Statistical Analysis Plan I4V-MC-JAHH

A Randomized, Double-Blind, Placebo-Controlled, Parallel- Group, Phase 2 Study of Baricitinib in Patients with Systemic Lupus Erythematosus (SLE)

NCT02708095

Approval Date: 26-Jul-2017

1.   Statistical Analysis Plan
I4V-MC-JAHH: A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Phase 2 Study of Baricitinib in Patients with Systemic Lupus Erythematosus (SLE)

Baricitinib (LY3009104) Systemic Lupus Erythematosus

Study I4V-MC-JAHH is a Phase 2, randomized, double-blind, placebo-controlled study of baricitinib 4-mg and 2-mg for the treatment of moderately to severely active systemic lupus erythematosus during a 24-week treatment period.

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Protocol I4V-MC-JAHH
Phase 2

Statistical Analysis Plan 1 electronically signed and approved by Lilly on date provided below.

Approval Date: 26-Jul-2017 GMT

2. Table of Contents

LY3009104

LY3009104

LY3009104

Table of Contents

Table of Contents

Table of Contents

3. Revision History

Statistical Analysis Plan (SAP) Version 1 was approved prior to first unblinding.

4. Study Objectives

4.1. Objectives

Objectives	Endpoints
Primary	
To evaluate the effect of baricitinib 4-mg QD or 2-mg QD compared to placebo on remission of SLE arthritis and/or rash in patients with SLE receiving concomitant standard therapy over 24 weeks.	Change in the proportion of patients who achieve remission of arthritis and/or rash as defined by SLEDAI-2K at Week 24 compared to baseline.
Secondary	

To evaluate the effect of baricitinib 4-mg QD or 2-mg QD compared to placebo on overall SLE disease activity in patients with SLE receiving concomitant standard therapy over 24 weeks.	• Change in the proportion of patients achieving SRI-4 response at Week 24 compared to baseline. SRI-4 response is defined as:
	   ○ Reduction of ≥4 points from baseline in SLEDAI-2K score
	   ○ No new BILAG A or no more than 1 new BILAG B disease activity scores
	   ○ No worsening (defined as an increase of ≥0.3 points [10 mm] from baseline) in the Physician's Global Assessment of Disease Activity
	• Change in the proportion of patients achieving a reduction of ≥4 points from baseline in SLEDAI-2K score at Week 24 compared to baseline.
	• Change in SLEDAI-2K total score at Week 24 compared to baseline.

To evaluate the effect of baricitinib 4-mg QD or 2-mg QD compared to placebo on the Patient's Global Assessment of Disease Activity in patients with SLE receiving concomitant standard therapy over 24 weeks.	• Change in Patient's Global Assessment of Disease Activity at Week 24 compared to baseline.
To characterize the PK of baricitinib 4-mg QD or 2-mg QD in patients with SLE receiving concomitant standard therapy over 24 weeks.	• Plasma baricitinib concentrations will be analyzed using a popPK approach.

Abbreviations: BILAG = British Isles Lupus Assessment Group; PK = pharmacokinetics; PopPK = population pharmacokinetics; QD = once daily; SLE = systemic lupus erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000; SRI = SLE Responder Indexer.

4.2. Exploratory Objectives

Exploratory Objectives	Endpoints
To evaluate the effect of baricitinib 4-mg QD or 2-mg QD compared to placebo on SLE arthritis and mucocutaneous disease activity in patients with SLE receiving concomitant standard therapy over 24 weeks.	- The proportion of patients who achieve remission of arthritis and/or rash as defined by the SLEDAI-2K at Week 12 compared to baseline. - Continuous and categorical endpoints based on swollen joint counts and tender joint counts in patients with joint disease activity at baseline at Weeks 12 and 24. - Continuous and categorical endpoints based on the CLASI.

To evaluate the effect of baricitinib 4-mg QD or 2-mg QD compared to placebo on overall SLE disease activity in patients with SLE receiving concomitant standard therapy over 24 weeks.	- Time to and proportion of patients experiencing first flare (any severity) and first severe flare defined by the SFI. - Change in SLICC/ACR Damage index score at Week 24 compared to baseline. - Change at Weeks 12 and 24 in Physician Global Assessment of Disease Activity. - Proportion of patients who achieve a LLDAS. The operational definition of LLDAS is fulfilled when all of the following criteria are met: (i) SLEDAI-2K total score ≤4, - with no activity in major organ systems (renal, CNS, cardiopulmonary, vasculitis, fever). Defined as no activity in any items (all items not present) in the SLEDAI-2K CNS, Vascular, Renal, Cardiovascular and Respiratory, or Constitutional Organ domains. Therefore, possible activity in Musculoskeletal, Mucocutaneous, Immunologic, and/or Hematologic only. - and no hemolytic anemia. Per SFI, hemolytic anemia defined as hemoglobin <7 g/dL or decrease in hemoglobin >3 g/dL. - and no gastrointestinal activity. Per BILAG, no gastrointestinal activity defined as BILAG Gastrointestinal Organ domain score of D or E. (ii) no new features of lupus disease activity compared to the previous assessment (per SLEDAI-2K) (iii) PGA, scale 0–3, 0-100 mm) ≤1 (33 mm) (iv) a current prednisone (or equivalent) dose ≤7.5 mg daily (v) well tolerated standard maintenance doses of immunosuppressive drugs and approved biological agents, excluding investigational drugs. Defined as no increases in concomitant immunosuppressant and antimalarial drugs above baseline dose.
To evaluate the corticosteroid-sparing effect of baricitinib 4-mg QD or 2-mg QD compared to placebo in patients with SLE receiving concomitant standard therapy over 24 weeks.	The proportion of patients receiving ≥10 mg/day prednisone at baseline able to reduce prednisone (or equivalent) dose to ≤7.5-mg for 12 consecutive weeks between Week 12 and Week 24.

LY3009104

To evaluate the effect of baricitinib 4-mg QD or 2-mg QD compared to placebo on serologic markers of SLE in patients with SLE receiving concomitant standard therapy over 24 weeks.	• In patients with elevated anti-dsDNA at baseline, change in anti-dsDNA level at Week 12 and Week 24 compared to baseline. • In patients with low C3 and C4 at baseline, change in C3 and C4 levels at Week 12 and Week 24 compared to baseline.
To evaluate the effect of baricitinib 4-mg QD or 2-mg QD compared to placebo on (PROs in patients with SLE receiving concomitant standard therapy over 24 weeks.	• Change in Worst Pain Numeric Rating Scale (NRS) at Weeks 12 and 24 compared to baseline. • Change in Worst Joint Pain Numerical Rating Scale (NRS) at Weeks 12 and 24 compared to baseline. • Change in Worst Fatigue Numeric Rating Scale (NRS) at Weeks 12 and 24 compared to baseline. • Change in MCS, PCS, and domain scores in the SF-36v2 acute at Weeks 12 and 24 compared to baseline.
To explore dose/exposure response relationships with key efficacy (such as SLEDAI-2K) and safety (such as hematologic parameters) endpoints of interest.	• The relationship between baricitinib PK and key efficacy and safety endpoints will be characterized using popPK/ (popPK/PD) modeling.

Abbreviations: BILAG = British Isles Lupus Assessment Group; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; CNS = central nervous system; dsDNA = double-stranded deoxyribonucleic acid; MCS = mental component score; LLDAS = Lupus Low Disease Activity State; PGA = Physician Global Assessment; PCS = physical component score; PD = pharmacodynamics; PK = pharmacokinetics; PopPK = population pharmacokinetics; PROs = patient-reported outcomes; QD = once daily; SF-36v2 = Short-Form 36-item health survey version 2; SLE = systemic lupus erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000; SFI = SLEDAI Flare Index; SLICC/ACR = Systemic Lupus International Collaborating Clinics/American College of Rheumatology.

5. Study Design

5.1. Summary of Study Design

Study I4V-MC-JAHH is a Phase 2, multicenter, randomized, double-blind, placebo-controlled, parallel-group, outpatient, 24-week study evaluating the efficacy and safety of baricitinib 4-mg and 2-mg in patients with systemic lupus erythematosus (SLE) receiving standard therapy.

The study consists of 3 periods:

1. **Screening Period:** Patients meeting entry criteria begin the 3- to 42-day screening period with signing of the Informed Consent Form. Patients will be screened for study enrollment eligibility at Visit 1, and eligible patients will be randomized at Visit 2 (Week 0).

2. **Double-Blinded Treatment Period:** Randomized patients begin the 24-week double-blind, placebo-controlled treatment period at Visit 2 (Week 0) and administer study drug daily through Visit 9 (Week 24).

   All patients will continue on stable background standard therapy consisting of corticosteroids, nonsteroidal anti-inflammatory drugs (NSAIDs [for which the NSAID use is intended for treatment of signs and symptoms of SLE]), a single antimalarial (such as hydroxychloroquine or chloroquine), and/or a single immunosuppressant (such as methotrexate, azathioprine, or mycophenolate). Nonsteroidal anti-inflammatory drugs, antimalarial, and immunosuppressant doses should remain stable throughout the double-blinded treatment period. Decreases in corticosteroid dose, as well as subsequent increases to less-than-or-equal-to the baseline dose will be permitted between Week 0 and Week 16. No changes in corticosteroid dose (increases or decreases) will be permitted between Week 16 and Week 24.

   Initiation or increase in dose of NSAIDs (for which the NSAID use is intended for treatment of signs and symptoms of SLE), corticosteroids (above the baseline dose through Week 16, or any increase after Week 16), antimalarials, and/or immunosuppressants is not permitted after randomization. After randomization, patients requiring the above-mentioned initiation or increase in dose of SLE standard-of-care medications compared to baseline, as well as those meeting discontinuation criteria, will be discontinued from study drug and complete the study procedures for an early termination visit (ETV).

3. **Follow-up Period:** Patients who complete the study through Visit 9 (Week 24), as well as those who discontinue study treatment early, will have a post-treatment follow-up visit (Visit 801) approximately 4 weeks after the last dose of study drug.

Figure JAHH.5.1 illustrates the study design. The 3 dosing regimens are described in Section 6.1. As this is a double-blind study, the blinding procedure described in Section 7 will be followed.
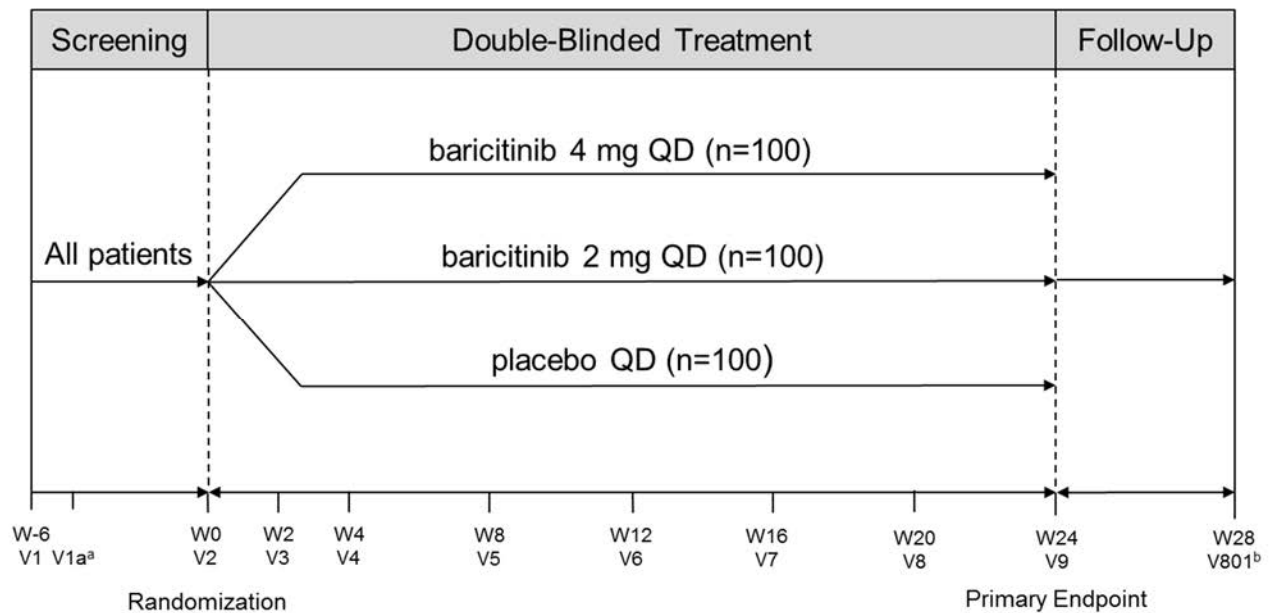
Abbreviations: n = number of patient per treatment group; PPD = purified protein derivative; QD = once daily; V = visit; W = week.

[a] For those patients with a PPD skin test for tuberculosis placed at Visit 1, Visit 1a must occur 48 to 72 hours post-Visit 1 for PPD to be read.

[b] All patients should return for a posttreatment follow-up visit (V801) 4 weeks after the last dose of investigational product.

Figure JAHH.5.1.   Illustration of study design for Clinical Protocol I4V-MC-JAHH.

5.2.   Determination of Sample Size

Approximately 300 patients will be randomized 1:1:1 to receive baricitinib 4-mg once daily (QD), baricitinib 2-mg QD, or placebo. With 100 patients per treatment group, this study will have approximately 81% power to detect a difference between baricitinib 4-mg or 2-mg and placebo of ≥20% in remission rates of rash and/or arthritis at Week 24. The sample size was determined based on the chi-square test with 2-sided type I error of 5% and 40% remission rate of rash and/or arthritis in placebo treatment group.

The above sample size and power estimates were obtained from nQuery® Advisor 7.0.

5.3.   Method of Assignment to Treatment

Patients who meet all criteria for enrollment will be randomized in a 1:1:1 ratio (baricitinib 4-mg, baricitinib 2-mg, or placebo) to double-blind treatment at Visit 2 (Week 0). Randomization will be stratified by disease activity (Systemic Lupus Erythematosus Disease Activity Index 2000 [SLEDAI-2K] <10; SLEDAI-2K ≥10), anti-dsDNA (double-stranded deoxyribonucleic acid) status (positive; negative), and region (United States, Europe, Asia, rest of world). Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS). The IWRS will be used to assign 2 bottles (Bottle A and Bottle B), each containing 36 double-blind investigational product (IP) tablets to each patient, starting at Visit 2 (Week 0), and at each visit through Visit 8 (Week 20). Site

personnel will confirm that they have located the correct bottles by entering a confirmation number found on the bottles into the IWRS. Patients will be instructed to take 1 tablet from Bottle A and 1 tablet from Bottle B each day.

This study will be conducted internationally at approximately 90 sites. Table JAHH.5.1 describes how each region will be defined for the statistical analyses and summaries.

Table JAHH.5.1.   Countries and Their Geographical Regions

Geographical Region	Country or Countries
United States	United States (including Puerto Rico)
Europe	Austria, France, Poland, Romania, Spain
Asia	Japan, South Korea, Taiwan
Rest of World	Argentina, Mexico

5.3.1. Treatment Administered

This study involves a comparison of baricitinib 4-mg and baricitinib 2-mg administered orally QD with placebo. Table JAHH.5.2 shows the treatment regimens.

Table JAHH.5.2.   Treatment Regimens

Regimen	Dose Week 0 through Week 24
Baricitinib 4-mg QD	1 baricitinib 4-mg tablet and 1 placebo tablet matching baricitinib 2-mg
Baricitinib 2-mg QD	1 baricitinib 2-mg tablet and 1 placebo tablet matching baricitinib 4-mg
Placebo	2 placebo tablets: 1 matching baricitinib 4-mg and 1 matching baricitinib 2-mg

Abbreviation: QD = once-daily administration.

The investigator or the investigator's designee is responsible for the following:

1. Explaining the correct use of the IP to the patient or legal representative.

2. Verifying that instructions are followed properly.

3. Maintaining accurate records of IP dispensing and collection.

4. At the end of the study, returning all unused medication to Lilly, or its designee, unless the sponsor and sites have agreed that all unused medication is to be destroyed by the site, as allowed by local law.

5.4. Handling of Dropouts or Missing Data

The general methods for missing data imputation are outlined in Sections 5.4.1 and 5.4.2.

5.4.1. Non-responder Imputation (NRI)

Analysis of categorical efficacy and health outcomes variables will be assessed using non-responder imputation (NRI). The following patients will be considered nonresponders at each visit for all categorical analyses:

- Patients who are not responders at that visit.
- Patients who permanently discontinue study treatment at any time prior to that visit, for any reason.
- Patients who require initiation or increase in dose compared to baseline of NSAIDs for 30 or more consecutive days, corticosteroids, antimalarials, or immunosuppressants after randomization. Patients will be analyzed as nonresponders from the day of initiation or increase in medication.
- Patients without a response at that visit.

5.4.2. Mixed-Effects Model Repeated Measures

Per protocol, patients who discontinue study treatment early should complete the early termination (ET) visit and proceed to the post-treatment follow-up period. An ET visit is considered a follow-up visit if it occurs 28 days or more after treatment discontinuation. Therefore, only data within 28 days after treatment discontinuation will be considered in the double-blinded treatment period. However, if a patient has a treatment interruption and never resumes treatment, the date of the treatment interruption becomes the date of treatment discontinuation. Patients with this type of interruption may have several months of data after treatment discontinuation. For consistency purposes, only data from visits that occur less than 28 days after treatment discontinuation will be used in the analysis of continuous measures. A likelihood-based Mixed-effects Model Repeated Measures (MMRM) is the primary analysis method for analyzing the continuous efficacy and health outcomes endpoints. In this method, the missing data are assumed to be missing at random.

6.   A Priori Statistical Methods

Statistical analysis of this study will be the responsibility of Eli Lilly and Company. The statistical analyses will be performed using SAS® Version 9.2 or higher.

6.1.   General Considerations

Efficacy and safety analyses will be conducted on the modified intent-to-treat (mITT) and safety populations, respectively. These populations are defined in Section 6.4.

Not all displays described in this SAP will necessarily be included in the clinical study report (CSR). Not all displays will necessarily be created as a "static" display. Some may be incorporated into interactive display tools instead of or in addition to a static display. Any display described in this SAP and not provided would be available upon request.

Continuous data will be summarized using the sample size, mean, standard deviation (SD), minimum (min), maximum (max), and median. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean and median will be reported to 1 more decimal place than the raw data recorded in the database. The SD will be reported to 2 more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be 4 for any summary statistic.

Categorical data will be summarized in terms of the sample size, the number of patients providing data at the relevant time point, frequency counts, and the percentages corresponding to the appropriate method. Percentages will be presented to 1 decimal place.

All tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, unless otherwise stated. The p-values will be rounded up to 3 decimal places. For example, any p-value strictly >0.049 and ≤0.05 will be displayed as 0.050. This guarantees that on any printed statistical output, the unrounded p-value will always be less than or equal to the displayed p-value. A displayed p-value of 0.001 will always be understood to mean ≤0.001. Likewise, any p-value displayed as 1.000 will be understood to mean >0.999 and ≤1.

The primary analysis method for treatment comparisons of categorical efficacy and health outcome variables will be made using a logistic regression analysis with treatment group, baseline disease activity (SLEDAI-2K <10; SLEDAI-2K ≥10), baseline anti-dsDNA status (positive; negative), and region in the model. For each treatment comparison, an estimate of the odds ratio, corresponding Wald 95% confidence interval (CI), and p-value will be presented. Each baricitinib treatment group will be compared to the placebo group. The p-value for all other explanatory variables will also be presented. When logistic regression sample size requirements are not met (<5 responders in any category for any factor), the p-value from the Fisher's exact test is produced instead of the odds ratio and 95% CI.

The primary analyses for continuous efficacy and health outcome variables will be made using MMRM. When MMRM is used, the model will include treatment, baseline score, baseline disease activity (SLEDAI-2K <10, SLEDAI-2K ≥10), baseline anti-dsDNA status (positive; negative), region, visit, and the interaction of treatment-by-visit as fixed factors. An

LY3009104

unstructured covariance matrix will be used to model the within-patient variance-covariance errors. If this analysis fails to converge, the following structures will be tested in this pre-specified order: 1) heterogeneous Toeplitz (TOEPH), 2) heterogeneous autoregressive (ARH[1]), and 3) heterogeneous compound symmetry (CSH). The first covariance structure that converges using this pre-specified order will be used. The Kenward-Roger method will be used to estimate the degrees of freedom. Type III sums of squares for the least squares (LS) means will be used for the statistical comparison. The LS mean for each treatment group along with the estimate of the difference between treatments (difference between each baricitinib dose group and placebo), standard error (SE), p-value, and the 95% CIs will be reported at each visit along with p-values.

When analysis of covariance (ANCOVA) is used for efficacy/Health Outcomes measures, the model will include treatment, baseline score, baseline disease activity (SLEDAI-2K <10; SLEDAI-2K ≥10), baseline anti-dsDNA status (positive; negative), and region fitted as explanatory variables. Type III sums of squares will be used. Differences in LS means between treatment groups will be displayed, with the p-value associated with the LS mean comparison to placebo (for each baricitinib dose group) along with the 95% CI of the LS mean difference also provided. In addition to the LS means and tests, mean, SD, minimum, first quartile, median, third quartile, and maximum will be displayed.

Fisher's exact test will be used for all adverse events (AE), discontinuation, and other categorical safety data. Continuous vital signs and laboratory values will be analyzed by an ANCOVA with treatment and baseline values in the model (see Sections 6.12.4 and 6.12.5. Statistics to be presented are the same as those for ANCOVA analyses stated above.

The following is a description of the analysis groups planned for this study.

6.1.1. Analysis Groups for the Efficacy, Health Outcomes, and Safety Data

Summaries of efficacy, health outcomes, and safety data will be presented according to the following treatment groups:

- Patients randomized to baricitinib 4-mg
- Patients randomized to baricitinib 2-mg
- Patients randomized to placebo

When provided, summaries of efficacy data collected during the posttreatment follow-up period will be presented separately from the main summaries according to the treatment group that the patients were randomized to initially.

6.1.2. Definition of Baseline and Postbaseline Measures

Baseline will be defined as the last available value before the first dose of study drug for both efficacy and safety analyses, unless otherwise specified. In most cases, this will be the measure

recorded at Week 0 (Visit 2). The treatment period starts after the first study drug administration at Visit 2 (Week 0) and ends on one of the following:

1. The date of Visit 9 (Week 24) for patients that completed treatment
2. For patients that discontinued treatment early, the date of ETV, if this visit occurs less than 28 days after treatment discontinuation. Otherwise, it would be the date of the last Visit that occurs less than 28 days after treatment discontinuation.

Change from baseline will be calculated as the visit value of interest minus the baseline value.

Postbaseline measurements are collected after study drug administration through Visit 9 (Week 24) or an early discontinuation visit. For electronic patient-reported outcomes (ePROs) related to efficacy assessments, unscheduled postbaseline visits that fall within the protocol-defined visit windows will be summarized in the by-visit analyses if there is no scheduled visit available. If more than 1 value is reported for the same scheduled visit, then the first value will be summarized in the by-visit analyses.

The Follow-up Period includes all visits that are 28 days or more after treatment discontinuation for patients that discontinue treatment early. For patients that complete treatment, the Follow-up Period includes only Visit 801. The baseline value for the efficacy analysis of the posttreatment follow-up period is defined as the last nonmissing assessment prior to entering the posttreatment Follow-up Period, that is Week 24 (Visit 9) for treatment completers or last visit that occurred less than 28 days after treatment discontinuation for patients that discontinued treatment early. When the Follow-up Period includes more than 1 visit, any efficacy analyses that are done for this period will only include the first visit of this period.

6.2. Multicenter Studies

Geographic region will be included as a covariate in all analyses, unless otherwise specified. Geographic region will be categorized as follows:

- USA
- Asia
- Europe
- Rest of World

6.3. Multiple Comparisons/Multiplicity

Comparisons between the baricitinib 4-mg and baricitinib 2-mg groups versus the placebo group will be performed for all analyses for the double-blinded treatment period. No adjustment for multiplicity will be done for these comparisons.

6.4. Definition of Populations

Modified Intent-to-treat (mITT) Population: All efficacy and health outcome analyses will be done using the mITT population. The mITT population is defined as all randomized patients who received at least 1 dose of study drug, even if the patient did not receive the correct

treatment or, otherwise, did not follow the protocol. Patients will be analyzed according to the treatment to which they were assigned.

Safety Population: This analysis set is defined as all randomized patients who received at least 1 dose of the study drug and did not discontinue the study for the reason "Lost to Follow-up" at the first postbaseline visit. This definition excludes patients with no safety assessments postbaseline so that incidence rates are not underestimated. Safety analyses will be performed on the safety population.

Follow-up Population: Efficacy analyses for the posttreatment follow-up period may be conducted on the follow-up population, defined as all randomized patients who received at least 1 dose of study drug and:

 a) Have completed treatment and have entered the posttreatment follow up period, or
 b) Have discontinued treatment early and have at least 1 visit that occurred 28 days or more after treatment discontinuation

Patients will be analyzed according to the dosing regimen to which they were assigned in the treatment period.

6.5. Patient/Subject Disposition

The following patient disposition summaries, frequency counts of screen failures, will be provided. Patient disposition will be summarized using the mITT population. Frequency counts and percentages of patients who complete or discontinue early from the study treatment will be summarized separately by treatment group along with the reason for study treatment discontinuation. Treatment groups will be compared using the Fisher's exact test. Frequency counts and percentages of patients who complete or discontinue both study treatment and follow-up and those who complete or discontinue the posttreatment follow-up period will also be summarized separately by treatment group along with the reason for discontinuation. However, no treatment comparison will be done. Reason for study discontinuation is defined as reason for study treatment discontinuation if the patient discontinues study treatment early. Otherwise, reason for study discontinuation would be defined as the reason for discontinuing the posttreatment follow-up visit. A listing of patient disposition will be provided for all randomized patients, with the extent of their participation in the study and the reason for discontinuation.

6.6. Treatment Compliance

Patient compliance with study medication will be assessed from the randomization visit (Week 0) to Visit 9 (Week 24) during the treatment period. Compliance will be summarized from randomization until end of treatment using the mITT population. A patient is considered noncompliant if >20% of the prescribed doses during the study are missed, not including doses withheld by the investigator. A patient who takes >120% of the prescribed doses during the treatment period is also considered noncompliant.

Compliance to study drug in the period of interest will be calculated as follows:

$$\frac{\text{actual total \# of tablets used}}{\text{Expected total \# of tablets used}} * 100$$

where the actual total number of tablets used equals the total number dispensed minus the total number returned in the period; and the expected total number of tablets used equals the number of days (multiplied by 2) in the period of interest.

If a patient has a dose temporarily interrupted by the investigator during the period, the total number of days that drug was withheld will be deducted from the total number of days in the calculation of the expected total number of tablets used.

The summary statistics of the percent of compliance and noncompliance rate will be summarized by treatment group. The percent of compliance for Week 0 through Week 24 will be presented, along with the associated noncompliance rates.

6.7. Patient/Subject Characteristics

Patient characteristics, including demographics, will be summarized using the mITT population by treatment group. The summary will include descriptive statistics such as the number of patients (n), mean, SD, median, min, and max for continuous measures, and frequency counts and percentages for categorical measures. No formal statistical comparisons will be made among treatment groups, unless otherwise stated.

The following continuous demographic and baseline characteristic variables will be summarized using descriptive statistics. Baseline is defined, as in Section 6.1.2, unless otherwise specified.

- Age (in years) will be calculated using an imputed date of birth of July 1st in the year of birth collected in the electronic case report form (eCRF). It will be calculated as:

    Age (in years) = (date of first dose – imputed date of birth)/365.25

- Height (cm)

- Weight (kg)

- Body mass index (BMI) (kg/m^2) at Visit 2.

    BMI (kg/m^2) = Weight [kg] / (Height [m])^2

- SLEDAI-2K score

- Patient's Global Assessment of Disease Activity score

- Physician's Global Assessment of Disease Activity score

- Total Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index score

- Time since onset of lupus (years). Note that time since onset of lupus will be calculated using the date of onset of lupus (as recorded on the SLE History eCRF page) as follows:

    Time since onset of lupus (years) =
    (date of first dose - date of onset of lupus + 1) / 365.25

- Number of tender joints (from 28-tender joint count)

- Number of swollen joints (from 28-swollen joint count)

- Daily dose of corticosteroid (mg/day). Patients with a non-zero dose at Visit 2 will be included in this baseline summary, see Section 6.8.1 for details of prednisone (or equivalent) baseline dose.

- Worst level of fatigue over the past 7 days (as measured by the Worst Fatigue Numeric Rating Scale [NRS] at baseline)

- Worst level of pain over the past 7 days (as measured by the Worst Pain NRS) at baseline

- Worst level of joint pain (as measured by the Worst Joint Pain NRS) at baseline

- Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI) Total Activity Score

- CLASI Total Damage Score

- Complement C3 and C4 levels

- Anti-dsDNA level

- Proteinuria as assessed by the urine protein to urine creatinine ratio

- Serum immunoglobulin (Ig) A, IgG, and IgM concentrations

- Estimated glomerular filtration rate (eGFR)

The following categorical variables will be summarized using frequency counts and percentages:

- Age group (<65, 65 to <75, 75 to 85 and ≥85)

- Sex

- Race

- Country

- Region (as defined in Section 5.3)

- Ethnicity

- Anti-dsDNA status (positive defined as >= 30 IU/mL)

- SLEDAI-2K status (<10 or ≥10)

British Isles Lupus Assessment Group (BILAG) organ system involvement at baseline (yes or no for each organ system domain). Involvement requires a baseline BILAG disease activity score of A or B.

- BILAG A organ system involvement at baseline (yes or no for each organ system domain). Involvement requires a baseline BILAG disease activity score of A.

- SLEDAI-2K organ system involvement at baseline (yes or no for each organ system domain).

- C3 status (less than the lower limit of normal [LLN] (<90.0 mg/dL)

- C4 status (less than LLN [<10.0 mg/dL])

- Anti-Sm+ antinuclear antibodies ($\geq$30 AU/mL) (yes or no)

- Anti-RNP+ antinuclear antibodies ($\geq$30 AU/mL) (yes or no)

- Anti-Sjögren's-syndrome-related antigen A (also called anti-Ro [Anti-SSA/Ro+]) antibodies ($\geq$30 AU/mL) (yes or no)

- Anti-Sjögren's-syndrome-related antigen B (also called anti-La [Anti-SSB/La+]) antibodies ($\geq$30 AU/mL) (yes or no)

- Anti-phospholipid antibody +, overall, and for cardiolipin IgA+ (>11 APL), Cardiolipin IgG+ (>14 GPL) and Cardiolipin IgM+ (>12 MPL) (yes or no)

- ANA positive (titer >= 1:80)

- Corticosteroid use (yes or no), and within those taking corticosteroids, <10 mg/day or $\geq$10 mg/day. See Section 6.8.1 for details of prednisone (or equivalent) baseline dose.

- Corticosteroid use (yes or no), and within those taking corticosteroids, <7.5 mg/day or $\geq$7.5 mg/day. See Section 6.8.1 for details of prednisone (or equivalent) baseline dose.

- Immunosuppressant use (yes or no)

- Mycophenolate mofetil use (yes or no)

- Azathioprine use (yes or no)

- Methotrexate use (yes or no)

- Antimalarial use (yes or no)

- NSAID use (yes or no)

Note that instructions for selecting medications are given in Appendix 1.

Historical illnesses and preexisting conditions will be classified using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA). The number and percentage of patients with historical illnesses and preexisting conditions will be provided by treatment group,

overall and by system organ class (SOC) and preferred term (PT) using the mITT population. Note that conditions with a partial or missing start date will be assumed to be "not preexisting" unless there is evidence, through comparison of partial dates, to suggest otherwise. Patients will only be counted once in each PT.

6.8. Concomitant Therapy

Medication start and stop dates will be compared to the date of the first dose of study treatment (recorded on the Study Drug Administration page of the eCRF) to allow medications to be classified as "previous" or "concomitant."

Medications that started on or before treatment end date and are ongoing or ended during the treatment period will be classified as "concomitant" medication. Medications that start and end before the first dose date will be classified as "previous" medication. Note that medications with partial or missing start and/or stop dates will be assumed to be "previous" unless there is evidence, through comparison of partial dates, to suggest otherwise.

For Zoster and tuberculosis (TB) immunizations, since only start date is reported, if start date is missing, these medications will be considered to be prior therapy.

Concomitant medication use will be summarized by treatment, organized according to preferred medication name within Anatomical Therapeutic Chemical (ATC) Level 2, with ATC Level 2 and preferred medication names sorted by frequency in the baricitinib 4-mg group. The summaries of concomitant medications will be provided separately for concomitant medications used to treat SLE and for statin concomitant medications. Previous medications used to treat SLE will be summarized in a similar fashion. Note that a patient will only be counted once, regardless of how many times medication included under the same PT was taken.

A by-patient listing of all concomitant medications will be provided.

Patients with a prohibited increase of concomitant medications of special interest will be summarized by type of increase (eg, increase of steroids, NSAIDs, immunosuppressants, or antimalarials). Prohibited increases of concomitant medications will be defined as follows:

- For corticosteroids, all doses of steroids will be converted to an equivalent dose of prednisone (see Section 6.8.1). Using the prednisone-equivalent dose, a daily dose of corticosteroid will be derived for each study day from baseline to the treatment discontinuation day for patients that discontinue treatment early or to the day before for treatment completers. If any postbaseline daily dose of corticosteroid is greater than that of baseline, then the patient will have a prohibited increase of corticosteroids.
- For immunosuppressant and antimalarial medications, a daily dose will be calculated for each medication by PT. If 1 or more medications increase dose from baseline or a new medication is started during the treatment period, then the patient will have a prohibited increase of immunosuppressant or antimalarial medications.
- For NSAIDs, a daily dose will be calculated for each medication by PT. If 1 or more medications increase dose from baseline or a new medication is started during the treatment period and lasts 30 or more, then the patient will have a prohibited increase of NSAIDs.

6.8.1. Corticosteroids

To allow for assessments of changes in doses of various corticosteroids, it was necessary to standardize all corticosteroid doses to an equivalent prednisone dose. Table JAHH.5.3 provides a summary of frequent corticosteroids and their prednisone-equivalent dose. The dose of the drug listed in Column 1 of Table JAHH.5.3 is multiplied by the conversion factor in Column 2 to provide the prednisone equivalent dose (in milligrams). This dose of corticosteroid will be referred to as "prednisone (or equivalent)" throughout this document.

Table JAHH.5.3.   **Conversion Factors for Calculating Prednisone (or Equivalent) Doses**

Column 1 Corticosteroid Preferred Term	Column 2 Conversion Factor for Converting to an Equivalent Prednisone Dose
Prednisone	1
Prednisolone	1
Methylprednisolone	1.25
Triamcinolone	1.25
Cortisone	0.2
Hydrocortisone	0.25
Betamethasone	6.25
Dexamethasone	6.25
Paramethasone	2.5
Deflazacort	0.83

See Appendix 1 for a complete table showing conversion factors for each corticosteroid medication identified during the study, instructions for selecting corticosteroids, and the manual review process. References for this table can also be found in Appendix 1.

Baseline prednisone (or equivalent) dose will be the total daily dose of all corticosteroids being taken by a patient at Visit 2. This baseline dose will be used for baseline summaries and for comparisons to later visits. A daily dose of prednisone (or equivalent) will be calculated for each day between baseline and treatment discontinuation. The daily dose of prednisone (or equivalent) at each visit date will be used as the corticosteroid dose for that visit.

The following analyses will be performed. For analysis of observed data, a patient will be included in a visit only if the patient had not discontinued treatment at that time:

1. For patients with a baseline corticosteroid dose ≥ 10 mg/day, the number and percentage of patients who were able to decrease their dose to ≤ 7.5 mg/day for 12 consecutive weeks between Week 12 and Week 24 will be presented using observed data and NRI.

2. For patients with a baseline corticosteroid dose ≥10 mg/day, the number and percentage of patients who were able to decrease their dose to ≤7.5 mg/day at each visit will be presented using observed data and NRI.

3. The number and percentage of patients with a dose of corticosteroids (prednisone or equivalent) in the following mutually exclusive categories at Visit 9 (Week 24) will be presented by treatment group using observed data only.

   - Daily dose >7.5 to ≤20 mg/day

   - Daily dose ≤7.5 mg/day

   - No corticosteroid use

4. The daily dose of corticosteroids and change from baseline (as defined above) in dose will also be summarized by treatment group and visit using the MMRM model described in Section 6.1.

5. In addition, the number and percentage of patients falling into the following corticosteroid dose categories will be presented by treatment group at Weeks 12 and 24 using observed data and NRI. The following categories will be assessed for those patients with a baseline corticosteroid dose >0 mg/day and also for those patients with a baseline corticosteroid dose ≥10 mg/day:

   - Any decrease (actual daily dose at postbaseline visit decreased by >0 mg/day, compared to baseline)

   - Proportion of patients with at least a 50% decrease (actual daily dose at postbaseline visit whose decrease is greater than or equal to 50%)

   - Proportion of patients with at least a 25% decrease (actual daily dose at postbaseline visit whose decrease is greater than or equal to 25%)

   - No decrease (patients who complete the study who had no decrease)

6.8.2. Prior Antimalarials and Immunosuppressants

The number and percentage of patients who received prior therapy for SLE (as recorded on the Prior Therapy of Interest - SLE eCRF page) will be presented by treatment group, overall, and by ATC class. The number and percentage of patients with the following reasons for discontinuing previous therapy will also be presented by treatment group and ATC class: Inadequate response, AE, Cannot Afford Medication, and Other.

6.9. Efficacy Analyses

All efficacy measures will be summarized on the mITT population and will be performed through Week 24 (Visit 9) for treatment completers and through the last visit that occurs less than 28 days after treatment discontinuation for patients that discontinue treatment early. Treatment group comparisons will be made between each of the baricitinib dose groups and the placebo group.

6.9.1. Primary Analyses

The primary efficacy measure is the proportion of patients who achieve remission of arthritis and/or rash, as defined by the SLEDAI-2K at Week 24. A remission is defined as the following:

Either arthritis OR rash (as defined by the SLEDAI-2K) are required to be present at baseline. If only arthritis is present at baseline, then arthritis must be absent at Week 24 to meet the primary endpoint. If only rash is present at baseline, then rash must be absent at Week 24 to meet the primary endpoint. If both arthritis and rash are present at baseline, then the primary endpoint is met if either arthritis, or rash, or both arthritis and rash are absent at Week 24.

The number and percentage of responders and nonresponders at each visit and at Week 24 (NRI) will be presented by treatment group. The primary efficacy analyses compare remission of arthritis and/or rash as defined by the SLEDAI-2K at Week 24 between baricitinib 2-mg versus placebo and 4-mg versus placebo using the mITT population. The proportion of responders at each visit and Week 24 (NRI) will be analyzed using a logistic regression model as described in Section 6.1.

6.9.2. Secondary Efficacy Analyses

Secondary efficacy analyses will be performed on the mITT population. The analytical details of the logistic regression and MMRM analyses stated below are specified in Section 6.1.

6.9.2.1. SLE Responder Index-4 (SRI-4)

The proportion of patients who meet criteria for response as defined by SLE Responder Index (SRI)-4 at Week 24 will be compared between treatment groups. Specifically, response is defined as follows:

- Reduction of ≥ 4 points from baseline in the SLEDAI-2K score

- No new BILAG A and ≤ 1 new BILAG B disease activity scores (the details of BILAG A and B disease activity scores are available in Appendix 2)

- No worsening (with worsening defined as an increase of ≥ 0.3 points [10mm] from baseline) in Physician's Global Assessment

The number and percentage of responders and nonresponders at each visit (NRI) will be presented by treatment group. The proportion of responders will be compared using a logistic regression model as specified in Section 6.1.

A table for reasons of nonresponse, based on NRI, will also be tabulated at Week 24. Reasons are study treatment noncompleters, concomitant medication violation (increase), SLEDAI not reduced by ≥4 points, BILAG criteria not met, Patient Global Assessment (PGA) criteria not met, or missing data. If a patient is a nonresponder by more than 1 reason, then count the patient in the reason listed above first. For example, if a patient didn't meet SLEDAI criteria AND BILAG criteria, then the patient would be counted in the SLEDAI count.

6.9.2.2. Reduction of ≥4 Points in SLEDAI-2K Score

The proportion of patients achieving a reduction of ≥4 points in SLEDAI-2K score at Week 24 will be compared between treatment groups using the logistic regression model as specified in Section 6.1.

6.9.2.3. Change in SLEDAI-2K Total Score

Comparison between treatment groups for the change from baseline in SLEDAI-2K total score at Week 24 will be analyzed using MMRM method as specified in Section 6.1.

6.9.2.4. Change in Patient's Global Assessment of Disease Activity

Comparison between treatment groups for the change from baseline in Patient's Global Assessment of Disease Activity at Week 24 will be analyzed using MMRM method as specified in Section 6.1.

6.9.3. Additional Analyses of the Primary Outcome

In addition to the primary endpoint, the proportion of patients who achieve remission of arthritis and/or rash as defined by the SLEDAI-2K at Week 24, the pattern of each remission type (arthritis, rash, or both remitted) will be summarized by treatment groups at Week 24. For patients who develop new rash or arthritis but improve, the other will also be summarized by treatment groups.

To ensure that improvement in disease activity is not accompanied by worsening of other disease manifestations, the population of patients that achieve the primary endpoint of remission of arthritis and/or rash as defined by the SLEDAI-2K at Week 24 will be evaluated for worsening in other organ systems using SLEDAI-2K and BILAG. This will be done by calculating the proportion of patients who meet the following criteria (each done separately):

1. Rash/arthritis went from present at baseline to absent at endpoint but patient went from absent to present on any other item on SLEDAI. This is to be done by organ systems and by individual descriptors. See Appendix 2 for definition of organ systems and individual items. Treatment groups will be compared using Fisher's exact test for the analysis by organ systems.
2. Rash/arthritis went from present at baseline to absent at endpoint but patient went from C, D, or E on BILAG organ system to A or B. This analysis will be done overall and by

organ system. Treatment groups will be compared using Fisher's exact test for the overall comparison.

3. Rash/arthritis went from present at baseline to absent at endpoint but patient went from B, C, D, or E on BILAG organ system to A. This analysis will be done overall and by organ system. Treatment groups will be compared using Fisher's exact test for the overall comparison.

6.9.4. Sensitivity Analyses

Baseline SLICC score ($\geq$2 and <2) will be included as an additional exploratory variable into the logistic regression model defined in Section 6.1 for the categorical primary endpoint.

6.10. Health Outcomes/Quality-of-Life Analyses

Categorical variables will be analyzed using logistic regression analyses whereas MMRM will be the primary method of analysis for continuous endpoints. The analyses will be based on the mITT population, unless otherwise specified.

6.10.1. Worst Fatigue Numeric Rating Scale

The Worst Fatigue NRS asks the patient to rate the patient's worst level of fatigue over the past 7 days.

The score will be summarized using descriptive statistics (n, mean, SD, median, minimum, and maximum), by treatment group and visit, including Week 24. The change from baseline to each visit will also be summarized by treatment group. Only patients with a baseline and postbaseline result will be included in the change from baseline summaries.

The change from baseline at all postbaseline visits for the Worst Fatigue NRS score will be analyzed using a MMRM as specified in Section 6.1.

6.10.2. Worst Pain Numeric Rating Scale

The Worst Pain NRS asks the patient to rate the patient's worst level of pain over the past 7 days. This will be measured at all visits starting at Visit 1.

The score will be summarized using descriptive statistics (n, mean, SD, median, minimum, and maximum), by treatment group and visit, including Week 24. The change from baseline to each visit will also be summarized by treatment group. Only patients with a baseline and postbaseline result will be included in the change from baseline summaries.

The change from baseline to all postbaseline visits for the Worst Pain NRS-will be analyzed using a MMRM as specified in Section 6.1.

6.10.3. Worst Joint Pain Numeric Rating Scale

The Worst Joint Pain NRS is a patient-administered, single-item, 11-point horizontal scale anchored at 0 and 10, with 0 representing "no joint pain" and 10 representing "joint pain as bad as you can imagine." Patients rate their joint pain by selecting the number that describes the patient's worst level of joint pain during the past 7 days using an ePRO tablet.

The score will be summarized using descriptive statistics (n, mean, SD, median, minimum, and maximum), by treatment group and visit, including Week 24. The change from baseline to each visit will also be summarized by treatment group. Only patients with a baseline and postbaseline result will be included in the change-from-baseline summaries.

The change from baseline to all postbaseline visits in the Worst Joint Pain NRS over the past 7 days will be analyzed using MMRM.

6.10.4. Short-Form 36-Item Health Survey v2

Summaries of domain scores and summary scores will be provided by visit and change from baseline to postbaseline visits.

An MMRM model will be used to analyze change from baseline in the physical component score (PCS) and the mental component score (MCS) through Week 24, as specified in Section 6.1. For this study, the change in MCS, PCS, and domain scores will be analyzed at each postbaseline visit compared to baseline.

Analyses will be repeated for the 8 health domains: physical functioning, bodily pain, role limitations due to physical problems, role limitations due to emotional problems, general health perceptions, mental health, social function, and vitality.

Summary statistics per treatment group will be tabulated for the transformed scale scores (physical functioning, role limitations/physical health, bodily pain, general health perceptions, vitality, social functioning, role limitations/emotional, mental health) and the 2 summary scores (PCS and MCS).

The summary and analyses of the Short-Form 36-item health survey version 2 (SF-36 v2) score will include the number of patients achieving a Minimal Clinical Important Difference (MCID), which is defined for each summary score, PCS, and the MCS, as follows:

> MCID $\geq$2.5-point improvement in PCS

> MCID $\geq$2.5-point improvement in MCS

The proportion of patients achieving SF-36 v2 score MCID will also be analyzed using a logistic regression model as specified in Section 6.1. The NRI method described in Section 5.4.1 will be used to impute missing data.

6.11. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

All details involving pharmacokinetic/pharmacodynamic (PK/PD) analyses will be documented in a separate SAP.

6.12. Safety Analyses

All safety data will be descriptively summarized by treatment groups and analyzed based on the safety population. The safety population is defined as those patients who received at least 1 dose of study drug and did not discontinue the study for the reason "lost to follow-up" at the first postbaseline visit. The safety analyses include AEs, safety in special groups and circumstances,

including Adverse Events of Special Interest (AESI), laboratory analytes, the 16-item Quick Inventory of Depressive Symptomatology Self-Report (QIDS-SR16), and vital signs. The duration of exposure will also be summarized. The categorical safety measures will be summarized using incidence rates and analyzed by Fisher's exact test. The mean change in the continuous safety measures including vital signs, QIDS-SR16, physical characteristics, and laboratory values will be summarized by visits and analyzed by ANCOVA, with treatment and baseline values in the model. More details are provided in subsequent sections.

For specific events of special interest (see Section 6.12.3 for more details), an incidence rate, IR per 100 patient-years of observation (PYO), will be provided. Patient-years of observation will be calculated as the sum of all patient observation time in the treatment group. For a patient with an event, the observation time will be censored at the event date; for a patient without the event, the observation time will be counted until the patient's last treatment dose date plus 30 days or the patient's last visit, whichever occurs first. Only events that occur within 30 days after the patient's treatment discontinuation date will be considered.

See formula as follows:

$$PYO = \sum_{pt\ w\ event} \frac{event\ start\ date - first\ trt\ dose\ date + 1}{365.25}$$
$$+ \sum_{pt\ w/o\ event} \frac{last\ observation\ date - first\ trt\ dose\ date + 1}{365.25}$$

Incidence rate will be calculated as follows:

$$IR = \frac{unique\ number\ of\ patients\ with\ event}{PYO} \times 100$$

For each IR provided, a 95% CI will be calculated based on the Poisson distribution. Treatment group comparisons based on IR will be provided based on the incidence rate difference (IRD) together with its 95% CI.

6.12.1. Extent of Exposure

Duration of exposure to study drug will be summarized for the safety population by treatment group. Exposure will be calculated as the date of last dose of study drug (or date of discontinuation) minus the date of first dose of study drug plus 1 day. Total patient-years (PYs) of exposure will be reported for each treatment group for overall duration of exposure. Descriptive statistics (n, mean, SD, minimum, first quartile, median, third quartile, and maximum) will be provided for patient-days of exposure and the frequency of patients falling into different exposure ranges will be summarized. Exposure ranges are as follows:

- ≥4 weeks, ≥12 weeks, and ≥24 weeks
- >0 to <4 weeks, ≥4 weeks to <12 weeks, ≥12 weeks to <24 weeks, and ≥24 weeks

Overall exposure will be summarized in total PYs, which are calculated according to the following:

> Exposure in PYs = sum of duration of exposure in days (for all patients in treatment group) / 365.25

No p-values will be reported in these tables as they are intended to describe the characteristics of the study sets.

6.12.2. Adverse Events

6.12.2.1. Adverse Events

Adverse events are recorded in the eCRF. Each AE will be coded to SOC and PT, using the MedDRA version that is current at the time of database lock. Severity of AEs is recorded as mild, moderate, or severe.

Treatment-emergent adverse events (TEAEs) are defined as events that either first occurred or worsened in severity after the first dose of study drug and the earliest of the visit study drug disposition date or the last visit date during the treatment period, whichever occurred first, and up to 30 days after study treatment discontinuation. The MedDRA Lowest Level Term (LLT) will be used in defining which events are treatment-emergent. The maximum severity for each LLT during the baseline period until the first dose of the study medication will be used as baseline. If an event is preexisting during the baseline period, but it has missing severity, and the event persists during the treatment period or up to 30 days after treatment discontinuation, then the baseline severity will be considered mild for determining any postbaseline treatment-emergence (ie, the event is treatment-emergent unless the severity is coded mild at postbaseline). If an event occurring postbaseline has a missing severity rating, then the event is considered treatment-emergent. Should there be insufficient data for an AE start date to make this comparison (eg, the AE start year is the same as the treatment start year, but the AE start month and day are missing), the AE will be considered treatment-emergent. For events occurring on the day of the first dose of study treatment, the day and time of the onset of the event will both be used to distinguish between pretreatment and posttreatment in order to derive treatment-emergence.

In general, summaries will include the number of patients in the safety population (N), frequency of patients experiencing the event (n), and relative frequency (ie, percentage; n/N*100).

In an AE overview table, the number and percentage of patients in the safety analysis set who experienced death, a serious adverse event (SAE), any TEAE, permanent discontinuation from study drug due to an AE, temporary interruption of study drug due to an AE or laboratory abnormality, or a severe TEAE will be summarized by treatment group.

The number and percentage of patients with TEAEs will be summarized by treatment group in 2 formats. For events that are gender specific, the denominator and computation of the percentage will only include patients from the given gender.

- by MedDRA PT nested within SOC with SOCs ordered alphabetically, and events ordered within each SOC by decreasing frequency in the baricitinib 4-mg group
- by MedDRA PT with events ordered by decreasing frequency in the baricitinib 4-mg group

Adverse events leading to permanent discontinuation of study drug and AEs leading to temporary interruption of study drug will also be summarized by treatment group using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the baricitinib 4-mg group within SOC.

A summary of temporary interruptions of study drug will also be provided, showing the number of patients who experienced at least 1 temporary interruption and the number of temporary interruptions per patient with an interruption. Further, the duration of each temporary interruption (in days) and the cumulative duration of dose interruption (in days) using basic descriptive statistics (n, mean, SD, minimum, first quartile, median, third quartile, and maximum) will be displayed.

Common TEAEs are defined as TEAEs that occurred in ≥2% (before rounding) of patients in any treatment group including placebo. The number and percentage of patients with common TEAEs will be summarized by treatment using MedDRA PT ordered by decreasing frequency in the baricitinib 4-mg group.

The number and percentage of patients with TEAEs will be summarized by maximum severity by treatment using MedDRA PT ordered by decreasing frequency in the baricitinib 4-mg group for the common TEAEs. For each patient and TEAE, the maximum severity for the MedDRA level being displayed is the maximum postbaseline severity observed from all associated LLTs mapping to that MedDRA PT.

An individual listing of all AEs including preexisting conditions will be provided. A separate listing will include AEs that led to permanent discontinuation from the study drug. In addition, a listing of AEs that occur more than 30 days after study treatment discontinuation will be provided.

6.12.2.2. Serious Adverse Events

Consistent with the International Conference on Harmonisation (ICH) E2A guideline, a SAE is any AE that results in 1 of the following outcomes:

- death

- initial or prolonged inpatient hospitalization

- a life-threating experience (that is, immediate risk of dying)

- persistent or significant disability/incapacity

- congenital anomaly/birth defect

- considered significant by the investigator for any other reason

The number and percentage of patients who experienced any ICH-defined SAE will be summarized by treatment group during the treatment and follow-up periods using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the baricitinib 4-mg group within SOC. In addition, the SAEs will be summarized by treatment group using MedDRA PT without SOC. An individual listing of all SAEs will be provided.

6.12.3. Adverse Events of Special Interest

6.12.3.1. Infections

Infections will be defined using the all PTs from the MedDRA Infections and Infestations SOC. Serious infection will be defined as all the infections that meet the SAE criteria.

The number and percentage of patients with TEAEs of infections, serious infections, and infections resulting in study drug discontinuation will be summarized by treatment group using MedDRA PTs.

The number and percentage of patients with TEAEs of infections by maximum severity will be summarized by treatment group using MedDRA PTs.

For infections of special interest (serious infections, potential opportunistic infections [POIs], herpes zoster, and herpes simplex), the IR (for detail, see Section 6.12) and 95% CI will be calculated.

Treatment-emergent infectious events will be reviewed in context of other clinical and laboratory parameters. A listing of patients experiencing treatment-emergent infectious AEs will be provided. The listing will include patient demographics, treatment group, treatment start and stop dates, infectious event, event start and stop dates, total leukocytes, total lymphocytes, absolute neutrophils, event seriousness, and event outcome.

The infectious TEAE will be further analyzed in terms of potential opportunistic infection, herpes zoster, and herpes simplex. A summary of hepatitis B virus (HBV) deoxyribonucleic acid (DNA) monitoring results and association between infection and neutropenia/lymphopenia will also be provided in the context of infections.

Potential Opportunistic Infections

Potential opportunistic infections will be identified according to 2 different approaches.

First, POIs will be identified from TEAEs based on a Lilly-defined list of MedDRA PTs shown in the Product Safety Analysis Plan (PSAP), version 5 (Appendix 6 of PSAP). These PTs are a subset of terms from the Infections and Infestations SOC.

Second, a list of all the infection details captured from the infection-specific eCRF (including primary/secondary infecting organism and infection site) will be provided. Of note, the infecting organism that was entered as free text by the investigator (instead of as a selection from the pull-down list) will also be provided.

The summary analysis of POIs identified using the 2 approaches above will be provided. Each case meeting the case definition for an opportunistic infection will be summarized by PT nested under infection pathogen. Events will be ordered by decreasing frequency of pathogen nested under pathogen species (mycobacteria, bacteria, fungal, viral, parasites). The order of frequency will be determined using the baricitinib 4-mg group.

Potential opportunistic infections identified through these approaches will be combined in 1 list for medical assessment for final classification of whether the case definition was met according to the consensus paper after database lock (i.e., Winthrop 2015). An additional summary may be conducted based on medical assessment.

Herpes Zoster

A summary table of herpes zoster will be provided. Herpes zoster will be defined based on the MedDRA PTs as listed under Herpes zoster (any form) (II) in Appendix 6 of the PSAP, version 5, excluding Varicella virus text (10070444). The summary table will also include event maximum severity, seriousness, whether resulting in temporary study drug interruption, whether resulting in study drug discontinuation, whether treated with antiviral medication, and event outcome.
If a patient has more than 1 event of herpes zoster, the event with the maximum severity will be used in these summary tables. If more than 1 event of herpes zoster occurs with the same severity, the event with the longest duration will be used in the summary table.

Herpes Simplex

A summary analysis of herpes simplex will be provided. Herpes simplex will be defined based on the MedDRA PTs as listed under Herpes simplex (invasive disease only) (IV) in Appendix 6 of the PSAP, version 5. The summary table will include event maximum severity, seriousness, whether resulting in temporary study drug interruption, whether resulting in study drug discontinuation, and whether treated with antiviral medication. Antiviral medication will be selected based on ATC code level 2 "antiviral for systemic use."
If a patient has more than 1 event of herpes simplex, the event with the maximum severity will be used in these summary tables. If more than 1 event of herpes simplex occurs with the same severity, the event with the longest duration will be used in the summary table.

HBV DNA

A listing of patients with detectable HBV DNA will be provided.

HBV DNA status (not detectable, detectable but not quantifiable [ie, <29 IU/mL], quantifiable [ie, ≥29 IU/mL]) will be summarized by treatment group stratified by baseline HBV serology status, specifically:

- HBsAb- / HBcAb-
- HBsAb+ / HBcAb-
- HBsAb+ / HBcAb+
- HBsAb- / HBcAb+

Association between Infection and Neutropenia/Lymphopenia

To evaluate the association between infection and neutropenia and also between infection and lymphopenia, the frequency of infections will be provided by the worst Common Terminology Criteria for Adverse Events (CTCAE) grades of neutropenia and lymphopenia, respectively. Infection outcomes considered for this analysis are any treatment-emergent infection, serious infection, and herpes zoster. For this analysis, no statistical comparison will be provided.

In addition, a summary table will be provided for treatment-emergent infections that were preceded or accompanied by neutropenia/lymphopenia. For this analysis, neutropenia is defined as CTCAE Grade 2 or greater. Infection events with onset date ≤14 days before or after the Grade 2 neutrophil/lymphocyte count collection date will be considered as infections preceded or accompanied by neutropenia.

6.12.3.2. Allergic Reactions and Hypersensitivities

A search will be performed using the current MedDRA version 19.1 Standardised MedDRA Queries (SMQs) to search for relevant events, using the following queries:

- Anaphylactic reaction SMQ (20000021)
- Hypersensitivity SMQ (20000214)
- Angioedema SMQ (20000024)

Events that satisfy the queries will be listed, by temporal order within patient ID, and will include SOC, PT, SMQ event categorization including detail on the scope (narrow or broad), reported AE term, AE onset and end dates, severity, seriousness, outcome, etc.

6.12.3.3. Malignancies

Malignancies will be identified using terms from the malignant tumors SMQ (SMQ 20000194). Malignancies excluding nonmelanoma skin cancers (NMSC) and NMSC will be reported separately.

A listing including all malignancy cases will be provided. An NMSC flag will be provided using the following MedDRA PTs (the list will be updated depending on the MedDRA version used for analysis):

- Squamous cell carcinoma of skin (10041834)
- Bowen's disease (10006059)
- Basal cell carcinoma (10004146)
- Basosquamous carcinoma (10004178)
- Basosquamous carcinoma of skin (10004179)
- Squamous cell carcinoma (10041823)
- Skin squamous cell carcinoma metastatic (10077314)
- Skin cancer (10040808)
- Carcinoma in situ of skin (10007390)

The number and percentage of patients with TEAE-associated malignancies excluding NMSC and NMSC will be summarized by treatment group. In addition, the IR (for detail, see

Section 6.12) and 95% CI will be calculated for the overall observation time. All cases identified by malignant tumors SMQ will be assessed after database lock by the medical team to determine (1) confirmed NMSC cases and (2) symptom and date that triggered the malignancy investigation or diagnosis. An additional listing based on medical review may also be provided if deemed necessary. All cases reported in the study database or by Lilly Safety System (LSS) report, disregarding the length of gap between the last treatment dose date and the event date will be included.

6.12.3.4. Gastrointestinal Perforations

Treatment-emergent adverse events potentially related to gastrointestinal (GI) perforations will be analyzed using reported AEs. Identification of these events will be based on the PTs of the MedDRA Gastrointestinal Perforations SMQ (SMQ 20000107); note that this SMQ holds only narrow terms and has no broad terms. Potential GI perforations identified by the above SMQ search will be provided as a listing for internal review by the medical safety team. Each case will be assessed to determine whether it is GI perforation. Frequency and relative frequency for each PT will be provided, ordered by decreasing frequency in the baricitinib 4-mg treatment group. Comparisons between each baricitinib treatment and placebo will be made using Fisher's exact test.

6.12.4. Clinical Laboratory Evaluation

All laboratory tests will be presented using the Système International of units (SI) and conventional (CN) units. For topics of safety in special groups and circumstances, laboratory test units will be specified for each analysis.

Lilly Large Clinical Trial Population Based (LCTPB) reference limits will be used to define the low and high limits because it is generally desirable for limits used for analyses to have greater specificity (identify fewer false positive cases) than reference limits used for individual patient management. When Lilly LCTPB reference ranges are unavailable, then central laboratory (Covance) reference ranges will be used. For the 4 key hepatic laboratory assessments (alanine aminotransferase [ALT], aspartate aminotransferase [AST], total bilirubin, and alkaline phosphatase [ALP]), central laboratory reference ranges (Covance) will be used and all results pertaining to these assessments will be included as a separate analysis to address the risk of liver injury as a special safety topic (see Section 6.12.4.1). Central laboratory reference ranges (Covance) will also be used to evaluate immunoglobulins and lymphocyte cell subsets (see Sections 6.12.4.3 and 6.12.4.7). See the PSAP version 5 (Appendix 2) for details of the reference range by laboratory analytes.

The low-density lipoprotein/high-density lipoprotein (LDL/HDL) ratio will be derived as the ratio of LDL cholesterol to HDL cholesterol. There are no Lilly LCTPB reference ranges or central lab reference ranges for the LDL/HDL ratio. The transferrin saturation will be derived as the ratio of total iron to total iron binding capacity (TIBC) expressed as a percent (multiplied by 100). Central lab reference ranges will be applied to transferrin saturation values for determining treatment-emergent high/low transferrin saturation as a safety outcome. The gender-specific reference ranges are 20% to 50% for males and 15% to 50% for females.

The following will be conducted for laboratory analyte measurements collected quantitatively:

- Box plots for observed values: Values at each visit (starting at randomization) and change from baseline to each visit and to last postbaseline measure will be displayed in box plots for patients who have a baseline and at least 1 postbaseline visit. For visits included in the treatment period, patients will be included only if the visit occurs on or before the date of treatment discontinuation/completion. Follow-up visit will be the first visit that occurred during the Follow-up period. Individual measurements outside of reference limits will also be displayed using distinct symbols overlaying the box plot. Original-scale data will be used for the display but for some analytes (eg, immunoglobulins) a logarithmic scale will be used to aid in viewing the measures of central tendency and dispersion. Unplanned measurements will be excluded. Descriptive summary statistics will be included in a table below the box plot. These box plots will be used to evaluate trends over time and to assess a potential impact of outliers on central tendency summaries. A p-value for change from baseline to endpoint will be provided using an ANCOVA model with explanatory term for treatment and the baseline value as a covariate. Endpoint will be the last observation where patient is on treatment.
- Treatment-emergent high/low analyses: The number and percentage of patients with treatment-emergent high and low laboratory results at any time will be summarized by treatment group. Planned and unplanned measurements will be included. A treatment-emergent **high** result is defined as a change from a value less than or equal to the high limit at all baseline visits to a value greater than the high limit at any time during the treatment period and up to 60 days after treatment discontinuation. A treatment-emergent **low** result is defined as a change from a value greater than or equal to the low limit at all baseline visits to a value less than the low limit at any time during the treatment period and up to 60 days after treatment discontinuation. The Fisher's exact test will be used for the treatment comparisons.

A listing of abnormal findings will be provided. The listing will include, but not be limited to, patient ID, treatment group, laboratory collection date, analyte name, and analyte finding.

6.12.4.1. Abnormal Hepatic Tests

Analyses for abnormal hepatic tests involve 4 laboratory analytes: ALT, AST, total bilirubin, and ALP. Analyses for the change from baseline to last visit that occurred on or before the date of treatment discontinuation and shift tables are described in Section 6.12.4. This section describes additional analyses for the topic. The central laboratory reference ranges (Covance) will be used for ALT, AST, total bilirubin, and ALP hepatic laboratory assessments.

The number and percentage of patients with the following abnormal elevations in hepatic laboratory tests at any time up to 60 days after treatment discontinuation will be summarized by treatment group. Baricitinib groups will be compared to placebo using Fisher's exact text:

- The percentages of patients with an ALT measurement ≥3×, 5×, and 10× the central laboratory upper limit of normal (ULN) during the treatment and follow-up periods will be summarized for all patients with a postbaseline value and for subsets based on various levels of baseline.

  - The analysis of 3× ULN will contain 4 subsets: patients whose nonmissing maximum baseline value is ≤1× ULN; patients whose maximum baseline is >1× ULN, but <3× ULN; patients whose maximum baseline value is ≥3× ULN; and patients whose baseline values are missing.

  - The analysis of 5× ULN will contain 5 subsets: patients whose nonmissing maximum baseline value is ≤1× ULN; patients whose maximum baseline is >1× ULN, but <3× ULN; patients whose maximum baseline is ≥3× ULN, but <5× ULN; patients whose maximum baseline value is ≥5× ULN; and patients whose baseline values are missing.

  - The analysis of 10× ULN will contain 6 subsets: patients whose nonmissing maximum baseline value is ≤1× ULN; patients whose maximum baseline is >1× ULN, but <3× ULN; patients whose maximum baseline is ≥3× ULN, but <5× ULN; patients whose maximum baseline is ≥5× ULN, but <10× ULN; patients whose maximum baseline value is ≥10× ULN; and patients whose baseline values are missing.

- The percentages of patients with an AST measurement greater than or equal to 3×, 5×, and 10× the central laboratory ULN during the treatment and follow-up periods will be summarized for all patients with a postbaseline value and for subsets based on various levels of baseline. Analyses will be constructed as described above for ALT.

- The percentages of patients with a total bilirubin measurement greater than or equal to 2 × the central laboratory ULN during the treatment period will be summarized for all patients with a postbaseline value and subset into 4 subsets: patients whose nonmissing maximum baseline value is ≤1× ULN; patients whose maximum baseline is >1× ULN but < 2× ULN; patients whose maximum baseline value is ≥2× ULN; and patients whose baseline values are missing.

- The percentages of patients with an ALP measurement ≥1.5× the central laboratory ULN during the treatment and follow-up periods will be summarized for all patients with a postbaseline value and subset into 4 subsets: patients whose nonmissing maximum baseline value is ≤1× ULN; patients whose maximum baseline is >1× ULN but <1.5× ULN; patients whose maximum baseline value is ≥1.5× ULN; and patients whose baseline values are missing.

Second, to further evaluate potential hepatotoxicity, an Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will be created for all patients whether treated with baricitinib and/or other treatment using the whole study and follow-up periods. Each patient with at least 1

postbaseline ALT and total bilirubin will be included in the eDISH. The points correspond to maximum total bilirubin and maximum ALT, even if not obtained from the same blood draw. A listing of patients potentially meeting Hy's rule will be provided (defined as greater than or equal to 3× ULN for ALT or AST, and greater than or equal to 2× ULN for total bilirubin, not necessarily at the same time).

Third, a listing will be provided to the medical safety team for internal review according to the following SMQs:

- Broad and narrow terms in the Liver-related investigations, signs and symptoms SMQ (SMQ 20000008)
- Broad and narrow terms in the Cholestasis and jaundice of hepatic origin SMQ (SMQ 20000009)
- Broad and narrow terms in the Hepatitis non-infectious SMQ (SMQ 20000010)
- Broad and narrow terms in the Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions SMQ (SMQ 20000013)
- Narrow terms in the Liver-related coagulation and bleeding disturbances SMQ (SMQ 20000015)

6.12.4.2. Hematologic Changes

Hematologic changes will be defined based on clinical laboratory assessments. Cutaneous Lupus Erythematosus Disease Area and Severity Index will be applied for laboratory tests potentially related to myelosuppressive events (refer to Table 5.2 in PSAP, version 5 (Appendix 2).

Treatment-emergent laboratory abnormalities potentially related to myelosuppression occurring at any time during the treatment and follow-up periods and shift tables of baseline to maximum grade during the treatment and follow-up periods will be tabulated. Planned and unplanned measurements will be included. Treatment emergence will be characterized using the following 5 criteria (as appropriate to the grading scheme):

1. any increase in postbaseline CTCAE grade from worst baseline grade
2. increase to Grade 1 or above at worst postbaseline
3. increase to Grade 2 or above at worst postbaseline
4. increase to Grade 3 or above at worst post-baseline
5. increase to Grade 4 at worst post-baseline

Shift tables will show the number and percentage of patients based on baseline to maximum during the treatment and follow-up periods, with baseline depicted by the most extreme grade during the baseline period. With each shift table, a shift table summary displaying the number and percentage of patients with maximum postbaseline results will be presented by treatment group for each treatment period within the following categories:

- Decreased; post-baseline category < baseline category
- Increased; post-baseline category > baseline category
- Same; postbaseline category = baseline category

LY3009104

A laboratory-based treatment-emergent outcome related to increased platelet count will be summarized in a similar fashion. Treatment-emergent thrombocytosis as a laboratory-based abnormality will be defined as an increase in platelet count from a maximum baseline value ≤600 billion/L to any postbaseline value >600 billion/L. Planned and unplanned measurements will be included. A listing of patients with treatment-emergent thrombocytosis will be provided for safety review.

6.12.4.3. Lymphocyte Subset Cell Counts

The following lymphocyte subsets will be analyzed:

CD3+ T Cells – %

CD3+ T Cells – Absolute

CD3+CD8+ T cells (CD8) – %

CD3+CD8+ T cells (CD8) – Absolute

CD3+CD4+ T cells (CD4) – %

CD3+CD4+ T cells (CD4) – Absolute

CD56+/CD16+ NK cells – %

CD56+/CD16+ NK cells – Absolute

CD19+ B cells – %

CD19+ B cells – Absolute

CD4/CD8 Ratio – Calculated

CD3+4+8+ %

CD3+4+8+ Abs

CD3/CD19 Ratio - Calculated

For each type of cells, both the absolute count and the relative count (ie, as a percentage of the total lymphocyte population) will be analyzed. In addition, the ratio of CD4 cell counts to CD8 cell counts will be analyzed.

The analyses for these parameters will be performed using the same approaches as described for analysis of clinical laboratory measurements in Section 6.12.4. For determining treatment-emergent abnormal, high, or low lymphocyte subset cell counts, central laboratory reference ranges (Covance) will be used when available; note that reference ranges are available only for some of the lymphocyte subset cell count analytes.

6.12.4.4. Lipids Effects

Analyses for the change from baseline to last observation, and shift tables in total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides are described in Section 6.12.4.

Treatment-emergent adverse events potentially related to hyperlipidemia will also be analyzed, based on reported AEs. The target surveillance term "Hyperlipidemia" is a Lilly-defined MedDRA search criteria list that is a subset of the PTs in the MedDRA SMQ "Dyslipidemia" that are related to elevated or increased lipids. MedDRA PTs, each with a narrow scope from the SMQ, for the target surveillance term are shown in the PSAP version 5 (Appendix 3 of PSAP). Frequency and relative frequency for each PT will be provided, ordered by decreasing frequency in baricitinib4-mg.

6.12.4.5. Renal Function Effects
Effects on renal function will be assessed through analyses of creatinine, which are described in Section 6.12.4.

6.12.4.6. Elevations in Creatine Phosphokinase
Analyses of creatine phosphokinase are described in Section 6.12.4.

6.12.4.7. Serum Immunoglobulin Concentrations
Each serum Ig concentration (IgA, IgG, and IgM) will be analyzed. The analyses for these parameters will be performed using the same approaches as described for analysis of clinical laboratory measurements in Section 6.12.4. For determining treatment-emergent abnormal, high, or low serum Ig concentrations for IgA, IgG, and IgM, central laboratory reference ranges (Covance) will be used.

6.12.5. Vital Signs and Other Physical Findings
Vital signs and physical characteristics include systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse, weight, BMI, and waist circumference. Original-scale data will be analyzed. When these parameters are analyzed as continuous numerical variables, unplanned measurements will be excluded. When these parameters are analyzed as categorical outcomes, planned and unplanned measurements will be included.

The planned analyses described for the laboratory analytes in Section 6.12.4 will be used to analyze the vital signs and physical characteristics.

Table JAHH.5.4 defines the low and high baseline values as well as the criteria used to define treatment-emergence based on postbaseline values. Postbaseline values include all values after baseline in the treatment and follow-up periods. The blood pressure and pulse rate criteria are consistent with the document *Selected Reference Limits for Blood Pressure, Orthostasis, and ECG Numerical Parameters (Including Heart Rate) for Use in Phase II-IV Clinical Trials Version 1.1* approved on 8 March 2013 as recommended by the Lilly Cardiovascular Safety Advisory Committee (CVSAC).

Table JAHH.5.4.   **Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurement, and Categorical Criteria for Weight Changes for Adults**

Parameter (Units of Measure)	Low	High
Systolic Blood Pressure (mm Hg)	≤90 (low limit) and decrease from lowest value during baseline ≥20 if >90 at each baseline visit	≥140 (high limit) and increase from highest value during baseline ≥20 if <140 at each baseline visit
Diastolic Blood Pressure (mm Hg)	≤50 (low limit) and decrease from lowest value during baseline ≥10 if >50 at each baseline visit	≥90 (high limit) and increase from highest value during baseline ≥10 if <90 at each baseline visit
Pulse (beats per minute)	<50 (low limit) and decrease from lowest value during baseline ≥15 if ≥50 at each baseline visit	>100 (high limit) and increase from highest value during baseline ≥15 if ≤100 at each baseline visit
Weight (kilograms)	(Loss) decrease ≥7% from lowest value during baseline	(Gain) increase ≥7% from highest value during baseline

Abbreviation: mm Hg = millimeters of mercury.

6.12.6. Additional Safety Sections

6.12.6.1. Symptoms of Depression (QIDS-SR16)

The QIDS-SR16 is a 16-item, self-report instrument intended to assess the existence and severity of symptoms of depression as listed in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition (DSM-IV; APA 1994). Patients are asked to consider each statement as it relates to the way they have felt for the past 7 days. There is a unique 4-point ordinal scale for each item with scores ranging from 0 to 3 reflecting increasing depressive symptoms as the item score increases. Additional information and the QIDS-SR16 questions may be found on the University of Pittsburgh Epidemiology Data Center website (http://www.ids-qids.org/index.html). A key reference for this instrument covers its psychometric properties for use in patients with chronic major depression (Rush et al. 2003). Quick Inventory of Depressive Symptomatology was developed at UT Southwestern.

The QIDS-SR16 total score is derived as the sum of the scores across the 9 scale domains. For scale domains that contain more than 1 item, the domain score is the highest item rating given across the items within that domain:

- Sleep: the highest score on any 1 of the 4 sleep items (Items 1 to 4)
- Depressed mood: Item 5
- Weight/appetite change: the highest score on any 1 of the 4 weight items (Items 6 to 9)
- Psychomotor changes: the highest score on either of the 2 psychomotor items (Items 15 and 16)
- Concentration: Item 10
- Worthlessness/Guilt: Item 11
- Suicidal ideation: Item 12
- Decreased interest: Item 13

- Decreased energy: Item 14

In the presence of missing data, the following rules will be employed to derive the total score. Firstly, considering the 3 multi-item domains (sleep, weight/appetite change, psychomotor change), the domain score should be based on the maximum value across the appropriate items, and it should be missing only if each item is missing. Further, considering the 9 domain scores, the total score should be derived as missing if there are 3 or more domains that are missing; if 1 or 2 domain scores are missing, then the total score should be derived using a total score that is prorated to the full scale range (0 to 27) based on the available domain scores, retaining 1 decimal place in the total score derived in the presence of missing data.

The QIDS-SR16 total scores will also be categorized in the following severity classes as shown in Table JAHH.5.5.

Table JAHH.5.5.   **QIDS-SR16 Severity of Depressive Symptoms Categories based on the QIDS-SR16 Total Score**

QIDS-SR16 Severity of Depressive Symptoms Category	QIDS-SR16 Total Score
0 = None	0-5
1 = Mild	6-10
2 = Moderate	11-15
3 = Severe	16-20
4 = Very Severe	21-27

Abbreviation: QIDS-SR16 = Quick Inventory of Depressive Symptomatology Self-Rated.

Treatment differences in mean change QIDS-SR16 total score will be analyzed using the MMRM model described in Section 6.1. Only visits that occur on or before the date the patient discontinued treatment will be included in this analysis.

Using the QIDS-SR16 Severity of Depressive Symptoms Categories shown in Table JAHH.5.5, shift tables will show the number and percentage of patients with total score in each category based on baseline to maximum category during the treatment period and up to 60 days after treatment discontinuation, with baseline depicted by the most extreme category during the baseline period, with further summarization of change from baseline in severity using categories of any improvement, no change, and any worsening, by treatment. Similarly, shift tables will be created for the QIDS-SR16 suicidal ideation item (Item 12) responses.

Treatment-emergent changes in $QIDS-SR_{16}$ total score severity categories will be characterized as follows:

- Increase (from None) to Mild, Moderate, Severe, or Very Severe
- Increase (from None or Mild) to Moderate, Severe, or Very Severe
- Increase (from None, Mild, or Moderate) to Severe or Very Severe
- Increase (from None, Mild, Moderate, or Severe) to Very Severe

Treatment-emergent changes in $QIDS-SR_{16}$ suicidal ideation item (Item 12) responses will be characterized as follows:

- Increase (from 0) to 1-3
- Increase (from 0-1) to 2-3
- Increase (from 0-2) to 3

6.13. Subgroup Analyses

Subgroup analyses comparing baricitinib 4-mg or 2-mg to placebo will be performed on the mITT population for the following:

- Primary outcome: the proportion of patients who achieve remission of arthritis and/or rash as defined by the SLEDAI-2K
- Change from baseline in SLEDAI-2K
- Change from baseline in Patient's Global Assessment of Disease Activity

- SRI-4 response

The following subgroups (but not limited to only these) will be evaluated:

- Gender: (Male; Female)
- Baseline anti-dsDNA status: (positive; negative)
- Baseline INF signature (positive; negative)
- Baseline Complement: (low C3 and/or low C4; normal/high C3; and normal/high C4)
- Race: (Asian; Black/African American; White; or Other)
- Ethnicity: (Hispanic or Latino; and Not Hispanic or Latino)
- Disease severity at baseline (SLEDAI-2k): (<10; and ≥10)
- Region: (US; Asia; Europe; and Rest of World)
- Japan only
- Age (cut by median)

Descriptive statistics will be provided for each treatment and stratum of a subgroup as outlined, regardless of sample size. Each categorical variable will be analyzed individually with a logistic model that contains the treatment, the subgroup variable, and subgroup by treatment. The treatment-by-subgroup interaction tests will determine whether treatment differences are the same for each subgroup category. For the continuous variables, data will be analyzed with an MMRM model. Baseline, Treatment, Time, subgroup, Treatment*subgroup, subgroup*Time, Treatment*Time and Treatment*subgroup*Time will be included as fixed effect in the model. An unstructured variance-covariance structure will be fitted. If this analysis fails to converge, other structures will be tested following the method described in Section 6.1. The LS mean and SE within subgroups will be presented from MMRM model. For the categorical variables, the number of responders and response rate will be reported for each subgroup.

6.14. Protocol Deviations

Protocol deviations will be tracked by the clinical team, and the importance will be assessed by key team members during the protocol deviation review meetings. Important protocol deviations are listed below.

- Protocol inclusion or exclusion criteria
  - General
    - Are unable to read, understand, and give written informed consent
    - Are pregnant or nursing at the time of study entry
    - Do not have a positive ANA (HEp-2 ANA titer ≥1:80) and/or a positive anti-dsDNA (≥30 IU) as assessed by a central laboratory at screening.
    - Do not have a SLEDAI-2K score ≥4 based on clinical symptoms (not including lab values) at randomization.
    - Do not have active arthritis nor active rash as defined by the SLEDAI-2K at randomization.
  - Medical Conditions
    - Have a history of lymphoproliferative disease; or have signs or symptoms suggestive of possible lymphoproliferative disease, including

lymphadenopathy or splenomegaly (other than primarily due to SLE); or have active primary or recurrent malignant disease; or have been in remission from clinically significant malignancy for <5 years

  ❖ Patients with cervical carcinoma in situ that has been resected with no evidence of recurrence or metastatic disease for at least 3 years may participate in the study.

  ❖ Patients with basal cell or squamous epithelial skin cancers that have been completely resected with no evidence of recurrence for at least 3 years may participate in the study.

- Have a current or recent (<4 weeks prior to screening) clinically serious viral, bacterial, fungal, or parasitic infection

- Have had symptomatic herpes zoster infection within 12 weeks prior to study entry

- Have a history of disseminated/complicated herpes zoster (multi-dermatomal involvement, ophthalmic zoster, central nervous system [CNS] involvement, or post-therapeutic neuralgia)

- Have active or chronic HBV, hepatitis C virus (HCV), or human immunodeficiency virus (HIV)

- Have had household contact with a person with active (TB) and did not receive appropriate and documented prophylaxis for TB

- Have evidence of active TB or have previously had evidence of active TB and did not receive appropriate and documented treatment

- Have severe active lupus nephritis, including urine protein/creatinine ratio >300 mg/mmol (as an estimate of approximate proteinuria >3 g/day) or active urinary sediment with red blood cell cast(s), or histological evidence (if available) of diffuse proliferative glomerulonephritis within the 12 weeks prior to screening

- Have active severe CNS lupus including aseptic meningitis, cerebral vasculitis, demyelinating syndrome, myelopathy, acute confusional state, psychosis, acute inflammatory demyelinating polyradiculoneuropathy, mononeuropathy (single/multiplex), cranial neuropathy, plexopathy, status epilepticus, or cerebellar ataxia

- Have evidence of active TB as documented by a purified protein derivative (PPD) test (≥5-mm induration between approximately 2 and 3 days after application, regardless of vaccination history), medical history, clinical symptoms, and abnormal chest x-ray at screening

  ❖ The QuantiFERON®-TB Gold test or T-SPOT®.TB test (as available and if compliant with local TB guidelines) may be used instead of the PPD test.

- Have evidence of latent TB (as documented by a positive PPD, no clinical symptoms consistent with active TB, and a normal chest x-ray at screening, or as outlined below), unless patients complete at least 4 weeks of appropriate

treatment prior to randomization and agree to complete the remainder of treatment while in the trial

- ❖ If the PPD test is positive and the patient has no medical history or chest x-ray findings consistent with active TB, the patient may have a QuantiFERON®-TB Gold test or T-SPOT®.TB test (as available and if compliant with local TB guidelines). If the test results are not negative, the patient will be considered to have latent TB (for purposes of this study) and will be excluded.
- ❖ The QuantiFERON®-TB Gold test or T-SPOT®.TB test (as available and if compliant with local TB guidelines) may be used instead of the PPD test. If the test results are positive, the patient will be considered to have latent TB and will be excluded. If the test is not negative, the test may be repeated once within 2 weeks of the initial value. If the repeat test results are again not negative, the patient will be considered to have latent TB (for purposes of this study) and will be excluded.
- ❖ Exceptions include patients with a history of active or latent TB who have documented evidence of appropriate treatment and with no history of re-exposure since their treatment was completed. (Such patients would not be required to undergo the protocol specific TB testing for PPD, QuantiFERON®-TB Gold test or T-SPOT®.TB test, but would still require a baseline chest x-ray.)
- ▪ Have a positive test for HBV defined as:
  - ❖ positive for hepatitis B surface antigen (HBsAg), or
  - ❖ positive for anti-hepatitis B core antibody (HBcAb) but negative for hepatitis B surface antibody (HBsAb), or
  - ❖ positive for HBcAb and positive for HBV DNA
  - ❖ Patients in Japan (or elsewhere if required) are excluded if they are positive for HBsAg or are positive for HBcAb and/or HBsAb and positive for HBV DNA
- ▪ Have HCV (positive for anti-hepatitis C antibody with confirmed presence of HCV)
- ▪ Have evidence of HIV infection or positive HIV antibodies
- o Prior/Concomitant Therapy
  - ▪ Have been exposed to a live vaccine within 12 weeks prior to planned randomization or are expected to need/receive a live vaccine during the course of the study (with the exception of herpes zoster vaccination
  - ▪ Have been exposed to herpes zoster vaccination within 4 weeks of planned randomization
  - ▪ Are currently receiving oral corticosteroids at doses >20-mg per day of prednisone (or equivalent) or have adjusted the dose of corticosteroids within 2 weeks of planned randomization
  - ▪ Have received oral or parenteral [intravenous (IV) or intramuscular (IM)] corticosteroids at doses >40-mg per day of prednisone (or equivalent) within 6

weeks of screening or within 8 weeks of planned randomization, or are anticipated to require parenteral injection of corticosteroids during the study
- Have started treatment with or adjusted the dose of NSAIDs (for which the NSAID use is intended for treatment of signs and symptoms of SLE) within 2 weeks of screening or within 4 weeks of planned randomization
- Have started treatment with or adjusted the dose of an antimalarial (such as hydroxychloroquine, chloroquine, quinacrine) within 10 weeks of screening or within 12 weeks of planned randomization
- Have started treatment with or adjusted the dose of an immunosuppressant (such as MTX, azathioprine, mycophenolate) within 10 weeks of screening or within 12 weeks of planned randomization
- Have received cyclophosphamide (or any other cytotoxic agent) within 12 weeks prior to screening
- Have received etanercept, infliximab, certolizumab, adalimumab, golimumab, or anakinra within 12 weeks of screening; tocilizumab, abatacept, ustekinumab, rituximab, belimumab, or any other B-cell targeted therapies (approved or investigational) within 24 weeks of screening; or any other biologic therapy within 4 weeks or 5 half-lives of screening, whichever is longer
  - o Laboratory Criteria
    - ALT or AST $>2\times$ ULN
    - total bilirubin $\geq1.5\times$ ULN
    - hemoglobin <9 g/dL (90.0 g/L)
    - total white blood cell (WBC) count <2500 cells/μL ($<2.50 \times 10^3$/μL or <2.50 GI/L)
    - neutropenia (absolute neutrophil count [ANC] <1200 cells/μL) ($<1.20 \times 10^3$/μL or <1.20 GI/L)
    - lymphopenia (lymphocyte count <500 cells/μL) ($<0.50 \times 10^3$/μL or <0.50 GI/L)
    - thrombocytopenia (platelets $<50,000$ cells/μL) ($<50 \times 10^3$/μL or <50 GI/L)
    - eGFR (Modification of Diet in Renal Disease [MDRD]) <50 mL/min/1.73 m2.
- Use of Prohibited Concomitant Medication during Treatment Period. For the purposes of defining protocol deviations, medication changes on the date of the Week 24 or ETV are not flagged as protocol deviations:
  - o Use of any new or increase in dose of corticosteroids above baseline dose after randomization without discontinuation
  - o Use of any new or increase in dose of NSAIDs intended for the treatment of SLE after randomization without discontinuation
  - o Use of any new or increase in dose of antimalarials after randomization without discontinuation
- Use of any new or increase in dose of immunosuppressants after randomization without discontinuation Study drug administration and compliance errors:

- o Lack of treatment compliance as defined in Section 6.6 during the treatment
period
- o Drug-dispensing error (eg, patient dispensed wrong package and received
incorrect dose or drug)
- o Patient received drug that was declared "not fit for use"
- o Use of expired clinical trial (CT) Material
- o IWRS data entry errors that impact patient stratification at randomization
- Protocol noncompliance
  - o Patient met temporary or permanent IP discontinuation criteria and IP not
discontinued per protocol
  - o Inadvertent unblinding
  - o Fraud
  - o Other significant Good Clinical Practices (GCP) issues

A summary of the number and percentage of patients with an important protocol deviation by treatment group, overall, and by type of deviation will be provided.

6.15. Exploratory Analyses

The planned exploratory analyses are summarized in Table JAHH.5.6. Details of all exploratory analyses are described from Sections 6.15.1 to 6.15.8.

Table JAHH.5.6.   Study JAHH Exploratory Analyses

Efficacy/Safety Measures	Variable	Analysis
SLEDAI-2K	The proportion of patients who achieve remission of arthritis and/or rash at Week 12	Logistic Regression with NRI
	By organ system and by individual descriptor: Proportion of patients who go from present at baseline to absent at week 24. (See Appendix 2 for definitions of organ systems and individual descriptors).	Fisher's exact test
	By organ system and by individual descriptor: Proportion of patients who go from absent at baseline to present at week 24. (See Appendix 2 for definitions of organ systems and individual descriptors).	
BILAG	By organ system: Proportion of patients who go from A or B at baseline to C or D at endpoint (week 24).	Fisher's exact test
28 Tender/Swollen Joint Count	Change in number of joints affected by SLE arthritis at Week 12 and Week 24, compared to baseline	MMRM
	The proportion of patients with ≥50% reduction in number of joints affected by SLE arthritis at Week 12 and Week 24, compared to baseline	Logistic Regression with NRI
	Proportion of patients with ≥ 50% reduction in number of involved joints in the subset of patients with severe joint involvement at baseline, defined as ≥ 8 swollen and ≥ 8 tender joints at baseline	
CLASI Activity Score	Change at Week 12 and Week 24 (for both activity and damage index), compared to baseline	MMRM
	The proportion of patients with ≥50% reduction from baseline (for both activity and damage index) at Week 12 and Week 24	Logistic Regression with NRI
	The proportion of patients with ≥50% reduction from baseline (for both activity and damage index) at Week 12 and Week 24 in the subset of patients with CLASI >=10 at baseline	Logistic Regression with NRI
Physician Global Assessment of Disease Activity	Change at Week 12 compared to baseline	MMRM
SFI	Proportion of patients experiencing severe and any severity flare	Logistic Regression
	Time to first severe and any severity flare	Cox regression
SLICC/ACR Damage Index Score	Change at Week 24, compared to baseline	MMRM
LLDAS	Proportion of patients who achieve a LLDAS	Logistic Regression

Corticosteroid Sparing	Change in proportion of patients receiving ≥10 mg/day prednisone at baseline able to reduce prednisone (or equivalent) dose ≤7.5 mg/day for 12 consecutive weeks between Week 12 and Week 24	Logistic Regression with NRI
Serologic Markers	Change in anti-dsDNA level at Week 12 and Week 24 compared to baseline	MMRM
	In patients with low C3 at baseline: proportion of those who become normal/high at endpoint (Week 12 and Week 24)	Fisher's exact test
	In patients with low C4 at baseline: proportion of those who become normal/high at endpoint (Week 12 and Week 24)	Fischer's exact test
	Change from baseline in C3 and C4 levels at Week 12 and Week 24	MMRM
	In patients with elevated anti-dsDNA at baseline: proportion of patients with normal at Week 12 and Week 24	Fisher's exact test

Abbreviations: ACR = American College of Rheumatology; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; dsDNA = double-stranded deoxyribonucleic acid; LLDAS = Lupus Low Disease Activity State; MMRM = mixed-effects models for repeated measures; NRI = non-responder imputation; SFI = SLEDAI Flare Index; SLE = systemic lupus erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000; SLICC = Systemic Lupus Erythematosus International Collaborating Clinics.
Note: Additional methods may be used for analysis, as deemed appropriate.

For the categorical measures, a logistic regression model as described in Section 6.1 will be used to test the treatment difference between baricitinib dose regimens (4-mg and 2-mg) versus placebo. The NRI approach will be used to impute response status for patients that qualify based on the approach specified in Section 5.4.1.

For the continuous exploratory analyses, a MMRM method as described in Section 6.1 will be performed. Type III sums of squares for the LS means will be used for the statistical comparison.

For the time-to-event analyses, the Kaplan-Meier estimate of the proportion of patients experiencing severe/mild/moderate flare will be presented. Estimates and 95% CIs for the 25th percentile, median, and 75th percentile will be provided by treatment group.

6.15.1. SLEDAI-2K & BILAG

Imputation rules for partially missing questionnaire data for SLEDAI-2K and BILAG are described in Appendix 2.

Logistic regression analysis, as specified in Section 6.1 will be used to test the treatment difference between each baricitinib dose group (4-mg or 2-mg) and placebo in the proportion of patients who achieve remission of arthritis and/or rash at Week 12.

Within each organ system and each individual item (as detailed in Appendix 2), the proportion of patients who go from "present" at baseline to absent at Week 24 will be compared between

treatment groups with a Fisher's exact test. No statistical comparison will be performed for the summaries by individual descriptors.

Within each organ system and each individual item (as detailed in Appendix 2) the proportion of patients who go from "absent" at baseline to "present" at Week 24 will be compared between treatment groups with a Fisher's exact test. No statistical comparison will be performed for the summaries by individual items.

Within each organ system using the BILAG, the proportion of patients who go from A or B at baseline to C or D at endpoint (Week 24) will be compared between treatment groups with a Fisher's exact test.

6.15.2. Twenty-Eight Tender and Swollen Joint Counts

The change from baseline in the number of 28 tender and swollen joint counts affected by SLE arthritis at Week 12 and Week 24 will be summarized and analyzed by MMRM as described in Section 6.1. The proportion of patients with ≥50% reduction in the number of tender/swollen joints affected by SLE arthritis at Week 12 and Week 24 will also be performed by logistic regression as described in Section 6.1. This analysis (proportion of patients with ≥50% reduction) will be repeated for those patients who had a baseline count of tender joints ≥8 and a baseline count of swollen joints ≥8.

The number of tender and swollen joints will be determined by examination of 28 joints (14 joints on each side of the patient's body). These 28 joints will be assessed and classified as no symptoms, tender only, swollen only, tender and swollen, or not evaluable. The number of tender and swollen joints ranges from 0 to 28, respectively.

The following subset will be assessed (both right and left side):

- shoulder
- elbow
- wrist
- metacarpophalangeal (MCP) I, II, III, IV, and V
- thumb interphalangeal
- proximal interphalangeal (PIP) II, III, IV, and V
- knee

Joints will be assessed for tenderness by pressure and joint manipulation on physical examination. Any positive response on pressure, movement, or both will then be translated into a single tender-versus nontender dichotomy. Joints will be classified as either swollen or not swollen. Swelling is defined as palpable fluctuating synovitis of the joint. Swelling secondary to osteoarthrosis will be assessed as not swollen, unless there is unmistakable fluctuation.

Imputation rules for joint count are described in Appendix 2.

LY3009104

6.15.3. CLASI Activity Score

The CLASI consists of the disease activity and damage scores. Activity is scored on the basis of erythema, scale/hyperkeratosis, mucous membrane involvement, acute hair loss, and nonscarring alopecia. Damage is scored in terms of dyspigmentation and scarring, including scarring alopecia. The scores are calculated by simple addition based on the extent of the symptoms. Note that when calculating the total damage score, the dyspigmentation score is multiplied by 2 if dyspigmentation usually lasts >12 months. Both disease activity and damage scores will be summarized by treatment and visit separately.

Change from baseline in the activity and damage scores will be analyzed at Week 12 and Week 24 by MMRM method as described in Section 6.1. In addition, the change from baseline in proportion of patients with ≥50% reduction in the disease activity score and damage scores at Week 12 and Week 24 will be analyzed by a logistic regression model, as described in Section 6.1. This analysis (proportion of patients with ≥50% reduction) will be repeated for those patients who had a baseline CLASI activity score of 10 or higher.

6.15.4. Physician Global Assessment of Disease Activity

Physician's Global Assessment scores will be summarized by treatment group and visit using descriptive statistics (n, mean, SD, median, minimum, and maximum). Similarly, the change from baseline in Physician's Global Assessment scores will be summarized at all scheduled postbaseline visits.

The change from baseline score to Week 12 will be analyzed using an MMRM model as described in Section 6.1.

6.15.5. SLEDAI Flare Index (SFI)

The SLEDAI Flare Index (SFI) uses the SLEDAI-2K score, disease activity scenarios, treatment changes, and Physician's Global Assessment to define mild/moderate and severe flares. Two sets of analyses will be performed: any flare and a separate analysis for severe flares. The number and percentage of patients with an event and the time to SLE flare (severe and any) will be performed.

The time to the first SLE flare (SFI) in the treatment period is defined as:

Time to first SLE flare (SFI) (days) =
Date of SLE flare (SFI) – Date of randomization +1

If a patient has not experienced a flare by study completion or study discontinuation, the patient will be censored at the date of the last postbaseline SFI assessment in the treatment period.

The number and percentage of patients with an event and the number and percentage of patients censored will be presented by treatment group. The Kaplan-Meier estimate of the proportion of patients experiencing a flare will be presented. Estimates and 95% CIs for the 25th percentile, median, and 75th percentile will be provided by treatment group. A cumulative incidence plot of the time to first flare (SFI) by treatment group will also be provided.

LY3009104

The time to first SLE flare (SFI) will be analyzed using a Cox proportional hazards model with treatment group, baseline disease activity (SLEDAI-2K <10 versus SLEDAI-2K ≥10), baseline anti-dsDNA status (positive or negative), and region fitted as explanatory variables. For each treatment comparison performed, the estimated hazard ratio, corresponding 95% CI, and p-value will be presented. The p-value for all other explanatory variables will also be presented.

Whether a patient has a flare at all during treatment period will be compared between treatment groups using logistic regression as described in Section 6.1. The number of patients experiencing an SFI flare in the treatment period will be presented along with an overall count of SFI flares by severity (mild/moderate and severe) and a count by visit. Patients can appear in the overall summary more than once (no percentage will be presented for this) and at multiple visits.

Finally, the annualized flare rate will be calculated for each treatment group.

6.15.6. SLICC/ACR Damage Index Score

The change from baseline to Week 24 will be analyzed using MMRM model, as described in Section 6.1.

6.15.7. Lupus Low Disease Activity Score (LLDAS)

Number of patients that are responder based on the m-Lupus Low Disease Activity Score will be summarized and compared across treatment groups at Week 24 using the logistic regression model described in Section 6.1 or Fisher's exact text if the size of the cells is too small.

6.15.8. Corticosteroid Sparing

Refer to Section 6.8.1 for a description of analyses in corticosteroids.

6.15.9. Serologic Markers

The following exploratory endpoints will be analyzed by a MMRM model, as described in Section 6.1:

- The change from baseline in anti-dsDNA level at Week 12 and Week 24
- The change from baseline in C3 and C4 levels at Week 12 and Week 24

Using the logistic regression model (or Fisher's exact text if number of response in any category <5) as described in Section 6.1 the following outcomes will be analyzed:

- In patients with low C3 at baseline: proportion of those who become normal/high at endpoint (Week 12 and Week 24)
- In patients with low C4 at baseline: proportion of those who become normal/high at endpoint (Week 12 and Week 24)
- In patients with elevated anti-dsDNA at baseline: proportion of patients with normal at Week 12 and Week 24

6.16. Annual Report Analyses

Annual report analyses will be stated in a separate document.

6.17. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

Summary of AEs, provided as a dataset that will be converted to an XML file. Both SAEs and "other" AEs are summarized by treatment group, by MedDRA PT.

- An AE is considered "serious," whether or not it is a TEAE.
- An AE is considered in the "other" category if it is both a TEAE and is not serious. For each SAE and "other" AE, for each term and treatment group, the following are provided:
  - the number of participants at risk of an event
  - the number of participants who experienced each event term
  - the number of events experienced
- Consistent with www.ClinicalTrials.gov requirements, "other" AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

7. Unblinding Plan

An early PK/PD unblinded snapshot is planned when all patients complete Week 8. Only a small group of personnel within the PK/PD function will be unblinded. The name of each person that is unblinded to the data, as well as the date when unblinding occurs will be documented. A first database lock is planned when all patients complete the treatment period (ie, all patients complete Week 24 or discontinue study treatment). The primary endpoint and key secondary endpoints for this study will be evaluated based on this database lock. The final database lock is planned when all patients complete the follow-up visit. All investigators, study site personnel, and patients will remain blinded to treatment assignments until the final database lock.

8.   References

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Arlington (VA): American Psychiatric Association; 1994.

Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305(6846):160-164.

Gladman D, Ginzler E, Goldsmith C, Fortin P, Liang M, Urowitz M, Bacon P, Bombardieri S, Hanly J, Hay E, Isenberg D, Jones J, Kalunian K, Maddison P, Nived O, Petri M, Richter M, Sanchez-Guerrero J, Snaith M, Sturfelt G, Symmons D, Zoma A. The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index for systemic lupus erythematosis. *Arthritis Rheum*. 1996;39(3):363-369.

Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med*. 1998;17(8):873-890.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003;54(5):573-583.

Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473-483.

Winthrop KL, et al. Ann Rheum Dis 2015;0:1-10. doi:10.1136/annrheumdis-2015-207841.

Yee CS, Cresswell L, Farewell V, Rahman A, Teh LS, Griffiths B, Bruce IN, Ahmad Y, Prabu A, Akil M, McHugh N, D'Cruz D, Khamashta MA, Isenberg DA, Gordon C. Numerical scoring for the BILAG-2004 index. *Rheumatology (Oxford)*. 2010;49(9):1665-1669.

9. Appendices

Appendix 1. Selection of Medications

The following categories of medication are required. For each category, instructions for selecting the correct medications are given.

Corticosteroids

The following ATC codes will be used to select all possible systemic corticosteroids:

H02 Corticosteroids for systemic use (except H02AA Mineralocorticoids), specifically

H02AB Glucocorticoids

H02BX Corticosteroids for systemic use, combinations

H02CA Anticorticosteroids

and additionally,

M01BA Anti-inflammatory/antirheumatic agents in combination with corticosteroids

All unique preferred terms in the database falling under the above ATC code will be reviewed by the Lilly Medical Group in order to determine which ones should be included.

For patients who have at least 1 corticosteroid medication with a start date or stop date during the treatment period (ie, from the date of randomization to the day of completion/early termination), all corticosteroid medications for this patient will be reviewed and the patient will be classified as having an increase or no increase during the treatment period. Note: "Increase" is defined as any dose greater than that taken at baseline, for any period of time.

All corticosteroid doses need to be converted to prednisone equivalent doses (as detailed in Section 6.13). If additional conversion factors are required, these will be added to the table below in a SAP amendment prior to database lock.

The following table should be used for converting nonprednisone medications to prednisone equivalent:

Multiply the dose of the corticosteroid taken by the patient (in milligrams) in Column 1 by the conversion factor in Column 2 to get the equivalent dose of prednisone (in milligrams).

Example: Patient is taking 16 mg of methylprednisolone po daily. To convert to prednisone: 16 mg methylprednisolone × 1.25 = 20 mg prednisone. 16 mg of methylprednisolone po daily is equivalent to 20 mg of prednisone po daily.

Column 1	Column 2
Corticosteroid Preferred Term	Conversion factor for converting to an equivalent prednisone dose
Prednisone	1
Prednisone acetate	1
Prednisolone	1
Prednisolone acetate	1
Prednisolone sodium phosphate	1
Methylprednisolone	1.25
Methylprednisolone acetate	1.25
Methylprednisolone sodium succinate	1.25
Triamcinolone	1.25
Triamcinolone acetonide	1.25
Triamcinolone hexacetonide	1.25
Cortisone	0.2
Cortisone acetate	0.2
Hydrocortisone	0.25
Hydrocortisone acetate	0.25
Hydrocortisone sodium succinate	0.25
Betamethasone	6.25
Betamethasone acetate	6.25
Betamethasone dipropionate	6.25
Betamethasone sodium phosphate	6.25
Dexamethasone	6.25
Dexamethasone acetate	6.25
Dexamethasone phosphate	6.25
Dexamethasone sodium phosphate	6.25
Paramethasone	2.5
Deflazacort	0.83
Celestona bifas	6.25
Depo-medrol med lidokain	1.25
Diprospan	6.25
Fluocortolone	1
Meprednisone	1.25

Antimalarials

The following ATC codes will be used to select all possible antimalarials:

M01AX Other antiinflammatory and antirheumatic agents, non-steroids

P01B Antimalarials, specifically

P01BA Aminoquinolines

P01BB Biguanides

P01BC Methanolquinolines

P01BD Diaminopyrimidines

P01BE Artemisinin and derivatives, plain

P01BF Artemisinin and derivatives, combinations

P01BX Other antimalarials

P01AX Other agents against amoebiasis and other protozoal diseases

M09AA Quinine and derivatives

All unique preferred terms in the database falling under the above ATC code will be reviewed by the Lilly Medical Group in order to determine which ones should be included.

For patients who have at least 1 antimalarial medication with a start date or stop date during the treatment period (ie, from the date of randomization to the day of completion/early termination), all antimalarial medications for this patient will be reviewed and the patient will be classified as having an increase or no increase during the treatment period. Note: "Increase" is defined as any dose greater than that taken at baseline, for any period of time.

Immunosuppressants

The following ATC codes will be used to select all possible immunosuppressants:

M01C Specific Antirheumatic Agents, specifically

M01CA Quinolines

M01CB Gold preparations

M01CC Penicillamine and similar agents

M01CX Other specific antirheumatic agents

L01AA Nitrogen mustard analogues

L01BA Folic acid analogues

L04AA: Selective immunosuppressants

L04AD Calcineurin inhibitors

L04AX Other immunosuppressants

A07EC Aminosalicylic acid and similar agents

J04BA Drugs for Treatment of Lepra

All unique preferred terms in the database falling under the above ATC code will be reviewed by the Lilly Medical Group in order to determine which ones should be included.

For patients who have at least 1 immunosuppressant medication with a start date or stop date during the treatment period (ie, from the date of randomization to the day of completion/early termination), all immunosuppressant medications for this patient will be reviewed and the patient will be classified as having an increase or no increase during the treatment period. Note: "Increase" is defined as any dose greater than that taken at baseline, for any period of time.

Intravenous immunoglobulin

The following ATC code will be used to select all immunoglobulins:

J06B Immunoglobulins (excluding J06BB as unlikely to be IV)

Biologics

The following ATC code will be used to select all biologics

L01XC Monoclonal antibodies

L01XX Other antineoplastic agents

L04AA Selective immunosuppressants

L04AB Tumor necrosis factor alpha (tnf-) inhibitors

L04AC Interleukin inhibitors

M01CX Other specific antirheumatic agents

M05BX Other drugs affecting bone structure and mineralization

V98 Investigational drug

All unique preferred terms in the database falling under the above ATC code will be reviewed by the Lilly Medical Group in order to determine which ones should be included.

NSAIDs

The following ATC codes will be used to select all NSAIDs:

M01 Anti-inflammatory and Antirheumatic Products, specifically

M01AA Butylpyrazolidines

M01AB Acetic acid derivatives and related substances

M01AC Oxicams

M01AE Propionic acid derivatives

M01AG Fenamates

M01AH Coxibs

M01BX Other antiinflammatory/antirheumatic agents in combination with other drugs

M09AX Other drugs for disorders of the musculo-skeletal system

Contraceptive Drugs

The following ATC codes will be used to select all possible contraceptive drugs:

G03A Hormonal Contraceptives for Systemic Use

G02B Contraceptives for Topical Use

All unique preferred terms in the database falling under the above ATC codes will be reviewed by the Lilly Medical Group in order to determine which ones should be included.

Live Vaccines

All medications falling under the following ATC code should be selected and reviewed by the Lilly medical group so that they can confirm which are live vaccines:

J07   Vaccines

Appendix 2.   BILAG 2004 and Additional Efficacy Scales

This appendix contains instructions for calculating scores for the BILAG scale, Joint assessment, SF-36, SLEDAI-2K, SLEDAI Flare Index (SFI), SLICC/ARC Damage Index, and CLASI.

1.   BILAG 2004

BILAG2004 assesses 97 clinical signs, symptoms and laboratory parameters across nine organ system domains: constitutional, mucocutaneous, neuropsychiatric, musculoskeletal, cardiorespiratory, gastrointestinal, ophthalmic, renal and hematological.

For BILAG, any item that appears 'Not done' will be assumed to be 'Not present'. Any missing data will be assumed to also be 'Not present' if at least there is a non-missing item in the whole questionnaire.

Calculation of the BILAG A and B Disease Activity Scores for Each Domain

Constitutional

Category A:

Pyrexia recorded as 2 (same), 3 (worse), or 4 (new) and

Any 2 or more of the following recorded as 2 (same), 3 (worse), or 4 (new):
> Weight loss
> Lymphadenopathy/splenomegaly
> Anorexia

Category B:

Pyrexia recorded as 2 (same), 3 (worse), or 4 (new) or

Any 2 or more of the following recorded as 2 (same), 3 (worse), or 4 (new):
> Weight loss
> Lymphadenopathy/splenomegaly
> Anorexia

but do not fulfill criteria for Category A

Category C

Pyrexia recorded as 1 (improving) or

One or more of the following recorded as >0:
> Weight loss

LY3009104

      Lymphadenopathy/Splenomegaly
      Anorexia

but does not fulfill criteria for category A or B

Category D
Previous involvement

Category E
No previous involvement

<u>MUCOCUTANEOUS</u>

Category A
Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
      Skin eruption - severe
      Angio-oedema - severe
      Mucosal ulceration - severe
      Panniculitis/Bullous lupus - severe
      Major cutaneous vasculitis/thrombosis

Category B
Any Category A features recorded as 1 (improving) or

Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
      Skin eruption - mild
      Panniculitis/Bullous lupus - mild
      Digital infarcts or nodular vasculitis
      Alopecia - severe

Category C
Any Category B features recorded as 1 (improving) or

Any of the following recorded as >0:
      Angio-oedema - mild
      Mucosal ulceration - mild
      Alopecia - mild
      Periungual erythema/chilblains
      Splinter haemorrhages

Category D
Previous involvement

Category E
No previous involvement

Neuropsychiatric

Category A
Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
> Aseptic meningitis
> Cerebral vasculitis
> Demyelinating syndrome
> Myelopathy
> Acute confusional state
> Psychosis
> Acute inflammatory demyelinating polyradiculoneuropathy
> Mononeuropathy (single/multiplex)
> Cranial neuropathy
> Plexopathy
> Polyneuropathy
> Status epilepticus
> Cerebellar ataxia

Category B
Any Category A features recorded as 1 (improving) or

Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
> Seizure disorder
> Cerebrovascular disease (not due to vasculitis)
> Cognitive dysfunction
> Movement disorder
> Autonomic disorder
> Lupus headache - severe unremitting
> Headache due to raised intracranial hypertension

Category C
Any Category B features recorded as 1 (improving)

Category D
Previous involvement

Category E
No previous involvement

MUSCULOSKELETAL

Category A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):
     Severe Myositis
     Severe Arthritis

Category B
Any Category A features recorded as 1 (improving) or

Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
     Mild Myositis
     Moderate Arthritis/Tendonitis/Tenosynovitis

Category C
Any Category B features recorded as 1 (improving) or

Any of the following recorded as >0:
     Mild Arthritis/Arthralgia/Myalgia

Category D
Previous involvement

Category E
No previous involvement

CARDIORESPIRATORY

Category A
Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
       Myocarditis/Endocarditis + Cardiac failure
       Arrhythmia
       New valvular dysfunction
       Cardiac tamponade
       Pleural effusion with dyspnoea

Pulmonary haemorrhage/vasculitis
Interstitial alveolitis/pneumonitis
Shrinking lung syndrome
Aortitis
Coronary vasculitis

Category B
Any Category A features recorded as 1 (improving) or

Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
Pleurisy/Pericarditis
Myocarditis - mild

Category C
Any Category B features recorded as 1 (improving)

Category D
Previous involvement

Category E
No previous involvement

GASTROINTESTINAL

Category A
Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
Peritonitis
Lupus enteritis/colitis
Intestinal pseudo-obstruction
Acute lupus cholecystitis
Acute lupus pancreatitis

Category B
Any Category A feature recorded as 1 (improving) or

Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
Abdominal serositis and/or ascites
Malabsorption
Protein losing enteropathy
Lupus hepatitis

Category C
Any Category B features recorded as 1 (improving)

Category D
Previous involvement

Category E
No previous involvement

OPHTHALMIC

Category A
Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
      Orbital inflammation/myositis/proptosis
      Keratitis - severe
      Posterior uveitis/retinal vasculitis - severe
      Scleritis - severe
      Retinal/choroidal vaso-occlusive disease
      Optic neuritis
      Anterior ischaemic optic neuropathy

Category B
Any Category A features recorded as 1 (improving) or

Any of the following recorded as 2 (same), 3 (worse), or 4 (new):
      Keratitis - mild
      Anterior uveitis
      Posterior uveitis/retinal vasculitis - mild
      Scleritis - mild

Category C
Any Category B features recorded as 1 (improving) or

Any of the following recorded as >0:
      Episcleritis
      Isolated cotton-wool spots (cytoid bodies)

Category D
Previous involvement

Category E
No previous involvement

RENAL

Category A
Two or more of the following providing 1, 4, or 5 is included:

1. Deteriorating proteinuria (severe) defined as

    (a) urine dipstick increased by ≥2 levels (used only if other methods of urine protein
estimation   not available); or
    (b) 24 hour urine protein >1 g that has not decreased (improved) by 25%; or
    (c) urine protein-creatinine ratio >100 mg/mmol that has not decreased (improved) by 25%;
or
    (d) urine albumin-creatinine ratio >100 mg/mmol that has not decreased (improved) by 25%

2. Accelerated hypertension

3. Deteriorating renal function (severe) defined as
    (a) plasma creatinine >130 mol/L and having risen to >130% of previous value; or
    (b) GFR <80 mL/min per 1.73 m^2 and having fallen to <67% of previous value; or
    (c) GFR <50 mL/min per 1.73 m^2 and last time was > 50 ml/min per 1.73 m2 or was not
        measured.
4. Active urinary sediment

5. Histological evidence of active nephritis within last 3 months

6. Nephrotic syndrome

Category B
One of the following:

1. One of the Category A feature

2. Proteinuria (that has not fulfilled Category A criteria)
    (a) urine dipstick which has risen by 1 level to at least 2+ (used only if other methods of
urine
        protein estimation not available); or
    (b) 24 hour urine protein ≥0.5 g that has not decreased (improved) by 25%; or

(c) urine protein-creatinine ratio ≥50 mg/mmol that has not decreased (improved) by 25%; or

(d) urine albumin-creatinine ratio ≥50 mg/mmol that has not decreased (improved) by 25%

3. Plasma creatinine > 130 mol/L and having risen to ≥115% but ≤130% of previous value

Category C

One of the following:

1. Mild/Stable proteinuria defined as
    (a) urine dipstick ≥1+ but has not fulfilled criteria for Category A and B (used only if other methods of urine protein estimation not available); or
    (b) 24 hour urine protein >0.25 g but has not fulfilled criteria for Category A and B; or
    (c) urine protein-creatinine ratio >25 mg/mmol but has not fulfilled criteria for Category A and B; or
    (d) urine albumin-creatinine ratio >25 mg/mmol but has not fulfilled criteria for Category A and
        B

2. Rising blood pressure (providing the recorded values are >140/90 mm Hg) which has not fulfilled criteria for Category A and B, defined as
    (a) systolic rise of ≥30 mm Hg; and
    (b) diastolic rise of ≥15mm Hg

Category D

Previous involvement

Category E

No previous involvement

Note: although albumin-creatinine ratio and protein-creatinine ratio are different, we use the same cut- off values for this index

HEMATOLOGICAL

Category A

TTP recorded as 2 (same), 3 (worse), or 4 (new) or

Any of the following:

Haemoglobin	<8 g/dL
White cell count	$<1.0 \times 10^9$/L
Neutrophil count	$<0.5 \times 10^9$/L
Platelet count	$<25 \times 10^9$/L

Category B
TTP recorded as 1 (improving) or

Any of the following:

Haemoglobin	8 - 8.9 g/dL
White cell count	1 - 1.9 × 10^9/L
Neutrophil count	0.5 - 0.9 × 10^9/L
Platelet count	25 - 49 × 10^9/L
Evidence of active haemolysis	

Category C
Any of the following:

Haemoglobin	9 - 10.9 g/dL
White cell count	2 - 3.9 × 10^9/L
Neutrophil count	1 - 1.9 × 10^9/L
Lymphocyte count	<1.0 × 10^9/L
Platelet count	50 - 149 × 10^9/L
Isolated Coombs' test positive	

Category D
Previous involvement

Category E
No previous involvement

Numerical Categorization for BILAG A- E

For each organ system domain, BILAG 2004 numeric score of A to E will be assigned according to the following rule (Yee et al. 2010)

- BILAG A disease activity score = 12 points
- BILAG B disease activity score = 8 points
- BILAG C disease activity score = 1 point
- BILAG D disease activity score = 0 points
- BILAG E disease activity score = 0 points

For each complete BILAG assessment, the points score from each of the nine BILAG organ system domains will be summed to obtain a BILAG 2004 numeric score. The numeric score will only be calculated if each of the 9 disease activity scores are nonmissing.

2. Joint Assessment

The 28 joints will be evaluated for tenderness and 28 joints will be evaluated for swelling (hips are excluded for swelling) at the specified visits as shown in the schedule of events of the protocol. The 28 joint count will be performed at Week 0 (baseline), Week 2 to Week 24, end of treatment, and at follow-up visit.

The following joint count imputation rules will be applied.

1. Joints that had any of the following procedures or disease conditions that occur at the screening (Visit 1), up to and including baseline (Visit 2), will be imputed as follows:
   - Arthroplasty, fusion, synovectomy, ankylosis, amputation, injury, fracture, infection and other condition: these will be considered "nonevaluable" from baseline up to end of the study. Joints that include "nonevaluable" joints will be prorated from baseline up to end of the study. For example, the swollen joint count score for 6 swollen plus 2 "nonevaluable" joints is calculated as [6/(28-2)]*28 = 6.46.
   - Injection and other procedure: joints that have received intra-articular and bursal injections will be collected in the data but set to painful/ tender or swollen after the injection for 6 months (up to Week 24 [Visit 9]).
2. Joints that had any of the following procedures or disease conditions that occur postbaseline (any time during treatment with the study drugs) will be imputed as follows:
   - Amputation: amputation that occurs postbaseline will be considered "nonevaluable" from the visit after amputation up to end of the study. Joints that include "nonevaluable" joints will be pro-rated from the visit after amputation up to end of the study. For example, the swollen joint count score for 6 swollen plus 2 "nonevaluable" joints is calculated as [6/(28-2)]*28 = 6.46.
   - Arthroplasty, fusion, synovectomy, and ankylosis: if any of these procedures occur postbaseline, joints will be set to tender and swollen from the visit after the procedure up to the end of the study.
   - Infection, injury, and fracture or other condition: if any of these conditions occur postbaseline, joints will be set to tender and swollen from the visit after the condition until the condition is resolved.
   - Injection and other procedure: joints that have received intra-articular and bursal injections postbaseline will be collected in the data but set to tender and swollen after the injection for 6 months or the end of the study (whichever comes first).

The number of tender and swollen joints will be calculated by summing all joints respectively. For patients who have an incomplete set of joints evaluated, the joint count will be adjusted to a 28-joint count for tenderness and a 28-joint count for swelling by dividing the number of affected joints by the number of evaluated joints and multiplying by 28 for tenderness and 28 for swelling.

3. <u>SLEDAI-2K</u>

The total SLEDAI score is a weighted sum of those questions that have been answered as "present" on the CRF. The weights are listed below. For example if a patient's CRF was marked present on psychosis, arthritis and rash then their score would be 8+4+2=14.

For the SLEDAI questionnaire, if any, but not all, of the 24 item scores is missing then impute a score of 0 for the missing item(s) when calculating the total score. In addition, any item with missing data will be imputed as 'not present' for any other analyses if at least one of the 24 items is non-missing. If all 24 item scores are missing, then the total score will be missing, as well as any other analysis by organ system or individual descriptor.

Weight	SLEDAI SCORE	Descriptor	Definition
8	_____	Seizure	Recent onset, exclude metabolic, infectious or drug causes.
8	_____	Psychosis	Altered ability to function in normal activity due to severe disturbance in the perception of reality. Include hallucinations, incoherence, marked loose associations, impoverished thought content, marked illogical thinking, bizarre, disorganized, or catatonic behavior. Exclude uremia and drug causes.
8	_____	Organic brain syndrome	Altered mental function with impaired orientation, memory, or other intellectual function, with rapid onset and fluctuating clinical features, inability to sustain attention to environment, plus at least 2 of the following: perceptual disturbance, incoherent speech, insomnia or daytime drowsiness, or increased or decreased psychomotor activity. Exclude metabolic, infectious, or drug causes.
8	_____	Visual disturbance	Retinal changes of SLE. Include cytoid bodies, retinal hemorrhages, serous exudate or hemorrhages in the choroid, or optic neuritis. Exclude hypertension, infection, or drug causes.
8	_____	Cranial nerve disorder	New onset of sensory or motor neuropathy involving cranial nerves.
8	_____	Lupus headache	Severe, persistent headache; may be migrainous, but must be nonresponsive to narcotic analgesia.
8	_____	CVA	New onset of cerebrovascular accident(s). Exclude arteriosclerosis.
8	_____	Vasculitis	Ulceration, gangrene, tender finger nodules, periungual infarction, splinter hemorrhages, or biopsy or angiogram proof of vasculitis.
4	_____	Arthritis	≥ 2 joints with pain and signs of inflammation (i.e., tenderness, swelling or effusion).
4	_____	Myositis	Proximal muscle aching/weakness, associated with elevated creatine phosphokinase/aldolase or electromyogram changes or a biopsy showing myositis.
4	_____	Urinary casts	Heme-granular or red blood cell casts.
4	_____	Hematuria	>5 red blood cells/high power field. Exclude stone, infection or other cause.
4	_____	Proteinuria	>0.5 gram/24 hours
4	_____	Pyuria	>5 white blood cells/high power field. Exclude infection.
2	_____	Rash	Inflammatory type rash.
2	_____	Alopecia	Abnormal, patchy or diffuse loss of hair.
2	_____	Mucosal ulcers	Oral or nasal ulcerations.
2	_____	Pleurisy	Pleuritic chest pain with pleural rub or effusion, or pleural thickening.
2	_____	Pericarditis	Pericardial pain with at least 1 of the following: rub, effusion, or electrocardiogram or echocardiogram confirmation.
2	_____	Low complement	Decrease in CH50, C3, or C4 below the lower limit of normal for testing laboratory
2	_____	Increased DNA binding	Increased DNA binding by Farr assay above normal range for testing laboratory.
1	_____	Fever	>38° C. Exclude infectious cause.
1	_____	Thrombocytopenia	<100,000 platelets / x10⁹/L, exclude drug causes.
1	_____	Leukopenia	< 3,000 white blood cells / x10⁹/L, exclude drug causes.

TOTAL SLEDAI SCORE _____

For the purposes of analyses by organ system, the following list shows what descriptors are part of each organ system:.

1. **CNS**: Seizure, Psychosis, Organic Brain Syndrome, Visual Disturbance, Cranial Nerve Disorder, Lupus Headache, CVA
2. **Vascular**: Vasculitis
3. **Musculoskeletal**: Arthritis, Myositis
4. **Renal**: Urinary Casts, Hematuria, Proteinuria, Pyuria
5. **Mucocutaneous**: Rash, Alopecia, Mucosal Ulcers
6. **Cardiovascular and Respiratory**: Pleurisy, Pericarditis

7. **Immunologic**: Low complement, Increased DNA Binding
8. **Constitutional**: Fever
9. **Hematologic**: Thrombocytopenia, Leukopenia

4.   SLEDAI Flare Index (SFI)

A mild or moderate flare is present if any of the CRF variables are selected under Mild/moderate flare index criteria. There are 14 possible criteria that will indicate a mild/moderate flare. If the box "not applicable" is checked, then the patient did not experience a flare. The variable is a simple yes/no as to whether the patient had a mild/moderate flare.

Similarly, for severe flares: if any of the 14 criteria are marked, then the patient has experienced a severe flare.

Additionally, the variable of either a mild/moderate or severe flare will be analyzed. This variable will be "yes" if the patient has experienced either a mild/moderate flare or a severe flare.

5.   SLICC/ARC Damage Index

The total score is a simple sum of all 39 items on the CRF. Note that while most items are scored either 0 (not present) or 1 (present), there are some items that are weighted heavier. Note for example that the following items can have a response of either 0, 1 or 2 on the CRF. If "2" is selected that that is what should be included in the sum

- Cerebrovascular accident,
- myocardial infarction,
- significant tissue loss
- infarction or resection of bowel
- avascular necrosis
- malignancy

Also note for end stage renal failure, it is either 0 (not present) or 3 (present). If it is present, then the score of 3 is added to the total**.**

6.   CLASI

There are two scores derived from the CLASI that will be analyzed. The Total Activity score and the Total Damage score. Both of these have been entered by the site and there has been an edit check against the calculations when the site enters the score. No additional calculation is necessary**.**

Leo Document ID = 851adfd9-d949-41ae-87d4-82a67691ccd7

Approver: PPD
Approval Date & Time: 26-Jul-2017 17:51:03 GMT
Signature meaning: Approved