I4V-MC-JAIA Statistical Analysis Plan Version 2

A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Phase 3 Study of Baricitinib in Patients with Systemic Lupus Erythematosus

NCT03616964

Approval Date: 01-Nov-2021

1 Statistical Analysis Plan: I4V-MC-JAIA: A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Phase 3 Study of Baricitinib in Patients with Systemic Lupus Erythematosus

Baricitinib (LY3009104) Systemic Lupus Erythematosus

Study I4V-MC-JAIA is a randomized, double-blind, placebo-controlled, parallel-group, Phase 3 study of baricitinib, designed to investigate the efficacy and safety of baricitinib 2-mg and 4-mg in patients with systemic lupus erythematosus.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I4V-MC-JAIA
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly on 4 June 2019.
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly on date provided below.

Approval Date: 01-Nov-2021 GMT

2 Table of Contents

Section **Page**

List of Tables

Table **Page**

List of Figures

Figure **Page**

List of Appendices

Appendix **Page**

3 Revision History

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

Statistical Analysis Plan Version 2 was approved prior to the primary outcome database lock and included the following key changes:

- Added clarification on summarization of patients randomized to 4 mg with decreased renal function at screening.
- Added clarifications for the time to first severe Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) Flare Index (SFI) flare analysis and annualized severe SFI flare rate analysis as major secondary and secondary endpoints, respectively.
- Removed the time to first mild/moderate flare analysis because the analysis treats patients with severe flares as being the same as patients with no flares.
- Added the other secondary objectives from protocol to Table JAIA.4.1.
- Added randomized and per protocol populations in the analysis populations.
- Added clarification of the definition of baseline and postbaseline.
- Added the definition of intercurrent events.
- Updated the order of covariance structure for mixed model for repeated measures (MMRM) analysis.
- Removed pattern mixture model for handling missing data.
- Added hybrid imputation (MI) for handling missing data due to COVID-19.
- Updated the figure for graphical testing scheme.
- Added the summary of pre-existing conditions in the patient characteristics section.
- Updated the treatment compliance criteria.
- Updated the protocol-specified concomitant medication rules and clarified the definition of baseline concomitant medication.
- Modified the derivation of efficacy/health outcomes measures and endpoints including the missing imputation approach.
- Added organ domain improvement/worsening defined by SLEDAI-2K (SLEDAI 2000) and British Isles Lupus Assessment Group (BILAG), Systemic Lupus Erythematosus Responder Index (SRI)-4 components, BICLA (British Isles Lupus Assessment Group), BICLA components, time to first severe SFI flare, and discontinuation of corticosteroid in the derivation of efficacy/health outcomes measures and endpoints.
- Added Per Protocol, COVID free (PPCV), Per Protocol (PP), and As Collected (AC) datapoint sets for efficacy/health outcomes analyses.
- Added analysis on PPCV and AC datapoint sets for BICLA responder and time to first severe flare analysis in modified intent-to-treat (mITT) population.
- Added analysis on PPCV datapoint sets with hybrid imputation for SRI-4 in per protocol population, and for each component of SRI-4 and BICLA in mITT population.
- Added analysis on PP datapoint sets for organ domain improvement/worsening defined by SLEDAI-2K and BILAG, and for change from baseline of Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SLICC/ACR) damage score.

- Added analysis on PP and PPCV datapoint sets for additional corticosteroid sparing endpoints.
- Modified tipping point analysis to provide 2-dimensional results along with a modified analysis framework.
- Added exposure adjusted analyses for adverse events.
- Added the summary of treatment-emergent potential hepatic disorders based on Medical Dictionary for Regulatory Activities (MedDRA) Standardized MedDRA Queries (SMQs).
- Removed the analysis of Framingham risk score.
- Added subgroup analysis section for SRI-4 responder rate at the Week 52 visit by demographic and clinical characteristic subgroups.
- Modified the analyses of adverse events and treatment-emergent lab tests and vital signs by defining the follow-up data as all data occurring up to and including 30 days after last dose of treatment.
- Added the summaries and listings of patients with COVID-19 impact.
- Updated the anatomical therapeutic chemical (ATC) codes and medical review process of prohibited medication in Appendix 1.
- Moved the table of patient characteristics to Appendix 2, and updated a few categories.
- Added Appendix 3 of Efficacy Laboratory Analytes.
- Revised the detailed instructions for calculating joint assessment scores and added to Appendix 4.
- Added SFI severe flare identification in Appendix 5.
- Removed publicly available efficacy scale derivations from appendix.
- Made other minor typographical corrections and clarifications not affecting content.

4 Study Objectives

Table JAIA.4.1 provides the protocol defined primary and major secondary objectives and endpoints of the study. Additional secondary endpoints to provide supportive evidence of efficacy are described in Table JAIA.6.6.

The estimand (ICH E9 R1) associated with each endpoint/analysis is documented in the following places:

- The population of interest is described in the protocol inclusion/exclusion criteria and in this document in Table JAIA.6.1 and Section 6.2. The endpoints/variables may be found in Table JAIA.6.4 and Table JAIA.6.6.
- The handling of intercurrent events and missing data may be found in Section 6.4.2 and Table JAIA.6.5.
- Population summary measures are described in Table JAIA.6.6.

Table JAIA.4.1. Primary and Major Secondary Objectives and Endpoints

Objectives	Endpoints (Variables)[a]
Primary[b]	

To evaluate the effect of baricitinib 4-mg QD and background SoC therapy compared to placebo and SoC on SLE disease activity	• Proportion of patients achieving an SRI-4 response at Week 52, defined as:
	  • a reduction of ≥4 points from baseline in SLEDAI-2K score, and
	  • No new BILAG A and no more than 1 new BILAG B disease activity score, and
	  • No worsening (defined as an increase of ≥0.3 points [10 mm] from baseline) in the PGA of Disease Activity.

Major Secondary[c]	
To evaluate the effect of baricitinib 4-mg QD plus SoC compared to placebo plus SoC on SLE disease activity	• Proportion of patients achieving an SRI-4 response at Week 24 • Proportion of patients achieving a LLDAS response at Week 52
To evaluate the corticosteroid sparing effect of baricitinib 4-mg QD plus SoC compared to placebo plus SoC	• Proportion of patients receiving >7.5 mg prednisone (or equivalent) at baseline able to decrease dose by ≥25% to a prednisone equivalent dose of ≤7.5 mg/day maintained between Week 40 and Week 52
To evaluate the effect of baricitinib 4-mg QD plus SoC compared to placebo plus SoC on SLE flares	• Time to first severe flare over 52 weeks
To evaluate the effect of baricitinib 4-mg QD plus SoC compared to placebo plus SoC on PROs	• Change from baseline in Worst Pain NRS at Week 52 • Change from baseline in FACIT-Fatigue total score at Week 52

Primary and Major Secondary Objectives and Endpoints

Objectives	Endpoints (Variables)[a]
To evaluate the effect of baricitinib 2-mg QD plus SoC compared to placebo plus SoC on SLE disease activity	• Proportion of patients achieving an SRI-4 response at Week 52 • Proportion of patients achieving an SRI-4 response at Week 24 • Proportion of patients achieving a LLDAS response at Week 52
To evaluate the corticosteroid sparing effect of baricitinib 2-mg QD plus SoC compared to placebo plus SoC	• Proportion of patients receiving >7.5 mg prednisone (or equivalent) at baseline able to decrease dose by ≥25% to a prednisone equivalent dose of ≤7.5 mg/day maintained between Week 40 and Week 52
To evaluate the effect of baricitinib 2-mg QD plus SoC compared to placebo plus SoC on SLE flares	• Time to first severe flare over 52 weeks
To evaluate the effect of baricitinib 2-mg QD plus SoC compared to placebo plus SoC on PROs	• Change from baseline in Worst Pain NRS at Week 52 • Change from baseline in FACIT-Fatigue total score at Week 52
Other Secondary (Non-Multiplicity Controlled)	

To evaluate the effect of baricitinib 4-mg or 2-mg QD plus SoC compared to placebo plus SoC on SLE disease activity	• Change from baseline in total SLEDAI-2K scores over time through Week 52
	• Change from baseline in PGA score over time through Week 52
	• Proportion of patients achieving an SRI-4 response over time through Week 52
	• Proportion of patients achieving an SRI-5, -6, -7, and -8 response over time through Week 52.
	• Proportion of patients achieving reduction of ≥4 points from baseline in SLEDAI-2K score at Week 52
	• Proportion of patients with no new BILAG A and no more than 1 new BILAG B disease activity score at Week 52
	• Proportion of patients with no worsening (defined as an increase of ≥0.3 points [10 mm] from baseline) in the PGA of Disease Activity at Week 52
	• Proportion of patients achieving a BICLA response at Week 52, defined as:
	  • Reduction of all baseline BILAG A to B/C/D and baseline BILAG B to C/D, and no BILAG worsening in other organ systems, where worsening is defined as ≥1 new BILAG A or ≥2 new BILAG B
	  • No worsening from baseline in SLEDAI-2K, where worsening is defined as an increase from baseline of >0 points in SLEDAI-2K
	  • No worsening from baseline in participants' lupus disease activity, where worsening is defined by an increase ≥0.30 points on a 3-point PGA VAS

Primary and Major Secondary Objectives and Endpoints

Objectives	Endpoints (Variables)[a]
To evaluate the corticosteroid sparing effect of baricitinib 4-mg or 2-mg QD plus SoC compared to placebo plus SoC	• Change from baseline in prednisone dose at Week 52 • Proportion of patients taking corticosteroids at baseline able to discontinue use at Week 52
To evaluate the effect of baricitinib 4-mg or 2-mg QD plus SoC compared to placebo plus SoC on SLE flares	• Time to first mild/moderate flare over 52 weeks* • Time to first flare (any severity) over 52 weeks • Annualized mild/moderate flare rate • Annualized severe flare rate • Annualized flare rate (any severity)
To evaluate the effect of baricitinib 4-mg QD or 2-mg QD plus SoC compared to placebo plus SoC on mucocutaneous manifestations of SLE	• Proportion of patients with CLASI total activity score ≥10 at baseline with ≥50% reduction in CLASI total activity score at Week 52
To evaluate the effect of baricitinib 4-mg QD or 2-mg QD plus SoC compared to placebo plus SoC on musculoskeletal manifestations of SLE	• Change from baseline in tender joint count at Week 52 • Change from baseline in swollen joint count at Week 52
To evaluate the effect of baricitinib 4-mg QD or 2-mg QD plus SoC compared to placebo plus SoC on individual organ system disease activity	• Proportion of patients with improvement in each SLEDAI-2K organ system versus baseline at Week 52 • Proportion of patients with worsening in each SLEDAI-2K organ system versus baseline at Week 52
To evaluate the effect of baricitinib 4-mg QD or 2-mg QD plus SoC compared to placebo plus SoC on PROs	• Change from baseline in Worst Fatigue NRS at Week 52 • Change from baseline in Worst Joint Pain NRS at Week 52 • Change from baseline in Joint Stiffness Duration at Week 52 • Change from baseline in Joint Stiffness Severity NRS at Week 52 • Change from baseline in Patient Global Impression of Severity at Week 52 • Change from baseline in Patient Global Impression of Change at Week 52 • Change from baseline in mental component score (MCS), physical component score (PCS), and domain scores in the Short-Form 36-item health survey version 2 (SF-36v2) acute at Week 52 • Change from baseline in the EQ-5D-5L at Week 52 • Change from baseline in the WPAI-Lupus at Week 52

Primary and Major Secondary Objectives and Endpoints

Objectives	Endpoints (Variables)[a]
To measure baricitinib PK exposure and assess the relationship between exposure and efficacy	• Population pharmacokinetics of baricitinib in patients with SLE • Proportions of patients achieving SRI-4 by exposure quartile

Abbreviations: BILAG = British Isles Lupus Assessment Group; BICLA = British Isles Lupus Assessment Group Based Composite Lupus Assessment; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; EQ-5D-5L = 5-level EQ-5D version; FACIT = Functional Assessment of Chronic Illness Therapy; LLDAS= Lupus Low Disease Activity State; MCS = mental component score; NRS = numeric rating scale; PGA = Physician Global Assessment; PK = pharmacokinetics; PRO = patient-reported outcome; 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; SoC = standard of care; SRI = Systemic Lupus Erythematosus Responder Index; VAS = visual analogue scale; WPAI-Lupus = Work Productivity and Activity Impairment Lupus.

[a] Additional analyses for other secondary (non-multiplicity controlled) endpoints are included in Table JAIA.6.6.

[b] All primary and major secondary endpoint analyses will utilize a multiplicity control approach based on a graphical multiple testing procedure to control the overall family-wise type I error rate at a 2-sided alpha level of 0.05. The graphical multiple testing procedure described in Bretz et al. 2009 will be used.

[c] The order of testing of the major secondary endpoints is determined from the results of statistical simulations before study team is unblinded to efficacy results.

* This objective is written as it is in protocol amendment (b), but will be removed because the analysis treats patients with severe flares as being the same as patients with no flares.

5 Study Design

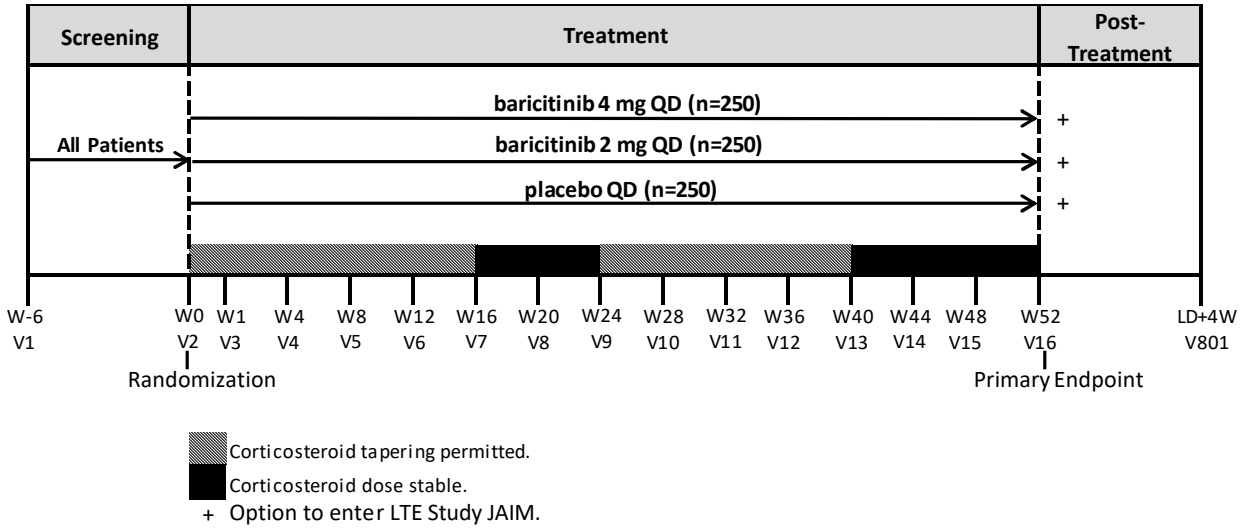
5.1 Summary of Study Design

Study I4V-MC-JAIA (JAIA) is a Phase 3, multicenter, randomized, double-blind, parallel-group, placebo-controlled, outpatient, 52-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 has a screening period followed by a 52-week double-blinded treatment period and a 4-week post-treatment follow-up period.

Patients will be randomized in a 1:1:1 ratio between baricitinib 4-mg, 2-mg, and placebo. Study treatment will be administered as 2-mg and 4-mg tablets once daily (QD). Patients with documented renal impairment (defined as screening estimated glomerular filtration rate [eGFR] ≥40 to <60 mL/min/1.73 m^2), who are randomized to active treatment (either to the baricitinib 4-mg arm or the baricitinib 2-mg arm) will receive a baricitinib 2-mg QD dose by the interactive web-response system (IWRS).

Investigators are encouraged to taper corticosteroids according to clinical practice with a goal of ≤ 7.5 mg/day by Week 40. A single corticosteroid burst is permitted for patients with increased disease activity early in the treatment period. Complete details of corticosteroid rules in relation to the clinical trial endpoints can be found in JAIA Protocol (b) Section 7.7.1.

Patients who were unable to attend Visit 16 within the original protocol-defined window (Week 52/Day 365 ± 4 days) due to COVID-19 were permitted to delay completion of the Visit 16 by up to 1 month. Patients who complete treatment in this study through Visit 16 may be eligible to participate in a study extension, if enrollment criteria are met.



Abbreviations: LD = last dose of investigational product; LTE = long-term extension; n = number of patients in each group; QD = once daily; V = visit; W = week.

Figure JAIA.5.1. Illustrates the study design.

5.2 Determination of Sample Size

Study JAIA will randomize approximately 750 patients with a 1:1:1 ratio between baricitinib 4-mg, 2-mg, and placebo. The proposed sample size will ensure >90% power to detect an absolute difference of 17% between a baricitinib treatment group and the placebo treatment group assuming a 40% placebo response rate for the primary endpoint using a 2-sided α of 0.05. The anticipated effect size represents a clinically relevant difference.

Sample size and power estimates were obtained from nQuery® Advisor 7.0 using the χ^2 test of equal proportions.

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; placebo) to double-blind treatment at Visit 2 (Week 0). Randomization will be stratified by disease activity (total Systemic Lupus Erythematosus Disease Activity Index 2000 [SLEDAI-2K] <10; ≥10), corticosteroid dose (<10 mg/day; ≥10 mg/day prednisone or equivalent at baseline), and region as given in Table JAIA.5.1.

Assignment to treatment groups will be determined by a computer-generated random sequence using an IWRS. The IWRS will be used to assign packages containing double-blind investigational product to each patient, starting at Visit 2 (Week 0), and at each visit subsequent through Visit 15 (Week 48).

Table JAIA.5.1. Countries and Associated Geographical Regions

Geographical Region	Country or Countries
North America	United States
Central, South America and Mexico	Argentina, Chile, Colombia
Europe	France, Italy, Poland, Romania, Spain
Asia	Japan, Korea, Philippines
Rest of World	India, Serbia, South Africa

6 A Priori Statistical Methods

6.1 General Considerations

This SAP is intended to describe the analyses of all primary and secondary objectives, as well as safety assessments, for Study JAIA. Sensitivity and supplementary analyses intended to support the primary and major secondary analyses are also included. Exploratory analyses will be described in a separate document.

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly) or its designee.

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 such as Spotfire instead of, or in addition to, a static display. Any display described in this SAP and not included in the CSR would be available upon request.

Throughout this document when analyses for corticosteroids are described, only systemic corticosteroids specified in Appendix 1 are intended for inclusion, unless otherwise specified. In addition, unless otherwise specified, accommodations for COVID-19 in analyses, summaries, and listings are specific to non-COVID-19-infection-related events.

6.2 Definition of Populations

Analysis populations are defined in Table JAIA.6.1 along with their associated purpose. The treatment groups and inferential comparisons described in this table will be used unless otherwise specified. Also, unless otherwise specified patients will be analyzed according to the treatment to which they were assigned.

Table JAIA.6.1. Analysis Populations

Population	Description
Screened Population	**Definition:** Includes all patients who signed informed consent form. **Purpose:** Used for summarizing screen failures and reason for screen failures. **Treatment Groups:** None **Inferential Comparisons:** None
Randomized Population	**Definition:** Includes all randomized patients. **Purpose:** Used for listing patient disposition and treatment assignment. **Treatment Groups:** baricitinib 2-mg QD, baricitinib 4-mg QD, placebo QD **Inferential Comparisons:** None
mITT Population	**Definition:** All randomized patients who receive at least 1 dose of study treatment (regardless of if the patient does not receive the correct treatment, or otherwise does not follow the protocol). **Purpose:** Used for efficacy and health outcomes related analyses. **Treatment Groups:** baricitinib 2-mg QD, baricitinib 4-mg QD, placebo QD **Inferential Comparisons:** baricitinib 4-mg QD vs. placebo QD; baricitinib 2-mg QD vs. placebo QD
Per Protocol Population	**Definition:** All randomized patients who received at least 1 dose of study treatment without important protocol deviations that impacted the analysis (defined in a separate study deviation rules document). **Purpose:** Used for the sensitivity analysis for the primary efficacy. **Treatment Groups:** baricitinib 2-mg QD, baricitinib 4-mg QD, placebo QD **Inferential Comparisons:** baricitinib 4-mg QD vs. placebo QD; baricitinib 2-mg QD vs. placebo QD
Safety Population	**Definition**: The safety population is defined as all randomized patients who receive at least 1 dose of investigational product and who did not discontinue from the study for the reason "Lost to follow-up" at the first postbaseline visit. **Purpose**: Used for all safety analyses. **Treatment Groups**: baricitinib 2-mg QD, baricitinib 4-mg QD, placebo QD **Inferential Comparisons**: baricitinib 4-mg QD vs. placebo QD; baricitinib 2-mg QD vs. placebo QD

Abbreviations: mITT = modified intent-to-treat; QD = once daily.

Patients who are randomized to 4 mg, but who have decreased renal function, identified by an eGFR measurement of <60 mL/min/1.73m^2 at screening, will receive a dose of 2 mg QD. These patients will be summarized within the 4-mg treatment arm for all summaries unless otherwise noted. A list of these patients who had this dose adjustment will be provided. A list of patients who had a dose correction due to incorrect eGFR entry at randomization will also be provided.

6.3 Definition of Baseline and Postbaseline Periods

For efficacy analyses (excluding daily diaries), baseline is defined as the last non-missing value prior to the first dose of study drug, unless otherwise specified. In most cases, this will be the measurement recorded at Week 0 (Visit 2).

Baseline for the safety analyses is defined as:

- The last non-missing scheduled (planned) measurement prior to the first dose of the study drug for continuous measures by-visit analyses, and
- All non-missing measurements prior to the first dose of the study drug for all other analyses.

In general, for efficacy and safety (except adverse events) analyses, when the baseline measurement is collected on the same day (where time is not collected) as the first dose date, this measurement will be used as the baseline value for data analysis.

Postbaseline measures for the safety analyses are defined as the non-missing scheduled (planned) measurements after the first dose of the study drug for continuous measures by-visit analyses and all non-missing measurements after the first dose of the study drug for all other analyses.

Postbaseline data collection begins after the first dose of study drug administration at Visit 2 (Week 0) and will continue through Visit 16 (Week 52), early study termination visit, or the follow-up visit.

The baseline value for the daily diary assessments (Worst Pain Numeric Rating Scale [NRS], Worst Joint Pain NRS, Worst Fatigue NRS, Joint Stiffness Duration, Joint Stiffness Severity NRS, and Patient Global Impression of Severity [PGI-S]) is defined as the average of the non-missing assessments in the last 7 days prior to the date of first study drug administration (expected at Week 0, Visit 2). There must be at least 4 non-missing daily measurements in the 7 days indicated. If there are less than 4 non-missing daily assessments collected prior to the date of first study drug administration, then the baseline value will be designated as missing. If there are multiple assessments collected on the same day, only the first collected record will be used.

Postbaseline values of daily diary assessments collected after first study drug administration through Visit 16, or early study termination visit, will be defined as the average of the last 7 days prior to the date of the study visit. If less than 4 non-missing daily measurements are present, the visit will be set to missing.

Protocol-defined visit windows are defined as ±3 days for Visit 3 (Week 1) and ±4 days, from Visit 4 (Week 4) until Visit 16 (Week 52). If there is data within the protocol-defined window, that data will be used in by-visit analyses. If a protocol-defined visit is missing, then the nearest unscheduled visit data that exists will be mapped to that visit and included in by-visit analyses. If 2 unscheduled visits are equidistant to a missing visit, the earlier unscheduled visit will be used.

6.4 Analysis Methods

The family-wise type I error rate (FWER) will be controlled at a 0.05 significance level for primary and major secondary endpoints using a prespecified graphical procedure (see Section 6.6). Unless otherwise specified, all other hypothesis testing will be conducted at a nominal significance level of 0.05.

For assessments of the primary endpoint and other binary efficacy and health outcomes endpoints, the following will be provided unless otherwise specified:

- Observed percentage for each treatment group along with the corresponding 2-sided 95% confidence interval (CI)
- The difference in percentages and 95% CI using the Newcombe-Wilson method without continuity correction, used for descriptive purposes
- The p-value and 95% (unless otherwise specified) CI for the odds ratio from the logistic regression model, used for primary statistical inference. The model will include the following explanatory variables: treatment, baseline disease activity (total SLEDAI-2K <10; ≥10), baseline corticosteroid dose (<10 mg/day; ≥10 mg/day prednisone or equivalent) and region (North America, Central/South America/Mexico, Europe, Asia and Rest of World). A Firth penalized likelihood will be used (Firth 1993), and
- If deemed necessary, additional analyses of categorical efficacy variables will be conducted to address sparse data or small sample sizes. The Fisher's exact test will be utilized if necessary.

Treatment comparisons of continuous efficacy and health outcome variables, except for SLICC/ACR Damage Index, Patient's Global Impression of Change and Joint Stiffness Duration, will use a restricted maximum likelihood-based MMRM analysis. The model will include the following explanatory variables as fixed factors: treatment, baseline disease activity (total SLEDAI-2K <10; ≥10), baseline corticosteroid dose (<10 mg/day; ≥10 mg/day prednisone or equivalent), region (North America, Central/South America/Mexico, Europe, Asia and Rest of World), visit (as categorical variable), baseline value, treatment-by-visit interaction, and baseline value-by-visit interaction. The response variable will be change from baseline across postbaseline visits. Within-patient errors will be modeled using an unstructured covariance structure. If the model fails to converge, the following covariance structures will be implemented in the following order: heterogeneous Toeplitz, heterogeneous autoregressive (1) heterogeneous compound symmetry, and autoregressive. The least squares (LS mean) difference, standard error, p-value, and CIs will be reported. Type III tests for the LS means will be used for the statistical comparisons.

Treatment comparisons of SLICC/ACR Damage Index will be made using analysis of covariance (ANCOVA) with baseline value as covariate. For Patient's Global Impression of Change, the analysis of variance (ANOVA) will be applied. For both models, treatment, baseline disease activity (total SLEDAI-2K <10; ≥10), baseline corticosteroid dose (<10 mg/day; ≥10 mg/day prednisone or equivalent), and region (North America, Central/South America/Mexico, Europe,

Asia and Rest of World) will be included as the categorical factors. Type III tests for LS means will be used for statistical comparison between treatment groups. The LS mean difference, standard error, p-value, and 95% CI may also be reported.

Treatment comparisons for duration of joint stiffness will be analyzed using a nonparametric method using the Hodges-Lehmann estimator. For the comparison between each baricitinib group and placebo group, the median difference of change from baseline score and its 95% CI from the Hodges-Lehmann estimator will be presented. The p-value for the median difference of change from baseline will be calculated using the Wilcoxon rank-sum test. A 95% CI for the median change from baseline score within each treatment group will be calculated using distribution-free confidence limits.

Time to first flare analyses will be analyzed using a Cox proportional hazards model with treatment group, with disease activity (total SLEDAI-2K <10; ≥10), baseline corticosteroid dose (<10 mg/day; ≥10 mg/day prednisone or equivalent), and region (North America, Central/South America/Mexico, Europe, Asia and Rest of World) fitted as explanatory variables. Hazard ratios, median time to flare, 95% CI and p-values will be presented for treatment comparisons to placebo. Only if flares occurred during the flare exposure time period (defined below for each datapoint set, respectively) will be counted as an event. For severe SFI flare, only the eligible severe flares will be counted (See 6 criteria for identifying eligible severe SFI flares in Appendix 5). Patients who do not have an event during the flare exposure time period will be censored at the end of the flare exposure time. The flare exposure time varies with different datapoint sets (Table JAIA.6.5).

- For the PPCV datapoint set, while considering the treatment effect without potentially confounding influences of COVID-19, the flare exposure time is defined as the time after the date of the first dose of study drug to the prohibited medication start date, treatment discontinuation date, study discontinuation date or date of the Week 52 visit, whichever is earlier, and excluding the time period of treatment interruption due to COVID-19 from the interruption start date plus 20 days to the interruption end date plus 20 days, both inclusive, if the interruption duration is 20 days or more.
- For the PP datapoint set, which ignores treatment interruption due to COVID-19, the flare exposure time is defined as the time after the date of the first dose of investigational product to the prohibited medication start date, treatment discontinuation date, study discontinuation date or date of the Week 52 visit, whichever earlier.
- For the As Collected (AC) datapoint set, which considers all collected data, the flare exposure time is defined as the time after the date of the first dose of investigational product to the study discontinuation date or date of the Week 52 visit, whichever earlier.

The annualized flare rate will be calculated as the number of flares divided by the flare exposure time in days multiplied with 365.25, where the flare exposure time is defined above and varies with the different datapoint sets. The number of flares includes the flare events that happened during the flare exposure time period. All collected flares, including mild/moderate and severe flares will be considered eligible for the analysis.

For comparing the flare rate in the baricitinib treatment group versus the placebo, the negative binomial regression model will be applied. The response variable in the model will be the number of flares during the flare exposure time with baseline disease activity (total SLEDAI-2K <10; ≥10), baseline corticosteroid dose (<10mg/day; ≥10 mg/day prednisone or equivalent), region (North America, Central/South America/Mexico, Europe, Asia and Rest of World) as explanatory variables. The natural logarithm of the flare exposure time will be used as an offset variable in the model to adjust for patients having different exposure times. The estimated rate ratio and the corresponding 95% CI, as well as the 2-sided p-value will be presented. When the negative binomial regression does not converge or the parameters cannot be estimated, an exact Poisson regression will be applied to calculate the rate ratio, 95% CI and the p-value.

Fisher's exact test will be used for patient disposition, adverse events (AEs), discontinuations, and other categorical safety data for between-treatment group comparisons. Continuous vital signs, body weight, and other continuous safety variables including laboratory variables will be analyzed by ANCOVA with treatment and baseline value in the model. Shift tables for select categorical safety analyses (for example, "high" or "low" laboratory results) will also be produced.

For efficacy, if study drug is temporarily interrupted (suspended) during a treatment period due to non-COVID related interruptions, such as, AEs or abnormal laboratory measures, etc., the measurements taken during the temporary interruption will be included in the analysis. For COVID related interruptions, Table JAIA.6.5 provides details of inclusion/exclusion in the different analysis.

For safety evaluations, if study drug is temporarily interrupted (suspended) during a treatment period due to either COVID or non-COVID related interruptions, the measurements taken during the temporary interruption will be included in the analysis.

Missing data methods for the above analysis methods are specified in Section 6.4.2.

6.4.1 Adjustments for Covariates

Randomization is stratified according to disease activity (total SLEDAI-2K at screening <10; ≥10), corticosteroid dose at baseline (<10 mg/day; ≥10 mg/day prednisone or equivalent), and region (North America, Central/South America/Mexico, Europe, Asia and Rest of World). Therefore, these factors will be adjusted for as described in Section 6.4.

6.4.2 Handling of Dropouts or Missing Data

Depending on the estimand being addressed, different data cutoffs and imputation methods will be used to handle intercurrent events. Table JAIA.6.5 and Table JAIA.6.6 specify how data will be included and the associated imputation method for each analysis. Intercurrent events (ICEs) for Study JAIA include:

- Initiation/increase of prohibited medication
- Permanent study drug discontinuation (COVID or non-COVID reasons)
- Study discontinuation (COVID or non-COVID reasons)

- Intermittent missing visits or assessment (COVID or non-COVID reasons), and
- Temporary treatment interruptions due to COVID reasons.

6.4.2.1 Non-responder Imputation (NRI)

The non-responder imputation (NRI) method may be used when the estimand of interest uses the composite strategy (EMA 2020) for handling intercurrent events. In this strategy, patients will be considered non-responders for the NRI analysis if they do not meet the clinical response criteria or have missing clinical response data (based on each set of Table JAIA.6.5) at the analysis time point.

6.4.2.2 Mixed Model for Repeated Measures

For continuous variables, the primary analysis will be an MMRM analysis with a missing at random assumption for handling missing data. This likelihood-based analysis takes into account both missingness of data and the correlation of the repeated measurements.

The MMRM method may be justified when the estimand of interest uses the hypothetical strategy (EMA 2020) for handling intercurrent events. In this strategy, the scientific question of interest is to assess the effect of study treatment in a hypothetical trial where all patients have complete data and continue to take study treatment as directed without dropping out of the study.

6.4.2.3 Modified Last Observation Carried Forward (mLOCF)

Modified last observation carried forward will be used to impute missing data for analyses implementing ANCOVA or Wilcoxon rank-sum in the PP datapoint sets (Table JAIA.6.5), unless otherwise specified. For patients who receive prohibited medication, the last nonmissing observation at or before prohibited use will be carried forward to subsequent time points for evaluation. For all other patients discontinuing from the study or permanently discontinuing the study treatment for any reason, the last nonmissing postbaseline observation on or before discontinuation will be carried forward to subsequent time points for evaluation. If a patient has a missing record that cannot be imputed by other means for a postbaseline visit, the last postbaseline record prior to the missed visit will be used for this postbaseline visit.

6.4.2.4 Hybrid Imputation

Hybrid imputation will be conducted to adjust for the effect of missing data due to the COVID-19 pandemic on the clinical trial. Missing data due to the COVID-19 pandemic includes: 1) data collected during drug interruption due to COVID-19 while a patient is considered off-drug; or 2) data which were not collected due to the COVID-19 pandemic (that is, missing visit data, in part or in whole). For the selected binary and continuous endpoints (see Table JAIA.6.6), multiple imputation (MI) will be used for imputing data that is missing due to COVID-19 in the PPCV datapoint sets (see Table JAIA.6.6, PPCV definition), whereas other missing data not due to COVID-19 will be imputed by NRI for the binary endpoints and by MMRM for the continuous endpoints.

This hybrid imputation procedure addresses the hybrid estimand assuming that the effects of treatments would be the same as the PP analysis, had patients not experienced any intercurrent

event related to COVID-19 (see Section 6.4.2). The algorithm for hybrid imputation on binary and continuous endpoints (see Table JAIA.6.4) is as follows:

1. Identify all missing data, either due to COVID-19 or otherwise. Note: data following ICEs described in Section 6.4.2, will be set to missing prior to Step 2. Any data after prohibited medication or treatment/study discontinuation due to non-COVID-19 reasons will be all considered as missing due to non-COVID-19 reasons.
2. Implement the MI via PROC MI FCS REG or PROC MI FCS LOGISTIC to impute all continuous or binary missing data, respectively with treatment arm and stratified variables as covariables and generate *m* imputed complete data sets.
3. For each of these *m* imputed complete data sets from Step 2, the imputed data for missing not due to COVID-19 (as in Step 1) will be set to missing and the imputation method specified in Table JAIA.6.6 will be implemented. For all other data including imputed or observed will be used for the analysis or to derive the binary outcomes.
4. Single estimates will be combined using Rubin's rules (Rubin 1987) via PROC MIANALYZE.

The number of imputed data sets will be m=25 and a 6-digit initial seed value prespecified for each analysis as 123456.

For the composite binary endpoints such as SRI-4, BICLA and Lupus Low Disease Activity State (LLDAS) (see Table JAIA.6.4), the hybrid imputation algorithm will be applied on each binary component first. For each binary component, multiple imputation may be applied on the continuous variable and then dichotomized if needed.

For the corticosteroid sparing endpoint in the PPCV datapoint set, mLOCF will be conducted to handle the missing data for corticosteroid daily dose due to COVID-19 whereas missing data due to other reasons will be handled by NRI. The algorithm for hybrid imputation on Corticosteroid Sparing is as follows:

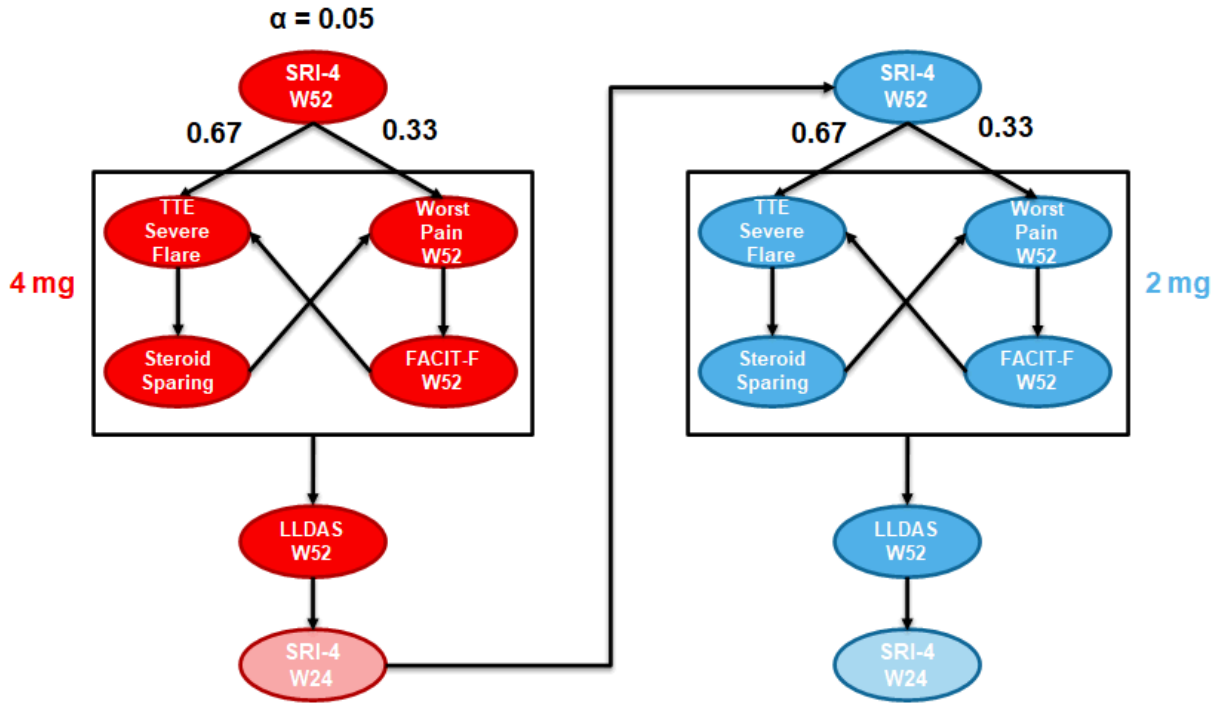
1. Identify all missing data, either due to COVID-19 or otherwise. Note: data following ICEs described in Section 6.4.2, will be set to missing prior to Step 2.
2. Implement the mLOCF to impute the missing data due to COVID-19.
3. The missing data not due to COVID-19 (as in Step 1) will be set to missing and the NRI method will be implemented. For all other data including imputed or observed will be used to derive the binary outcomes.

6.5 Multicenter Studies

For the analysis of the primary endpoint, assessment of the impact of regions will be specified within the integrated efficacy analysis plan.

6.6 Multiple Comparisons/Multiplicity

A prespecified graphical multiple testing approach (Bretz et al. 2009, 2011) will be implemented to control the overall Type I error rate at 2-sided alpha of 0.05, for all primary and major secondary endpoints. The graphical approach is a closed testing procedure; hence, it strongly controls the familywise error rate across all endpoints (Alosh et al. 2014). Details of the graphical testing scheme (including testing order, interrelationships, Type I error allocation, and the associated propagation) are included below (see Figure JAIA.6.1). Each hypothesis is represented as a node in a graph. Directed arrows between the nodes with associated weights represents how alpha is passed from its initial allocation to other nodes. The testing scheme was finalized before the first unblinding of the study team to efficacy data. Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses.



Abbreviations: FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; LLDAS = Lupus Low Disease Activity State; SRI-4 = Systemic Lupus Erythematosus Responder Index-4; TTE = time to event; W = week.

Note: All tests are compared to placebo. All hypothesis tests contained within a box must be rejected before alpha is passed on.

Figure JAIA.6.1. Illustration of graphical multiple testing procedure with initial α allocation and weights.

6.7 Patient Disposition

Patient flow, screen failures and reason for screen failure will be summarized for the entered population. Treatment and study disposition will also be summarized for the mITT population. Included in the reasons for treatment or study discontinuation will be a category for

discontinuations due to COVID-19 related logistical issues (i.e., quarantine or travel restrictions, non-illness related). For patients who discontinued the treatment or study due to COVID-19 infection, the reason for treatment discontinuation will be classified as "due to adverse events (AEs)" even when the discontinuation was due to the requirement of quarantine.

Patient disposition will be summarized with reasons for discontinuation by treatment group. Treatment groups will be compared using the Fisher's Exact test. Summary of patient disposition will include all periods.

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.

Patient allocation by region, country, and center/site will be summarized along with disposition information. A listing of randomization/treatment assignment will be provided.

6.7.1 Patient Characteristics

Patient demographics and baseline characteristics will be summarized using the mITT population, by treatment group. No formal statistical comparisons will be made among treatment groups, unless otherwise stated. A listing of patient demographics will be provided.

Appendix 2 itemizes the specific baseline measures and patient characteristics to be presented and how they will be summarized. Changes to Appendix 2, including the summary of additional patient characteristics, will not require an amendment to the SAP.

Pre-existing conditions are defined as conditions with a start date prior to the first dose of the study drug and stop dates that are at or after the informed consent date or have no stop date (that is, are ongoing). The number and percentage of patients with pre-existing conditions using MedDRA preferred term (PT) nested within system organ class (SOC) will be summarized with treatment comparisons.

6.8 Treatment Compliance

Compliance with investigational product treatment for baricitinib and placebo will be assessed through counts of returned investigational product tablets. A patient will be considered significantly noncompliant if he or she takes <80% of the prescribed doses. If a patient had his/her treatment temporarily interrupted by the investigator due to any reason, the number of days that drug was withheld or unavailable will be deducted from the total number of days in computing the expected total number of tablets used. Compliance for patients without available data due to early termination or lost to follow-up will be based on last completed visit with available tablet count.

$$\text{Compliance} = \frac{\text{total number of tablets dispensed} - \text{total number of tablets returned}}{\text{expected number of total tablets used}}$$

Patient treatment compliance will be summarized for the mITT population at the Week 52 visit.

6.9 Concomitant Therapy

Concomitant medications will be coded and descriptively summarized by treatment group in terms of frequencies and percentages using the mITT population.

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 electronic case report form [eCRF]) to allow medications to be classified as "previous" or "concomitant."

Medications that started before the last date of treatment period 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 of study treatment will be classified as "previous" medication.

For Zoster and tuberculosis (TB) vaccine, 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 name, sorted by frequency in baricitinib 4-mg group. The summaries of concomitant medications will be provided overall and separately for concomitant medications used to treat SLE and for statin concomitant medications. Previous medications used to treat SLE will also be summarized. Note that a patient will only be counted once, regardless of how many times medication included under the same preferred name was taken.

For anti-malarials, immunosuppressants, corticosteroids, and other medications, baseline is defined as the corresponding medication and dose taken on the day prior to the date of the first dose of study drug administration.

Patients who have any increase in antimalarials, immunosuppressants, or corticosteroids during the treatment period will have their concomitant medication records reviewed by the study team (blinded to study treatment) to assess clinical significance of change in medication. The final determination of clinical meaningfulness will be documented and will be incorporated into the analysis datasets.

Patients who, after medical review, have medication increases that are deemed to be non-clinically significant will not be considered automatic non-responders for PPCV and PP data points set analyses.

Patients with prohibited use of concomitant medications will be summarized by type of prohibited change (e.g., addition or increase in dose of corticosteroids, immunosuppressants, or antimalarials).

A summary of permitted steroid bursts will be provided by treatment arm.

Patients are expected to maintain their usual medication regimen for SLE and for any other concomitant disease throughout the study unless otherwise instructed in the protocol as specified in JAIA Protocol Section 6.2. Table JAIA.6.2 provides an abbreviated summary of the protocol-specified concomitant medication rules.

Table JAIA.6.2. Abbreviated Concomitant Medication Rules

Permitted Concomitant Medication			
Timing	**Drug**	**Dose**	**Rule**
At baseline	Single antimalarial	Stable per protocol	At least 8 weeks prior to screening
	Single immunosuppressant	Stable per protocol	At least 8 weeks prior to screening
	Oral corticosteroid	Stable per protocol	Initiated at least 4 weeks prior to screening (V1), stable at (≤ 40 mg/day) for at least 2 weeks prior to screening through baseline
		≥ 7.5 mg/day	Required if no antimalarial or immunosuppressant
Post baseline	Corticosteroid	Steroid Burst	For Oral: A single maximum increase (or initiation) of 40 mg/day prednisone over baseline dose; total maximum daily dose not to exceed 60 mg/day prednisone or equivalent; for up to 7 consecutive days during week 0 to 12. Patient must be at or below baseline dose by the end of the 7-day period or end of 12 week period (Visit 6), whichever is earlier. • Alternatively to this increase, patients may receive a single intramuscular dose of methylprednisolone 80-mg or less, or equivalent, between Week 0 and Week 11.
	Corticosteroid	Decrease from baseline dose	Allowed between Week 0 and Week 16 and from Week 24 and Week 40 visit. • Increase allowed during taper by maximum of 10 mg/day or up to baseline dose, whichever is lower. Corticosteroid may not exceed baseline dose in this case.
Not permitted[a]			
Post-baseline	Corticosteroids	Increase from baseline	Any increase from baseline during the treatment period other than the steroid burst
	Antimalarial	Initiation or increase	Any increase or initiation is prohibited during treatment period
	Immunosuppressant	Initiation or increase	Any increase or initiation is prohibited during treatment period
	Other biologics or protocol-prohibited concomitant medications[b]	Initiation	Any initiation is prohibited during treatment period

Abbreviations: PPCV = Per Protocol, COVID free; PP = Per Protocol.

[a] Use of these medications will be reviewed by the study team (blinded to study treatment) on a patient-by-patient basis to assess the clinical significance of the change in medication. After review, if deemed to be non-clinically significant, the patient will not be considered a non-responders for PPCV and PP data points set analyses.

[b] Defined from the ATC codes of L01XC, L01XX, L04AB, L04AC, V98, or J06B (excluding J06BB).

For corticosteroids, all doses of steroids will be converted to an equivalent dose of prednisone (see Section 6.9.1). Using the prednisone-equivalent dose, a daily dose of corticosteroid will be derived for each study day from baseline to the study discontinuation day for patients who discontinue study early or to the day of the Week 52 visit for study completers.

For immunosuppressant and antimalarial medications, a daily dose will be calculated for each medication by preferred name.

6.9.1 Corticosteroid Dosing Conversion

To allow for assessments of changes in doses of various corticosteroids, all corticosteroid doses will be standardized to an equivalent prednisone dose. Table JAIA.6.3 provides a summary of frequent corticosteroids and their prednisone-equivalent dose. Given intramuscular injections (IM) are slow release, such injections will be divided by 7 and distributed across 7 days. This dose of corticosteroid will be referred to as "prednisone (or equivalent)" throughout this document.

Table JAIA.6.3. Conversion Factors for Calculating Prednisone (or Equivalent) Doses

Column 1	Column 2
Corticosteroid Preferred Name	**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.

Baseline prednisone (or equivalent) dose will be the total daily dose of all corticosteroids being taken by a patient on the day prior to the date of the first dose of study drug administration. This baseline dose will be used for baseline summaries and for comparisons to later visits. The corticosteroid dose for a postbaseline visit will be the daily dose of prednisone (or equivalent) at each visit date -1.

6.10 Efficacy Analyses

6.10.1 General Considerations

Table JAIA.6.4 includes the description, derivation and definition of missingness for efficacy and health outcomes measures and endpoints.

Table JAIA.6.5 summarizes the data to be included/excluded in each analysis specified in Table JAIA.6.6.

Table JAIA.6.6 includes the description of the analysis method, population and primary time point(s) of analysis for efficacy and health outcomes measures. Assessments collected at multiple visits will be analyzed at each visit, in addition to the primary time point. Details of each analysis will follow the general principles described under Analysis Methods in Section 6.4.

Table JAIA.6.4. Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
SRI-4	SRI-4 is a composite index to measure overall improvement in disease activity (SLEDAI-2K) while ensuring there is no worsening in other organ systems (BILAG and PGA)	SLE Responder Index 4	• A decrease in SLEDAI-2K ≥4 (from baseline), and • No new BILAG A and no more than 1 new BILAG B disease activity score / organ domain (both compared with baseline), and • No worsening in PGA (defined as an increase of 0.3 points [10 mm] from baseline).	Missing if any component (SLEDAI, BILAG or PGA) remains missing after each instrument level imputation rule is applied.
SLEDAI-2K	The SLEDAI-2K is a global disease activity instrument that focuses on high-impact disease manifestations across 9 organ domains: constitutional, mucocutaneous, musculoskeletal, vascular, cardiorespiratory, central nervous system, immunologic, renal, and hematologic.	SLEDAI-2K Total Score	Calculated by summing the weighted organ manifestation. For Renal domain, when the urinary casts, hematuria and pyuria items are "not done" or "not accessed", they will be coded as "not present".	If any or all item scores are missing (not done or not assessed), but the visit occurred, data for that item can be carried forward if obtained within previous 36 days of visit.
		Individual Organ Domain Improvement Defined by SLEDAI-2K	Patients with SLEDAI-2K score >0 within the organ domain at baseline, and able to decrease SLEDAI-2K (from baseline) within each organ domain score, separately.	Missing if any SLEDAI-2K item for that organ domain remains missing after instrument level imputation rules are applied
		Individual Organ Domain Worsening Defined by SLEDAI-2K	Among patients with at least one SLEDAI-2K item =0 in the organ domain score at baseline and able to increase SLEDAI-2K (from baseline) within each organ domain score, separately.	Missing if any SLEDAI-2K item for that organ domain remains missing after instrument level imputation rules are applied.
		No worsening from baseline in SLEDAI-2K	No increment from baseline of >0 points in SLEDAI-2K	Missing if any SLEDAI-2K item for that organ domain remains missing after instrument level imputation rules are applied.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
BILAG-2004	BILAG-2004 assesses 97 clinical signs, symptoms and laboratory parameters across nine organ system domains: constitutional, mucocutaneous, neuropsychiatric, musculoskeletal, cardiorespiratory, gastrointestinal, ophthalmic, renal and hematological	BILAG	A, B, C, D, or E score will be used in analyses for each of the 9 individual organ systems.	Within each organ domain (except renal and hematological), any missing data will be assumed to be 'Not present' if there is at least 1 non-missing item in that organ. If all items in one organ domain are completely missing but the visit occurred, then the letter score of that organ from the previous visit will be pulled forward, provided data were obtained within 36 days of visit; otherwise missing except when the letter score from the last non-missing visit is E, then E score will be pulled forward. For renal and hematological domain, if any item is missing but there is at least 1 non-missing item in the organ, the missing data will be assumed to be "No". If any vital/lab item with "Yes" but vital and lab value is missing, then vital/lab values will be pulled forward, provided data were obtained within 36 days of visit; otherwise missing. After the value imputation, if the A/B/C score for the organ domain still cannot be derived, then the letter score of that organ from the previous visit will be pulled forward, provided data were obtained within 36 days of visit; otherwise missing except when the letter score from the last non-missing visit is E, then E score will be pulled forward.
		BILAG improvement	Reduction of all baseline BILAG A to B/C/D and baseline BILAG B to C/D and no BILAG worsening in other organ systems, where worsening is defined as ≥1 new BILAG A or ≥2 new BILAG B	Missing if any of 9 domains at baseline or at the visit remains missing after instrument level imputation rules are applied except when the missing is at the baseline (any one or all domains) and the value at the visit is not A or B, then the other non-missing organ domains will be used to determine the response status.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
		No BILAG worsening	No new BILAG A and no more than 1 new BILAG B disease activity score (both compared with baseline), where worsening is defined as ≥1 new BILAG A or ≥2 new BILAG B, both compared with the baseline. The baseline BILAG A improved to BILAG B at the visit will not be considered as the new BILAG B.	Missing if any of 9 domains at baseline or at the visit remains missing after instrument level imputation rules are applied except when the baseline (any one or all domains) is BILAG A and the value at the visit is missing, then the other non-missing organ domains will be used to determine the response status.
		Individual Organ Domain Improvement Defined by BILAG	Among patients with BILAG A or B at baseline and able to reduce the baseline BILAG A to B/C/D and BILAG B to C/D for each organ domain, separately.	Missing if baseline or value remains missing after instrument level imputation rules are applied.
		Individual Organ Domain Worsening Defined by BILAG	Among patients without BILAG A at baseline, an increment of baseline BILAG B/C/D/E to A or baseline BILAG C/D/E to B for each organ domain, separately.	Missing if baseline or value remains missing after instrument level imputation rules are applied.
PGA	The PGA is the physician's assessment of the patient's overall disease activity due to SLE. It is scored using a visual analog scale where 0 mm indicates no disease activity and 100 mm indicates the most severe disease activity possible. There are benchmarks of 0 (0 mm), 1 (33 mm), 2 (67 mm), and 3 (100 mm) on the line corresponding to no, mild, moderate, and severe SLE disease activity, respectively.	PGA score PGA category	Permitted range of values is from 0 to 100 mm. PGA categories are defined as: None (0) = '0 mm', Mild (>0 and <1.5) = ' >0 to < 50 mm', Moderate (≥1.5 to ≤2.5) = '≥ 50 mm to ≤83mm', Severe (>2.5) = '>83mm'.	If the visit occurred, data can be carried forward if obtained within 36 days of visit; otherwise missing.
		No worsening in PGA	Worsening is defined as an increase of ≥0.3 points (10 mm) from baseline. Therefore, no worsening is defined as any decrease, no change, or <0.3 points (10 mm) increase from baseline.	Missing if baseline or value remains missing after instrument level imputation rules are applied.
		PGA ≤1	PGA <=33 mm	Missing if value remains missing at the visit after instrument level imputation rules are applied.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
BICLA	BICLA is a composite index to measure overall improvement in disease activity (BILAG) while ensuring there is no worsening in other organ systems (SLEDAI and PGA).	BILAG Based Composite Lupus Assessment	• Reduction of all baseline BILAG A to B/C/D and baseline BILAG B to C/D and no BILAG worsening in other organ systems, where worsening is defined as ≥1 new BILAG A or ≥2 • New BILAG B • No worsening from baseline in SLEDAI-2K, where worsening is defined as an increase of >0 points from baseline in SLEDAI-2K • No worsening from baseline in participants' lupus disease activity, where worsening is defined by an increase ≥0.30 points on a 3-point • PGA VAS.	Missing if any component (BILAG, SLEDAI or PGA) remains missing after each instrument level imputation rule is applied.
CLASI	The CLASI is a validated instrument to assess cutaneous manifestations of SLE consisting of 2 scores: • Activity • Damage	Total Activity Score	Calculated within tablet system with edit checks; no additional derivation.	Within tablet system: If any item scores, but not all are missing then impute a score of 0 for the missing item(s). For completely missing questionnaires, data can be carried forward if obtained within 36 days of visit; otherwise missing.
		Total Damage Score	Calculated within tablet system with edit checks; no additional derivation.	Within tablet system: if any item scores, but not all are missing then impute a score of 0 for the missing item(s). For completely missing questionnaires, data can be carried forward if obtained within 36 days of visit; otherwise missing.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
LLDAS	The LLDAS is a composite measure designed to identify patients achieving a state of low disease activity.	LLDAS	• SLEDAI-2K ≤ 4, with no activity in SLEDAI-2K major organ systems (CNS, Vascular, Renal, Cardiorespiratory and Constitutional), where "no activity" is defined as all items of SLEDAI-2K within these major organ systems equal to 0. • No new features of Lupus disease activity in SLEDAI-2K compared to previous occurred visit, where the "new feature" is defined as any of the SLEDAI-2K 24 items changed from 0 to greater than 0. • PGA ≤1 • Current prednisone or equivalent ≤7.5 mg/day	Missing if any component remains missing after each instrument level imputation rule is applied.
SFI Flare	The SFI uses the SLEDAI score, disease activity scenarios, treatment changes, and PGA to define mild/moderate and severe flares.	SFI Flare	Not derived; used as entered.	The absence of a flare record, or 'Not Applicable', both are indicative of no occurrence of flare.
		Annualized Flare Rate	Calculated as the number of SFI flares divided by the flare exposure time in days multiplied with 365.25.	No minimum data requirement.
		Time to First Flare	The time to first flare will be derived as the first date of most recent flare minus date of first administration of investigational product. If the subject did not have a flare, the time to first flare will be censored at the end of the flare exposure time	No minimum data requirement.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
Tender/Swollen Joint Count (28 Joints)	28 joints are assessed as tender only, swollen only, tender and swollen, not evaluable.	Tender Joint Count 28/ Swollen Joint Count 28	The number of tender and swollen joints will be calculated by summing all joints respectively. Other rules apply. (See Appendix 4).	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.
SLICC/ACR	SLICC/ACR Damage Index assesses damage to 12 organ systems regardless of its cause.	SLICC/ACR Damage Index Score	Calculated as the sum of all available entries with a maximum score of 47.	"Missing" item is coded as 0.
Worst Pain NRS	This Worst Pain NRS is a single-item patient-administered 11-point horizontal scale anchored at 0 (no pain) and 10 (pain as bad as you can imagine).	Worst Pain NRS score	Calculated by averaging data for all available daily diary entries for a 7-day period prior to the visit, not inclusive of the clinic visit day	Missing if fewer than 4 available daily measurements in the relevant 7 days
FACIT-Fatigue	The FACIT-Fatigue scale (Cella and Webster 1997) is a 13-item symptom-specific questionnaire that assesses self-reported severity of fatigue and its impact upon daily activities and functioning.	Facit-Fatigue total score	The FACIT-Fatigue uses 0 ("not at all") to 4 ("very much") numeric rating scales to assess fatigue and its impact in the past 7 days. Scores range from 0 to 52 with higher scores indicating less fatigue. The FACIT-Fatigue Scoring Guidelines (Version 4) will be used to calculate the Total Score. Reversals are needed for all items except An5 and An7, as described in scoring manual. http://www.ser.es/wp-content/uploads/2015/03/FACIT-F_INDICE.pdf.	Missing items are acceptable as long as more than 50% of the items are answered (that is, a minimum of 7 out of 13 items), the sum of available items will be multiplied by 13, then divided by the number of items answered to obtain the total score. If less than 7 items are answered, the FACIT-Fatigue total score will be set to missing.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
Steroid sparing	This measures the proportion of patients who, having a high enough dose at baseline, are able to reduce steroids to a certain level.	Receiving >7.5 mg prednisone (or equivalent) at baseline and able to decrease dose by ≥25% to a prednisone equivalent dose of ≤7.5 mg/day maintained between the Week 40 visit and the Week 52 visit	Dose reduction must be greater than or equal to baseline dose x .25 and achieve a dose of less than or equal to 7.5 mg per day perpetually between the Week 40 visit and Week 52 visit.	No minimum data requirement.
		Receiving >7.5 mg prednisone (or equivalent) at baseline and able to decrease dose by ≥25% by the Week 52 visit	Dose reduction must be greater than or equal to baseline dose x .25 per day perpetually between the Week 40 visit and the Week 52 visit.	No minimum data requirement.
		Receiving >7.5 mg prednisone (or equivalent) at baseline and able to decrease to a prednisone equivalent dose of ≤7.5 mg/day by the Week 52 visit	Dose reduction must be less than or equal to 7.5 mg per day perpetually between the Week 40 visit and the Week 52 visit.	No minimum data requirement.
		Receiving prednisone (or equivalent) at baseline and able to discontinue and maintain through the Week 52 visit	Among patients with baseline prednisone (or equivalent) dose > 0mg and able to reduce the dose to 0mg at some time point and then maintained at 0mg from that time through the Week 52 visit.	No minimum data requirement.
Worst Joint Pain NRS	This Worst Joint Pain NRS is a single-item, patient-administered 11-point horizontal scale anchored at 0 (no joint pain) and 10 (joint pain as bad as you can imagine)	Worst Joint Pain NRS Score	Calculated by averaging data for all available daily diary entries for a 7-day period prior to the visit, not inclusive of the clinic visit day.	Missing if fewer than 4 available daily measurements in the relevant 7 days

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
Worst Fatigue NRS	This Worst Fatigue NRS is a patient-administered 11-point horizontal scale anchored at 0 (no fatigue) and 10 (fatigue as bad as you can imagine)	Worst Fatigue NRS Score	Calculated by averaging data for all available daily diary entries for a 7-day period prior to the visit, not inclusive of the clinic visit day.	Missing if fewer than 4 available daily measurements in the relevant 7 days
Joint Stiffness Duration	This patient-administered scale captures information on the self-reported length of time, in minutes, that a patient's joint stiffness lasted each day.	Joint Stiffness Duration	Calculated by averaging data for all available daily diary entries for a 7-day period prior to the visit, not inclusive of the clinic visit day.	Missing if fewer than 4 available daily measurements in the relevant 7 days
Joint Stiffness Severity NRS	This Joint Stiffness Severity NRS is a single-item patient-administered 11-point horizontal scale anchored at 0 (no joint stiffness) and 10 (joint stiffness as bad as you can imagine)	Joint Stiffness Severity NRS Score	Calculated by averaging data for all available daily diary entries for a 7-day period prior to the visit, not inclusive of the clinic visit day.	Missing if fewer than 4 available daily measurements in the relevant 7 days
Patient Global Impression of Severity	The Patient's Global Impression of Severity is a single-item question asking the patient how they would rate their overall lupus symptoms over the last 24 hours.	Patient Global Impression of Severity Score	Possible values are: (0) No symptoms, (1) Very mild, (2) Mild (3) Moderate (4) Severe	Missing if fewer than 4 available daily measurements in the relevant 7 days
Patient's Global Impression of Change - Lupus	The Patient's Global Impression of Change is a single-item question asking the patient how they would rate their overall lupus symptoms since the start of the study.	Patient's Global Impression of Change Score	Possible values are: (1) Very much better (2) Much better (3) A little better (4) No change (5) A little worse (6) Much worse (7) Very much worse	Single item; missing if missing.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
SF-36 v2 Acute	The SF-36 Version 2 acute (1-week recall) is a 36-item, patient-completed measure designed to be a short, multipurpose assessment of health (The SF Community – SF-36 Health Survey Update). The summary scores range from 0 to 100, with higher scores indicating better levels of function and/or better health. Items are answered on Likert scales of varying lengths. The SF-36 comprises 8 domain scores and 2 overarching component scores. SF-36 domain scores are: (1) Physical functioning, (2) Role-physical, (3) Role-emotional, (4) bodily pain, (5) vitality, (6) social functioning, (7) mental health and (8) general health. The component scores are: (1) the Physical Component Summary (PCS) and (2) Mental Component Summary (MCS).	SF-36 Domain scores and SF-36 Component Scores	Per copyright owner, the Quality Metric Health Outcomes™ Scoring Software will be used to derive SF-36 domain and component scores. SF-36V2 Normed-Based Scoring will be used for the domain and component scores and includes 2 steps: 1/ z-score standardization of SF-36v2 scales; and 2/ norm-based transformation of the SF-36v2 z scores. 2009 norms will be used. (Maruish, 2011).	Missing data handling offered by SF-36 software will be used. Maximum Data Recovery will be selected for Missing Score Estimator in the software.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
EQ-5D-5L	The European Quality of Life–5 Dimensions–5 Level (EQ-5D-5L) is a standardized measure of health status used to provide a simple, generic measure of health for clinical and economic appraisal. The EQ-5D-5L consists of 2 components: a descriptive system of the respondent's health and a rating of his/her current health state using a 0- to 100-mm VAS. The descriptive system comprises the following 5 dimensions: Item 1: mobility Item 2: self-care Item 3: usual activities Item 4: pain/discomfort Item 5: anxiety/depression The respondent is asked to indicate his/her health state by ticking (or placing a cross) in the box associated with the most appropriate statement in each of the 5 dimensions. It should be noted that the numerals 1 to 5 have no arithmetic properties and should not be used as an ordinal score. The VAS records the respondent's self-rated health on a vertical VAS where the endpoints are labeled "best imaginable health state" and "worst imaginable health state." This information can be used as a quantitative measure of health outcome. The EQ-5D-5L health states, defined by the EQ-5D-5L descriptive system, may be converted into a single summary index by applying a formula that essentially attaches values (also called weights) to each of the levels in each dimension.	EQ-5D-5L Items	Five health profile dimensions, each dimension has 5 levels: 1 = no problems 2 = slight problems 3 = moderate problems 4 = severe problems 5 = extreme problems It should be noted that the numerals 1 to 5 have no arithmetic properties and should not be used as a primary score.	Each dimension is a single item, missing if missing. Note: score of 9 is missing.
		EQ-5D-5L UK/US Population-based index scores	Uses the concatenation of the value of each EQ-5D-5L dimension score in the order: Item 1, Item 2, Item3; Item 4; Item 5. Derive EQ-5D-5L UK/US Population-based index scores according to the link by using the UK/US algorithm (Szende et al. 2006) to produce a patient-level index score between -0.59 and 1.0 (continuous variable): https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/crosswalk-index-value-calculator/	If any of the items is missing or equal to 9, the index score is missing
		EQ-5D VAS	Single item. Range from 0 = "worst imaginable health state" to 100 = "best imaginable health state". Note: higher value indicates better health state.	Single item, missing if missing

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation/Comment	Definition of Missing
WPAI-Lupus	The Work Productivity and Activity Impairment – Lupus (WPAI:Lupus) is a patient-reported instrument developed to measure the impact on work productivity and regular activities attributable to SLE. It contains 6 items that measure: 1) employment status, 2) hours missed from work due to the specific health problem, 3) hours missed from work for other reasons, 4) hours actually worked, 5) degree health affected productivity while working, and 6) degree health. Scores are calculated as impairment percentages (Reilly et al. 1993), with higher scores indicating greater productivity impairment.	Employment Status	Yes/No	Missing if answer is missing
		Absenteeism Score (%)	$\frac{Q2}{(Q2+Q4)} \times 100$	Missing if Q2 or Q4 are missing. Also missing if Employment Status is No.
		Presenteeism Score (%)	$\frac{Q5}{10} \times 100$	Missing if Q5 is missing. Also missing if Employment Status is No.
		Work productivity Loss Score (%)	$\left[\frac{Q2}{Q2+Q4} + \left(1 - \frac{Q2}{Q2+Q4}\right)\frac{Q5}{10}\right] \times 100$	Missing if Q2, Q4 or Q5 is missing. Also missing if Employment Status is No.
		Activity Impairment Score (%)	$\frac{Q6}{10} \times 100$	Missing if Q6 is missing. May still be present and non-missing if patient is unemployed.

Abbreviations: ACR = American College of Rheumatology Damage Index; BILAG = British Isles Lupus Activity Group; BICLA = British Isles Lupus Assessment Group Based Composite Lupus Assessment; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; CNS = Central Nervous System; EQ-5D-5L = European Quality of Life–5 Dimensions–5 Level; FACIT = Functional Assessment of Chronic Illness Therapy; LLDAS = Lupus Low Disease Activity State; MCS= Mental Component Score; NRS = Numeric Rating Scale; PCS= Physical Component Summary; PGA = Physician's Global Assessment of Disease Activity; SF-36 = Short Form 36; SFI = SLEDAI Flare Index; SLE = Systemic Lupus Erythematosus; SLICC = Systemic Lupus International Collaborating Clinics; SRI-4 = SLE Responder Index; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000; UK = United Kingdom; US = United States; VAS = visual analogue scale; WPAI:Lupus = Work Productivity and Activity Impairment – Lupus.

Note: Additional supplemental and sensitivity analyses are described in Section 6.10.3.3.

The datapoint sets are defined to estimate the estimands in Table JAIA.6.5. The mITT population will be included in the analysis according to the treatment they were randomized, unless otherwise specified in Table JAIA.6.6.

Patients who are imputed as nonresponders due to a qualifying intercurrent event for binary categorical efficacy variables, will be defined as such from the time of qualifying event, plus 1 day, onward.

Table JAIA.6.5. Datapoint Sets for Use in Efficacy/Health Outcomes Analyses

Datapoint sets	Definition of Missing Data Due to ICEs	Estimand of Interest
PPCV	For patients experiencing the ICEs of treatment discontinuation, study discontinuation, or prohibited medication use, observations after such ICEs will not be included in analysis (that is, will be considered as missing). Data from patients experiencing treatment interruptions due to non-COVID related interruptions, such as, adverse events or abnormal labs, etc., will be included and used as is, per treatment regimen. Data from patients experiencing off-drug interruptions 20 or more days due to COVID-19 will not be included (that is, will be considered as missing) until treatment resumption of 20 days or more.	Addresses the clinically relevant question comparing the number of subjects able to both complete the study treatment and to achieve adequate response without further medication being required in a pandemic-free world. Targets treatment effect without potentially confounding influences of COVID-19.
PP	For patients experiencing the ICEs of treatment discontinuation, study discontinuation, or prohibited medication use, observations after such ICEs will not be included in analyses (that is, will be considered as missing). Data from patients experiencing treatment interruptions due to non-COVID related interruptions, such as, adverse events or abnormal labs, etc., will be included and used as is, per treatment regimen. Data from patients experiencing off-drug interruptions due to COVID-19 will be included and used as collected.	Addresses the clinically relevant question comparing the number of subjects able to both complete the study treatment and to achieve adequate response without further medication being required.
AC	For patients experiencing the ICEs of treatment discontinuation, prohibited medication use, or COVID-19 off-drug assessments, observations after such ICEs will be included in analyses until the study discontinuation or the date of the Week 52 visit, whichever earlier. The post-treatment follow-up period will not be included in the efficacy analysis.	Addresses the clinically relevant question comparing the number of subjects able to achieve adequate response, regardless of adherence and treatment compliance, in resemblance to real-world expectations.

Abbreviation: AC = as collected; ICE = intercurrent events; PP = per protocol; PPCV = per protocol COVID free.

Table JAIA.6.6. Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Analysis and Imputation Method	Datapoint Sets	TPI*	Analysis Type
SRI-4	Proportion of patients achieving SRI-4 response	Logistic Regression, NRI + MI	PPCV	Week 24 visit, Week 52 visit	Primary endpoint (Week 52); major secondary (Week 24)
		Logistic Regression, NRI	PP	Week 24 visit, Week 52 visit	Supplemental
		Logistic Regression, NRI	AC	Week 24 visit, Week 52 visit	Supplemental
		Logistic Regression, NRI + MI	PPCV, Per Protocol population	Week 52 visit	Supplemental
		Logistic Regression, NRI + MI	PPCV	Week 52 visit	Sensitivity (Tipping point analysis)
SRI-x (where x is 5, 6, 7, 8)	Proportion of patients achieving SRI-x response[a]	Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint
SLEDAI-2K	Change from baseline in total SLEDAI-2K through Week 52 visit	MMRM	PP	Week 52 visit	Secondary endpoint
	Proportion of patients achieving reduction of ≥4 points from baseline	Logistic Regression, NRI + MI	PPCV	Week 52 visit	Secondary endpoint
		Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint
	Proportion of patients with no worsening in SLEDAI-2K compared with baseline	Logistic Regression, NRI + MI	PPCV	Week 52 visit	Supplemental
		Logistic Regression, NRI	PP	Week 52 visit	Supplemental
	Proportion of patients with improvement in each SLEDAI-2K individual organ systems versus baseline[b]	Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint
	Proportion of patients with worsening in each SLEDAI-2K individual organ systems versus baseline[c]	Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint

Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Analysis and Imputation Method	Datapoint Sets	TPI*	Analysis Type
BILAG	Proportion of patients with no new BILAG A and no more than 1 new BILAG B disease activity score (compared with baseline)	Logistic Regression, NRI + MI	PPCV	Week 52 visit	Secondary endpoint
		Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint
	Proportion of patients with reduction of all baseline BILAG A to B/C/D and baseline BILAG B to C/D and no BILAG worsening in other organ systems, where worsening is defined as ≥1 new BILAG A or ≥2 new BILAG B	Logistic Regression, NRI+MI	PPCV	Week 52 visit	Supplemental
		Logistic Regression, NRI	PP	Week 52 visit	Supplemental
	Proportion of patients with improvement in each BILAG individual organ systems versus baseline[d]	Logistic Regression, NRI	PP	Week 52 visit	Exploratory endpoint
	Proportion of patients with worsening in each BILAG individual organ systems versus baseline[e]	Logistic Regression, NRI	PP	Week 52 visit	Exploratory endpoint
PGA	Proportion of patients with no worsening in PGA (compared with baseline)	Logistic Regression, NRI+MI	PPCV	Week 52 visit	Secondary endpoint
		Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint
	Change from baseline in PGA score over time through Week 52 visit	MMRM	PP	Week 52 visit	Secondary endpoint
BICLA	Proportion of patients achieving BICLA response	Logistic Regression, NRI + MI	PPCV	Week 52 visit	Secondary endpoint
		Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint
		Logistic Regression, NRI	AC	Week 52 visit	Secondary endpoint

Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Analysis and Imputation Method	Datapoint Sets	TPI*	Analysis Type
28 Tender/ Swollen Joint Count	Change from baseline in tender joint count	MMRM	PP	Week 52 visit	Secondary endpoint
	Change from baseline in swollen joint count	MMRM	PP	Week 52 visit	Secondary endpoint
CLASI Activity Score	The proportion of patients with ≥ 50% reduction from baseline in total activity score[f]	Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint
SFI	Annualized severe flare rate	Negative binomial regression	PPCV	Through Week 52 visit	Secondary endpoint
		Negative binomial regression	PP		Supplemental
		Negative binomial regression	AC		Supplemental
	Annualized mild/moderate flare rate	Negative binomial regression	PP	Through Week 52 visit	Secondary endpoint
	Annualized flare rate (any severity)	Negative binomial regression	PP	Through Week 52 visit	Secondary endpoint
	Time to first severe flare	Cox regression	PPCV	Through Week 52 visit	Major secondary (Principal inference)
			PP		Supplemental
			AC		Supplemental
	Time to first flare (any severity)	Cox regression	PP	Through Week 52 visit	Secondary endpoint
LLDAS	Proportion of patients who achieve a LLDAS	Logistic Regression, NRI + MI	PPCV	Week 52 visit	Major secondary (Principal inference)
		Logistic Regression, NRI	PP		Supplemental
		Logistic Regression, NRI	AC		Supplemental

Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Analysis and Imputation Method	Datapoint Sets	TPI*	Analysis Type
Corticosteroid Sparing	Proportion of patients receiving > 7.5 mg/day prednisone at baseline able to reduce prednisone (or equivalent) dose by ≥ 25% to a prednisone equivalent dose of ≤ 7.5 mg/day maintained between Week 40 and Week 52 visit[g]	Logistic Regression, NRI + mLOCF	PPCV	Week 52 visit	Major secondary (Principal inference)
		Logistic Regression, NRI	PP		Supplemental
		Logistic Regression, NRI	AC		Supplemental
	Proportion of patients receiving > 7.5 mg/day prednisone at baseline able to reduce prednisone (or equivalent) dose by ≥ 25% maintained between Week 40 and Week 52 visit[g]	Logistic Regression, NRI + mLOCF	PPCV	Week 52 visit	Supplemental
		Logistic Regression, NRI	PP		Supplemental
	Proportion of patients receiving > 7.5 mg/day prednisone at baseline able to reduce prednisone (or equivalent) dose to ≤ 7.5 mg/day maintained between Week 40 and Week 52 visit[g]	Logistic Regression, NRI + mLOCF	PPCV	Week 52 visit	Supplemental
		Logistic Regression, NRI	PP		Supplemental
	Change from baseline in prednisone dose	MMRM	PP	Week 52 visit	Secondary endpoint
	Proportion of patients able to discontinue corticosteroids and maintained through Week 52 visit[h]	Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint
Worst Pain NRS	Change from baseline in Worst Pain NRS	MMRM+MI	PPCV	Week 52 visit	Major secondary (Principal inference)
		MMRM	PP		Supplemental
		MMRM	AC		Supplemental

Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Analysis and Imputation Method	Datapoint Sets	TPI*	Analysis Type
FACIT-Fatigue	Change from baseline in FACIT-Fatigue total score	MMRM+MI	PPCV	Week 52 visit	Major Secondary (Principal inference)
		MMRM	PP		Supplemental
		MMRM	AC		Supplemental
Worst Fatigue NRS	Change from baseline in Worst Fatigue NRS	MMRM	PP	Week 52 visit	Secondary endpoint
Worst Joint Pain	Change from baseline in Worst Joint Pain NRS	MMRM	PP	Week 52 visit	Secondary endpoint
Joint Stiffness Duration	Change from baseline in Joint Stiffness Duration	Wilcoxon rank-sum with mLOCF	PP	Week 52 visit	Secondary endpoint
Joint Stiffness Severity	Change from baseline in Joint Stiffness Severity	MMRM	PP	Week 52 visit	Secondary endpoint
Patient Global Impression of Severity	Change from baseline in Patient Global Impression of Severity	MMRM	PP	Week 52 visit	Secondary endpoint
Patient Global Impression of Change	Change from baseline in Patient Global Impression of Change	ANOVA with mLOCF	PP	Week 52 visit	Secondary endpoint
SLICC/ACR	Change from baseline in SLICC/ACR	ANCOVA with mLOCF	PP	Week 52 visit	Exploratory endpoint
SF-36 V2 acute	Change from baseline in MCS	MMRM	PP	Week 52 visit	Secondary endpoint
	Change from baseline in physical component score (PCS)	MMRM	PP	Week 52 visit	Secondary endpoint
	Change from baseline in domain scores SF-36 health survey version 2 acute	MMRM	PP	Week 52 visit	Secondary endpoint
EQ-5D-5L	Change from baseline in EQ-5D-5L US/UK population-based index scores and VAS	MMRM	PP	Week 52 visit	Secondary endpoint
	Proportion of patients with no problem in each EQ-5D-5L item score	Logistic Regression, NRI	PP	Week 52 visit	Secondary endpoint

Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Analysis and Imputation Method	Datapoint Sets	TPI*	Analysis Type
WPAI-Lupus	Change from baseline in WPAI-Lupus[i]	MMRM	PP	Week 52 visit	Secondary endpoint

Abbreviations: AC = as collected; ACR = American College of Rheumatology Damage Index; ANCOVA = analysis of covariance; ANOVA = analysis of variance; BICLA = British Isles Lupus Assessment Group Based Composite Lupus Assessment; BILAG = British Isles Lupus Activity Group; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; EQ-5D-5L = European Quality of Life–5 Dimensions–5 Level; FACIT = Functional Assessment of Chronic Illness Therapy; LLDAS = Lupus Low Disease Activity State; MCS = mental component score; MI = multiple imputation; mITT = modified intent-to-treat; mLOCF = modified last observation carried forward; MMRM = mixed model for repeated measures; NRI = non-responder imputation; NRS = Numeric Rating Scale; PCS = Physical Component Summary; PGA = Physician's Global Assessment; PP = per protocol; PPCV = per protocol COVID free; SFI = SLEDAI Flare Index; SLE = Systemic Lupus Erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000; SF-36 = Short Form 36; SRI-4 = SLE Responder Index; SLICC = Systemic Lupus International Collaborating Clinics; TPI = time point of interest; UK = United Kingdom; US = United States; VAS = visual analogue scale; WPAI-Lupus = Work Productivity and Activity Impairment Questionnaire: Lupus V2.0.

[a] Population is a subset of mITT which includes patients with SLEDAI-2K score ≥ x at baseline (where x is 5, 6, 7, 8) for each SRI-x responder analysis.

[b] Population is a subset of mITT which includes patients with SLEDAI-2K score >0 within the organ domain at baseline.

[c] Population is a subset of mITT which includes patients with at least one SLEDAI-2K item = 0 in the organ domain at baseline.

[d] Population is a subset of mITT which includes patients with BILAG A or B in the organ domain at baseline.

[e] Population is a subset of mITT which includes patients without BILAG A in the organ domain at baseline.

[f] Population is a subset of mITT which includes patients with CLASI total activity score ≥10 at baseline.

[g] Population is a subset of mITT which includes patients with > 7.5 mg/day prednisone at baseline. The logistic regression model covariates include treatment groups, baseline disease activity (total SLEDAI-2K <10; ≥10) and region (North America, Central/South America/Mexico, Europe, Asia and Rest of World).

[h] Population is a subset of mITT which includes patients with > 0 mg/day prednisone at baseline.

[i] For absenteeism score, presenteeism score, and work productivity loss score, population is a subset of mITT which includes patients who had employment status of Yes at baseline.

* Assessments collected at multiple visits will be analyzed at each visit, in addition to the primary time point listed in the table.

6.10.2 Primary Outcome and Methodology

Analysis of the primary endpoint (SRI-4 at the Week 52 visit) is described in Table JAIA.6.4 and Table JAIA.6.6. The primary analysis will utilize logistic regression (as described in Section 6.4) with NRI and multiple imputation (as described in Section 6.4.2.1 and Section 6.4.2.4). The estimand of interest to assess the population-average effect (odds ratio) attributable to the medication to which patients are initially randomized, if no 'rescue' medications were available, in a pandemic-free environment. This estimand is free of confounding effects of increase in concomitant medication use (namely, increases in use of corticosteroids above the baseline dose other than the permitted corticosteroid burst/post-taper increase, and/or initiation of or an increase above baseline dose in immunosuppressant or antimalarial treatment).

6.10.3 Secondary Efficacy Outcomes & Methodology

6.10.3.1 Secondary Efficacy Analyses

The analysis of the major secondary endpoints is described in Table JAIA.6.4 and Table JAIA.6.6 and Section 6.4. The list of major secondary endpoints may be found Table JAIA.4.1. A multiple testing procedure will be utilized to control the FWER at the 0.05 significance level for principal inference of the primary endpoint and all major secondary endpoints (see Section 6.7).

6.10.3.2 Other Secondary Efficacy Analyses

The analysis of the other secondary endpoints is described in Table JAIA.6.6.

6.10.3.3 Supplementary and Sensitivity Analyses

Supplemental analyses based on all observed data will be provided for all primary and major (multiplicity controlled) endpoints:

- AC analyses will include all data collected from randomization through the Week 52 visit, regardless of treatment discontinuation, prohibited medication use or treatment interruption; this analysis does not address a pandemic-free estimand.
- PP analyses are planned to target the estimand not free from the confounding effects of the pandemic.

6.10.3.3.1 Tipping Point Analyses

Planned tipping point analyses will be conducted provided the associated nominal p-value is significant. An interpretation of clinical plausibility of the assumption underlying the tipping point will be provided. This analysis will be performed on SRI-4 at the Week 52 visit to assess the robustness of the primary analysis.

Based on the evaluation of baricitinib SLE Phase 2 data (Study JAHH), the adjusted odds ratios are expected to be similar to unadjusted odds ratios, and therefore the tipping point analyses will be unadjusted for covariates. In this analysis, patients who have not had a treatment failure, but are missing data at the Visit 16 (Week 52) time point, will have their response status changed from not achieving response to achieving response, in an iterative manner.

At each step of the analysis, one subject will be switched from not achieving SRI-4 response to achieving SRI-4 response and the analysis re-run. The results (statistical significance) will be presented in a grid where the x-axis and y-axis represent the number of subjects assumed to achieve SRI-4 in the placebo group and the given baricitinib treatment group, respectively. The region where the conclusion changes will be considered as the tipping point.

The grid will be divided in 3 regions, limited in the top by the expected number of subjects from the baricitinib treated arm that could have achieved SRI-4 had they had sufficient data and could have achieved SRI-4 response, and limited from the right by the number of subjects in the equivalent placebo group.

6.10.3.4 Exploratory Efficacy Endpoints

Analyses of exploratory efficacy endpoints will be described in a separate SAP.

6.11 Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

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

6.12 Health Outcomes/Quality-of-Life Analyses

Analysis of the health outcomes/quality-of-life measures are described in Table JAIA.6.6 and Section 6.4. The descriptions and high-level derivations are provided in Table JAIA.6.4. Additional psychometric analyses will be documented in a separate analysis plan.

6.13 Safety Analyses

The definition of baseline and post-baseline are described in Section 6.3.

Safety analyses will include data from first dose of the study treatment and patients will be analyzed according to the investigational product which they were randomized at Week 0 (Visit 2). Safety analyses will use the safety population defined in Table JAIA.6.1.

Safety topics that will be addressed include the following: AEs including TEAEs and SAEs, clinical laboratory evaluations, vital signs and physical characteristics, Columbia Suicide Severity Rating Scale (C-SSRS), the Self-Harm Supplemental Form, Quick Inventory of Depressive Symptomatology Self Report (QIDS-SR16), and safety in special groups and circumstances, including adverse events of special interest and investigational product interruptions.

For tables that summarize events (such as AEs, categorical lab abnormalities, shift to maximum severity), post-last dose follow-up data will be included. Follow-up data is defined as all data occurring up to and including 30 days after last dose of treatment. In general, for laboratory analytes and vital signs, planned (scheduled visits) and unplanned (unscheduled visits) measures are used for categorical analyses and planned (scheduled visits) measures are used for continuous by-visit analyses including change from last baseline to last postbaseline.

The planned safety analyses are consistent with compound level safety standards that are based on multiple sources, including company standards, internal and external subject matter experts,

and cross-industry initiatives (e.g., white papers produced by a PhUSE Computational Science Working Group [a collaboration with FDA and PhUSE], published in the PhUSE Deliverables Catalog at https://www.phuse.eu/white-papers). Descriptions of the safety analyses are provided in this SAP, however, additional details (such as handling of unscheduled visits, missing data, etc.) are in compound level safety standards.

The following statistical methods will be used for safety unless otherwise noted:

- Fisher's exact test will be used for treatment comparisons of proportions, and odds ratios with corresponding 95% CIs, where specified.
- Treatment differences in mean change for continuous measurements will be assessed using an ANCOVA model fitting "baseline" as a covariate. Type 3 sums of squares will be used.

Two-sided p-values are often reported yet should not be over-interpreted. For safety analyses, they are used primarily as a flagging mechanism.

Exposure-adjusted incidence rates (EAIR) will be calculated by treatment group. Incidence rate difference and ratio along with 95% CI may be provided for SAEs and AESIs.

The EAIR evaluating the incidence of a first event per 100 patient-years at risk (PYR) will be provided. Exposure will be calculated based on the analysis period defined as the treatment period plus up to 30 days off-drug follow-up time. Exposure time for a patient with an event will be counted up to the time of the start of event. Exposure time for a patient without an event will be censored at the end of the analysis period.

6.13.1 Extent of Exposure

Duration of exposure (in days) will be calculated as follows:

- Duration of exposure to investigational product is date of last dose of study drug – date of first dose of study drug +1.

Last dose of study drug is calculated as last date on study drug. See the compound-level safety standards for more details including imputation of missing last date on study drug.

Duration of exposure to study drug will be summarized using descriptive statistics (n, mean, standard deviation [SD], minimum, first quartile, median, third quartile, maximum). Cumulative exposure and duration of exposure will be summarized in terms of frequency counts and percentages by category and treatment group.

Exposure ranges will generally be reported in weeks using the following as a general guide:

- ≥4 weeks, ≥16 weeks, ≥ 24 weeks, ≥ 52 weeks
- >0 to <4 weeks, ≥4 weeks to <16 weeks, ≥16 weeks to <24 weeks, ≥24 to <52 weeks, and ≥52 weeks

Total patient-years (PY) of exposure to study drug will be reported for each treatment group for overall duration of exposure. Overall exposure will be summarized in total PY which is calculated according to the following formula:

Total patient-years of exposures (PYE) = sum of duration of exposure in days (for all patients in treatment group) / 365.25

6.13.2 Adverse Events

Adverse events are recorded in the eCRFs. Each AE will be coded to SOC and PT using the MedDRA version 24.0. Severity of AEs is recorded as mild, moderate, or severe.

A treatment-emergent adverse event (TEAE) is defined as an event that either first occurred or worsened in severity after the first dose of study treatment during the analysis period. The analysis period is defined as the treatment period plus up to 30 days off-drug follow-up time. Refer to the compound-level safety standards for more details.

In general, baseline measurements are those taken prior to starting study medication. Measurements taken during the treatment period are those taken after starting study medication, unless otherwise noted. If study drug is temporarily interrupted and subsequently restarted during the treatment period, the measurements taken during the temporary interruption will be included in the analysis. Where applicable, the time elapsed during the temporary interruption will also be included in analyses.

For events that are gender-specific (as defined by the MedDRA), the denominator and computation of the percentage will include only patients from the given sex.

Planned summaries are specified below and are described more fully in compound level safety standards and in the AE-related PhUSE white paper (Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated Summary Document [PhUSE 2017]), available at https://www.phuse.eu/white-papers. All analyses will be conducted on the Safety Population.

Table JAIA.6.7. Summary Tables Related to Adverse Events

Analysis
An overview table, with the number and percentage of patients in the safety set with death, a SAE, any TEAE, discontinuation from the study due to an AE, permanent discontinuation from study drug due to an AE, or a severe TEAE
The number and percentage of patients with TEAEs using MedDRA PT nested within SOC
The number and percentage of patients with TEAEs using MedDRA PT (without regard to SOC)
The number and percentage of patients with TEAEs by maximum severity using MedDRA PT
The number and percentage of patients who experienced a serious adverse event (including deaths) using MedDRA PT nested within SOC
A listing of SAEs, deaths, AEs led to permanent study drug discontinuation during the study and listing of all temporary study drug interruptions, including interruptions for reasons other than AEs (abnormal laboratory)
The number and percentage of patients who permanently discontinued from study drug due to an adverse event (including adverse events that led to death) using MedDRA PT nested within SOC
The number and percentage of patients who temporarily interrupted study drug due to an adverse event using MedDRA PT nested within SOC
A summary of the number and percentage of patients who temporarily interrupted study drug, to include the number of interruptions per patient, and the duration and cumulative duration of interruptions
Listing of COVID-19 AE based on SMQ=20000237 using the narrow term classification
Summary of COVID-19 TEAE using MedDRA PT based on SMQ=20000237 with the narrow term classification

Abbreviations: AE = adverse event; MedDRA = Medical Dictionary for Regulatory Activities; PT = preferred term; PY = patient-year; SAE = serious adverse event; SMQ = Standardized MedDRA Query; SOC = System Organ Class; TEAE = treatment emergent adverse event.

In general, the number and percentage of patients with events will be summarized by treatment group in 2 formats:

- by MedDRA PT nested within SOC with decreasing frequency in SOC, 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.

6.13.3 Deaths, Other Serious Adverse Events, and Other Notable Adverse Events

The planned table and listing for SAEs are included in the previous section. A listing of deaths, regardless of when they occurred during the study, will also be provided.

6.13.4 Clinical Laboratory Evaluations

The planned summaries for clinical laboratory evaluations are provided in Table JAIA.6.8 and are described more fully in compound level safety standards and in the laboratory-related PhUSE white papers (PhUSE 2013; PhUSE 2015).

All laboratory tests will be presented using the International System (SI) and United States conventional (CN) units.

There is one special circumstance for laboratory values to be derived. The low-density lipoprotein (LDL)/high-density lipoprotein (HDL) ratio will be derived as the ratio of LDL cholesterol to HDL cholesterol. Similarly, the ratio of HDL to LDL will be derived. There are no central laboratory reference ranges for the LDL/HDL or HDL/LDL ratio.

For the categorical laboratory analyses (shift and treatment-emergent), the analysis period is defined as the treatment period plus up to 30 days off-drug follow-up time. The analysis period for the continuous by-visit laboratory analyses including change from baseline by visit and to last observation is defined as the treatment period excluding off-drug follow-up time.

Table JAIA.6.8. Summary Tables Related to Clinical Laboratory Evaluation

Analysis
Box plots for observed values by visit and change from baseline by visit and at last observation excluding the follow-up period.
Tables with the number and percentage of patients who shift from normal/high to low (i.e., treatment-emergent low) and percentages of patients who shift from normal/low to high (i.e., treatment-emergent high).
Listing of abnormal findings for laboratory analyte measurements, including qualitative measures.

6.13.5 Vital Signs and Other Findings

For the treatment-emergent categorical analyses, the analysis period is defined as the treatment period plus up to 30 days off-drug follow-up time. The analysis period for the continuous by-visit laboratory analyses including change from baseline by visit and to last observation is defined as the treatment period excluding off-drug follow-up time.

The planned summaries for vital signs (systolic blood pressure [BP], diastolic BP, pulse, weight, BMI, temperature) are provided in Table JAIA.6.10, and are described more fully in compound-level safety standards and in the vitals-related PhUSE white papers (PhUSE 2013; PhUSE 2015).

Table JAIA.6.9. Parameter Definitions Related to Vital Signs

Parameter	Low	High
Systolic BP (mm Hg) (Supine or sitting – forearm at heart level)	≤ 90 and decrease from baseline ≥ 20	≥ 140 and increase from baseline ≥ 20
Diastolic BP (mm Hg) (Supine or sitting – forearm at heart level)	≤ 50 and decrease from baseline ≥ 10	≥ 90 and increase from baseline ≥ 10
Pulse (bpm) (Supine or sitting)	< 50 and decrease from baseline ≥ 15	> 100 and increase from baseline ≥ 15
Weight (kg) (Consistent clothing and timing in relationship to meals and voiding)	(Loss) decrease ≥ 7%	(Gain) increase ≥ 7%
Temperature	< 96 degrees F and decrease ≥ 2 degrees F	≥ 101 degrees F and increase ≥ 2 degrees F

Table JAIA.6.10. Summary Tables Related to Vital Signs

Analysis
Box plots for observed values by visit and change from baseline values by visit and at last observation excluding the follow-up period.
Tables with the number and percentage of patients who shift from normal/high to low (i.e., treatment-emergent low) and percentages of patients who shift from normal/low to high (i.e., treatment-emergent high). For weight, the number and percentage of patients are based on a percentage gain/loss that exceeds defined limits. The limits are based on literature defined in Table JAIA.6.9 and in the compound level safety standards.

6.13.6 Safety Topics of Interest

Safety topics of interest include adverse events of special interest, potential toxicities that all products should consider (e.g., hepatic-related events), potential findings based on drug class, or topics anticipated to be requested by a regulatory agency for any reason.

6.13.6.1 Abnormal Hepatic Tests

Hepatic labs include alanine aminotransferase (ALT) and aspartate transaminase (AST), total bilirubin (TBL) and serum alkaline phosphatase (ALP). When criteria are met for hepatic evaluations, investigators will complete a follow-up hepatic safety eCRF. The planned summaries are provided in Table JAIA.6.11.

.

Table JAIA.6.11. Summary Tables Relate to Hepatic Safety

Analysis
ALT and AST: The number and percentage of patients with a measurement greater than or equal to 3 times (3X), 5 times (5X), and 10 times (10X) the lab ULN during the analysis period for all patients with a post-baseline value and for subsets based on various levels of baseline value.
TBL: The number and percentage of patients with a measurement greater than or equal to 2 times (2X) the lab ULN during the analysis period for all patients with a post-baseline value and for subsets based on various levels of baseline value.
ALP: The number and percentage of patients with a measurement greater than or equal to 1.5 times (1.5X) the lab ULN during the analysis period for all patients with a post-baseline value and for subsets based on various levels of baseline value.
Lab Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot maximum postbaseline ALT divided by ULN vs. maximum postbaseline total bilirubin divided by ULN.
Patient profiles including demographics, disposition, information collected on the hepatic-safety CRF (where applicable) and a display of study drug exposure, adverse events, medications, blood pressure, heart rate, and the liver-related measurements over time will be provided for patients with information collected on the hepatic-safety CRF and any additional patients meeting ALT or AST measurement greater than or equal to 5X ULN (on a single measurement), or ALP measurement greater than or equal to 2X ULN (on a single measurement).

Summary Tables Relate to Hepatic Safety

Analysis
Treatment-Emergent Potential Hepatic Disorders Based on MedDRA SMQs: treatment-emergent potentially drug-related hepatic disorders are defined by using the MedDRA preferred terms contained in any of the following SMQs: • Broad and narrow terms in the Liver related investigations, signs and symptoms SMQ (20000008) • Broad and narrow terms in the Cholestasis and jaundice of hepatic origin SMQ (20000009) • Broad and narrow terms in the Hepatitis noninfections SMQ (20000010) • Broad and narrow terms in the Hepatic failure, fibrosis and cirrhosis, and other liver damage SMQ (20000013) • Narrow terms in the liver-related coagulation and bleeding disturbances SMQ (20000015)

Abbreviations: ALP = alkaline phosphatase; ALT = alanine transaminase; AST = aspartate aminotransferase; CRF = case report form; MedDRA = Medical Dictionary for Regulatory Activities; SMQ = Standardized MedDRA Query; TBL = total bilirubin; ULN = upper limit of normal.

6.13.6.2 Hematologic Changes

Hematologic changes will be assessed through analysis of hemoglobin, white blood cell (leukocyte) count, absolute neutrophil count, lymphocyte count, and platelet count. The planned summaries are provided in Table JAIA.6.12.

.

Table JAIA.6.12. Summary Tables Related to Hematologic Changes

Analysis
Shift tables showing the number and percentage of patients based on baseline to maximum during the analysis period will be created, with baseline depicted by the most extreme CTCAE grade during the baseline period. With each shift table, a summary displaying the number and percentage of patients who decreased, increased, or stayed the same in CTCAE grade category will be presented.
The number and percentage of patients with treatment-emergent shifts at any time during the analysis period, based on any increase to CTCAE Grade 1 or above, Grade 2 or above, Grade 3 or above, Grade 4.
Listing of patients with treatment-emergent thrombocytosis.

Abbreviations: CTCAE = Common Terminology Criteria for Adverse Events.

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 (Lengfelder et al. 1998). Similar analysis will use a cut-off of 400 billion/L. Planned and unplanned measurements will be included.

6.13.6.3 Lipid Effects

Lipid effects will be assessed through analysis of elevated total cholesterol, elevated low-density lipoprotein cholesterol, decreased and increased high-density lipoprotein cholesterol, and elevated triglycerides. The planned summaries are provided in Table JAIA.6.13.

Table JAIA.6.13. Summary Tables Related to Lipid Effects

Analysis
Shift tables showing the number and percentage of patients based on baseline to maximum during the analysis period will be created, with baseline depicted by the most extreme NCEP-based level during the baseline period. With each shift table, a summary displaying the number and percentage of patients who decreased, increased, or stayed the same in NCEP-based level will be presented.
The number and percentage of patients with treatment-emergent shifts at any time during the analysis period, based on increases to various levels of NCEP-based categories.
The number and percentage of patients with treatment-emergent potential hyperlipidemia using a predefined MedDRA list of PTs that is a subset of the narrow scope PTs in the MedDRA SMQ "Dyslipidemia" (code 200000026) (see compound-level safety standards).

Abbreviations: MedDRA = Medical Dictionary for Regulatory Activities; NCEP =National Cholesterol Education Program; PT = preferred term; SMQ = Standardized MedDRA Query.

6.13.6.4 Renal Function Effects

Effects on renal function will be assessed through analysis of elevated creatinine. The planned summaries are provided in Table JAIA.6.14.

.

Table JAIA.6.14. Summary Tables Related to Effects on Renal Function

Analysis
Shift tables showing the number and percentage of patients based on baseline to maximum during the treatment period will be created, with baseline depicted by the most extreme CTCAE grade during the baseline period. With each shift table, a summary displaying the number and percentage of patients who decreased, increased, or stayed the same in CTCAE grade category will be presented.
The number and percentage of patients with treatment-emergent shifts at any time during the treatment period, based on any increase to CTCAE Grade 1 or above, Grade 2 or above, Grade 3 or above, Grade 4 and above.

Abbreviation: CTCAE = Common Terminology Criteria for Adverse Events.

6.13.6.5 Evaluations in Creatine Phosphokinase (CPK)

The planned summaries are provided in Table JAIA.6.15.

.

Table JAIA.6.15. Summary Tables Related to Effects on CPK

Analysis
Shift tables showing the number and percentage of patients based on baseline to maximum during the analysis period will be created, with baseline depicted by the most extreme CTCAE grade during the baseline period. With each shift table, a summary displaying the number and percentage of patients who decreased, increased, or stayed the same in CTCAE grade category will be presented.
The number and percentage of patients with treatment-emergent shifts at any time during the analysis period, based on any increase to CTCAE Grade 1 or above, Grade 2 or above, Grade 3 or above, Grade 4 and above.
Treatment-emergent adverse events potentially related to muscle symptoms may also be analyzed based on reported AEs. The Muscle Symptoms special search category is a predefined MedDRA search criteria list that contains the narrow scope terms from the Rhabdomyolysis/myopathy SMQ (code 20000002) plus selected terms from the Musculoskeletal SOC.

Abbreviations: AE = adverse event; CPK = creatine phosphokinase; CTCAE = Common Terminology Criteria for Adverse Events, MedDRA = Medical Dictionary for Regulatory Activities; SMQ = Standardized MedDRA Query; SOC = System Organ Class.

6.13.6.6 Infections

Infections will be defined using all the PTs from the MedDRA Infections and Infestations SOC. The MedDRA terms used to identify infections considered to be opportunistic infections (OIs) are based on Winthrop et al. 2015 and are listed in the compound level safety standards. The list contains narrow (more specific) and broad (less specific) PTs.

The planned summaries are provided in Table JAIA.6.16.

Table JAIA.6.16. Summary Tables Related to Infections

Analysis
The number and percentage of patients with treatment-emergent infections, serious infections, and infections resulting in permanent study drug discontinuation using MedDRA PTs
The number and percentage of patients with TEAEs of infections by maximum severity using MedDRA PTs
Listing of patients experiencing TEAE infections will be provided. The listing will include patient demographics, treatment group, treatment start and stop dates, infectious PT event, infecting organism(s), event start and stop dates, total leukocytes, total lymphocytes, absolute neutrophils, event seriousness, and event outcome.
Summary of Opportunistic Infections based on MedDRA PTs after the potential opportunistic infections are reviewed by medical and confirmed as opportunistic infections.
Listing of Opportunistic Infections based on MedDRA PTs during the study.
The summary table of herpes simplex will 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.
A summary table of herpes zoster will be provided, including 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. The incidence rate will also be provided.
A listing of patients with detectable HBV DNA include patient demographics, treatment group, visit, lab sample dates, baseline serology.
Hepatitis B virus DNA status (not detectable, detectable but not quantifiable [i.e., < lower limit of detection (LLOD)], quantifiable [i.e., ≥LLOD]) stratified by applicable baseline HBV serology status (HBsAb- / HBcAb-, HBsAb+ / HBcAb-, HBsAb+ / HBcAb+, HBsAb- / HBcAb+)

Abbreviations: HBV = hepatitis B virus; MedDRA = Medical Dictionary for Regulatory Activities; PT = preferred term; TEAE = treatment emergent adverse event.

6.13.6.7 Allergic Reactions and Hypersensitivities

A search for relevant events related to allergic reaction and hypersensitivity will be performed using the following SMQs:

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

The planned summaries are provided in Table JAIA.6.17.

The anaphylactic reaction SMQ consists of a narrow search containing PTs that represent core anaphylactic reaction terms, a broad search that contains additional terms (signs and symptoms possibly indicative of anaphylactic reaction) that are added to those included in the narrow search, and an algorithm.

The algorithmic approach (Algorithm 1), which is similar to the algorithm approach defined in Sampson, et al. 2006, comprises 1 or more events associated with an individual administration of study drug, where the events include:

- A narrow term from the anaphylactic reaction SMQ (Category A of the SMQ) or
- Paired terms from the anaphylactic reaction SMQ, comprising terms from at least 2 of the following 3 categories from the SMQ:
  - Category B - (Upper Airway/Respiratory signs and symptoms)
  - Category C - (Angioedema/Urticaria/Pruritus/Flush signs and symptoms)
  - Category D - (Cardiovascular/Hypotension signs and symptoms).

Within the paired terms approach, it is important to recognize that occurrence of these events should be nearly coincident and develop rapidly after exposure to an antigen; a window wherein onset or severity change of the events occur within the same calendar day will be used.

In addition, a second algorithmic approach (Algorithm 2) will be calculated similarly to the algorithm approach defined above (Algorithm 2) but includes an additional category, Category E. The paired terms according to Algorithm 2 will comprise terms from at least 2 of 4 categories (Categories B, C, D, and E). Categories B, C and D are already defined, and Category E includes any of the Gastrointestinal preferred term events (Nausea, Vomiting, Diarrhoea, and Abdominal pain). A patient's listing will be generated based on Algorithm 2 and individual cases will be examined to determine if the cases suggest anaphylaxis. Those cases suggestive of anaphylaxis will be described in CSR.

Table JAIA.6.17. Summary Tables Related to Allergic Reactions/Hypersensitivities

Analysis
Two listings for Allergic Reaction and Hypersensitivities for events that satisfy the queries defined in this section will be listed, by temporal order within patient ID, and will include SOC, PT, SMQ event categorization including detail on the scope (narrow, algorithmic, or broad), reported AE term, AE onset and end dates, severity, seriousness, outcome, etc.
The number and percentage of patients reporting a TEAE for the following will be analyzed: • Any narrow or algorithmic term (Algorithm 1) from any one of the 3 SMQs indicated above (i.e., combined search across narrow portions of all 3 SMQs, and anaphylaxis algorithm for Anaphylactic reaction SMQ) • Any narrow scope term within each SMQ, separately (i.e., narrow SMQ search) • Any term within each SMQ, separately (i.e., broad SMQ search)

Abbreviations: AE = adverse event; ID = identification; PT = preferred term; SOC = system organ class; SMQ = Standardized MedDRA Query; TEAE = treatment emergent adverse event.

6.13.6.8 Major Adverse Cardiovascular Events (MACE) and Other Cardiovascular Events

Major adverse cardiovascular events (MACE) and other cardiovascular events will be adjudicated by an independent, external adjudication committee. All confirmed events after adjudication will be used for the analysis.

The planned summaries are provided in Table JAIA.6.18.

Table JAIA.6.18. Summary Tables Related to MACE and Other Cardiovascular Events

Analysis
The number and percentage of patients with MACE, other cardiovascular events, non-cardiovascular death, and all-cause death, as positively adjudicated, based on the categories and subcategories as defined in compound level safety standards
A listing of the MACE and other CV sent for adjudication will be provided to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result.

Abbreviations: CV = cardiovascular; MACE = major adverse cardiovascular event; MedDRA = Medical Dictionary for Regulatory Activities; PT = preferred term.

6.13.6.9 Thromboembolic Events

Venous thromboembolisms will be adjudicated by an independent, external adjudication committee. Venous and pulmonary artery thromboembolic events will be classified as deep vein thrombosis (DVT), pulmonary embolism (PE), or other peripheral venous thrombosis. Additionally, arterial thromboembolic (ATE) events will be adjudicated by an independent, external adjudication committee. All confirmed events after adjudication will be used for the analyses.

The planned summaries for VTE related events are provided in Table JAIA.6.19.

Table JAIA.6.19. Summary Tables Related to VTEs

Analysis
The number and percentage of patients with a positively adjudicated VTE, DVT, PE, and other peripheral venous thrombosis.
A listing of the VTEs sent for adjudication, to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result

Abbreviations: DVT = deep vein thrombosis; MedDRA = Medical Dictionary for Regulatory Activities; PE = pulmonary embolism; PT = preferred term; T = Treatment period; VTE = venous thromboembolism.

The planned summaries for ATE related events are provided in Table JAIA.6.20.

.

Table JAIA.6.20. Summary Tables Related to ATEs

Analysis
The number and percentage of patients with a positively adjudicated ATE
A listing of the ATEs sent for adjudication to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result

Abbreviations: ATE = arterial thromboembolic; MedDRA = Medical Dictionary for Regulatory Activities; PT = preferred term.

6.13.6.10 Malignancies

Malignancies will be identified using terms from the malignant tumors SMQ. Malignancies excluding nonmelanoma skin cancers (NMSC) and NMSC will be reported separately. All the cases identified by the malignant tumors SMQ (SMQ code = 20000194) will be assessed through medical review to determine *confirmed* NMSC cases. Refer to the compound-level safety standards for more details including the list of the planned NMSC terms.

The planned summaries are provided in Table JAIA.6.21.

Table JAIA.6.21. Summary Tables Related to Malignancies

Analysis
The number and percentage of patients with treatment-emergent malignancies excluding NMSC, and NMSC
Listing of all malignancy cases, with an NMSC flag

Abbreviation: NMSC = nonmelanoma skin cancers.

6.13.6.11 Gastrointestinal Perforations

Potential GI perforations identified by the SMQ search (SMQ code = 20000107) will be provided as a listing for internal review by the medical safety team. Each case will be assessed to determine whether it is "confirmed" GI perforation. All confirmed events after medical review will be used for the analysis. The planned summaries are provided in Table JAIA.6.22.

Table JAIA.6.22. Summary Tables Related to Gastrointestinal Perforations

Analysis
The number and percentage of patients with treatment-emergent gastrointestinal perforations using MedDRA PTs.
Listing of all treatment-emergent gastrointestinal perforations during the study.

Abbreviations: MedDRA = Medical Dictionary for Regulatory Activities; PT = preferred term.

6.13.6.12 Depression and Suicide

During the study, suicidal ideation and behavior, and depression will be assessed prospectively by the investigator via signs and symptoms and through the use of the C-SSRS and the QIDS-SR16. For self-harm events, the Self-Harm follow-up form is used to collect answers to a series of questions that provides a more detailed description of self-harm cases. The QIDS-SR16 total scores will also be categorized into severity classes (none, mild, moderate, severe, very severe). Refer to the compound level safety standards, and to Appendix 2 of the integrated safety analysis plan for more details for the QIDS-SR16 including documentation of errors by country and items needed to set to missing by statistical programming.

The planned summaries are provided in Table JAIA.6.23.

.

Table JAIA.6.23. Summaries Related to Depression and Suicide

Analysis
A listing of the C-SSRS answers for patients with any "yes" answer
A listing of the responses give on the Self-Harm follow-up form
Number and percentage of C-SSRS binary responses and composite endpoints
Number and percentage of patients reporting suicidal ideation, suicidal behavior, and non-suicidal self-injurious behavior based on the C-SSRS during treatment
Shift analysis and Treatment-emergent changes in QIDS-SR16 total score

Abbreviations: C-SSRS = Columbia-Suicide Severity Rating Scale; QIDS-SR16 = Quick Inventory of Depressive Symptomatology Self Report.

6.14 Subgroup Analysis

6.14.1 Efficacy Subgroup Analyses

Subgroup analyses will be conducted for the primary endpoints SRI-4 at the Week 52 visit in the mITT population with PPCV datapoint set (Table JAIA.6.5) using hybrid imputation approach as in primary analysis (Section 6.4.2.4). Additional subgroup analyses may be conducted for the other important secondary endpoints if needed.

A logistic regression analysis with treatment, subgroup, and treatment-by-subgroup interaction as factors will be applied on each of 25 imputed datasets after the hybrid imputation. The treatment-by-subgroup interaction will be tested using the Firth correction (Firth 1993) at the 10% significance level. Treatment group differences will be evaluated within each subgroup using the Chi-square test on each of 25 imputed datasets after the hybrid imputation as well, regardless of whether the interaction is statistically significant. Then the single estimates from both logistic regression and Chi-square test for each imputed dataset will be combined using Rubin's rules (Rubin 1987) via PROC MIANALYZE, respectively.

If any group within the subgroup (for example, yes, no) is <10% of the total population, only descriptive statistics will be provided for that subgroup (that is, no inferential testing).

Forest plots may be created to illustrate the treatment differences with 95% CIs between each of the baricitinib-2mg and 4mg treatment groups and placebo group, by each subgroup category.

The following subgroups will be analyzed:

- Disease severity at baseline (SLEDAI-2K): (<10; ≥10)
- SLEDAI-2K organ system involvement at baseline (yes or no for each organ system domain)
- Gender: (Male; Female)
- Baseline anti-dsDNA status: (positive; negative)
- Baseline Complement: (low C3 and/or low C4 levels; high C3 and C4 levels)
- Baseline Low complement/anti-dsDNA positive: (low C3 and/or C4 levels, and anti-dsDNA positive; Others)
- Race: (American Indian/Alaska Native; Asian; Black/African American; Native Hawaiian or other Pacific Islander; White; Multiple)

- Ethnicity (US only): (Hispanic or Latino; Not Hispanic or Latino)
- Region: (North America; Central/South America/Mexico; Europe, Asia; Rest of World)
- Age (<40; ≥40 years old)
- Age (<65; ≥65 years old)
- Baseline corticosteroid dose (<10 mg/day; ≥10 mg/day prednisone or equivalent)

6.14.2 Safety Subgroup Analyses

Subgroup analysis for safety related endpoints will be performed within the context of the integrated safety analysis. No safety subgroup analysis will be performed specifically for this study unless there is a potentially relevant finding during the periodic study safety reviews.

6.15 COVID-19 Trial Impact

Patients who experience an impact to their trial participation due to quarantine and/or travel restrictions related to COVID-19 will have their type of impact summarized. COVID-19 specific impacts for summarization include protocol deviations, which contains out of window visits, treatment interruptions, treatment and/or study discontinuations, and missed visits, regardless of whether or not the patient has a COVID-19 infection documented as an adverse event. The proportion of patients impacted in each category will be summarized. This summary will be provided for the overall mITT patients as well as by region. Number of patients who missed visit related to COVID-19 will be summarized by visit.

The following by-patient listings will be provided:

- Listing of study and treatment disposition related to COVID-19
- Listing of important/non-important protocol deviations due to COVID-19
- Listing of COVID-19 adverse events (specified in Table JAIA.6.7.)
- Listing of temporary treatment interruption related to COVID-19
- Listing of missed visits related to COVID-19

6.16 Protocol Deviations

Protocol deviations will be identified throughout the study. Important protocol deviations (IPDs) are defined as those deviations from the protocol that would potentially compromise patient safety, data integrity, or study outcome.

A separate study deviation rules document describes the categories and subcategories of important protocol deviations, whether or not these are considered IPDs and how the IPDs would be identified.

A by-patient listing of IPDs will be provided.

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.

6.17 Interim Analyses and Data Monitoring

A data monitoring committee (DMC) will oversee the conduct of all the Phase 3 clinical trials evaluating baricitinib in patients with SLE. The DMC will consist of members external to Lilly. This DMC will follow the rules defined in the DMC charter, focusing on assessing safety data for baricitinib. A pre-specified interim efficacy analysis is planned, to include unblinded efficacy data. The same DMC which monitors data for other ongoing studies of baricitinib in other indications will be used.

Access to the unblinded data will be limited to the DMC and statisticians providing the data. In addition, a group of internal review committee members (IRC), who are distinct from the study team members, will be incorporated into those reviews at the recommendation of the DMC. All unblinded personnel will be independent from the study team. The study team will not have access to the unblinded data. The study sites will receive information about interim results ONLY if they need to know for the safety of their patients. The DMC may request to review efficacy data to investigate the benefit/risk relationship in the context of safety observations for ongoing patients in the study. However, the study will not be stopped for positive efficacy results. Hence, there will be no α adjustment for these interim analyses.

Information that may unblind the study during the analyses will not be reported to study sites or blinded study team until the study has been unblinded.

6.18 Annual Report Analyses

Based on regulatory requirements for the Development Safety Update Report (DSUR), reports will be produced (if not already available from the CSR) for the reporting period covered by the DSUR.

6.19 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 which will be converted to an XML file. Both SAEs and 'Other' AEs are summarized by treatment group and 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).

7 Unblinding Plan

A separate unblinding plan was approved prior to the primary outcome database lock and study team unblinding.

8 References

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med*. 2014;33(4):693-713. https://doi.org/10.1002/sim.5974

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Sta. Med*. 2009;28(4):586-604. https://doi.org/10.1002/sim.3495

Bretz F, Posch M, Glimm E, et al. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J*. 2011;53(6):894-913. https://doi.org/10.1002/bimj.201000239

Cella D, Webster K. Linking outcomes management to quality-of-life measurement. *Oncology*. (Willeston Park). 1997;11(11A):232-235. https://pubmed.ncbi.nlm.nih.gov/9430193/

[EMA] European Medicines Agency. ICH E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Published February 18, 2020. Accessed July 16, 2021. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf.

[FDA] United States Food and Drug Administration. Collection of race and ethnicity data in clinical trials: guidance for industry and Food and Drug Administration staff. Available at: https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126396.pdf. October 2016. Accessed January 02, 2019.

[FDA] United States Food and Drug Administration. Systemic lupus erythematosus – developing medical products for treatment: guidance for industry. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072063.pdf. June 2010. Accessed January 02, 2019.

[FDA] United States Food and Drug Administration. Collection of race and ethnicity data in clinical trials: guidance for industry and Food and Drug Administration staff. Available at: https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126396.pdf. October 2016. Accessed January 02, 2019.

Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993;80(1):27-38. https://doi.org/10.1093/biomet/80.1.27

Lengfelder E, Hochhaus A, Kronawitter U, Höche D, Queisser W, Jahn-Eder M, Burkhardt R, Reiter A, Ansari H, Hehlmann R. Should a platelet limit of 600 x 10(9)/l be used as a diagnostic criterion in essential thrombocythaemia? An analysis of the natural course including early stages. *Br J Haematol*. 1998;100(1):15-23. https://doi.org/ doi: 10.1046/j.1365-2141.1998.00529.x

Maruish ME (Ed.). User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric Incorporated. 2011.

PhUSE. PhUSE Computational Science Standard Analyses and Code Sharing Working Group. Analysis and Display White Papers Project Team. Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated

Summary Documents. Available at: https://www.phuse.eu/documents/working-groups/cs-whitepaper-adverseevents-v10-4442.pdf. Accessed January 02, 2019.

PhUSE. Computational Science Symposium Development of Standard Scripts for analysis and Programming Working Group. Analysis and Display White Papers Project Team. Analysis and Displays Associated with Measures of Central Tendency - Focus on Vital Sign, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Summary Documents. (October, 2013). Accessed January 03, 2019. Available at: https://www.phuse.eu/documents//working-groups/deliverables/analyses-displays-associated-with-measures-of-central-tendancy-version-10-10-oct-13-11816.pdf

PhUSE. PhUSE Computational Science Standard Analyses and Code Sharing Working Group. Analysis and Display White Papers Project Team. Analysis and Displays Associated with Outliers or Shifts from Normal to Abnormal: Focus on Vital Sign, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Summary Documents. (September, 2015). Accessed January 03, 2019. Available at: https://www.phuse.eu/documents//working-groups/deliverables/analyses-displays-associated-with-outliers-or-shifts-from-normal-to-abnormal-version-10-10-sept-15-11815.pdf

Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *PharmacoEconomics* 1993;4(5):353-365. https://doi.org/10.2165/00019053-199304050-0000

Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc.; 1987.

Sampson HA, Muñoz-Furlong A, Campbell RL, et al. Second symposium on the definition and management of anaphylaxis: summary report--Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. *J Allergy Clin Immunol*. 2006;117(2):391-397. https://doi.org/10.1016/j.jaci.2005.12.1303

Szende A, Oppe M, Devlin N. EQ-5D value sets: Inventory, comparative review and user guide. EuroQol Group Monographs. Volume 2. Springer, 2006. Available at: https://link.springer.com/book/10.1007/1-4020-5511-0

Winthrop KL, Novosad SA, Baddley JW, et al. Opportunistic infections and biologic therapies in immune-mediated inflammatory diseases: consensus recommendations for infection reporting during clinical trials and postmarketing surveillance. *Ann Rheum Dis*. 2015;74(12):2107-2116. https://doi.org/10.1136/annrheumdis-2015-207841

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 codes 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 (i.e., 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 "prohibited" or "not prohibited" use during the treatment period. Note: Determination of prohibited use is based on a case-by-case assessment of new corticosteroid use or increase in dose greater than baseline, considering whether dose and/or duration of use are clinically relevant. All corticosteroid doses need to be converted to prednisone equivalent doses. If additional conversion factors are required, these will be added to the table below in a SAP revision 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 Name	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:

P01B Antimalarials, specifically

P01BA Aminoquinolines

P01BB Biguanides

P01BC Methanolquinolines

P01BD Diaminopyrimidines

P01BE Artemisinin and derivatives, plain

P01BF Artemisinin and derivatives, combinations

P01BX Other antimalarials

M09AA Quinine and derivatives

All unique preferred names 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 with a start date or stop date during the treatment period (i.e., 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 "prohibited" or "not prohibited" use during the treatment period. Note: Determination of prohibited use is based on a case-by-case assessment of new antimalarial use or increase in dose greater than baseline, considering whether dose and/or duration of use are clinically relevant (for example, switch from hydroxychloroquine to equivalent dose of chloroquine is not prohibited).

Immunosuppressants

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

M01AX Other anti-inflammatory and antirheumatic agents, non-steroids. Note, for M01AX, only if the medication preferred name is sulfasalazine

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

J04BA Drugs for Treatment of Lepra

All unique preferred names 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 (i.e., 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 "prohibited" or "not prohibited" during the treatment period. Note: Determination of prohibited use is based on a case-by-case assessment of new immunosuppressant use or increase in dose greater than baseline, considering whether dose and/or duration of use are clinically relevant (for example, switch from methotrexate oral daily dose to equivalent weekly injection is not prohibited).

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

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. Summary Tables Related to Patient Characteristics

Variable	Continuous Summary	Categorical Summary[a]
Age[b]	Yes	<65 years, ≥65 to <75 years, ≥75 to <85 years, ≥85 years; <40 years, ≥ 40 years; ≥65 years; ≥75 years
Sex	No	Male, Female
Ethnicity (US only)	No	Hispanic/Latino, Non-Hispanic/Non-Latino
Race	No	American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian or other Pacific Islander, White, or Multiple
Geographic Region	No	North America, Central/South America/Mexico, Europe, Asia, Rest of World
Country	No	By country
Height (cm)	Yes	None
Waist circumference (cm)	Yes	None
Weight (kg)	Yes	<80 kg, ≥80 kg
		<100 kg, ≥100 kg
BMI at Visit 2[c]	Yes	Underweight (<18.5 kg/m^2), Normal (≥18.5 and <25 kg/m^2), Overweight (≥25 and <30 kg/m^2), Obese (≥30 and <40 kg/m^2), Severely obese (≥40 kg/m^2)
Alcohol Use	No	Never, Current, Former
Tobacco Use	No	Never, Current, Former
Caffeine Use	No	Never, Current, Former
Prior SLE Therapy		
Prior use of immunosuppresants[d]	No	Yes, No
		0, 1, ≥2
Prior use of methotrexate[d]	No	Yes, No
Prior use of azathioprine[d]	No	Yes, No
Prior use of mycophenolate[d]	No	Yes, No
Prior use of corticosteroids[d]	No	Yes, No
Prior use of antimalarials[d]	No	Yes, No
Prior use of biologics[d]	No	Yes, No
Prior use of belimumab	No	Yes, No
Prior use of rituximab	No	Yes, No
Baseline Disease Characteristics		
Time since diagnosis of lupus (years)[e]	Yes	<1 year, ≥1 to <3 years, ≥3 year to <7 years, ≥7 years
Age at diagnosis of SLE[f]	Yes	<6, ≥6 to <10 year, ≥10 to <17 years, ≥17 year to <40 years, ≥40
Any flare at baseline, (SELENA-SLEDAI definition)	No	None, Mild/Moderate, Severe
SLEDAI-2K score	Yes	(<10 or ≥10)

Variable	Continuous Summary	Categorical Summary[a]
Physician's Global Assessment of Disease Activity score	Yes	Mild (>0 and <1.5), Moderate (≥1.5 to ≤2.5), Severe (>2.5)
Total systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index score	Yes	None
Joint Stiffness Duration	Yes	None
Joint Stiffness Severity	Yes	None
Patient's Global Impression of Severity	Yes	5 categories: no symptoms, very mild, mild, moderate, severe
Number of tender joints (from 28-tender joint count)	Yes	None
Number of swollen joints (from 28-swollen joint count)	Yes	None
CLASI Total Activity Score	Yes	CLASI (<10, ≥10)
CLASI Total Damage Score	Yes	None
Complement C3 level	Yes	Low (<90mg/dL); Normal or higher (≥90mg/dL)
Complement C4 level	Yes	Low (<10mg/dL); Normal or higher (≥10mg/dL)
Complement C3 and C4 levels	No	low C3 and/or low C4 levels; high C3 and C4 levels
Anti-dsDNA level (positive > 15 U/mL)	Yes	Yes, No
Complement and anti-dsDNA level	No	Low complement/anti-dsDNA positive (low C3 and/or C4 levels and anti-dsDNA positive); Others
Proteinuria as assessed by the urine protein to urine creatinine ratio	Yes	(<25, ≥25 mg/mmol) (<50, ≥50mg/mmol)
Serum immunoglobulin IgA, IgG, and IgM concentrations	Yes	Less than the lower limit of normal [LLN] (Yes, No)
Estimated glomerular filtration rate (eGFR)	Yes	eGFR (<60, ≥60 mL/min/1.73m^2)
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.	No	Yes, No
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.	No	Yes, No
BILAG organ domain with at least one A item at baseline	No	Yes, No
BILAG organ domain with no A item but at least two B items at baseline	No	Yes, No
SLEDAI-2K organ system involvement at baseline (yes or no for each organ system domain). Involvement is defined as SLEDAI-2K score within each organ system >0.	No	Yes, No
Anti-Sm+ antinuclear antibodies (positive >10 U/mL)	Yes	Yes, No

Variable	Continuous Summary	Categorical Summary[a]
Anti-RNP+ antinuclear antibodies (positive >10 U/mL)	Yes	Yes, No
Anti-Sjögren's-syndrome-related antigen A (SSA/Ro) antibodies (positive > 10U/mL)	Yes	Yes, No
Anti-Sjögren's-syndrome-related antigen B (also called anti-La [Anti-SSB/La+]) antibodies (positive > 10 U/mL)	Yes	Yes, No
Anti-phospholipid antibody, overall (any of cardiolipin IgA, IgG, IgM, beta-2-glycoprotein IgG, beta-2-glycoprotein IgM, or lupus anticoagulant are positive, then overall is positive, otherwise negative)	No	Positive, Negative
cardiolipin IgA	Yes	<14.00 U/mL = Negative, 14.00-20.00 U/mL = Indeterminate, >20.00 U/mL = Positive
cardiolipin IgG	Yes	<10.00 U/mL = Negative, 10.00-40.00 U/mL = Indeterminate, >40.00 U/mL = Positive
cardiolipin IgM	Yes	<10.00 U/mL = Negative, 10.00-40.00 U/mL = Indeterminate, >40.00 U/mL = Positive
Beta-2-glycoprotein IgG	Yes	<7.00 U/ML = negative, 7.00 - 10.00 U/ML = intermediate, >10.00 = positive
Beta-2-glycoprotein IgM	Yes	<7.00 U/ML = negative, 7.00 - 10.00 U/ML = intermediate, >10.00 = positive
Lupus anticoagulant	No	Positive, Negative
ANA positive (titer >=1:80)	No	Yes, No
Corticosteroid use[d]. See Section 6.9.1 for details of prednisone (or equivalent) baseline dose	Yes	Yes, No
		<10 mg/day or ≥10 mg/day
		≤7.5 mg/day or >7.5 mg/day
Immunosuppressant[d] use at baseline	No	Yes, No
Mycophenolate use at baseline	No	Yes, No
Azathioprine[d] use at baseline	No	Yes, No
Methotrexate[d] use at baseline	No	Yes, No
Antimalarial[d] use at baseline	No	Yes, No
Hydroxychloroquine use at baseline	No	Yes, No
NSAID[d] use at baseline	No	Yes, No
Statin use at baseline	No	Yes, No
Other Baseline Measures		
Baseline SF-36 PCS, MCS	Yes	None
Baseline WPAI Lupus – Employment status	No	Yes, No
Baseline WPAI Lupus – Absenteeism score, Presenteeism Score, Work productivity Loss Score, Activity Impairment Score.	Yes[g]	None
Baseline QIDS-SR16 score	Yes	<11, ≥11
		None (0 – 5), Mild (6 – 10), Moderate (11 – 15), Severe (16 – 20), Very severe (21 – 27)
EQ-5D-5L VAS and Item scores	Yes	None
FACIT-Fatigue Total score	Yes	None

Variable	Continuous Summary	Categorical Summary[a]
Worst Pain NRS	Yes	None
Worst Joint Pain NRS	Yes	None
Worst Fatigue NRS	Yes	None

Abbreviations: ACR = American College of Rheumatology Damage Index; ANA = antinuclear antibodies; anti-dsDNA = anti-double stranded DNA; BILAG = British Isles Lupus Assessment Group; BMI = body mass index; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; eGFR = estimated glomerular filtration rate; EQ-5D-5L = 5-level EQ-5D version; FACIT = Functional Assessment of Chronic Illness Therapy; IgA = immunoglobulin A; IgG = immunoglobulin G; IgM = immunoglobulin M; LLN = lower limit of normal; MCS = Mental Component Summary; NRS = Numeric Rating Scale; NSAID = nonsteroidal anti-inflammatory drug; PCS = Physical Component Summary; QIDS SR-16 = Quick Inventory of Depressive Symptomatology; SELENA = Safety of Estrogens in Lupus Erythematosus National Assessment; SF-36 = Short Form 36; SLE = Systemic Lupus Erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000; SLICC = Systemic Lupus International Collaborating Clinics; VAS = Visual Analog Scale; WPAI = Work Productivity and Activity Impairment.

[a] This column specifies the levels of the categorical variable to be summarized.

[b] Age in years will be calculated as length of the time interval from the imputed date of birth (July 1st in the year of birth collected in the electronic case report form [eCRF]) to the informed consent date. Formula below: age (in years) = (date of informed consent date – imputed date of birth)/365.25.

[c] Body Mass Index will be calculated as: $BMI\ (kg\ /\ m^2) = Weight\ (kg)\ /(Height\ (m))^2$.

[d] ATC codes for selecting medications are provided in Appendix 1.

[e] Time since diagnosis of lupus (years) = (date of first dose - date of diagnosis of lupus + 1) / 365.25.

[f] Age at diagnosis in years will be calculated as the time interval from the imputed date of birth (July 1st in the year of birth collected in the eCRF) to the date of diagnosis.

[g] Only applied for Employment Status at baseline is "Yes", except the Activity Impairment Score.

Appendix 3. Efficacy Laboratory Analytes

The following laboratory analytes are used in primary and major secondary analytes. The analytes must be confirmed prior to database lock and will not warrant an SAP update if changed.

Analyte	Lab codes
Complement C3	BL8
Complement C4	BL9
Creatinine	F77
Haemoglobin	A02
Lymphocytes Absolute	A14
Neutrophils Absolute	A13
Platelet count	A20
Protein / Creatinine Ratio Urine	G53
Urinalysis RBC	C11
Urinalysis WBC	C15
WBC	A10
anti-dsDNA	Q33
eGFR/1.73m2 (MDRD4)	BL3
Urine dipstick protein	C21

Note: 24 hour urine protein and urine albumin-creatinine ratio are not collected in this study.

Appendix 4. Joint Assessment

This appendix contains detailed instructions for calculating scores for the Joint Assessment.

Joints are evaluated and recorded as either nonevaluable, or if evaluable, then if tenderness is present or absent, and if swelling is present or absent. In total, 28 joints are assessed for tenderness and 28 are assessed for swelling. Hips are not assessed for swelling.

The number of tender/swollen joints will be calculated by summing all joints checked to have tenderness/swelling present. If at least half but not all of the joints are evaluable, then the observed prorated joint count will be calculated instead. The prorated scores for tender joint count (TJC) will be adjusted based upon the number of evaluable joints: the counted score will be multiplied by 28 then divided by the number of joints evaluated (excluding non-evaluable joints and any joints with a missing response). For example, if only 26 of the 28 joints are assessed to be evaluable at a visit, and 6 of those 26 are tender, the prorated joint count is $(6/26)\times28=6.46$ and be used in calculating the change from baseline in TJC. The same algorithm will be applied to the swollen joint count (SJC) as well. If less than half of the joints are evaluable, the number of tender/swollen joints is missing.

Appendix 5. Time to First Severe Flare

In the time to first severe SFI flare analysis, only severe flares in patients who were eligible for that particular flare, as described below, will be counted. For example, a severe flare based on an increase in SLEDAI score to greater than 12 would only be counted in patients with a SLEDAI score of 12 or less at baseline. This would avoid the issue of a patient with a SLEDAI score of greater than 12 at baseline requiring an improvement in SLEDAI score prior to being eligible for flare based on that criterion. Those patients with a baseline SLEDAI score greater than 12 would be eligible for counting as a flare from other criteria, but not from that particular criterion.

There are 6 criteria to identify severe flares that are eligible to be counted in the time to severe flare analysis as showed below:

1. *Increase in SLEDAI score to greater than 12*

Limit patients who can be counted as having a severe flare by this criterion to patients who have SLEDAI score ≤ 12 at baseline.

2. *New or worsening (requiring doubling the prednisone dose or increase in prednisone dose to more than 0.5 mg/kg per day or acute hospitalization) of one or more of the following: Central nervous system involvement, Glomerulonephritis, Hemolytic anemia (hemoglobin less than 70g/L or decrease of 30 g/L), Myositis, Thrombocytopenia (platelets less than 60,000/L), or Vasculitis (more than cutaneous involvement)*

A patient would be counted as having a flare by this criterion if they do not have a baseline flare on the same component at baseline. For example, if a patient's case report form indicates that they had a postbaseline flare due to glomerulonephritis, this flare would only count in the analysis if they didn't have a severe baseline flare based on glomerulonephritis. That is, if that patient had a baseline flare due to central nervous system involvement, but not due to glomerulonephritis, the post-baseline flare due to glomerulonephritis would be counted in the analysis.

3. *Increase in prednisone (or equivalent) to more than 0.5 mg/kg/day*

Limit patients who can be counted as having a severe flare by this criterion to patients who have a prednisone dose of ≤ 0.5 mg/kg/day at baseline.

4. *Initiation of immunosuppressive therapy*

Limit patients who can be counted as having a severe flare by this criterion to patients who were not in flare at baseline based on this criterion.

5. *Hospitalization for lupus activity*

Limit patients who can be counted as having a severe flare by this criterion to patients who were not in flare at baseline based on this criterion.

6. *Increase in PGA to greater than 2.5 points (83mm)*

Limit patients who can be counted as having a severe flare by this criterion to patients who have PGA $\leq$ 2.5 points (83mm) at baseline.

Leo Document ID = 13b7c832-052d-48b7-a4fc-d04b6530722d

Approver: PPD

Approval Date & Time: 01-Nov-2021 17:04:08 GMT

Signature meaning: Approved