missing data imputed as non-responder at all scheduled protocol visits subsequent to the point of discontinuation. The analysis will be preformed based on Efficacy Analysis Population analysis population. More details of imputation rule for classification of binary response status are provided in Sections 8.3.

A senstivity analysis of statistical testing may be preformed using the Cochram-Mantel-Haenszel(CMH) Chi-square test with stratification factors (prior treatment failure status and region). The CMH Chi-square p-value will be provided for each comparison between the active treatment and the placebo arm along with the relative risk and its 95% CI. Similar testing procedure as the primary analysis will be conducted. The absolute treatment difference and its 95% CI will be provided using the mimimum risk Weights[8]. In case in a given stratum, there are 0 responders (all non-responders) in the compared treatment groups, 0.5 will be added to (subtracted from) both treatment groups in order to avoid a variance of 0 in that stratum.

7.5.2 Secondary Efficacy Endpoints in Double-blind Period

For each of the binary endpoints (DAS28-CRP<2.6, CDAI<=2.8 etc.), the estimate and its corresponding 95% confidence interval will be calculated for the difference of the proportions

between each active treatment arm and the placebo arm for each applicable visit, similar to the primary analysis except no p-values will be provided.

For each of the continuous endpoints (DAS28-CRP, CDAI, etc.), the mixed effect model will be fit with treatment and visit as the fixed effects and measurements within each subject as the repeated measurements. The baseline value will be added into the model as a covariate if necessary. Based on the mixed effect model, the estimate and 95% confidence interval will be calculated for the difference between each active treatment and the placebo at each specified visit. No p-values will be provided. In addition, descriptive summary statistics of mean change from baseline and 95% confidence interval for each treatment arm will be provided. The missing data in continuous measures will be handled via utilization of mixed model procedure in SAS.

7.5.2.1 Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging will be performed at baseline, Week 4 and Week 12 (or at the time of discontinuation, e.g., early termination visit [ET]). The independent read of MRI image data will be performed by two primary readers who have no knowledge of the subject's identity, treatment arm or time point chronologic sequence. The readers will be presented with the image set(s) of the most clinically swollen hand/wrist acquired at baseline and of that same hand/wrist acquired at Week 4 and Week 12 for the assessment of four parameters of bone erosion, bone marrow edema (osteitis), synovitis, and cartilage loss (joint space narrowing). The scoring of four parameters will be assessed according to the OMERACT RAMRIS criteria[9].

The MRI score for each parameter is the sum of scores at joints/bone ends where applicable by each reader. The MRI score that is used in the analyses is calculated as the average of MRI scores provided by the two readers for each parameter per each subject at each visit. If a MRI score from one of the readers is missing, then MRI score from the other reader will be used. If the score is missing for both readers at any given time point, the average score is considered missing. If a case requires adjudication, a consensus read of all four parameters will be performed by the two readers working together as defined in the charter. The consensus score will be used in the analysis in this case for each of the four parameters.

If there are more than 20% of joints/bone ends with missing score, the MRI score of each parameter will be considered missing. If there are less than or equal to 20% of joints/bone ends with missing score, the MRI score of each parameter from the missing joints/bone ends will be carried forward from previous scan (e.g., baseline to Week 4, Week 4 to Week 12), or carried backward from later time point or early termination visit (ET) scan if previous scan is not available (e.g., Week 12 to Week 4). If subjects had missing complete MRI assessment at Week 4 or 12, their assessment will be imputed via linear in/extrapolation. Imputation will require a subject to have available scores at minimum 2 visits (including baseline, Week 4, Week 12 or early termination (ET) visit). Subjects with no baseline or only baseline scores will be excluded from the analysis.

The following analyses will be performed based on imputed data unless specified otherwise. In addition, the analysis of changes from baseline may be performed based on as-observed data to assess the robustness of results at the presence of missing data.

The summary of changes from baseline to Week 4 and 12 in each of four parameters will be provided by treatment group using a mixed model. The proportion of subjects with new erosion count (defined as erosion score =0 at baseline and erosion score >0 at post baseline) per hand/wrist will be also summarized at Week 4 and 12.

In addition, the proportion of responders who show no progression of damage based on each of four parameters will be summarized by treatment group at Week 4 and 12 using point estimate and 95% CI, and point estimate and 95% CI of the difference in proportions between the treatment groups. The responder at Week 4 and 12 is defined as the change score from baseline is less than negative value of the smallest detectable change (SDC)[10]. The SDC is calculated by $1.96 \times SD_{\Delta(change-scores)}/(\sqrt{2} \times \sqrt{k})$ in which $SD_{\Delta(change-scores)}/\sqrt{2}$ represents the measurement error of single change-scores and k=2 represents the number of readings or raters used for the actual analyses of the trial. The calculation of SDC will be based on the data from two readers excluding the cases that require adjudication.

7.5.3 Subgroup Analyses

Subgroup analyses of ACR 20/50/70 responses, and DAS28-CRP/DAS28-ESR change from baseline will be provided by point estimates and 95% CI by treatment group at Week 12. The subgroup analyses will be performed using Efficacy analysis population. Subgroup factors and breakdown are listed in Table 2.

Table 2: Subgroup Analyses Factors and Categories

Subgroup Factor	Categories
Age	< 65 Years Old
	≥ 65 Years Old
Gender	Male
	Female
Race	White
	Black
	Asian
	Other
Geographic Region Grouping Scheme A[a]	Latin America+Eastern Europe
	North America+Western Europe+ROW
Geographic Region Grouping Scheme B[b]	North America
	Latin America
	Europe
	ROW
Prior treatment	MTX-IR
	TNF-IR
# of TNFi failure	=1
	=2

Table 2: Subgroup Analyses Factors and Categories

Subgroup Factor	Categories
Duration of Rheumatoid Arthritis	≤ 10 Years
	> 10 Years
Baseline DAS28-CRP	Baseline DAS28-CRP > 5.1
	Baseline DAS28-CRP ≤ 5.1

[a] See region/country listin in Appendix 3.

7.5.4 Sensitivity Analyses

If more than 10% of total subjects in the efficacy analysis population discontinued due to any reasons other than Lack of Efficacy (LOE), AE or Unknown reason, a sensitivity analysis may be performed. The subjects who have missing response due to premature discontinuation with reasons other than LOE, AE or Unknown will be excluded from the analyses, hence defining the sensitivity analysis population. The sensitivity analysis will be carried out for the primary and selected secondary endpoints (ACR20/50/70 response, DAS28-CRP and DAS28-ESR change from baseline) based on the sensitivity analysis population.

7.5.5 Per-Protocol Analyses

If at least 10% of total subjects in efficacy analysis population demonstrate relevant protocol deviations, a per-protocol analysis may be performed. A Per-protocol (PP) analysis population will exclude all subjects with at least one relevant protocol deviation (defined in Appendix 1) from the efficacy analysis population. A per-protocol analysis will be carried out for the primary and selected secondary endpoints (as defined in section 7.5.4) based on the PP analysis population. The grouping scheme will be the same as in the primary efficacy analysis.

7.6 Safety

Analysis of all safety data will follow the BMS standard safety data conventions[11] [12] and supplements to the standard conventions as defined in this document.

The evaluation of drug safety is based primarily on clinical AE, laboratory abnormalities and vital signs reported during the study.

The summaries of AEs and laboratory marked abnormality during DB period will be provided by treatment groups. AEs and laboratory marked abnormalities will be included in the analysis of the DB period if the onset date is on or after first dose date and event occurs within 30 days post last dose date (DB period).

Frequency distribution and individual listings of all AEs will be generated. Laboratory marked abnormality using pre-defined abnormality criteria will also be descriptively summarized. There will be no statistical testing of group difference with respect to frequencies of AEs or laboratory marked abnormalities.

All recorded adverse events will be listed and tabulated by system organ class, preferred term, treatment and overall. Any physical examination findings will be listed. ECG, vital signs and clinical laboratory test results and corresponding change from baseline values will be listed and summarized by treatment. Values for ECG, vital signs and clinical laboratory test results outside the pre-specified criteria will also be listed and summarized. ECG readings will be evaluated by the Investigator and abnormalities, if present, will be listed.

Unless otherwise specified, baseline is defined as the last non-missing result with a collection date-time less than the date-time of the first dose of study medication, for laboratory results, vital signs and ECG.

Presentations for the DB period will be provided by "as treated" treatment group for as-treated analysis population, unless otherwise specified.

7.6.1 Adverse Events

The investigator will determine the intensity of each AE as mild, moderate, severe, or very severe. In addition, the investigator will determine the relationship of the AE to the administration of the study drug.

All AEs are coded and grouped into Preferred Terms (PT) by System Organ Class (SOC), using the latest approved version of the Medical Dictionary for Regulatory Activities (MedDRA) at the time of database lock. Listings and summaries will be based on the resulting SOCs and PTs.

AEs will be included in the frequency tabulations if they occur while a subject is on study medication up to and including 30 days after the last dose of study medication. These AE summaries will be based on proportions, which represent the number of subjects experiencing the AEs divided by the number of subjects received at least 1 dose of study medication in the current period.

All AE listings will indicate the unique subject identifier, age, gender, current treatment, the date of onset, the date of resolution, day of onset relative to the start of treatment, action taken, investigator's assessment of severity and relationship to study drug. Additional listings will be provided for adverse events leading to discontinuation and adverse events without recorded resolution. Summaries of adverse events will include adverse events, adverse events by intensity (based on CTC grade) and adverse events by relationship.

Where a subject has the same adverse event, based on preferred terminology, reported multiple times in a single analysis period, the subject will only be counted once at the preferred terminology level in adverse event frequency tables.

Where a subject has multiple adverse events within the same system organ class in a single analysis period, the subject will only be counted once at the system organ class level in adverse event frequency tables.

When a subject has the same adverse event, based on preferred terminology, reported multiple times in a single analysis period, the following criteria, in order of precedence, will be used to select the event to be included in summary tables:

- Relationship to study medication
- Intensity of event
- Onset date and time

When reporting adverse events by intensity, in addition to providing a summary table based on the event selection criteria detailed above, summary tables will also be provided based on the most intense event during the analysis period - independent of relationship to study medication. For these tables, the following criteria, in order of precedence, will be used to select the event to be included in summary tables:

- Intensity of event
- Onset date and time

Subjects will only be counted once in the 'Total' at their maximum intensity, regardless of SOC or PT.

Any reported AEs that occur more than 30 days after the last dose of study drug will be excluded from safety summaries. Any reported AEs that occur more than 30 days post the last dose of study drug will appear in the listing.

Summary information (the number and percent of subjects by treatment) will be tabulated for the following AEs:

- All AEs including clinical and laboratory AEs
- Most common AEs (reported in 5% of subjects or more in any treatment group)
- All related events
- Serious AEs (SAEs)
- Discontinuations due to AEs

Laboratory AEs are laboratory changes identified by the investigator as AEs and thus reported on the AE pages of the eCRF. The intensity of all related AE will be provided if enough data warrants the anlysis.

7.6.2 Adverse Events of Interest

The following adverse events of special interest will be listed and summarized by treatment groups during the DB period.

- Pre-specified liver function abnormalities
- Infections: All reported infections and infestations within the System Organ Class (SOC): *Infections and infestations.* The severity of serious infections will be summarized.
- Malignancy: All reported events defined in MedDRA Maintenance and Support Services Organization (MSSO) Structured MedDRA Query (SMQ) list.
- ~~Blood pressure related AE~~
- Hypertension: All reported events defined in pre-specified MedDRA code of hypertension.

- Neutropenia: All reported events defined in pre-specified MedDRA code of neutropenia.
- Thrombocytopenia: All reported events defined in pre-specified MedDRA code of thrombocytopenia.
- Demyelinating disorders: All reported events defined in MedDRA Maintenance and Support Services Organization (MSSO) Structured MedDRA Query (SMQ) list.
- Autoimmune diseases: All reported events defined in pre-specified MedDRA code of autoimmune disorders.

7.6.3 Deaths

All AEs with the outcome of death reported during the study will be listed together with a clinical narrative relating to the death of the subject. All reported deaths after a subject is enrolled will be listed separately by subject.

7.6.4 Other Serious Adverse Events

SAEs are captured on an eCRF page. All SAEs and related SAEs will be summarized by treatment group and listed together with a clinical narrative.

7.6.5 Adverse Events Leading to Discontinuation of Study Therapy

AEs leading to discontinuation of study drug are identified on the eCRF will be summarized and listed together with a clincial narrative.

7.6.6 Multiple Occurrences of Adverse Events

Several descriptive summaries of adverse events that takes into account the number of occurrences that an AE was reported by individual subjects will be provided. In order to prepare these summaries, the CRF data will be processed according to standard BMS algorithms to categorize each line of subject data as a new occurrence or a continuation of an existing event. This determination will be based upon onset and resolution dates. Each line of subject data will represent the maximum severity observed as well as the last known assessed relationship to study medication by the investigator.

The number and proportion of subjects who experienced an AE once or multiple times (0 Events, 1 Events, 2-3 Events, >=4 Events) will be summarized by treatment groups for most common AE (at least 5% subject in any treatment group) and event of special interest, infections.

Listing displaying the unique instances of all AEs will be provided.

Note: The unique instances of all AEs will be generated after duplicates have been eliminated and overlapping and contiguous occurrences of the same event have been collapsed.

7.6.7 Summaries of Laboratory Data

The results of all protocol-specified clinical laboratory tests, as well as laboratory results outside of the normal range, will be listed. Scheduled laboratory measurements and corresponding change from baseline values will be summarized by treatment and nominal visit for each laboratory test for the MITT analysis population.

The frequency of subjects with any marked laboratory abnormality will be presented by laboratory test for the DB period. The results are based on the As-treated analysis population. The criteria used for classifying laboratory test results as markedly abnormal will be listed.

7.6.8 Other Safety Considerations

7.6.8.1 ECG

Summary statistics (n, mean, standard deviation, median, minimum, and maximum) will be presented for each ECG parameters and the corresponding changes from baseline by treatment and visit. The baseline value is defined as the last measurement before the first dose. ECG readings will be evaluated by the Investigator and abnormalities, if present, will be listed.

The frequency distribution of subjects' maximum recorded post-dose QTcF, PR, QRS and ΔQTcF will be tabulated by treatment and summarized within the CSR text for the following ranges:

For QTcF: QTcF ≤ 450 msec, 450 msec < QTcF ≤ 480 msec, 480 msec < QTcF ≤ 500 msec, QTcF > 500 msec

For PR: PR ≤ 200 msec, PR > 200 msec

For QRS: QRS ≤ 120 msec, QRS > 120 msec

For ΔQTcF: ΔQTcF ≤ 30 msec, 30 msec < ΔQTcF ≤ 60 msec, ΔQTcF > 60 msec

Individual QTcF, PR, QRS or ΔQTcF values meeting these criteria will be flagged in the data listing.

7.6.8.2 Vital Sign

Vital signs parameters (systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate and body temperature) will be listed. Summaries of vital sign parameters will be provided for each vital sign parameter at corresponding visits by treatment and respective changes from baseline. Subjects with vital signs outside of a pre-specified range will also be listed.

The following criteria will be used to determine vital sign results that are outside of a pre-specified range, where changes from baseline are based on matched postural positions and are calculated as parameter value - baseline parameter value:

Table 3: Vital sign parameter cut-off values

Heart Rate(bpm)	Value > 100 and change from baseline > 30, or Value < 55 and change from baseline < -15
Systolic BP(mmHg)	Value > 140 and change from baseline > 20, or Value < 90 and change from baseline < -20
Diastolic BP(mmHg)	Value > 90 and change from baseline > 10, or Value < 55 and change from baseline < -10
Respiration(breaths/min)	Value > 16 or change from baseline > 10

Table 3: Vital sign parameter cut-off values

Heart Rate(bpm)	Value > 100 and change from baseline > 30, or Value < 55 and change from baseline < -15
Temperature (°C)	Value > 38.3°C or change from baseline > 1.6°C

7.6.8.3 Physical Examination Findings

All physical examination abnormal findings will be listed per subject by visit.

7.7 Pharmacokinetic Analysis

Summary statistics of Ctrough (mean, SD, geometric mean, % CV, median, min, and max) for the PK analysis population by treatment and study day.

The listing of Concentration of BMS-986142 by treatment and study day will be provided as well.

[redacted]

7.9 Interim Analyses

After about 20 subjects per treatment arm have completed at least 4 weeks of the treatment period or discontinued the treatment, an interim analysis will be conducted to assess the efficacy of the BMS-986142 dose arms and the overall safety. The interim analysis will use pre-specified efficacy, safety and PK endpoints. The results of the interim analysis will be reviewed by an unblinded sponsor team who is not involved in the study conduct and who will provide recommendations with regard to the adaptive design decisions to the blinded study team.

Population PK and exposure-response analysis may be performed to support the statistical Bayesian analysis and to facilitate decisions around dose selection. Population PK analysis to characterize the PK of BMS-986142 may be performed by pooling data from the present study with the Phase 1 study IMXXXXXX, in order to supplement the PK data as well as test for covariate effects, including but not limited to demographics, disease state, concomitant medications, etc. Exposure-response analysis may be performed using DAS28-CRP as the primary endpoint to characterize the relationship between measures of BMS-986142 exposure and reduction in DAS28-CRP. Details regarding the data evaluation and modeling methodology for interim analysis will be provided in a separate modeling and simulation analysis plan.

Efficacy analysis (DAS28) using Bayesian modeling and dose/exposure response analysis:

The interim analysis will be performed on all available data of DAS28 change from baseline up to Week 12 using a Bayesian predictive approach.[13] This analysis assumes that the DAS28 at Week 4 and at Week 12 in the subjects who have not yet been observed will be similar to what was observed for the subjects included at Week 4 and Week 12, respectively. As such the unobserved data will be simulated from the predictive distribution conditional on the interim data and the prior distribution of the treatment difference (using a non-informative prior). Under the Bayesian framework, the posterior distribution of the treatment difference (BMS-986142 active dose – Placebo) will be constructed to determine the predictive probability of ΔΔDAS28 improvement greater than a cutoff value at the planned end of the trial for each dose arm.

Note: ΔΔDAS28 improvement: The difference of change from baseline in DAS28-CRP between active dose arm and placebo arm. The lower is the DAS28 score, the better is the disease status.

The Bayesian predicative analysis and the dose/exposure response analysis will determine if new BMS-986142 dose(s) will be needed to fully characterize dose-efficacy relationship. Pre-specified efficacy endpoint (DAS28) will be assessed with exposure endpoints to guide the selection of new dose levels in different scenarios as described below. Considering the safety margin estimated from non-clinical studies, new dose levels based on the interim analysis will not be greater than 350 mg/day.

If the predictive probability of ΔΔDAS28 improvement being greater than 1.0 is high (eg, > 80%) in all active arm(s), the BMS team may recommend to maintain the study design and treatment arms. In this scenario, all enrolled subjects will continue the treatments without modifications and also be included in the final analysis for the existing dose arm. The BMS team may also recommend adding an additional dose arm at lower dose level than current active doses to explore the suboptimal dose. In this case, up to 41 subjects will be randomized into the suboptimal dose arm along with existing efficacious dose arms. The new suboptimal dose level may be selected based on dose/exposure response analysis. In support of adding new dose arm (suboptimal dose or middle dose(s) as descriped below), population exposure-response analysis may be conducted for the time-course of DAS28-CRP scores and/or selected safety endpoints collected until the interim analsyis using a mixed effects inhibitory Emax model, or other alternative models, with respect to time and BMS-986142 exposure.

If the predictive probability of ΔΔDAS28 improvement being greater than 1.0 is high (eg, >80%) in 2 BMS-986142 arms (mid and high dose arms), then the BMS team may recommend to maintain the study design and these 2 treatment arms. The enrolled subjects will be carried over in the existing dose arm and also be included in the final analysis for the existing dose arm. The subjects enrolled in the low dose arm will continue the treatment only if they have received at least 1 treatment and treatment is considered safe before the interim analysis is completed. However, no new subjects will be enrolled in this treatment arm. Based on the dose/exposure response analysis, new dose levels may be added to further characterize the steep portion of the dose-response relationship if the existing 2 consecutive dose levels show a relatively flat exposure response.

If the predictive probability of ΔΔDAS28 improvement being greater than 1.0 is high (eg, > 80%) in 1 BMS-986142 arm (high dose arm), then the BMS team may recommend to maintain the study design and this treatment arm. The enrolled subjects will be carried over in the existing dose arm and also be included in the final analysis for the existing dose arm. The subjects enrolled in the low and mid dose arms will continue the treatment only if they have received at least 1 treatment and treatment is safe before the interim analysis is completed. However, no new subjects will be enrolled in this treatment arm. Based on the dose/exposure response analysis, up to 2 new dose levels may be added to adequately characterize the dose-response relationship.

If the predictive probability of ΔΔDAS28 improvement being greater than 0.6 is low (eg, < 20%), that dose arm is deemed futile in terms of DAS28CRP change from baseline. However, the final futility will also incorporate the results from other efficacy endpoints as defined below.

The proposed stopping rules at the interim analysis, based on the futility assessment and the overall safety assessment, are as follows:

- Rule 1: If safety issues are identified in all dose arms or all dose arms are futile, stop the study
- Rule 2: Drop the dose arm(s) with safety issues identified
- Rule 3: Drop the futile dose arm(s) and the dose arm cannot be dropped until the lower dose arm has been dropped.

If the predictive probability of ΔΔDAS28 improvement being greater than 0.6 is lower than 20% in all active treatment arms, BMS team may recommend stopping further enrollment and continuing those subjects receiving therapy up to 12 weeks.

Descriptive efficacy analysis (other efficacy endpoints):

In addition to Bayesian predictive analysis above, the following as-observed clinical data up to database lock will be summarized at the time of Week 4 interim analysis to facilitate decision-making.

- ACR20/50/70 response over time
- DAS28-CRP < 2.6, DAS28-ESR < 2.6, CDAI ≤ 2.8, SDAI ≤ 3.3 and Boolean remission over time
- Change from baseline in DAS28-CRP, DAS28-ESR, CDAI, SDAI over time.

Descriptive safety analysis:

For the following pre-specified safety endpoints:

- AE/SAE (including but not limited to hypertension, diarrhea, neutropenia)
- Change from baseline in vitals including systolic blood pressure (SBP) and diastolic blood pressure (DBP).

The study will be stopped if intolerable safety events are identified in all treatment arms. If intolerable safety events are identified in 1 or 2 treatment arms, these arms will be dropped and

the study will be continued with caution. New dose levels may be added based on the exposure response analysis (not greater than 350 mg daily). If intolerable safety events show a clear relationship with higher exposure of BMS-986142, a lower dose will be selected where the predicted exposure in ~ 90% of subjects does not exceed the exposure ranges associated with safety findings.

Based on the totality of safety and efficacy considerations, unsafe dose arm will be stopped immediately. The dose arm(s) that are projected to be ineffective will continue till the end of double-blind period for subjects already taking at least one dose of study medication and no new subjects will be enrolled into ineffective dose arm(s).

[redacted]

8 CONVENTIONS

8.1 Calculations of Key Measures

8.1.1 Evaluation of ACR

ACR 20, the American College of Rheumatology (ACR) definition of 20% improvement is based on a 20% improvement (compared to baseline (Day 1) values) in tender and swollen joint counts and 20% improvement in 3 of the remaining 5 core set measures (subject global assessment of pain, subject global assessment of disease activity, physician global assessment of disease activity, subject assessment of physical function and one acute phase reactant value (CRP)).

If ACR 20 response based on the criteria above, was not met, because no 20% improvement in tender and swollen joints, then the patient is an ACR20 non-responder.

If ACR 20 response, based on the criteria above was not met, because not 3 of the remaining 5 core set measures had 20 % improvement then the number of components with 20% or more improvement (for the 5 remaining core components) will be tabulated in relation to the number

of missing components (in the remaining 5 core components) prior to the final calculation of ACR 20. If the sum of the number of components with 20% or more improvement and the number of missing components is greater than or equal to 3, then ACR 20 will be set to missing as there is insufficient amount of data to determine the correct status, otherwise the patient is an ACR20 non responder.

Higher values for each component of the ACR composite variable indicate greater severity of disease. Improvement from baseline in a component of the ACR composite variable will be computed as the difference between the baseline value and the value at a given post-baseline visit. A positive value for improvement from baseline for an individual component indicates lesser severity of disease.

Percent improvement will be computed as the ratio of improvement from baseline to the baseline value times 100 for an endpoint:

% Improvement = 100 * (Improvement from Baseline) / Baseline Value

= 100 * (Baseline Value - Post-baseline Value) / Baseline Value.

Based on the percent improvement from baseline in each of the ACR components, ACR 50 and ACR 70 will be similarly defined with 50% and 70% improvement, respectively.

The conventions for dealing with missing data, procedures or intra-articular injections of joints (has impact on tender and swollen joint counts), restricted medications and high dose of corticosteroids are provided in Section 8.2, 8.3, 8.5.1.

8.1.2 Evaluation of DAS28-CRP and DAS28-ESR

The DAS used in this protocol is the DAS using the 28-count subsets of tender/painful joints and swollen joints, together with either CRP or erythrocyte sedimentation rate ESR, to derive the, DAS28-CRP, and DAS28-ESR, which are calculated using the following formulae, respectively:

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

DAS28-ESR = 0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.70*ln(ESR) + 0.014*GH.

where TJC28 is number of painful joints out of 28 joints, SJC28 is number of swollen joints out of 28 joints, GH is the general health or patients' global assessment of disease activity on a 100 mm VAS, ln is the natural logarithm, ESR is in mm/hour, and CRP is in mg/L.

8.1.3 Evaluation of CDAI

Calculation of Clinical Disease Activity Index (CDAI) is the simple linear sum of the outcome parameters: tender joint count (TJC) and swollen joint count (SJC) based on a 28-joint assessment, patient global assessment of disease activity (PGA: VAS 0–10 cm), and physician global assessment of disease activity (EGA: VAS 0–10 cm):

CDAI = TJC28 + SJC28 + PGA + EGA.

CDAI Remission responder is defined as a CDAI score less than or equal to 2.8.

8.1.4 Evaluation of SDAI and Boolean-based Remission

Calculation of Simple Disease Activity Index (SDAI) is the simple linear sum of the outcome parameters: tender joint count (TJC) and swollen joint count (SJC) based on a 28-joint assessment, patient global assessment of disease activity (PGA: VAS 0–10 cm), physician global assessment of disease activity (EGA: VAS 0–10 cm) and C-reactive protein (CRP in mg/dL) as shown below:

SDAI = TJC28 + SJC28 + PGA + EGA+ CRP.

SDAI Remission responder is defined as a SDAI score less than or equal 3.3.

Remission by Boolean-based definition: Subject must satisfy all of the followings:

TJC28 <=1; SJC28<=1; PGA<=1; CRP<=1 mg/dL

8.1.5 Evaluation of HAQ Disability Index

Scoring conventions for the Standard Disability Index of HAQ will be followed. The Standard disability index (HAQ DI) takes into account the subject's use of aids or devices or assistance in the scoring algorithm for a disability category. The score for each category is the single response within the category with the highest score. For each of the eight disability categories these is an AIDS OR DEVICES companion variable(s) that is used to record the type of assistance, if any, a subject uses for his/her usual activities. If aids or devices and/or assistance from another person are checked for a disability category, the score for this category is set to 2 (much difficulty), if the original score is 0 (no difficulty) or 1 (some difficulty). The HAQ DI is then calculated by summing the adjusted categories scores and dividing by the number of categories answered. The HAQ DI cannot be calculated if the patient does not have scores for at least 6 categories. Details of the HAQ DI scoring conventions are documented in "The Health Assessment Questionnaire"[14].

A HAQ response for RA is defined as a subject with a reduction from baseline in HAQ DI of at least 0.22.

8.1.6 Evaluation of FACIT-Fatigue

The FACIT-Fatigue is a 13 item subject self-assessment tool that measures the subjects' usual daily activities over the post week. Responses of 'not at all', 'a little', 'somewhat', 'quite a bit', and 'very much' are available for each question, and correspond to scores of 0, 1, 2, 3, and 4, respectively for the item codes An5 and An7 and correspond to scores of 4, 3, 2, 1, 0 (reversal of response) for the other item codes.

The total score = (sum of scores) *13/ (number of items answered), resulting in a maximum of 52 and a minimum of 0. The total score cannot be calculated if the patient does not have answered at least 7 of the 13 items (more than 50 % of the items should be answered).

8.1.7 Evaluation of BRAF-MDQ

The BRAF MDQ (Bristol Rheumatoid Arthritis [RA] Fatigue Multi-Dimensional Questionnaire)[15] was developed to assess the overall experience and impact of RA fatigue, and its different dimensions.

Items scored 0–3, except for items 1 (scored 0–10), 2 (scored 0–7), and 3 (scored 0–2). A total fatigue score is obtained by summing the 20 item scores. Subscale items are summed to produce scores for physical fatigue, living with fatigue, cognitive fatigue, and emotional fatigue. Instructions for missing data are that only 3 questions may be omitted in total, questions 1 and 2 must be completed, and only 1 question may be omitted from each subscale (replaced with patient's average score for that subscale).

Higher scores reflect greater severity. Total fatigue score is 0–70; subscale scores are physical fatigue 0–22, living with fatigue 0–21, cognitive fatigue 0–15, and emotional fatigue 0–12.

Figure 2: BRAF-MDQ Score table

Dimension		Questions	Range	Score
Physical	1	NRS fatigue	0-10	
	2	How many days?	0-7	
	3	How long on average has each episode of fatigue lasted?	0-2	
	4	Have you lacked physical energy because of fatigue?	0-3	
			Physical total (0-22)	
Living	5	Has fatigue made it difficult to bath or shower?	0-3	
	6	Has fatigue made it difficult to dress yourself?	0-3	
	7	Has fatigue made it difficult to do your work or other daily activities?	0-3	
	8	Have you avoided making plans because of fatigue?	0-3	
	9	Has fatigue affected your social life?	0-3	
	10	Have you cancelled plans because of fatigue?	0-3	
	11	Have you refused invitations because of fatigue?	0-3	
			Living total (0-21)	
Cognition	12	Have you lacked mental energy because of fatigue?	0-3	
	13	Have you forgotten things because of fatigue?	0-3	
	14	Has fatigue made it difficult to think clearly?	0-3	
	15	Has fatigue made it difficult to concentrate?	0-3	
	16	Have you made mistakes because of fatigue?	0-3	
			Cognition total (0-15)	
Emotion	17	Have you felt you have less control because of fatigue?	0-3	
	18	Have you felt embarrassed because of fatigue?	0-3	
	19	Has being fatigued upset you?	0-3	
	20	Have you felt down or depressed because of fatigue?	0-3	
			Emotion total (0-12)	
			BRAF-MDQ Total score (0-70)	

Missing data is handled as follows:

- Questions 1 and 2 are compulsory.
- Only 1 question may be missing from each dimension (maximum of 3 in the overall BRAF-MDQ).
- Replace the missing question score with the average score for that dimension.
- For the Physical Fatigue dimension, a weighted average score is used to account for the varying item score ranges: Total the 3 completed scores, divide by the total max possible score for those 3 questions, then multiply by the maximum score possible for all 4 questions (22).

For example:

- Q1 is 10/10, Q2 is 6/7, Q3 is missing, and Q4 is 2/3.
- Total completed score =18.
- Divide by total max possible for those 3 questions (10+7+3) i.e. 18/20 = 0.9 weighted average.
- Multiply by the max possible score for all 4 questions i.e. 0.9 x 22 = 19.8.
- Physical Function score is thus imputed as 19.8.

8.1.8 Evaluation of SF-36 Summary Functions and Components

The SF-36 is composed of 36 items measuring 8 health concepts: Physical Functioning, Role-physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-emotional, and Mental Health. These 8 scales form 2 summary scores: (1) Physical Component Summary (PCS) and (2) Mental Component Summary (MCS). Scoring conventions for the SF-36 scales and components are documented in "SF Health Outcomes ™ Scoring Software User's Guide"[16].

8.2 Handling of Procedures, Conditions and Intra-articular Injections of Joints

Joints may have been undergone procedures (e.g., fusion/ankylosis, synovectomy, steroid injection, and amputation) due to worsening RA during the study. Joint status is adjusted according to the following rules:

- If a joint undergoes procedure on or before the start of study drug, then for this joint both tenderness and swelling will be considered as 'not present' for all visits.
- If a joint undergoes procedure due to RA involvement and occurs after the start of study drug, then for this joint both tenderness and swelling will be considered as 'present' for all visits on or after the procedure.
- If a joint undergoes procedure for reason other than RA involvement and occurs after the start of study drug, then for this joint both tenderness and swelling will be considered as 'not present' for all visits on or after the surgery procedure date.

In addition, joint status is also adjusted for conditions and IA injection as follows:

- If a joint undergoes other procedures or conditions at any visit, then for that joint both tenderness and swelling will be considered as 'not present' at that particular visit.
- If an intra-articular injection occurs, then for that joint both tenderness and swelling will be considered as 'present' at that particular visit and all subsequent visits.

8.3 Determination of Responder Data

Subjects who receive 1 or more IA or IM injection during the double-blind period of study will be classified as ACR non-responder on and after the time of the first IA or IM injection and up to the end of the double-blind period. Subjects who receive a high oral steroids course that is defined as the use of either any oral dose (equivalent of prednisone) higher than baseline (Day 1) use for >=1 day or any oral dose (equivalent of prednisone) >10mg/day for >=1 day will be counted as non-responder at all visits subsequent to the date of dose that is higher than the baseline dose.

Subjects who increase MTX use (versus baseline) during the double-blind period will be counted as non-responder at all visits subsequent to the earliest date of increased MTX dose.

Subjects who receive any DMARDs (other than MTX) during the double-blind period will be counted as non-responder at all visits subsequent to the earliest date of DMARDs.

These rules will apply to binary endpoints including ACR20/50/70, DAS28-CRP <2.6, DAS28-ESR <2.6 and remission status defined by CDAI, SDAI and Boolean definition.

8.4 Baseline Measures

The baseline value is the last assessment taken prior to the first dose of study medication. In general, baseline value is the assessment taken on study Day 1 before the administration of study medication. If a measurement on Day 1 is missing, the last assessment taken at the screening period prior to the first dose of study medication will be used.

8.5 Missing Measurements

8.5.1 Imputation of Missing Data for Binary Variables

This imputation is applicable for all binary variables (ACR20/50/70, DAS28-CRP <2.6, DAS28-ESR <2.6, and remission status defined by CDAI, SDAI and Boolean definition).

If the response cannot be assessed due to missing data or a subject's early discontinuation, the following conventions will be implemented in a sequence:

- Any subject who prematurely discontinues the trial during the double-blind period after receiving study medication will have data imputed as non-responder at all scheduled protocol visits subsequent to the point of discontinuation up to the end of the double-blind period.
- After any imputation above or none, if for some reason the binary variable still can't be determined, including the case where baseline data is missing, then its value will be set to non-responder.

For listings, missing responses will be presented as missing.

8.5.2 Imputation of Missing Data for Continuous Variables

For the analysis of changes from baseline values (e.g., individual components of ACR, DAS28-CRP, and DAS28-ESR etc), missing values will be dealt with via the longitudinal repeated measures analysis assuming MAR (missingness at random).

For all safety (including laboratory measurements), and PK measures, missing values will not be imputed.

8.6 Missing, Unknown or Partial Dates

The BMS safety guidelines for conventions relating to the handling of missing or partial dates and the determination of appropriate default values in such cases (in particular, for concomitant medication dose start-dates and end-dates and AE onset dates) will be utilized.

8.7 Day Ranges for Analysis of Time Points

Subjects who do not always adhere strictly to the visit schedule timing in the protocol. Therefore the designation of visits during the double-blind period of the study will be based on the day of

evaluation relative to the trial (day of first study medication = study Day 1) rather than the nominal visit recorded in the eCRF. Mutually exclusive relative day windows are defined below to provide derived visits that correspond to the post-baseline time points. If a visit falls outside of the pre-specified visit windows, then the data collected at that visit will not be assigned a derived visit but will remain in the derived data sets. Determination of baseline values is addressed in Section 8.2.

If a subject has more than 1 visit where a measurement is recorded within a window, the measurement closest to the target day will be used. In case of 2 visits being equidistant from the target, the earlier measurement will be used in analyses. For these safety indicators, the least favorable value (toward a positive response) in the window will be used.

Designation of visits for efficacy assessments, and other measures during the double-blind period is tabulated below.

Table 4: Days Ranges for Assessments at Every Scheduled Visits During the Double-blind Period

Visit	Target Day	ACR20/50/70,DAS28-CRP, DAS28-ESR, HAQ	MRI	Fatigue, SF-36, Stiffness, BRAF-MDQ	Biomarker
Baseline /Day 1	1	1	1	1	1
D15	15	8-21			8-21
D29	29	22-43	15-57		22-43
D57	57	44-71			44-71
D85	85	72-99	58-113	43-113	72-99

9 CONTENT OF REPORTS

The results of the study conducted by Protocol IM006-016 will be presented in a standard BMS CSR appendix. Key results and any unanticipated findings that are unusual for this study will be identified. A meeting for the initial dissemination of study results will be held after database locked. Attendees at this meeting will review all efficacy and safety summaries and listings and will identify key results that should be highlighted in the CSR.

1.0	[redacted]	12/04/2015	Original version