Statistical Analysis Plan for

Official Title of Study

A Phase 3 Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of Subcutaneous Abatacept in Adults with Active Primary Sjögrens Syndrome

NCT02915159

17-June-2019

STATISTICAL ANALYSIS PLAN
FOR CLINICAL STUDY REPORT

A PHASE 3 RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY TO EVALUATE THE EFFICACY AND SAFETY OF SUBCUTANEOUS ABATACEPT IN ADULTS WITH ACTIVE PRIMARY SJÖGRENS SYNDROME

PROTOCOL(S) IM101-603

VERSION # 2.0

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 [redacted]

[redacted]

[redacted]

2 STUDY DESCRIPTION

2.1 Study Design

This is a randomized, double-blind, placebo controlled study in subjects who have been diagnosed with active, moderate to severe, primary Sjögren's Syndrome based on the 2016 ACR/EULAR Classification Criteria for pSS and have an ESSDAI disease activity score of at least 5 at screening. The study consists of a 28-day screening period, followed by a 169-day Double-Blind Period, a 197-day open-label abatacept period and a subsequent 169-day follow-up period.

Screening period:

The screening phase will last for a minimum of 7 days and a maximum of 28 days. Subjects who experience an acute infection, initiate treatment for latent TB or require additional time for the completion of screening assessments may extend the screening period to 56 days. If the screening period extends beyond 28 days, a second screening visit will be required within 28 days prior to randomization. Eligible subjects will have been diagnosed with Sjögren's syndrome based on the 2016 ACR/EULAR Classification Criteria (see protocol Appendix 3) and will have an ESSDAI score of at least five points at screening (see protocol Appendix 6). Subjects must be anti-SSA/Ro positive at screening and at least 124 subjects (72%) must have a stimulated salivary flow of at least 0.1mL/min at screening and randomization.

Subjects taking hydroxychloroquine are permitted to continue this medication provided that the therapy has been taken for at least 12 weeks and that the dose is stable for at least 4 weeks prior to randomization and remains stable throughout the study. Subjects receiving oral corticosteroids must be on a stable dose and at the equivalent of ≤ 10mg prednisone daily for at least 4 weeks prior to randomization. Subjects may not have received an IM, IV or IA administration of a corticosteroid within 4 weeks prior to randomization. Stable doses of pilocarpine, cevimeline, glaucoma eye drops, autologous serum eye drops and cyclosporine eye preparations, as well as artificial tears and saliva are permitted. Cyclosporine, lifitegrast, corticosteroid eye preparations, hyaluraonate eye drops or autologous serum eye drops should not be used for 12 hours prior to Day 1 and prior to the quarterly eye assessments. Pilocarpine and cevimeline must not be used for 48 hours prior to Day 1 and prior to the day of quarterly salivary and eye assessments. Glaucoma

eye drops must not be used for 8 hours prior to Day 1 and prior to the day of quarterly eye assessments. The use of ophthalimic and oral lubricants should not be used on Day 1 or on the day of the quarterly salivary and eye assessments (prior to the assessments). Study procedures will occur as specified in protocol Table 5.1-1.

Double-Blind Treatment Period:

On day 1, subjects will be randomized to one of two blinded treatment arms in a 1:1 ratio:

- Abatacept 125mg for subcutaneous administration weekly
- Matching placebo for subcutaneous administration weekly

The duration of the Double-Blind Treatment Period is 169 days. Abatacept (125mg) or matching placebo will be administered subcutaneously (SC) once per week. During this period, the dose of oral corticosteroids (≤ 10mg/day or equivalent), hydroxychloroquine, pilocarpine, cevimeline, cyclosporine eye drops, lifitegrast, autologous serum eye drops, oral and ocular lubricants and/or NSAIDs should remain stable. Analgesics are permitted with certain restrictions (see protocol Section 3.4.1.2). Study procedures will occur as specified in protocol Table 5.1-2, 5.1-4, 5.1-5.

Subjects who discontinue treatment with study drug during the Double-Blind Treatment Period must continue to be assessed for all efficacy measures and for safety at all remaining double-blind study visits through Week 24 (Day 169). In addition, subjects who discontinue treatment with the study drug will require visits to assess safety 28, 85 and 169 days after the last dose of study drug, and immunogenicity 85 and 169 days after the last dose of study drug. If these visits will fall within ± 14 days of a scheduled study visit, only an additional laboratory assessment (immunogenicity) may need to be performed at the scheduled study visit. Otherwise, the follow-up visits 28, 85, and/or 169 days after last dose of study medication need to be performed at the appropriate time.

Open-Label period:

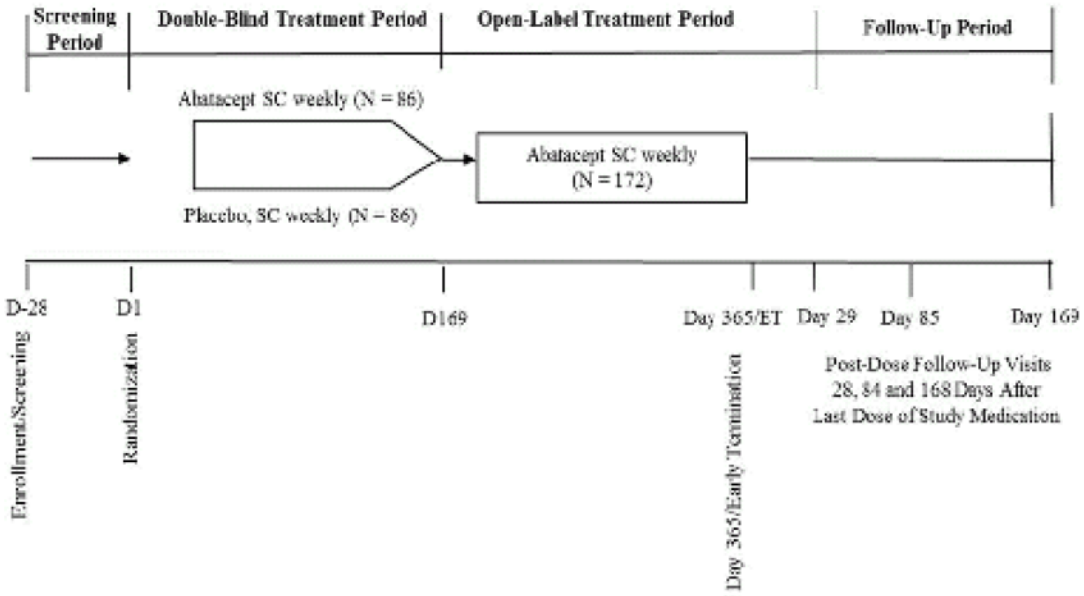
At the end of the double blind treatment period there will be a 197 day open label extension (OLE). On Day 169 all subjects continuing on treatment (including those subjects who received placebo during the double blind period) will receive open-label abatacept SC, 125mg weekly to Day 357. Subjects not wishing to continue on open-label abatacept have the option of discontinuing treatment at the end of the Double-Blind Period and proceeding to the post-treatment follow-up period. Study procedures will occur as specified in protocol Table 5.1-3.

Follow-Up Period:

Subjects who discontinue treatment of study drug during the double-blind or open-label periods or complete the study will have three follow-up visits 28, 85 and 169 days after the last treatment visit, to perform safety and laboratory assessments, including immunogenicity testing. Subjects who go into the post-study drug access program (i.e. switch from study drug to post-study abatacept) will be excluded from the Follow-up period.

The study design schematic is presented in Figure 2.1-1. On office visit days, SC injections should occur after all assessments (including efficacy, AE, blood draws for assessment of immunogenicity and drug concentrations) are done.

Figure 2.1-1: Study Design Schematic



The approximate duration of the study is up to a 4-week screening period (28 days), a 24-week Double-Blind Period (169 days), a 197-day open label abatacept period and a subsequent 24 week follow up period (169 days), for a total of approximately 80 weeks (561 days).

The start of the trial is defined as the date the first subject signs the informed consent. End of trial is defined as the last visit or scheduled procedure shown in the Time & Events schedule for the last subject. The study will continue beyond primary endpoint collection until the End of Trial.

2.2 Treatment Assignment

At the time of the screening visit, immediately after written informed consent is obtained and before performing any study-related procedures, the investigator or coordinator will contact IWRS designated by BMS for assignment of a 5 digit subject number that will be unique across all sites. Enrolled subjects, including those not dosed, will be assigned unique sequential subject numbers by the IWRS system starting with 00001, 00002, 00003, etc. for identification throughout the study. This subject number must not be reused for any other participant in the study.

After completion of all screening evaluations and concomitant adjustment or stabilization, all eligible subjects will be randomly assigned to 1 of 2 treatment arms (abatacept or placebo) in a 1:1 ratio. To randomize a subject, the system will be contacted and the randomization option of the IWRS will be used 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. Randomization will be stratified globally by current use of oral corticosteroids (yes/no).

Randomized schedules will be generated and kept by the Randomization Group within Drug Supply Management of Bristol-Myers Squibb.

At all study visits when study drug is dispensed, each subject will be assigned specific container numbers by the IWRS. Container numbers will be assigned non-sequentially and will correspond to the numbers printed on the containers and bottles containing study drug, and will be recorded on the appropriate eCRF.

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, i.e., 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 one of the active products. It is highly desirable that the decision to unblind treatment assignment be discussed with the BMS 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 the Central Randomization System (IWRS).

In cases of any accidental unblinding, the Study Director or Medical Monitor should be contacted and every attempt is made to preserve the blind.

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

The BMS Bioanalytical Science Department or its designee will be unblinded to the randomized treatment assignments in order to accurately perform sample analysis for the PK and immunogenicity samples.

2.4 Protocol Amendments

The protocol currently has 3 amendments and 1 administrative letter. The tables below summarize the main purpose of the amendments and the administrative letter.

Table 2.4-1: Amendments

Amendment No.	Amendment Date	Main Purpose of Amendment
1	25 Aug. 2016	• Changes to Protocol include: • To provide clear differentiate between discontinuation of study treatment from discontinuation from the study. • Subjects who discontinue investigational product during the Double-Blind Treatment Period should continue to comply with all protocol specified procedures during the Double-Blind Treatment Period. • Define the assessments required for the collection of follow-up data. • To clarify inclusion/exclusion criteria • To clarify testing for stimulated salivary flow • Add allowable window for obtaining the optional biopsy samples. • Modify laboratory testing to limit to the most critical tests and to adjust for feasibility issues. • Add additional secondary and exploratory endpoints. • Increase the total number of subjects targeted for enrollment in order to increase statistical power and the overall size of the safety database. • Revise appendices to include most current versions. • Corrections to minor typographical errors
6	13 Mar. 2017	• Changes to Protocol include: • Update contact information for the Medical Monitor, add Study Director and contact information. • Update the Study Design Section and Schematic to include Day 169. • Add stratification criteria for Japan. • Clarify potential reason for extended screening visit. • Clarify exclusion criteria for DMARDs and eye surgeries and clarify restrictions for corticosteroids • Allow for the use of glaucoma eye drops and autologous serum eye drops. • Clarify criteria for discontinuation of study therapy. • Update Study Assessments and procedure clarifications • Update the statistical section with stratification for Japan and HA request from Germany. • Add Appendix 14 for Topical corticosteroids (low potency/Class VI and VII) • Corrections to typographical errors and reconcile inconsistencies.
8	06 Jun. 2017	• Changes to the protocol include: • Clarify exclusion criterion for severe renal involvement.

Table 2.4-1: Amendments

Amendment No.	Amendment Date	Main Purpose of Amendment
		• Add exclusion criterion for herbal supplements and remedies. • Add restriction for use of contact lenses • Clarify corticosteroid use • Add restriction for use of hyaluronate eye drops • Add domain by domain guidance and/or clarification to non-protocol specified procedures and tests. • Add testing for complement (C3, C4 and CH50) to additional visits. • Add lab testing for serum protein electrophoresis and cryoglobulins to all visits at which the ESSDAI is assessed. • Corrections to typographical errors and reconcile inconsistencies.

Table 2.4-2: Administrative Letter

Administrative Letter No.	Administrative Letter Date	Main Purpose of Administrative Letter
1	22 Nov. 2016	Updated ACR EULAR Classification Criteria for Primary Sjögren's syndrome from proposed 2015 criteria to 2016 criteria. Corrected typographical error for stimulated salivary flow procedure Corrected Appendix 11 with correct example of PROMIS Fatigue questionnaire

2.5 Data Monitoring Committee and Other External Committees

A Data Monitoring Committee (DMC) will monitor overall safety data regularly to ensure that the benefits and risks of study participation remain acceptable. Based on the regular reviews of emerging data, the DMC may recommend to the Sponsor alteration and/or termination of the trial.

Data summaries and listings will be provided to the DMC to facilitate their safety assessment at the regularly scheduled times as well as on an ad hoc basis if needed. The DMC will review safety data including SAEs and events of special interest, focusing on early signal detection. Further details on the frequency, content, and methods of data reports to the DMC will be outlined in the Charter of that Committee along with the processes and procedures the committee will follow.

3 OBJECTIVES

3.1 Primary

To compare the mean change from baseline (Day 1) to Day 169 in ESSDAI of abatacept versus placebo in subjects with moderate to severe pSS.

3.2 Secondary

Key Secondary:

- To compare the mean change from baseline (Day 1) to Day 169 in EULAR Sjögren's Syndrome Patient Reported Index (ESSPRI) of abatacept versus placebo in subjects with moderate to severe pSS.
- To compare the mean changes from baseline (Day 1) to Day 169 in the stimulated whole salivary flow of abatacept versus placebo in subjects with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline.

Other Secondary:

- To assess the mean change from baseline at all measured time points up to Day 169 in DAS28-CRP among those with a tender joint count plus swollen joint count of at least 3 at baseline, among those with a tender joint count plus swollen joint count of less than 3 at baseline and in the full population.
- To assess the mean change from baseline at all measured time points up to Day 169 in the individual components of DAS28-CRP among those with a tender plus swollen joint count of at least 3 at baseline, among those with a tender joint count plus swollen joint count of less than 3 and in the full population.
- To assess the proportion of subjects who achieve a minimally clinically important improvement in ESSDAI (≥ 3) at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- To assess the proportion of subjects who achieve a clinically important improvement in ESSDAI (≥ 5) at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline
- To assess the proportion of subjects who achieve a minimally clinically important improvement in ESSPRI (≥ 1) at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- To assess the mean change from baseline at all measured time points up to Day 169 in ESSDAI in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min

at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

- To assess the mean change from baseline at all measured time points up to Day 169 in ESSPRI in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- To assess the mean change from baseline in the individual components of the ESSDAI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- To assess the mean change from baseline in the individual components of the ESSPRI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- To assess the mean change from baseline in Schirmer's test at all measured time points up to Day 169.
- To assess the mean change from baseline in ocular staining score at all measured time points up to Day 169.
- To assess the mean change from baseline in tear break-up time at all measured time points up to Day 169.
- To assess the mean change from baseline in unstimulated whole salivary flow at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- To assess the mean change from baseline in stimulated whole salivary flow at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- To assess the change from baseline in patient symptoms using the Numeric Rating Scale (NRS) for mouth dryness at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- To assess the mean change from baseline in patient symptoms using the Numeric Rating Scale (NRS) for eye dryness at all measured time points up to Day 169.
- To assess the mean change from baseline in patient symptoms using the patient global assessment of disease activity at all measured time points up to Day 169.
- To assess the mean change from baseline in the physician global assessment at all measured time points up to Day 169.
- To assess the mean change from baseline in patient fatigue using PROMIS Fatigue assessment at all measured time points up to Day 169.
- To assess the mean change from baseline in female sexual function using the Female Sexual Function Index (FSFI) at all measured time points up to Day 169.

- To assess the mean change from baseline in patient quality of life using the SF-36 (v2.0) at all measured time points up to Day 169.
- To assess safety and immunogenicity of abatacept.
- To assess the pharmacokinetics of abatacept.

[redacted]

4 ENDPOINTS

4.1 Primary Endpoint(s)

The primary endpoint is the mean change from baseline (Day 1) in ESSDAI at Day 169.

4.2 Secondary Endpoint(s)

Key Secondary Endpoints

- The mean change from baseline in ESSPRI at Day 169
- The mean change from baseline in the stimulated whole salivary flow at Day 169 among subjects with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline.

Other Secondary Endpoints

- The mean change from baseline at all measured time points up to Day 169 in the DAS28-CRP among those with a tender plus swollen joint count ≥ 3 at baseline, among those with a tender plus swollen joint count < 3 at baseline and in the full population.

- The mean change from baseline at all measured time points up to Day 169 in the individual components of DAS28-CRP among those with a tender plus swollen joint count ≥ 3 at baseline, among those with a tender plus swollen joint count < 3 at baseline and in the full population.
- Proportion of subjects who achieve a minimally clinically important change (of at least 3 points) in the ESSDAI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- Proportion of subjects who achieve a clinically important change (of at least 5 points) in the ESSDAI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- Proportion of subjects who achieve a minimally clinically important change (of at least 1 point) in the ESSPRI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline at all measured time points up to Day 169 in the ESSDAI in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline at all measured time points up to Day 169 in the ESSPRI in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in the individual components of the ESSDAI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in the individual components of the ESSPRI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in Schirmer's test at all measured time points up to Day 169.
- The mean change from baseline in ocular staining score at all measured time points up to Day 169.
- The mean change from baseline in tear break-up time at all measured time points up to Day 169.
- The mean change from baseline in unstimulated whole salivary flow at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

- The mean change from baseline in stimulated whole salivary flow at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in patient symptoms using the Numeric Rating Scale (NRS) for mouth dryness at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in patient symptoms using the Numeric Rating Scale (NRS) for eye dryness at all measured time points up to Day 169.
- The mean change from baseline in subject assessment of disease activity at all measured time points up to Day 169.
- The mean change from baseline in the physician global assessment of disease activity at all measured time points up to Day 169.
- The mean change from baseline in patient fatigue using PROMIS Fatigue assessment at all measured time points up to Day 169.
- The mean change from baseline in female sexual function using the FSFI at all measured time points up to Day 169.
- The mean change from baseline in patient function using SF-36 (v2.0) at all measured time points up to Day 169.
- Geometric mean of trough concentration (Cmin) of abatacept at all measured time points.
- Proportion of subjects with at least one positive immunogenicity response up to Day 169 and during 3 months follow up (for subjects who discontinue during the 6-months double-blind) and during the cumulative abatacept period and 3 months follow-up (for the cumulative abatacept population).
- Safety (proportion of subjects with adverse events, deaths, SAEs, and AEs leading to discontinuation and proportion of laboratory marked abnormalities) up to Day 169 and during the cumulative abatacept period.

[redacted]

5 SAMPLE SIZE AND POWER

A sample size of 172 patients (86 per arm) will achieve 98% power to detect a treatment difference of 3 in changes from baseline in ESSDAI at Day 169 between the abatacept and placebo group using a two-sided t-test with a significance level (alpha) of 0.05 and assuming a common SD of 4.8. The sample size of 86 patients per arm will achieve 90% power to detect a treatment difference of 1 in change from baseline in ESSPRI at Day 169 assuming a common SD of 2. Taking into account the hierarchical procedure (described in Section 7.5), a sample size of 86 patients per arm will achieve an overall power of at least 88.2% (98% multiplied by 90%) to detect both a treatment difference of 3 in mean change from baseline in ESSDAI and a difference of 1 in change from baseline in ESSPRI at Week 24 (Day 169).

The treatment differences of 3 for ESSDAI and 1 for ESSPRI correspond to the minimal clinically important improvement (MCII) for ESSDAI and ESSPRI, the key endpoints that address the major signs and symptoms of this disease. Estimates of the SD of mean change from baseline in ESSDAI derived from published results, range from 3.6 to 6.3. Estimates of the SD of mean change from baseline in ESSPRI derived from published results range from 1.7 to 2.2.

In addition, the power to detect a treatment difference in mean change from baseline of 0.165mL/min in salivary flow in subjects with stimulated whole salivary flow at baseline of at least 0.1mL/min is 91%, assuming a SD of 0.275 and for a sample size of 62 subjects per arm with a residual salivary flow of at least 0.1mL/min at screening and baseline. Taking into account the hierarchical procedure, a sample size of 86 patients per arm will achieve an overall power of at least 80.3 % (98% multiplied by 90% multiplied by 91%) to detect a treatment difference of 3 in mean change from baseline in ESSDAI, a difference of 1 in mean change from baseline in ESSPRI and a treatment difference of 0.165mL/min in mean change from baseline in salivary flow (in subjects with stimulated whole salivary flow at screening and at baseline of at least 0.1mL/min) at Week 24 (Day 169).

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

In this study, there are five study periods: Screening period, Double-Blind Period, Open-Label Period, Cumulative Abatacept Period and Follow-up Period. The following analysis periods defined below are for analysis purposes. The definitions below indicate the data to be included in the analysis of each specified period.

- **Screening period:** It covers the time period which starts from the day of enrollment and lasts until the initiation of the randomized Double-Blind Period.
- **Double-Blind Treatment Period:** It starts at the time of the first dose of blinded treatment and continues up to 56 days post last dose in the Double-Blind Period or prior to the first dose in the open-label period, whichever is earlier for safety (AE and laboratory summaries). Up to 42 days post last dose in the Double-Blind Period or prior to the first dose in the open-label period, whichever is earlier for efficacy analyses. Up to 21 days post last dose in the Double-Blind Period or prior to the first dose in the open-label period, whichever is earlier for on-treatment immunogenicity summaries.
- **Open-Label period:** It starts at the time of the first dose of open-label treatment and continues up to 56 days post last dose for safety (AE and laboratory summaries). Up to 42 days post last dose for efficacy analyses and up to 21 days post last dose for the on-treatment immunogenicity summaries.
- **Cumulative Abatacept Period:** It starts at the time of the first dose of abatacept (either in the Double-Blind Period for the subjects randomized to abatacept or in the open-label period for the subjects randomized to placebo) and continues up to 56 days post last dose of abatacept for safety (AE and laboratory summaries) and up to 21 days post last abatacept dose for the on-treatment immunogenicity summaries. The last abatacept dose can be last dose in Double-Blind Period for subjects not treated in the open-label period.
- **Follow-up Period**: immunogenicity summaries are provided for the follow-up period that starts at Day 22 post last dose and continues up to last assessment of immunogenicity in the study.

6.2 Treatment Regimens

During the Double-Blind Treatment Period all subjects will receive one of the following treatments:

- Abatacept 125mg for subcutaneous administration weekly
- Matching placebo for subcutaneous administration weekly

For analyses using the "as randomized" treatment group, a subject will appear in the treatment group to which the subject was randomized at the start of the Double-Blind Period (even if the treatment they received was different).

The "as treated" treatment group is the same as the "as randomized" treatment group, except in cases where information is available which indicates that a subject received a different treatment for the entire course of their participation in the study (or period). In this case, the "as treated" treatment group is set to the treatment the subject actually received.

During the open-label treatment period all subjects will receive abatacept 125mg for subcutaneous administration weekly.

6.3 Populations for Analyses

- Modified Intent-to-treat (ITT) analysis population: all randomized subjects who receive at least one dose of study medication. This population will exclude subjects who are randomized, but not treated. Given the blinded nature of this study, it is reasonable to assume that subjects who discontinue prior to the receipt of study medication do so for reasons unrelated to study medication and that excluding these patients will not bias the estimate of the treatment effect. Moreover, in previous studies, it happened very infrequently that subjects were randomized and not treated. Analyses using the Modified ITT analysis population will group the subjects according to the treatment group to which they are randomized. All efficacy analyses will use the Modified ITT population
- Intent-to-treat (ITT) analysis population: all randomized subjects. This population will also include randomized subjects who are never treated. Subjects who were randomized and never treated and did not return for any assessments after discontinuation (i.e. lost to follow-up or withdrew consent) will be excluded from the ITT analysis population. This population will be used for a sensitivity analysis of the primary and key secondary endpoints. Details are provided in Section 7.1. Analyses using the ITT analysis population will group the subjects according to the treatment group to which they are randomized.
- Per-protocol (PP) analysis population: all randomized subjects who receive at least one dose of study medication, excluding the subjects with relevant protocol deviations. Analyses using the PP analysis population will group the subjects according to the treatment group to which they are randomized.
- As-treated analysis population: all subjects who receive at least one dose of study medication. Analyses using the as treated analysis population will group the subjects on an as-randomized basis unless the subject received the incorrect medication for the entire Double-Blind Period. In that case, the subject will be analyzed in the treatment group associated with the incorrect medication they received ("as-treated").This population will be used for the safety analyses.
- Open-Label population: all randomized subjects who receive at least one dose of abatacept during the open-label period.
- Cumulative abatacept population: all randomized subjects who receive at least one dose of abatacept during the double-blind and/or the open-label period. All subjects of the 'As-treated population' will be included, except placebo subjects who never entered the open-label period. This population will be used for safety analysis.
- Immunogenicity analysis population: all subjects who receive at least one dose of study medication and who had at least 1 immunogenicity result reported after start of study medication.

- Full PK analysis population: all subjects who received at least one dose of abatacept study medication and who had at least 1 PK result reported after start of this study medication.
- Evaluable PK analysis population: this population is a subset of the PK analysis population and consists of the evaluable subjects for PK analysis. For all PK summaries and plots, a subject is evaluable for PK analysis at a specific day if the predose PK measurements were collected in the 4 to 10 days window after the previous SC abatacept dose and prior to the dose of the specific day.

7 STATISTICAL ANALYSES

7.1 General Methods

The stratification variables included in all the statistical models will be based on the IWRS stratification data used for the stratification of the randomization, while the subgroups used for the analyses (e.g., subgroup of subjects with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline) will be defined based on data entered on the CRF.

7.1.1 Change from Baseline

Change from baseline to any Visit t in Double-Blind Treatment Period is defined as follows:

$$C_{Visit\ t} = M_{Visit\ t} - M_{baseline}$$

where:

- $C_{\text{Visit t}}$ is the change from baseline at Visit *t*,
- $M_{\text{Visit t}}$ is the measurement at Visit *t*,
- M_{baseline} is the measurement at baseline.

The "Visit t" and baseline to which a measurement belongs is determined using the conventions described in Section 8.1

7.1.2 Percent Change from Baseline

Percent change from baseline to any Visit *t* in Double-Blind Treatment Period is defined as follows:

$$P_{Visit\ t} = 100 \times (M_{Visit\ t} - M_{baseline}) / M_{baseline}\ .$$

where:

- $P_{\text{Visit t}}$ is the percent change from baseline at Visit *t*,
- $M_{\text{Visit t}}$ is the measurement at Visit *t*,
- M_{baseline} is the measurement at baseline.

The "Visit t" and baseline to which a measurement belongs is determined using the conventions described in Section 8.1.

7.1.3 Longitudinal Repeated Measures Analysis

Primary Endpoint (ESSDAI)

Changes from baseline over time for the primary endpoint, ESSDAI, will be analyzed using a longitudinal repeated measures analysis. The model will include the fixed categorical effects of treatment, day (is a windowed time point), Baseline oral corticosteroid use (yes/no), baseline hydroxychloroquine use (yes/no), baseline stimulated salivary flow (< 0.1mL/min, ≥ 0.1mL/min)., Japan (yes/no), treatment-by-day interaction, baseline oral corticosteroid use-by-day interaction, baseline hydroxychloroquine use-by-day interaction, baseline stimulated salivary flow-by-day interaction, Japan-by-day interaction as well as, the continuous fixed covariate of baseline ESSDAI and baseline ESSDAI -by-day interaction. An unstructured covariance matrix will be used to represent the correlation of the repeated measures within each subject. The parameter estimations will be based on the assumption of data being missing at random (MAR) and using the method of restrictive maximum likelihood (REML). The degrees of freedom will be calculated using the Kenward-Rogers method. Adjusted mean changes (LSMeans) with corresponding SE and 95% CI per treatment arm will be provided. The difference in adjusted means (LSMEANS) between abatacept and placebo at the primary time point, Day 169, and corresponding 95% CI and p-value based on this longitudinal repeated measures model will be provided. The analysis is using the modified ITT analysis population and all efficacy measurements collected up to last dose in Double-Blind Period + 42 days (and prior to start of open-label treatment).

In case of non-convergence of the preferred longitudinal model, memory space issues, or other computational issues the following back-up models are defined and will be applied sequentially.

The first back-up model that resolves the computational issues will be used.

- In the first backup model, the repeated statement will specify a spatial power covariance matrix instead of the unstructured variance-covariance matrix; and intercept and slope will be specified as random effects on a subject level. All terms and interactions will be as for the primary model.
- In the second backup model, in addition to specifying a spatial power covariance matrix, the model will not include terms for the interactions between the stratification variables (baseline oral corticosteroid use, baseline hydroxychloroquine use, baseline stimulated salivary flow, Japan) and day. This means that the model will include the following terms: the fixed categorical effects of treatment, day (is a windowed time point), baseline oral corticosteroid use, baseline hydroxychloroquine use, baseline stimulated salivary flow, Japan, treatment-by-day interaction, and the continuous fixed covariate of baseline ESSDAI score and baseline ESSDAI score -by-day interaction.
- In the third backup model, the spatial power covariance structure will be used and the model will not include any terms for interactions between baseline variables and day. Therefore the model will include the following terms: the fixed categorical effects of treatment, day (is a windowed time point), baseline oral corticosteroid use, baseline hydroxychloroquine use,

baseline stimulated salivary flow, Japan, treatment-by-day interaction, and the continuous fixed covariate of baseline ESSDAI score.

Patterns of missing data from Day 1 up to Day 169 will be provided:

- Patterns of missing data for the primary endpoint of change from baseline in ESSDAI over time will be displayed separately for the two treatment groups up to Day 169. Data collected more than 42 days after the subject discontinued from the study will not be considered as missing for this summary.
- In addition, patterns of missing data including data collected more than 42 days after a subject discontinued study treatment up to Day 169 will be provided.

All patterns (patterns of monotone missing and non-monotone missing) observed will be summarized for each treatment group separately.

Key Secondary Endpoints (ESSPRI and Stimulated Salivary Flow)

Changes from baseline over time for the key secondary endpoint, ESSPRI, will be analyzed using a longitudinal repeated measures analysis. The model will include the fixed categorical effects of treatment, day (is a windowed time point), Baseline oral corticosteroid use (yes/no), baseline hydroxychloroquine use (yes/no), baseline stimulated salivary flow (< 0.1mL/min, ≥ 0.1mL/min), Japan (yes/no) treatment-by-day interaction, baseline oral corticosteroid use-by-day interaction, baseline hydroxychloroquine use-by-day interaction, baseline stimulated salivary flow-by-day interaction, Japan-by-day interaction as well as, the continuous fixed covariate of baseline ESSPRI and baseline ESSPRI -by-day interaction. An unstructured covariance matrix will be used to represent the correlation of the repeated measures within each subject. The parameter estimations will be based on the assumption of data being missing at random (MAR) and using the method of restrictive maximum likelihood (REML). The degrees of freedom will be calculated using the Kenward-Rogers Method. Adjusted mean changes (LSMeans) with corresponding SE and 95% CI per treatment arm will be provided. The difference in adjusted means (LSMEANS) between abatacept and placebo at the primary time point, Day 169, and corresponding 95% CI and p-value based on the longitudinal repeated measures above will be provided. The analysis is using the modified ITT analysis population and all efficacy measurements collected up to last dose in Double-Blind Period + 42 days (and prior to open-label treatment).

In case of computational or non-convergence issues of the model, the following back-up models will sequentially be applied and the first back-up model that resolves the issues will be used. First a spatial power covariance matrix will be used instead of an unstructured covariance. The second back-up model will use a spatial covariance matrix and remove the interactions with baseline stratification from the model. The third back up model will include the spatial power covariance matrix and remove all interactions between baseline and day from the model.

Changes from baseline over time for the key secondary endpoint, stimulated salivary flow, will be analyzed using a longitudinal repeated measures analysis. This analysis will be provided for the subgroup of subjects with stimulated whole salivary flow of at least 0.1mL/min at both screening

and baseline. The model will include the fixed categorical effects of treatment, day (is a windowed time point), Baseline oral corticosteroid use (yes/no), baseline hydroxychloroquine use (yes/no), Japan (yes/no), treatment-by-day interaction, baseline oral corticosteroid use-by-day interaction, baseline hydroxychloroquine use-by-day interaction, Japan-by-day interaction as well as, the continuous fixed covariate of baseline stimulated salivary flow and baseline stimulated salivary flow-by-day interaction. An unstructured covariance matrix will be used to represent the correlation of the repeated measures within each subject. The parameter estimations will be based on the assumption of data being missing at random (MAR) and using the method of restrictive maximum likelihood (REML). The degrees of freedom will be calculated using the Kenward-Rogers Method. Adjusted mean changes (LSMeans) with corresponding SE and 95% CI per treatment arm will be provided. The difference in adjusted means (LSMEANS) between abatacept and placebo at the primary time point, Day 169, and corresponding 95% CI and p-value based on the longitudinal repeated measures above will be provided. The analysis is using the modified ITT analysis population and all efficacy measurements collected up to last dose in Double-Blind Period + 42 days (and prior to open-label treatment).

In case of computational or non-convergence issues of the model, the following back-up models will sequentially be applied and the first back-up model that resolves the issues will be used. First, a spatial power covariance matrix will be used instead of an unstructured covariance. The second back-up model will use a spatial covariance matrix and remove the interactions with baseline stratification from the model. The third back-up model will include the spatial power covariance matrix and remove all interactions between baseline and day from the model.

Patterns of missing data for the change from baseline over time for the key secondary endpoints will be assessed using the same methods that were described for the primary endpoint in the previous section.

7.1.4 Binary Secondary Endpoints

The binary secondary efficacy endpoints, e.g., the proportion of subjects with a decrease in ESSDAI of at least 3 from baseline to Day 169 and the proportion of subjects with a decrease in the ESSPRI of at least 1 to Day 169, will be analyzed with a logistic regression including treatment, oral corticosteroid use, hydroxychloroquine use, stimulated salivary flow, Japan and baseline value in the model. The odds ratio along with its 95% confidence interval will be provided. The construction of CIs for differences in response rates between treatment groups will be based on minimum risk weights to account for randomization stratification factors, unless otherwise noted.

In case there are fitting issues with the minimum risk weights model including the 4 randomization stratification variables (too small strata), the first back-up model will exclude the Japan and stimulated salivary flow randomization stratification factors. The second back-up model will exclude all randomization stratification factors. The 95% CI within each treatment arm will be based on normal approximation.

A missing responder value at Day 169 due to discontinuation or for other reasons will be imputed as a non-responder. No p-values will be provided for the binary secondary efficacy endpoints

except for the proportion of subjects with a decrease in ESSDAI of at least 5 from baseline to Day 169 and the proportion of subjects with a decrease in the ESSPRI of at least 1 to Day 169.

7.1.5 Sensitivity Analyses

For the primary and key secondary endpoints the following sensitivity analyses will be provided:

1) To investigate the MAR assumption 2 sensitivity analyses will be provided:
   a) analysis with control-based pattern imputation
   b) analysis with tipping-point approach (SAS example provided in reference Section[1])

   The details of the MAR assumption sensitivity analyses are given below. P-values will be provided for these sensitivity analyses.
2) Longitudinal Repeated Measures analysis (as described in Section 7.1.3) removing the interaction terms for the stratification variables by time. P-values will be provided for these sensitivity analyses
3) Analyses based on the full ITT population including all randomized subjects and all observations obtained on or off study treatment. The protocol specifies to collect for all randomized subjects the planned efficacy measurements at each time point up to the end of the Week 24 Double-Blind Period also in the case a subject is off treatment. Therefore the expectation is that these measurements also be available for a randomized subject after the subject prematurely discontinues treatment and for a randomized subject who was never treated. The longitudinal repeated analysis model used for the primary analysis will be repeated, but for this sensitivity analysis the full ITT population will be used and all measurements collected on and off treatment up to Day 169 will be included in the analysis. P-values will be provided for these analyses.
4) The sensitivity analysis described in 3) including all efficacy measurements collected on and off treatment up to Day 169 will be repeated for all randomized and treated subjects (modified ITT population instead of ITT population described in bullet 3).

The sensitivity analyses described in bullet 1) will only be provided if there are at least 10% of subjects with missing data (for the corresponding efficacy measure) at Day 169 in either treatment group.

The sensitivity analyses including measurements on and off treatments (bullet 3) and 4)) will only be provided if:

- at Day 169 there are at least 10% of subjects with missing data (for the corresponding efficacy measure) in either treatment group for the primary analysis
- and on top, at Day 169 there are at least 10% more non-missing data (for the corresponding efficacy measure) in either treatment group using all collected data (including off treatment measures) compared to number of non-missing efficacy measurements in primary analysis

[redacted]

Each of the 100 imputed datasets will be analyzed using the same MMRM as the primary efficacy analysis. Results from the analyses of each imputed dataset, i.e. least squares (LS) mean treatment differences and their standard errors, will be combined using Rubin's imputation rules (using the SAS MIANALYZE procedure) to produce a pooled LS mean estimate of treatment difference, its 95% CI, and a pooled p-value for the test of null hypothesis of no treatment effect.

The above control-based imputation will also be done for the key secondary endpoints of ESSPRI and Stimulated Whole Salivary Flow.

Tipping Point Analysis

Sensitivity to departures from the MAR assumption will also be investigated using a tipping point analysis. In this analysis, departures from MAR in the abatacept group will be assessed assuming that subjects who discontinue the study have, on average, efficacy outcomes after discontinuation that are worse by some amount δ compared to other similar subjects with observed data (i.e., compared to a value which would have been assumed under a MAR model).

A series of analyses will be performed with increasing values of δ until the analysis conclusion of a statistically significant treatment effect no longer holds. The value of δ that overturns the primary results will represent a tipping point. An interpretation of clinical plausibility of the assumption underlying the tipping point will be provided.

Change from baseline in ESSDAI total score at Week 24 will be analyzed based on data observed while the subject remains on treatment as well as data imputed using multiple imputation methodology for time points at which no value is observed. Intermittent (non-monotone) missing data will be imputed first based on the MAR assumption and a multivariate joint Gaussian imputation model using MCMC method within each treatment arm, as described above for the pattern mixture model with control-based imputation.

The remaining monotone missing data will be imputed using sequential regression multiple imputation, where a separate regression model is estimated for imputation of each variable (i.e., measurement at each time point). Each regression model will include explanatory variables for treatment, stratification variables (current use of oral corticosteroids, current use of hydroxychloroquine, by stimulated salivary flow (< 0.1mL/min, ≥ 0.1mL/min) and Japan.) and all previous (Baseline, Weeks 4, 8, 12, 16, 20) values of ESSDAI total score.

Imputed data will consist of 100 imputed datasets. The random seed number for partial imputation with the MCMC method and for the sequential regression multiple imputation will be 12345.

After the MAR-based imputations have been generated for ESSDAI total score at each time point, a value of δ will be added to all imputed values in the abatacept group prior to analyzing the imputed data. This approach assumes that the marginal mean of imputed subject measurements is worse by δ at each time point after discontinuation compared to the marginal mean of subjects with observed data at the same time point.

Each of the 100 imputed and δ-adjusted datasets will be analyzed using the same MMRM as the primary efficacy analysis. Results from the analyses of each imputed dataset, i.e. LS mean treatment differences and their standard errors, will be combined using Rubin's imputation rules (using the SAS MIANALYZE procedure) to produce a pooled LS mean estimate of treatment difference, its 95% CI, and a pooled p-value for the test of null hypothesis of no treatment effect.

Analyses will be conducted with different values of δ at each visit, which present a percentage of the LS mean treatment difference at that visit, starting at 5% with 5% increments, until either the tipping point is identified (i.e. where a statistically significant treatment effect no longer holds).

In case there are issues with the multiple imputation analysis including the 4 randomization stratification variables, the first back-up model is to exclude the Japan and stimulated salivary flow randomization stratification factors. The second back-up model is to exclude all randomization stratification factors.

7.1.6 Outliers

The 'influence' option of the SAS procedure PROC MIXED will be included in the longitudinal repeated measures analysis to explore if there are subjects with outlying observations that have an impact on the changes from baseline at Day 169 for the primary and 2 key secondary endpoints. Subjects with a high restricted likelihood distance compared to other subjects will be identified. The impact of these outlying subjects on the result at Day 169 will be investigated by deleting the

subjects and submitting the analysis model without these subjects (sensitivity analysis). A plot of the restricted likelihood distance per subject and per treatment group will be provided.

7.2 Study Conduct

Relevant protocol deviations, which could have an impact on the primary 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 2.

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 with relevant protocol deviation will not be excluded from any analysis population described in Section 6.3. If at least 10% of the subjects in one of the treatment arms in the modified ITT analysis population demonstrate relevant protocol deviations, a per-protocol analysis will be performed for the primary endpoint, change from baseline in ESSDAI at Day 169. The per-protocol analysis will exclude data from subjects with relevant protocol deviations as specified in APPENDIX 2.

7.3 Study Population

7.3.1 Subject Disposition

The disposition of subjects will be summarized for the pre-treatment period, Double-Blind Period, open-label period and follow-up period. The disposition of subjects will also be represented graphically with a flow-chart for the same periods.

The summary of subject disposition for the pre-treatment period will be based on all enrolled subjects. It will include a summary of the:

- number of subjects enrolled
- number (percentage) of subjects randomized
- number of subjects randomized and treated by treatment group
- number of subjects randomized but not treated by treatment group
- number (percentage) of subjects per reason for not being randomized.
- number (percentage) of subjects per reason for being randomized, but not treated.

Percentages will be based on a denominator of all enrolled subjects. The reasons for not being randomized or being randomized but not treated will also be listed

The summary of subject disposition for the Double-Blind Treatment Period (as defined in Section 6.1, is on-treatment period) will be based on the modified ITT analysis population and will summarize by treatment group the:

- number (percentage) of subjects completing the treatment in the Double-Blind Treatment Period

- number (percentage) of subjects not completing the treatment the Double-Blind Treatment Period (i.e. early discontinued treatment subjects)
- number (percentage) of subjects per reason for not completing the treatment in the Double-Blind Treatment Period (i.e. reasons for early discontinuation of treatment)
- number (percentage) of subjects who completed treatment in the Double-Blind Treatment Period and treated in open-label period
- number (percentage) of subjects who completed treatment in the Double-Blind Period and not treated in open-label period
- number (percentage) of subjects per reason for completing the double-blind treatment and not treated in open-label period

Percentages will be based on a denominator of all randomized and treated subjects (modified ITT population). The reasons for not completing treatment in the Double-Blind Treatment Period or for not being treated in open-label will also be listed.

According to the protocol, patients who discontinue treatment in the Double-Blind Period should continue the study off treatment up to Day 169 to be assessed for efficacy. A CRF is collecting if a subject (on of off treatment) is continuing up to Day 169 or not. The summaries detailed below will provide the reasons for discontinuation on or off treatment prior to Day 169.The summary of subject disposition for the Double-Blind Period (on and off treatment up to Day 169) will be based on the modified ITT analysis population and will summarize by treatment group the:

- number (percentage) of subjects completing the Double-Blind Period (on and off treatment combined) up to Day 169
- number (percentage) of subjects who discontinued treatment and did not continue in the study up to Day 169
- number (percentage) of subjects per reason for not continuing in the study up to day 169 (after subject discontinued the treatment)

Percentages will be based on a denominator of all randomized and treated subjects (modified ITT population). The reasons for not completing in the Double-Blind Period (on or off treatment) or for not being treated in open-label will also be listed.

The summary of subject disposition for the open-label treatment period will be based on the Open-Label Population analysis population and will summarize by treatment group the:

- number (percentage) of subjects completing the open-label period
- number (percentage) of subjects ongoing in the open-label period (i.e. entering Open-label period but not completed or discontinued at the time of lock)
- number (percentage) of subjects not completing the open-label period (i.e. early discontinued subjects)
- number (percentage) of subjects per reason for not completing the open-label period (i.e. reasons for early discontinuation)
- number (percentage) of subjects continuing in the study (i.e. entering in the follow-up period)

- number (percentage) of subjects not continuing in the study (i.e. not entering in the follow-up period)
- number (percentage) of subjects per reason for not continuing in the study

Percentages will be based on a denominator of all subjects treated in the open-label period. The reasons for not completing the open-label period or for not continuing in the study will also be listed.

The summary of subject disposition for the follow-up period will be based on all subjects who discontinued in the double-blind or who discontinued in the open-label or who completed the double-blind but not treated in the open-label, or who completed the open-label (those subjects are expected to go into the follow-up period) and will summarize by randomized treatment group the:

- number (percentage) of subjects completing the follow-up period
- number (percentage) of subjects not completing the follow-up period
- number (percentage) of subjects per reason for not completing the follow-up period

Percentages will be based on a denominator of all subjects who discontinued in the double-blind or who discontinued in the open-label or who completed the double-blind but not treated in the open-label, or who completed the open-label (those subjects are expected to go into the follow-up period). The reasons for not completing the follow-up period will also be listed.

The number of subjects enrolled, randomized, and treated will be summarized by country and study site.

7.3.2 Demography and Baseline Characteristics

Demographic and baseline disease characteristics will be summarized by randomized treatment group and overall based on the modified ITT analysis population. For continuous variables, they will be summarized using means, standard deviations, median, Q1, Q3 and ranges (minimum and maximum), based on non-missing observations. The distribution of categorical variables will be summarized by treatment group using frequency and percentage. For categorical variables, percentages will be calculated out of the total number of subjects in the modified ITT population, overall and by treatment group (i.e., each denominator includes the number of subjects with missing/unknown values for the characteristic and missing is included as a category).

Table 7.3.2-1: Demographic and Baseline Characteristics

Characteristic	Summarized as	Categories
Age (years)	Categorical and Continuous	< 65 yrs., ≥ 65 yrs. ≤ 50 yrs., > 50 yrs.
Gender	Categorical	Male Female

Table 7.3.2-1: Demographic and Baseline Characteristics

Characteristic	Summarized as	Categories
Race	Categorical	White Black or African American Asian Other
Ethnicity (to be summarized only for USA subjects)	Categorical	Hispanic/Latino Non Hispanic/Latino
Region	Categorical	North America South America Europe Asia ROW (see APPENDIX 3)
Body Weight (kg)	Continuous and Categorical	< 60kg 60 - 100kg > 100kg
Randomization stratification factors as recorded on IWRS	Categorical	Baseline Hydroxychloroquine use (YES/NO) Baseline oral corticosteroid use (YES/NO) Screening and Baseline Stimulated Salivary Flow (< 0.1mL/min, ≥ 0.1mL/min) Japan (YES/NO) and all possible combinations of these 4 stratification factors
Randomization stratification factors as recorded on CRF	Categorical	Baseline Hydroxychloroquine use (YES/NO) Baseline oral corticosteroid use (YES/NO) Screening and Baseline Stimulated Salivary Flow (< 0.1mL/min, ≥ 0.1mL/min) Japan (YES/NO) and all possible combinations of these 4 stratification factors on CRF
2016 ACR-EULAR Classification Criteria met: • Labial salivary gland with focal lymphocytic sialadenitis and focus score ≥ 1 • Anti-SSA (Ro)+ • Ocular Staining score ≥ 5 (or van Bijsterfeld score ≥ 4) on at least one eye • Schirmer ≤ 5mm/5min on at least one eye • Unstimulated whole saliva flow rate ≤ 0.1ml/min	Categorical	Yes/No
ESSDAI Total Score	Continuous	

Table 7.3.2-1: Demographic and Baseline Characteristics

Characteristic	Summarized as	Categories
ESSDAI Total Score	Categorical	Low < 5
		Moderate 5 ≤ ESDDAI ≤ 13
		High ≥ 14
ESSDAI Constitutional Domain Score	Continuous	
ESSDAI Lymphadenopathy Domain Score	Continuous	
ESSDAI Glandular Domain Score	Continuous	
ESSDAI Articular Domain Score	Continuous	
ESSDAI Cutaneous Domain Score	Continuous	
ESSDAI Pulmonary Domain Score	Continuous	
ESSDAI Renal Domain Score	Continuous	
ESSDAI Muscular Domain Score	Continuous	
ESSDAI Peripheral Nervous System (PNS) Domain Score	Continuous	
ESSDAI Central Nervous System (CNS) Domain Score	Continuous	
ESSDAI Haematological Domain Score	Continuous	
ESSDAI Biological Domain Score	Continuous	
ESSPRI Total Score	Continuous	
ESSPRI - Pain Score	Continuous	
ESSPRI- Fatigue Score	Continuous	
ESSPRI - Dryness Score	Continuous	
Stimulated whole saliva secretion	Continuous	
DAS28-CRP at Day 1	Continuous	
DAS28-CRP at Day 1<	Categorical	DAS28-CRP ≤ 2.6
		DAS28CRP > 2.6
DAS28-CRP at Day 1<	Categorical	DAS28-CRP ≤ 3.2
		DAS28CRP > 3.2

7.3.3 General Medical History

The number (percent) of subjects with general medical history findings before the first study medication will be summarized by treatment group and overall using the modified ITT analysis population. A corresponding listing of medical history findings will be provided.

7.3.4 Prior Medication and Medication at Day 1

Prior medications are defined as medications with a start date prior to the first day of Double-Blind Treatment Period.

Prior medication of special interest and medication of special interest at Day 1 will be summarized for the 2 treatment arms. The results will be based on the modified ITT analysis population. The medication of special interest are defined in Section 7.4.4.

Missing and partial date handling of start and stop dates of prior medications, is described in Section 8.4. The World Health Organization (WHO) dictionary is used to code non-study medications.

7.4 Extent of Exposure

7.4.1 Study Therapy

The extent of exposure during the double-blind (for the As-treated Analysis Population) and the cumulative abatacept periods (for the cumulative abatacept population) will be summarized in the following ways:

- Number of abatacept/placebo injections
- Number of days a subject is known to be on study drug.

The number of injections subjects received will be based on the study medication page of the CRF when it is indicated that a subject received the study medication.

The number of days the subject is known to be on study drug (exposure to study drug) during the Double-Blind Period is calculated as:

For subjects who discontinued during the Double-Blind Period OR completed the Double-Blind Period but not treated in open-label period OR not treated in the open-label period within 56 days of the last injection in the Double-Blind Period:

> Exposure in days = (date of last injection in the Double-Blind Period - date of first injection in the Double-Blind Period + 1) + 56.

For subjects who were treated in the open-label within 56 days of the last injection of the Double-Blind Period:

> Exposure in days = (date of first injection in the open-label period - date of first injection in the Double-Blind Period).

Exposure in cumulative abatacept period for subjects who discontinued treatment in double-blind or open-label OR who completed double-blind and not treated in Open-label or who completed open-label is calculated as:

> Exposure in days = (date of last abatacept injection - date of first abatacept injection + 1) +56

In case a patient randomized to abatacept does not go into open-label within 56 days post last dose in double-blind, the exposure has to be adjusted for the gap exceeding 56 days.

Exposure in cumulative abatacept period for subjects who are on-going in open-label period is calculated as:

> Exposure in days = cut-off date - date of first abatacept injection with cut-off date = the date database lock date.

For subjects randomized to placebo, the date of first abatacept injection is date of first open-label abatacept treatment. For subjects randomized to abatacept, the date of first abatacept injection is the date of first double-blind abatacept treatment.

The date of last abatacept injection can be the last date of abatacept in double-blind if the subject is not treated in open-label or the date of last abatacept injection in open-label if subject is treated in open-label.

The offset of 56 days is the length of 2 regular dosing cycles and represents approximately 4 times the half-life of abatacept in humans.

Summaries of exposure to study drug during the Double-Blind Period and the cumulative abatacept period will show the distribution of the number of injections and days on drug, together with the means, standard deviations, medians, minimum and maximum. For the Double-Blind Period the presentation will be by treatment group and for the cumulative abatacept period all subjects will be presented together.

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. Section 3.5 of the protocol mentions the reasons for which a subject must discontinue treatment.

Discontinuation from study therapy during the Double-Blind Period is captured on the case report form (CRF) pages. For the analysis of discontinuations from study therapy refer to Section 7.3.1.

7.4.3 Treatment Compliance

The CRF for this study will capture information on injections of abatacept or matching placebo during the Double-Blind Period and injections of abatacept during the open-label period.

The number of subjects with missed injections (excluding missed injections due to premature discontinuation from the study) during the Double-Blind Period or in the cumulative abatacept period will be summarized by number of missed injections based on the As-treated analysis population by treatment groups. A corresponding listing for all subjects who skip any injection will be provided.

[redacted]

7.5 Efficacy

All efficacy analyses will be provided using the modified ITT analysis population. In addition, a sensitivity analysis will be provided for the primary and key secondary endpoints using the ITT population (see Section 6.3 for more details).

A hierarchical testing procedure will be applied for the primary endpoint (change from baseline in ESSDAI at Day 169) and the 2 key secondary endpoints (change from baseline in ESSPRI at Day 169 and change from baseline in the stimulated whole salivary flow at Day 169 among the subjects with stimulated whole salivary flow of at least 0.1 mL/min at both screening and baseline) to ensure the preservation of the overall type I error of 5%. The first key secondary endpoint will only be tested (at a significance level of 5%) if the test for the primary endpoint is statistically significant at a significance level of 5%. If both the test for the primary endpoint and the first key secondary endpoint are statistically significant (both at a significance level of 5%), then the second key secondary endpoint will be tested (at a significance level of 5%).

P-values will be presented for each of these 3 endpoints. However, endpoints should not be interpreted for significance if they have a rank lower than that endpoint whose null hypothesis was the first that could not be rejected. That is, the significance of endpoints should not be interpreted if they have a rank lower than that endpoint whose test was the first to be non-statistically significant (p-value > 0.05).

The proportion of subjects with a decrease in ESSDAI of at least 5 from baseline to Day 169 and the proportion of subjects with a decrease in the ESSPRI of at least 1 to Day 169 are not included in the hierarchical testing procedure, however a nominal p-value will be presented for this endpoint at Day 169

No other p-values will be presented, except for a few sensitivity analyses of the primary and key secondary endpoints (see Section 7.1.5 for more details).

7.5.1 Primary Efficacy Analysis

The EULAR Sjögren's Syndrome Disease Activity Index (ESSDAI) is a systemic disease activity index that was designed to measure disease activity in subjects with pSS (see Protocol Appendix 3). The ESSDAI parameters are assessed by the investigators and the results from each assessment are recorded in site source records as well as entered into the corresponding eCRF. The ESSDAI total score will be calculated as described in Section 8.1.1.

Details of the primary efficacy analysis including the model specifications and back-up models are given in Section 7.1.3 also provides the sensitivity analyses for the primary endpoint.

In addition, in case that more than 10% of subjects in any treatment group have a relevant protocol deviation, then the primary efficacy analysis will be repeated excluding efficacy data from subjects with relevant protocol deviations, as described in APPENDIX 2. A p-value will be provided.

The adjusted mean change from baseline in ESSDAI will also be presented graphically by treatment arm for the Double-Blind Treatment Period (from baseline up to Day 169), as well as the double-blind and open-label periods combined (from baseline up to Day 365).

7.5.2 Secondary and Exploratory Efficacy Analyses

7.5.2.1 Continuous Endpoints

The continuous key secondary endpoints defined in this study are:

- The mean change from baseline in ESSPRI at Day 169
- The mean change from baseline in the stimulated whole salivary flow at Day 169 among subjects with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline

The other continuous secondary endpoints defined in this study are:

- The mean change from baseline at all measured time points up to Day 169 in the DAS28-CRP among those with a tender plus swollen joint count ≥ 3 at baseline, among those with a tender plus swollen joint count < 3 at baseline and in the full population.
- The mean change from baseline at all measured time points up to Day 169 in the individual components of DAS28-CRP among those with a tender plus swollen joint count ≥ 3 at baseline, among those with a tender plus swollen joint count < 3 at baseline and in the full population.

- The mean change from baseline at all measured time points up to Day 169 in the ESSDAI in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline at all measured time points up to Day 169 in the ESSPRI in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in the individual components of the ESSDAI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in the individual components of the ESSPRI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in Schirmer's test at all measured time points up to Day 169.
- The mean change from baseline in ocular staining score at all measured time points up to Day 169.
- The mean change from baseline in tear break-up time at all measured time points up to Day 169.
- The mean change from baseline in unstimulated whole salivary flow at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in stimulated whole salivary flow at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- The mean change from baseline in stimulated whole salivary flow at all measured time points up to Day 169 among
  - subjects with non-zero value at baseline for ESSDAI glandular domain and with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline
  - subjects with non-zero value at baseline for ESSDAI glandular and with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
  - subjects with zero value at baseline for ESSDAI glandular domain and with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline
  - subjects with zero value at baseline for ESSDAI glandular and with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

- The mean change from baseline in unstimulated whole salivary flow at all measured time points up to Day 169 among
  - subjects with non-zero value at baseline for ESSDAI glandular domain and with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline
  - subjects with non-zero value at baseline for ESSDAI glandular and with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
  - subjects with zero value at baseline for ESSDAI glandular domain and with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline
  - subjects with zero value at baseline for ESSDAI glandular and with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

[redacted]

All these endpoints will also be assessed up to Day 365.

The details of the calculation of the total score for ESSDAI, total score for ESSPRI, salivary flow rate, DAS28-CRP, Schirmer's test, tear break-up time, ocular surface staining are provided in Sections 8.1.1, 8.1.2, 8.1.3, 8.1.4, 8.1.5, 8.1.6, 8.1.7, respectively.

The 3 analyses of eye dryness as measured by Schirmer's test (ST), tear break-up time (TBUT), ocular surface staining (OSS) will be

- for the "study eye" at baseline. The study eye at baseline is defined as follows: the eye with the highest total score for OSS at baseline, will be selected as the study eye. In case both eyes have the same total score for OSS at baseline, the eye with the lowest ST at baseline will be selected as the study eye. If both eyes also have an equal ST at baseline, then the eye with the lowest TBUT will be selected as the study eye. If both eyes also have an equal TBUT at baseline, then the right eye will be selected as the study eye. The same study eye will be used for ST, TBUT and OSS analyses.
- for the non-study eye at baseline. This is the eye that is not the study eye at baseline.
- based on the average of the measurement for the two eyes on the same visit.

Double-blind analysis for key secondary endpoints at Day 169

Details (model specifications, back-up models and sensitivity analyses) are provided in Section 7.1.

[redacted]

[redacted]

Table 7.5.2.1-1: Schirmer's Test Score

Schirmer's test score/5 minutes (mm)
0 - ≤ 2
3 - ≤ 5
6 - ≤ 10
> 10

[redacted]

Analysis from Day 1 up to Day 365

The continuous primary, secondary and exploratory endpoints in the list given above will be analyzed using the longitudinal repeated measures model given in Section 7.1. For the non-key secondary and exploratory endpoints, the stratification variable by day interactions will be removed from the model (the baseline variable and baseline-by -day interaction will be kept). All timepoints of the Double-Blind Treatment Period and the open-label period (at which endpoint is collected) will be included. The treatment variable in the model is the original randomized treatment. This analysis will include all efficacy measurements collected up to last dose +42. The baseline measurement in the model will correspond to the variable that is analyzed. The modified ITT population will be used. In addition, some of these endpoints (see Section 4.2 and 4.3) will also be analyzed for the following 2 subgroups:

- subgroup of subjects with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline
- subgroup of subjects with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

The subgroup analyses mentioned in the 2 bullets above will only be provided if there are at least 10 subjects in each treatment group in each of the subgroups

The longitudinal repeated measures model for these subgroup analyses will not include the stimulated salivary flow stratification variable. At each timepoints the adjusted changes for the 2 treatments will be provided. No treatment differences will be provided for the open-label timepoints.

Plots of the adjusted mean changes (95%) over time (from Day 29 up to Day 365) per treatment group will be provided for the primary and key secondary endpoints.

7.5.2.2 Categorical Endpoints

The binary secondary endpoints defined in this study are:

- Proportion of subjects who achieve a minimally clinically important improvement (of at least 3 points) in the ESSDAI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and

baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

- Proportion of subjects who achieve an improvement of at least 5 points in the ESSDAI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.
- Proportion of subjects who achieve a minimally clinically important improvement (of at least 1 point) in the ESSPRI at all measured time points up to Day 169 in the full population, among those with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline, and among those with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

[redacted]

Double-blind analysis over time

The binary endpoints given above at each timepoint (at which endpoint is collected) up to Day 169 will be analyzed as specified in Section 7.1.4. The modified ITT population will be used. Missing values will be imputed as non-responders. In addition, the secondary binary efficacy endpoints will also be analyzed for 2 subgroups:

- subgroup of subjects with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline
- subgroup of subjects with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

The subgroup analyses mentioned in the 2 bullets above will only be provided if there are at least 10 subjects in each treatment group in each of the subgroups.

The logistic model for these subgroup analysis will not include the stimulated salivary flow stratification variable.

The response rate (%) and 95% CI over time will be plotted for each treatment group up to Day 169 for the secondary binary endoints.

Analysis from Day 1 up to Day 365

The binary endpoints given above at each timepoint in double-blind and open-label period up to Day 365 will be summarized by a point estimate for the response rate and corresponding 95 % CI for the 2 original randomized treatment arms. No treatment differences will be provided for any timepoints during the open-label period.

The modified ITT population will be used. Missing values will be imputed as non-responders. In addition, the secondary binary efficacy endpoints will also be analyzed for 2 subgroups:

- subgroup of subjects with stimulated whole salivary flow of at least 0.1mL/min at both screening and baseline
- subgroup of subjects with stimulated whole salivary flow of less than 0.1mL/min at screening or baseline.

The subgroup analyses mentioned in the 2 bullets above will only be provided if there are at least 10 subjects in each treatment group in each of the subgroups.

The response rate (%) and 95% CI over time will be plotted for each treatment group during the double-blind and open-label periods up to Day 365 for the secondary binary endpoints.

7.5.2.3 Subgroup Analyses

All analyses described in this section will be performed on the modified ITT Analysis Population. All subgroups will be defined based on the data as entered on the CRF; no IWRS will be used for any subgroup analysis.

Table 7.5.2.3-1 shows the subgroups of interest for analyses of the primary efficacy endpoint of Change from Baseline at Day 169 (ESSDAI). The analysis will be using the methods described in Section 7.1.3 for each level of each subgroup, except that all stratification variables and the baseline ESSDAI day-by-day interaction will be removed from the model for this subgroup analysis. If the value of the grouping variable cannot be determined for a subject, the subject will be excluded from the corresponding subgroup analysis. Only subgroups with at least 10 patients in both treatment arms will be considered.

A forest plot showing per subgroup the estimates per treatment group and the treatment differences will be provided. The sample size for each treatment group in the subgroup will be presented along with the adjusted mean change from baseline (with 95% CI) and the treatment difference (with 95% CI).

Table 7.5.2.3-1: Subgroups of Interest

Subgroup factor	Categories
Age	< 65 Years Old
	≥ 65 Years Old
Baseline Weight	< 60kg
	60 - 100kg
	> 100kg
Gender	Male
	Female
Race	White
	Black
	Asian
	Other
Geographic Region	North America
	South America
	Asia
	Europe
	ROW
Baseline Hydroxychloroquine use as recorded in CRF	Yes
	No
Baseline oral corticosteroid use as recorded in CRF	Yes
	No
Baseline Stimulated Salivary Flow as recorded in CRF	< 0.1mL/min
	≥ 0.1mL/min

Table 7.5.2.3-1: Subgroups of Interest

Subgroup factor	Categories
Japan as recorded in CRF	Yes
	No
ESSDAI categories at Day 1	Low < 5
	Moderate 5 ≤ ESSDAI ≤ 13
	High ≥ 14

7.6 Safety

The safety analysis endpoints described in this section are considered as secondary endpoints. For their analysis, there will be no statistical testing of group differences.

Analysis of all safety data will follow the BMS standard safety data conventions[3] and abatacept program safety conventions[4].

All summaries of safety parameters during the Double-Blind Treatment Period will be provided by treatment groups. All safety assessments will be included in the summary tables if the onset date is on or after the first dose date and up to 56 days post the last dosing day in the Double-Blind Treatment Period or the first dosing date in the open-label period, whichever occurs first. Presentations for the Double-Blind Treatment Period will be provided by "as treated" treatment group for the as-treated analysis population, unless otherwise specified.

These safety endpoints will also be summarized during the cumulative abatacept period (from the first day of abatacept treatment in the study up to 56 days after the last abatacept treatment in the study) for the cumulative abatacept population. For the cumulative abatacept population all subjects will be combined in one group.

7.6.1 Adverse Events

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 during the Double-Blind Treatment Period will be included in the frequency tabulations if they occur while a subject is on study medication during the Double-Blind Treatment Period and up to 56 days post the last dosing day in Double-Blind Treatment Period or the first dosing date in the open-label period, whichever occurs first. These AE summaries will be based on proportions, which represent the number of subjects experiencing the AEs divided by the number of subjects who received at least 1 dose of study medication.

AEs during the cumulative abatacept period will be included in the frequency tabulations if they occur while a subject is on abatacept up to 56 days after the last abatacept treatment in the study. These AE summaries will be based on proportions, which represent the number of subjects experiencing the AEs divided by the number of subjects who received at least 1 abatacept dose during the study. For the cumulative abatacept summaries all subjects will be combined in one group.

For summaries by SOC and PT, SOCs will be sorted by decreasing frequency then alphabetically and within SOC, PTs will be sorted by decreasing frequency then alphabetical. For summaries by PT, PTs will be sorted by decreasing frequency then alphabetically.

All adverse events 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.

Laboratory AEs are laboratory changes identified by the investigator as AEs and thus reported on the AE pages of the eCRF and included in the summaries of AEs.

All serious and non-serious adverse events with onset during the Double-Blind Treatment Period will be summarized by SOC, PT, and treatment group. All serious and non-serious adverse events with onset during the cumulative abatacept period will be summarized by SOC and PT overall (in one treatment group). These summaries will be produced for:

- all adverse events
- adverse events related to the study drug, as determined by the investigator
- adverse events leading to discontinuation of study medication
- most common related adverse events (reported in 2% of subjects or more in any treatment group)
- most common adverse events (reported in 5% of subjects or more in any treatment group)
- all serious adverse events
- all serious adverse events related to the study drug
- adverse events by intensity

7.6.2 Adverse Events of Interest

The following adverse events of special interest will be summarized by PT and treatment groups during the Double-Blind Treatment and cumulative abatacept periods. A subject listing will also be provided.

7.6.2.1 Infections

The summaries for the SOC: *Infections and infestations* are included in the summaries of all adverse events detailed above. A by subject listing with the infections will also be provided

7.6.2.2 Malignancy

All events in the MedDRA Maintenance and Support Services Organization (MSSO) malignancies Structured MedDRA Query (SMQ) list occurring during each study period (Double-Blind Treatment Period, Cumulative Abatacept Period) will be reported separately by study period. These events will be summarized by preferred term within each study period separately and will also be listed.

7.6.2.3 Autoimmune Disorders

The frequency of pre-specified autoimmune disorders (defined using a BMS custom Autoimmune Disorder SMQ) occurring during each study period (Double-Blind Treatment Period, Cumulative Abatacept Period) will be provided separately for each study period. Autoimmune disorders will also be summarized by intensity and a listing of all reported autoimmune disorders will be provided.

7.6.2.4 Injection Reactions

1) *Systemic Injection Reactions* are defined as pre-specified systemic AEs (such as hypersensitivity reactions) occurring during the first 24 hours after SC injection.
2) *Local Injection Site Reactions (LISR)* are defined as those AEs that occur at the site of SC injection. A pre-specified list of MedDRA codes for Local Injection Site Reaction (LISR) events of interest will be used to identify the local injection site reactions.

The frequency of pre-specified systemic injections (1) and local injection site reactions (2) (defined using a BMS custom SMQ) occurring during each study period (Double-Blind Treatment Period, Cumulative Abatacept Period) will be provided for (1) and (2) separately and for each study period separately. These AEs will also be summarized by intensity and a listing of all these reported AEs will be provided

In additional to the summary of LISR, the following 5 AEs PTs will be summarized separately for the each study period (Double-Blind Treatment Period or Cumulative Abatacept Period): Injection Site *Erythema, Injection Site Pain, Injection Site Pruritus, Injection Site Haematoma, and Injection Site Swelling*.

7.6.2.5 Deaths

All deaths recorded on the status page, the AE page, or SAE page (with a death date, or cause of death, or outcome or SAE categorization present) of the CRF will be reported. Any deaths that occur during the study will be described in depth as narrative in the CSR. All Adverse events with the outcome of death reported during the study will be listed.

7.6.2.6 Other Serious Adverse Events

SAEs are captured on a special CRF page. Summaries of all SAEs, related SAEs, and SAEs leading to discontinuation will be provided for 2 periods (Double-Blind Treatment Period, Cumulative Abatacept Period). All SAEs will be listed.

7.6.2.7 Adverse Events Leading to Discontinuation of Study Therapy

AEs leading to discontinuation of study drug are identified as those AEs with an action code of 5 on the CRF page (5 = discontinued study drug). All such AEs will be summarized for the Double-Blind Treatment Period and cumulative abatacept period. All AEs leading to discontinuation will be listed.

7.6.2.8 Multiple Adverse Events

Several descriptive summaries of adverse events that takes into account the number of occurrences that an AE was reported by individual patients 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 patient 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 patient data will represent the maximum severity observed as well as the last known assessed relationship to study medication by the investigator.

This data will be presented as the rate per 100 years of patient exposure. Exposure to study medication will be calculated according to approved standard BMS algorithms as well.

As an example, if 5 patients report 7 unique episodes of headache and had a combined cumulative exposure of 20 years to study medication, the incidence rate is reported as 7 / 20 * (100) or 35 cases per 100 patient years of exposure.

The following summary tables will be provided:

- A table showing the total number and rate (exposure adjusted) of occurrences for all AEs occurring in at least 5% of the subjects treated will be presented for the double-blind treatment period and the cumulative abatacept period separately.

- For all AEs as well as for AEs of special interest:
  - A table showing total number of events and rate (exposure adjusted) will be presented for the Double-Blind Treatment Period (overall) and the cumulative abatacept period (by 6 months time intervals) separately.
  - A table showing the number of subjects experiencing an AE once or multiple times will be presented for the Double-Blind Treatment Period and the cumulative abatacept period separately.

- Listing displaying the unique instances of all AEs, i.e., after duplicates have been eliminated and overlapping and contiguous occurrences of the same event have been collapses.

No formal comparisons are made between treatments. No formal statistical testing will be performed, only summary statistics are provided.

7.6.3 Laboratory Evaluation

Unless otherwise specified, laboratory data obtained after the start of study medication dosing up to the first dose in the open-label period or up to and including 56 days after the last double-blind dosing date, whichever occurs first will be considered as obtained during Double-Blind Treatment Period. For the cumulative abatacept period, laboratory data obtained after the start of abatacept up to and including 56 days after the last abatacept dose will be considered as obtained during the cumulative abatacept period.

7.6.3.1 Marked Laboratory Abnormalities

Laboratory abnormalities will be reported for the same periods as for the AEs, i.e., Double-Blind Treatment Period and Cumulative Abatacept Period. Laboratory measurements will be included in the analysis for the Double-Blind Treatment Period if the measurement date is on or after the first day of dosing and up to first dose in the open-label period or 56 days post the last dosing day in the Double-Blind Treatment Period whichever is earlier. Laboratory measurements will be included in the analysis for the Cumulative Abatacept Period if the measurement date is after first abatacept dose date up to the 56 days post last abatacept dose date.

Laboratory abnormalities are identified using a pre-defined set of marked abnormality criteria. The criteria will be listed in the study report.

The frequency of subjects with any marked laboratory abnormality will be presented by laboratory test during the Double-Blind Treatment Period and separately, during the cumulative abatacept period. The results are based on the As-treated analysis population for the double-blind period and based on the cumulative abatacept population for the cumulative abatacept period.

The pre-specified criteria used for classifying laboratory test results as markedly abnormal will be listed.

7.6.3.2 Change from Baseline for Selected Laboratory Parameters

Scheduled laboratory measurements and corresponding change from baseline values will be summarized (mean baseline with standard deviation, mean post-baseline with standard deviation, mean change from baseline with standard error and 95% CI for change from baseline) by time point and laboratory test. For the Double-Blind Treatment Period (As-Treated Analysis Population) summaries will be provided by treatment group. For the cumulative abatacept period (cumulative abatacept population) all subjects' data will be combined into one treatment group.

Visit windows are provided in Section 8.5 in order to link each laboratory test to a scheduled visit. The following laboratory parameters will be analyzed:

- Hemoglobin
- Hematocrit
- Total leukocyte
- Basophils
- Eosinophils
- Lymphocytes
- Monocytes
- Neutrophils
- Platelet count
- Aspartate aminotransferase (AST)
- Alanine aminotransferase (ALT)
- Total bilirubin
- Alkaline phosphatase
- Creatinine
- Blood Urea Nitrogen (BUN)
- Bicarbonate
- GGT (Gamma-glutamyltransferase
- Glucose
- Total Protein
- Albumin
- Sodium
- Potassium
- Chloride
- Calcium
- Phosphorus
- Creatine kinase
- hsCRP
- Urine pH

Subjects who have laboratory measures at baseline and corresponding measure at the given on-treatment visit will be included in the laboratory analyte assessment. Note that not all subjects have laboratory determinations for all analytes at all visits, and therefore the sample size may vary from analyte to analyte at each time point. The 95% CI for the change from baseline within the treatment arm will be constructed based on t-test.

7.6.3.3 Pregnancy Test Results

A by-subject listing of positive pregnancy test results will be provided using the As-treated analysis population.

7.6.4 Vital Sign

Summaries of vital signs parameters (seated systolic blood pressure, seated diastolic blood pressure, heart rate, and body temperature) will be provided for each vital sign parameter at each timepoint. For the Double-Blind Treatment Period summaries (As-Treated Analysis Population) will be done by treatment group. For the cumulative abatacept period (cumulative abatacept population) all subjects' data will be combined into one treatment group. Visit windows are provided in Section 8.5 in order to link vital sign assessment to a scheduled visit

7.6.5 Pharmacokinetic Analyses

PK Samples will be drawn to determine serum abatacept concentrations at Days 1, Week 4 (Day 29), Week 8 (Day 57), Week 12 (Day 85), Week 16 (Day 113), Week 20 (Day 141), and Week 24 (Day 169) in the Double-Blind Treatment Period, at Week 52 (Day 365) in the Open-Label Period. For subjects who discontinue during the Double-Blind Treatment Period samples drawn 85 and 169 days after their last dose of study medication will be evaluated to determine serum abatacept concentration in the event there is a corresponding positive immunogenicity result.

Trough serum concentration of abatacept will be summarized by study visits listed above for the Double-Blind and Open-Label Periods separately (PK evaluable population). Geometric means and coefficients of variation will be presented for C_{min}. Statistics that will be provided are N, Mean, Standard Deviation, Geometric Mean, %CV, Median, Minimum and Maximum.

7.7 Outcomes Research Analysis

The outcome research analysis endpoints described in this section for the Double-Blind Treatment Period are considered as secondary endpoints and exploratory in the open-label period.

The analysis for the below outcome research endpoints will follow the analysis methods for the continuous secondary efficacy endpoints detailed in Section 7.5.2.1

- Subject global assessment of disease activity (SubGDA) and physician global assessment of disease activity (phyGDA) will be analyzed separately (see Section 8.1.9)
- Short Form - 36 (SF-36 v2.0) (see Section 8.1.10). Both the physical and mental scales as well as the individual sub-scales will be analyzed.
- Female Sexual Function Index (FSFI) (see Section 8.1.11). Both the full scale score and the individual domain scores will be analyzed.
- PROMIS Fatigue Short Form 8a (see Section 8.1.12).

The Numeric Rating Scale (NRS) scores for mouth and eye dryness (see Section 8.1.8) will be summarized by treatment and timepoint using descriptive statistics (median, interquartile range, minimum and maximum). Furthermore, all NRS scores available at baseline will be categorized based on tertiles (lowest third, middle third, and highest third). These same tertile cut points will be applied to all post-baseline data. The shift from baseline category to post-baseline category will be summarized at each post-baseline visit for each treatment group.

[redacted]

[redacted]

7.10 Interim Analyses

No interim analysis is planned for this study.

8 CONVENTIONS

8.1 Calculations of Key Measures

8.1.1 EULAR Sjögren's Syndrome Disease Activity Index (ESSDAI)

The EULAR Sjögren's Syndrome Disease Activity Index (ESSDAI) total score is calculated as the sum of scores for activity level for each domain. The scores and the domains are presented in Table 8.1.1-1. If all components are missing, total ESSDAI score will not be calculated; otherwise missing component will be imputed to zero (since it is assumed that active domains are not likely to be left blank).

The ESSDAI domain scores will be calculated using the values presented in Table 8.1.1-1 for each domain separately.

Table 8.1.1-1: The EULAR Sjögren's Syndrome Disease Activity Index (ESSDAI) Scoring Algorithm

Domain	Score for activity level
Constitutional	No=0, Low=3, Moderate=6
Lymphadenopathy	No=0, Low=4, Moderate=8, High=12
Glandular	No=0, Low=2, Moderate=4
Articular	No=0, Low=2, Moderate=4, High=6
Cutaneous	No=0, Low=3, Moderate=6, High=9
Pulmonary	No=0, Low=5, Moderate=10, High=15
Renal	No=0, Low=5, Moderate=10, High=15
Muscular	No=0, Low=6, Moderate=12, High=18
Peripheral Nervous System (PNS)	No=0, Low=5, Moderate=10, High=15
Central Nervous System (CNS)	No=0, Moderate=10, High=15
Haematological	No=0, Low=2, Moderate=4, High=6
Biological	No=0, Low=1, Moderate=2

[redacted]

8.1.3 Salivary Flow Rate

Salivary flow rate is assessed as described in the protocol Section 5.4.2. The CRF collects information on start and stop date/time of saliva collection as well as the pre-collection and post collection weight of the tube used to collect saliva.

For analysis, the salivary flow rate will be expressed in ml per min (ml/min) using the formula below:

Figure 8.1.3-1: Formula for Calculating Salivary Flow Rate

$$\text{Salivary flow rate (ml/min)} = \frac{(\text{Post - collection weight in grams}) - (\text{Pre - collection weight in grams})}{(\text{Stop date/time of collection}) - (\text{Start date/time of collection}) \text{ in minutes}}$$

In the formula above, it is assumed that 1 gram of saliva corresponds to 1ml of saliva. Round to 2 decimals.

The same formula will be used for the calculation of both stimulated and unstimulated salivary flow rate.

8.1.4 DAS28-CRP

The disease activity score DAS28-CRP is a continuous variable which is a composite of 4 variables: the 28 tender joint count, the 28 swollen joint count, CRP and subject assessment of disease activity measure on a visual analogue scale (VAS) of 100mm. The formula for calculation of DAS28-CRP is given below:

$$\text{DAS28-CRP} = 0.56 * \text{sqrt}(\text{tender28}) + 0.28 * \text{sqrt}(\text{swollen28}) + 0.36 * \ln(\text{hsCRP}+1) + 0.014 * \text{VAS} + 0.96.$$

The measurement of hsCRP should be in mg/L. If any component is missing, then DAS28-CRP will not be calculated.

8.1.5 Schirmer's Test

Schirmer's test will be performed as described in protocol Section 5.4.3. The CRF collects the length in millimeters that the strip wets during the 5 minute test period for each eye. Collection is done separately for each eye.

8.1.6 Tear Break-up Time

Tear break up time will be assessed as described in protocol Section 5.4.3. The CRF collects the time in seconds to first appearance of a random dry spot on the corneal surface for 3 repetitions in each eye. The average time will calculated for each eye averaging the 3 measurements for each eye separately. In case only 2 measurements are available, the average of the 2 measurements will be calculated. In case there is only 1 measurement, that measurement will be used for the analysis.

8.1.7 Ocular Surface Staining

Ocular surface staining will be assessed as described in protocol Section 5.4.3. The CRF collects the score for the following for each eye:

- MNBC (Medial Nasal Bulbar Conjunctiva) [maximum score = 3],
- CORN (Corneal Staining of Punctate Epithelial Erosions (PEE)) [maximum score = 3],
- LTBC (Lateral Temporal Bulbar Conjunctiva) [maximum score = 3],
- CONF (Patches of Confluent Staining) [maximum score = 1],

- PUPL (PEE observed in the pupil region, i.e. central 4mm diameter portion of the cornea) [maximum score = 1],
- FILA (one of more filaments seen anywhere on the cornea) [maximum score = 1]

The total score will be calculated as the sum of the score for these parameters for each eye.

8.1.8 Numeric Rating Scale (NRS) Scores

The oral and ocular dryness are each assessed by the patients with numeric rating scales from 0 to 10 with 0 representing no dryness and 10 representing maximal dryness (protocol Appendices 8 and 9). The analysis will be based on the scores as reported on the CRF.

8.1.9 Visual Analog Scale

The subject and physician global assessment of disease activity are assessed with visual analog scales presented in protocol Appendices 4 and 5. The CRF collects the distance in millimeters from the start of the scale that is marked as well as the length of the scale in millimeters. In cases that the length of the scale is in less or more than 100 millimeters, then the subject's or physician measurement will be rescaled to the equivalent of 100 millimeters using the formula below:

Figure 8.1.9-1: Formula to Rescale Visual Analog Scale

$$\text{Rescale measurement in mm} = \frac{\text{measurement as reported on CRF in mm}}{\text{length of the line on CRF in mm}} \bullet 100\,\text{mm}$$

8.1.10 Short Form-36 (SF-36 v2.0)

The SF-36 v2.0 is composed of 36 items (questions) measuring 8 health concepts (scales): Physical Function, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health. The 8 scales are used to form two summary scales: Physical Component Summary (PCS) and Mental Component Summary (MCS).

The QualityMetric Health Outcomes Scoring Software 5.0 will be used for scoring these scales

8.1.11 Female Sexual Function Index (FSFI)

For the FSFI, domain scores are calculated by summing the scores of the individual questions that make up the domain and multiplying the sum by the factor in the table below. The full scale score is the sum of the six domain scores.

For subjects without a partner, thus not able to answer item 15, the domain score for satisfaction and the full scale score will not be calculated.

Table 8.1.11-1: Scoring Algorithm for Female Sexual Function Index (FSFI)

Domain	Questions	Score Range	Factor	Minimum score	Maximum score
Desire	1, 2	1–5	0.6	1.2	6.0
Arousal	3, 4, 5, 6	0–5	0.3	0	6.0
Lubrication	7, 8, 9, 10	0–5	0.3	0	6.0
Orgasm	11, 12, 13	0–5	0.4	0	6.0
Satisfaction	14, 15, 16	0 (or 1)–5	0.4	0	6.0
Pain	17, 18, 19	0–5	0.4	0	6.0
Full Scale Score Range				2.0	36.0

8.1.12 PROMIS Fatigue Short Form 8a

As mentioned in the brief guide to the PROMIS Fatigue instruments "*each question has five response options ranging in value from one to five ... To find the total raw score for a short form with all questions answered, sum the values of the response to each question. For example, for the 8-item form, the lowest possible raw score is 8; the highest possible raw score is 40*". Table 8.1.12-1 presents the raw score values.

"A score can be approximated if a participant skips a question. If items are missing, first check how many items were answered. For short forms with at least 5 items, confirm that 4 or 50% of items, whichever is greater, were answered." If less than 4 items were answered, then the T-score cannot be calculated *"After confirming that enough responses were provided, sum the response scores from the items that were answered (not including any screening question). Multiply this sum by the total number of items in the short form. Finally, divide by the number of items that were answered. For example, if a respondent answered 5 of 8 questions and answered all items with the second lowest response option (2), you would sum all responses (10), multiply by the number of items in the short form (8) and divide by the number of items that were answered (5). Here (10x8)/5=16. If the result is a fraction, round up to the nearest whole number. This is a pro-rated raw score. Again, the formula is:*

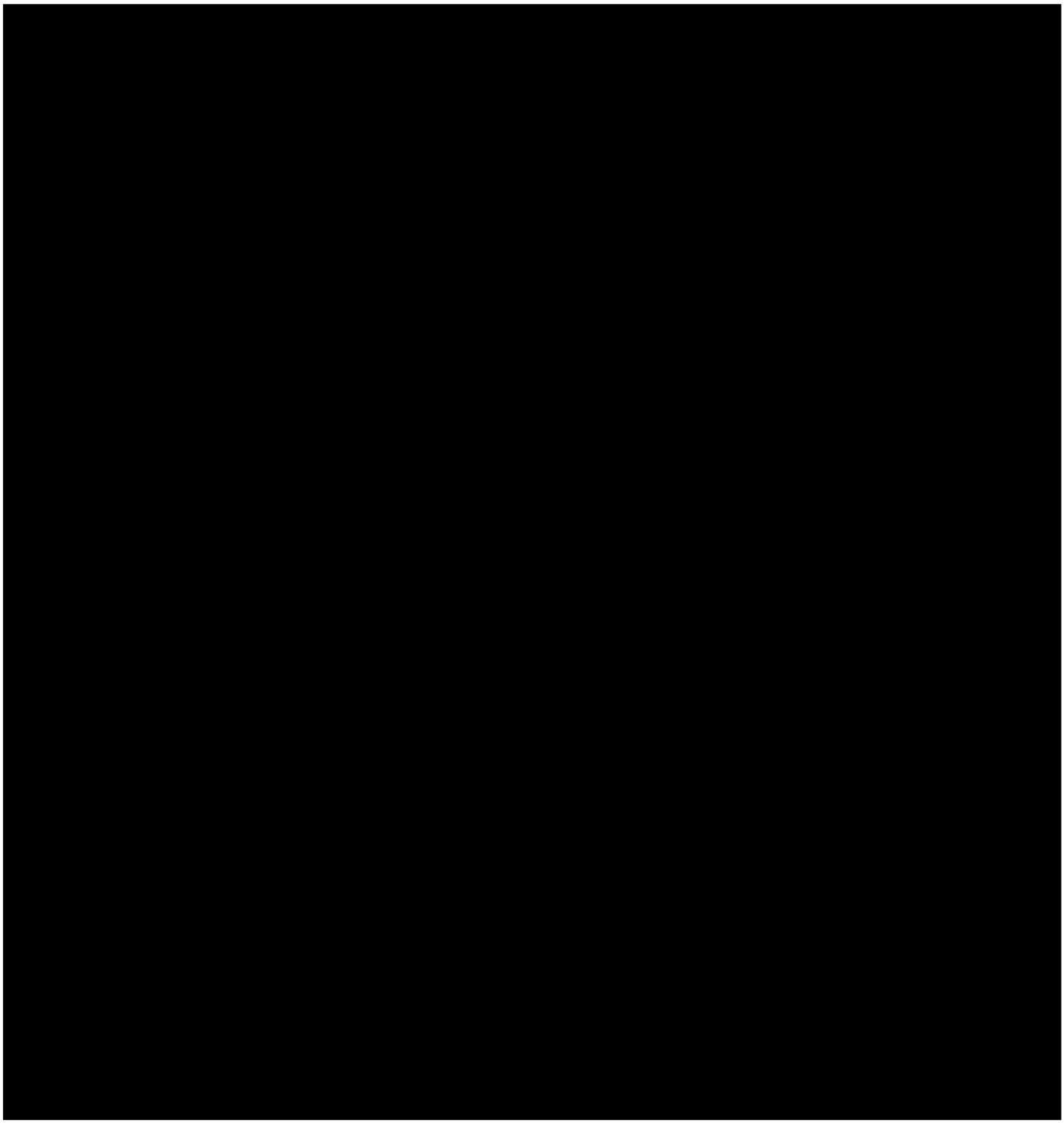
Figure 8.1.12-1: Formula for PROMIS Raw Score

$$\frac{(\text{Raw sum x number of items on the short form})}{\text{Number of items that were actually answered}}$$

Locate the applicable score conversion table in ... [Table 8.1.12-1] *and use this table to translate the total raw score or pro-rated score into a T-score for each participant. The T-score rescales the raw score into a standardized score with a mean of 50 and a standard deviation (SD) of 10.... The standardized T-score is reported as the final score for each participant. ... For pro-rated*

scores, this calculation assumes that responses are missing at random. This isn't always true. Therefore, use caution when interpreting the final pro-rated T-score."

The PROMIS brief guide refers to the calculation of 95% confidence intervals for subject's scores. Only the T-score will be used for every subject for the analyses described in this analysis plan.



8.2 Baseline Measures

For each subject, the baseline value of a parameter is defined as the last assessment of that parameter on or prior to the date/time of the first dose of the double-blind study medication.

If time of measurement of a parameter or time of first dose are missing, then the baseline value is defined as the last assessment of that parameter on or prior to the date of the first dose of the double-blind study medication.

8.3 Missing Measurements

For listings of efficacy measures, missing values will be represented as missing. For analyses involving binary secondary endpoints (the proportion of subjects with a decrease in ESSDAI of at least 3 from baseline and the proportion of subjects with a decrease in the ESSPRI of at least 1) a missing responder value due to discontinuation or for other reasons will be imputed as a non-responder.

8.5 Day Ranges for Analysis Time Points

Subjects do not always adhere strictly to the visit schedule timing in the protocol. Therefore, the designation of visits during the study period will be based on the day of evaluation relative to the Day 1 of the trial (day of first dosing = study Day 1) rather than the nominal visit recorded in the case report form (CRF). Tables below define visit windows to be used (note that the window for the Day 1 visit contains the first dosing day only). If a subject has more than one visit where a measurement is recorded within a window, the measurement closest to the target day will be used. In case of two visits being equally distant from the target day, the later measurement will be used in analyses. Exception to these rules applies to immunogenicity. For this, the least favorable value (toward a positive response) in the window will be used.

For subjects who discontinue from study therapy prematurely or completed the study, assessments performed after the last dose of study drug will be included in the efficacy data sets provided that the corresponding study visit is made within 42 days of the last dose. For some sensitivity analyses (see Section 7.1.5) all data available for the subject will be used, regardless of whether the visit is made within 42 days of the last dose.

For subjects who discontinued study drug, only assessments done within 42/56 (efficacy/safety) days after the last dose of study drug would be included for analysis purposes using the visit windows below.

Table 8.5-1: Day Ranges for Safety and Efficacy Analyses

Visit	Target Day	Safety Assessments(Vitals and, Laboratory (Hematology, Chemistry, Urinalysis)), and the following Efficacy Assessments: hsCRP, ESR, PK, ESSDAI, Joints, PhyGAD, SubGAD, ESSPRI, PROMIS Fatigue, Oral and Ocular Dryness	Efficacy Assessments: SF-36, FSFI, Unstimulated/Stimulated Salivary Flow, Schirmer Test, Tear Break-Up Time, Ocular Staining Score
Double-Blind Period*			
Day 1	1	≤ 1	≤ 1
Day 29 (Week 4)	29	2 - 43	
Day 57 (Week 8)	57	44 - 71	
Day 85 (Week 12)	85	72 - 99	2 - 127
Day 113 (Week 16)	113	100 - 127	
Day 141 (Week 20)	141	128 - 155	
Day 169 (Week 24/ET)	169	156 - 183[a]	128 - 212[a]
Open-Label Period*			
Day 197 (Week 28)	197	184[b] - 211	
Day 225 (Week 32)	225	212 - 239	
Day 253 (Week 36)	253	240 - 267	213[b] - 308
Day 281 (Week 40)	281	268 - 295	
Day 309 (Week 48)	309	296 - 323	
Day 365 (Week 52/ET)	365	324 - 351[c]	309 - 421[c]

* See Section 6.1 for the definition of double-blind and open-label periods

[a] The upper limit of the last visit window is 42/56 (efficacy/safety) days post the last study drug dose during the Double-Blind Period or the first dose date in the Open-Label period, whichever is earlier.

[b] The lower limit of the first visit window in the open-label period is the day after the first open-label dose.

[c] The upper limit of the last visit window is 42/56 (efficacy/safety) days post the last study drug dose during the open-label period.

Table 8.5-2: Day Ranges for Immunogenicity

Visit	Target Day	Day Range
Double-Blind Period*		
Day 1	1	≤ 1
Day 85 (Week 12)	85	2 - 127
Day 169 (Week 24/ET)	169	128 - 267[a]
Open-Label Period*		
Day 365 (Week 52/ET)	365	268[b] - 421[c]

Table 8.5-2: Day Ranges for Immunogenicity

Visit	Target Day	Day Range
Follow-Up Period**		
Follow-Up Day 85	85	22 - 127
Follow-Up Day 169	169	> 127

* See Section 6.1 above for the definition of double-blind and open-label periods
** Target Day and Day Range are relative to the last dose of study medication
[a] The upper limit of the last visit window is 21 days post the last study drug dose during the Double-Blind Period or the first dose date in the Open-Label period, whichever is earlier.
[b] The lower limit of the first visit window in the open-label period is after the first open-label dose.
[c] The upper limit of the last visit window is 21 days post the last study drug dose during the open-label period.

For all parameters, if a subject has more than one measurement included within a window, the assessment closest to the target day will be used. In case of 2 visits being equidistant from the target, the later measurement will be used in analyses. Exceptions to these rules apply to immunogenicity, ANA, dsDNA. For this safety indicator, the least favorable value (toward a positive) in the window will be used.

8.6 Multiple Assessments

When clinical assessments or lab samples are inadvertently taken or analyzed multiple times, this produces multiple measures for the same date (and time in the case of labs) for the same subject. In such cases, the average of such multiple assessments for labs or vitals will be used for this time point for this subject. For immunogenicity, ANA, dsDNA and the efficacy assessments, the least favorable value will be used.

8.7 Safety Data Conventions

Safety data will be handled according to the BMS safety data convention standards[3] and Supplement to Safety Guidelines for abatacept[4].

9 CONTENT OF REPORTS

The results of this study will be presented in a standard BMS Clinical Study Report (CSR).

CSR Primary analysis Month 6:

At the time of the primary analysis (i.e. after all subjects complete the Double-Blind Period), the final efficacy and safety analysis up to Day 169 will be provided (Double-Blind Period). In addition, all available safety data will be summarized for the cumulative abatacept period up to Year 1 (interim analysis). No efficacy analysis beyond Day 169 will be provided.

CSR Year 1:

This CSR will include the final analysis for Year 1 (efficacy and safety).

APPENDIX 1 DOCUMENT HISTORY

Table 9-1: Document History

Version	Statistician	Date	Notes/Revisions
1.0	[redacted]	13September2017	Original version
2.0	[redacted]	10 August 2018	• Subgroup analyses by stimulated salivary flow at screening and/or baseline < 0.1 (subgroup 1) and ≥ 0.1 (subgroup 2) will only be provided if there are at least 10 patients in each treatment group for both subgroups. • Subgroup analyes for mean change from baseline for ESSDAI at day 169 (by gender, race, region, ...) will only be provided if at least 10 subjects in each treatment group (instead of 10% currently in SAP v1.0). ESSDAI-by-time interaction removed from model. • Sensitivity analysis including the efficacy data after 42 days (post treatment up to day 169): only provided - if at Day 169 there are at least 10% of subjects with missing data in either treatment group for the primary analysis - and on top, at Day 169 there are at least 10% more non-missing measurement in either treatment groups using all collected data (including post treatment) compared to number of non-missing measurement in primary analysis • Sensitivity analysis for MAR (tipping point and control-based pattern imputation): only if at Day 169 there are at least 10% of subjects with missing data in either treatment group for primary analysis • Forest plot: for treatment difference instead of by treatment group • Change wording 'clinically important change' to 'clinically important improvement'. • Replacement of worst eye by study eye (and none study eye) • According to SAP v1.0 the minimal detectable change for 4 binary exploratory variables (stimulated and unstimulated salivary flow, Schirmer's test, ocular staining score and tear break-up time) will be explored to define a meaningful change in each endpoint. Implementation in SAP v2.0: Proportion of subjects with 25% improvement and proportion of subjects with 50% improvement • Added: proportion of subject with ESSDAI < 5 up at Day 169

Table 9-1: Document History

Version	Statistician	Date	Notes/Revisions
			• Added: proportion of subjects with ESSPRI < 5 up at Day 169 • Back-up models for minimal risk weight macro: - The back-up model 1: exclude the japan and salivary flow stratification variables. - The back-up model 2: exclude all stratification variables. • Back-up for proc MI (sensitivity analysis): - The back-up model 1: exclude the Japan and salivary flow stratification variables. - The back-up model 2: exclude all stratification variables. • Tear Break Up Time: In case only 2 measurements are available, the average of the 2 measurements will be calculated. In case there is only 1 measurement, that measurement will be used for the analysis. • Shift tables for ESSDAI components added • Proportion of improvement and proportion of worsening for ESSDAI components added • No longitudinal analysis for ESSDAI components in case changes from baseline = 0 for all subjects for all timepoints.

APPENDIX 2 RELEVANT PROTOCOL DEVIATIONS

- Eligibility Deviations:
  - Subjects who did not meet 2016 ACR/EULAR classification criteria at screening
  - Subjects with ESSDAI < 5 at screening
  - Subjects who were anti-SSA/Ro negative at screening
  - Subjects who have received IV, IM, SC or intra-articular corticosteroids within 4 weeks of first double-blind abatacept dose
  - Subjects who have taken MTX within 12 weeks of first double-blind abatacept dose
  - For subjects on hydroxychloroquine at first date of double-blind abatacept dose: subjects on hydroxychloroquine for < 12 weeks and/or the dose not stable for at least 4 weeks prior to Day 1
- Incorrect dosing:
  - Subjects received treatment different than randomized group
  - If more than 4 consecutive doses are missed just prior to the primary time point, Day 169
- Restricted and Prohibited medications:
  - Subjects receiving a single IM, IV, SC or oral course of high dose of corticosteroid (> 10mg/day for > 1 consecutive days) in the Double-Blind Treatment Period within 56 days of primary time point (i.e. the efficacy assessment done at Day 169).
  - Subjects receiving a non-biologic DMARD (except hydroxychloroquine already taken at Day 1) at any time during the Double-Blind Treatment Period up to the day prior to the efficacy assessment done at Day 169
  - Subjects receiving a biologic DMARD at any time during the Double-Blind Treatment Period up to the day prior to the efficacy assessment done at Day 169
  - Subjects adding hydroxychloroquine during the Double-Blind Treatment Period up to the day prior to the efficacy assessment done at Day 169 while not on this medication at Day 1.

Note: Day 169 is the time point used for the **primary analysis** of ESSDAI at Day 169.

APPENDIX 3 REGIONS

Table 9-2: Regions

Regions by Country	
North America	USA/Puerto Rico
	Canada
South America	Argentina
	Brazil
	Mexico
Europe	Czech Republic
	France
	Germany
	Italy
	Sweden
Asia	Japan
	Korea
ROW (Rest of World)	Australia