Statistical Analysis Plan

Protocol Title:	A Phase 2b Dose Ranging Study to Evaluate the Efficacy and Safety of Rozibafusp Alfa (AMG 570) in Subjects With Active Systemic Lupus Erythematosus (SLE) With Inadequate Response to Standard of Care (SOC) Therapy
Short Protocol Title:	Safety and Efficacy of Rozibafusp Alfa in Subjects With Active Systemic Lupus Erythematosus (SLE)
Protocol Number:	20170588
NCT Number:	04058028
Authors:	[redacted]
Sponsor:	Amgen Inc. One Amgen Center Drive, Thousand Oaks, CA, 91320, USA
SAP Date:	Document Version — Date Original — 27 MAY 2020 Amendment 1 — 18 JAN 2022 Amendment 2 — 03 NOV 2022 **Amendment 3 — 03 OCT 2023**

Version Number	**Date (DDMMMYYYY)**	**Summary of Changes, including rationale for changes**
Original (v1.0)	27MAY2020	
Amendment 1	18JAN2022	Incorporate changes in protocol superseding Amendment 2: • Updated the analyses schedule of interim analyses and primary analysis, and updated features at last interim • Added independent Data Monitoring Committee to monitor safety and interim data • Added language on using adjudicated data for primary, secondary and exploratory endpoints. • Revise primary endpoint to SRI-4 response rate at week 52, and the primary analysis method of primary endpoint using Bayesian Hierarchical Model • Change key secondary endpoints to secondary endpoints and add secondary disease flare endpoints • Revise the exploratory endpoints • Added derivation rules for LLDAS • Administrative, typographical, and formatting changes throughout
Amendment 2	03NOV2022	Incorporate changes in protocol amendment 3: [redacted] • Updated estimand definition. • Updated the imputation rules for missing data. • Updated the analyses of adverse events to include events of interests. • Administrative, typographical, and formatting changes throughout
Amendment 3	**03OCT2023**	**Incorporate changes in the scope of planned analyses due to early termination of the study:**

		• **Removed subsequent IAs following study early termination** • **Removed some exploratory endpoints** • **Removed statistical comparisons of treatment groups; descriptive statistics by treatment groups will be generated instead**

Table of Contents

Table of Contents 4

1. Introduction 9

2. Objectives, Endpoints and Hypotheses 9
   - 2.1 Objectives and Endpoints/Estimands 9
   - 2.2 Hypotheses and/or Estimations 11

3. Study Overview 12
   - 3.1 Study Design 12
   - 3.2 Sample Size 13
   - 3.3 Adaptive Design 13

4. Covariates and Subgroups 14
   - 4.1 Planned Covariates 14
   - 4.2 Subgroups 14

5. Definitions 14
   - 5.1 Basic Definition 14
   - 5.2 Study Points of Reference 16
   - 5.3 Study Dates 16
   - 5.4 Study Time Intervals 17
   - 5.5 Arithmetic Calculations 17
   - 5.6 Efficacy Assessments 17
   - 5.7 Patient-reported Outcome (PRO) Assessments 23

6. Analysis Sets 25
   - 6.1 Full Analysis Set 25
   - 6.2 Safety Analysis Set 25
   - 6.3 Pharmacokinetic/Pharmacodynamic Analyses Set(s) 25
   - 6.4 Interim Analysis Set(s) 25

[redacted]

7. Planned Analyses 25
   - 7.1 Interim Analysis and Early Stopping Guidelines 25
   - 7.2 Final Analysis 27

8. Data Screening and Acceptance 28
   - 8.1 General Principles 28
   - 8.2 Data Handling and Electronic Transfer of Data 28
   - 8.3 Handling of Missing and Incomplete Data 28
   - 8.4 Detection of Bias 31
   - 8.5 Outliers 31
   - 8.6 Distributional Characteristics 31
   - 8.7 Validation of Statistical Analyses 31

9. Statistical Methods of Analysis 31
9.1 General Considerations 31
9.2 Subject Accountability 32
9.3 Important Protocol Deviations 32
9.4 Demographic and Baseline Characteristics 32
9.5 Efficacy Analyses 35
9.6 Safety Analyses 38
9.6.1 Adverse Events 38
9.6.2 Laboratory Test Results 38
9.6.3 Vital Signs 39
9.6.4 Physical Measurements 39
9.6.5 Exposure to Investigational Product (IP) 39
9.6.6 Exposure to Other Protocol-required Therapy 39
9.7 Other Analyses 39
9.7.1 Analyses of Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints 39
[redacted]
10. Changes From Protocol-specified Analyses. 40
11. Literature Citations / References 40
12. Appendices 42
Appendix A. Reference Values/Toxicity Grades 42
Appendix B. SLE Assessment Forms/Instruments 42
Appendix C. Patient-reported Outcome Forms/Instruments 52
Appendix D. Analytical Windows 54
Appendix E. Handling of Dates, Incomplete Dates and Missing Dates 60

List of Table

Table 3-1. Planned Analyses Schedule: Decisions at each Interim Analysis, Primary Analysis, and Final Analysis 13
Table 9-1. Primary Efficacy Endpoint Summary Table 36
Table 9-2. Secondary Efficacy Endpoints Summary Table 37
Table 9-3. Exploratory Efficacy Endpoints Summary Table 38
Table 9-4 Selected Safety Laboratory Tests 40

List of Abbreviations and Definition of Terms

Abbreviation or Term	Definition/Explanation
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
[redacted]	
APPT	Activated partial thromboplastin time
AST	Aspartate aminotransferase
BHM	Bayesian Hierarchical Model
BICLA	BILAG-based Composite Lupus Assessment
BILAG	British Isles Lupus Assessment Group
BMI	Body mass index
BUN	Blood urea nitrogen
CLASI	Cutaneous Lupus Erythematosus Disease Area and Severity Index
CNS	Central nervous system
CPMS	Amgen Clinical Pharmacology Modeling and Simulation
CRF	Case report form
CTCAE	Common Terminology Criteria for Adverse Events
CVA	Cerebrovascular accident
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	Electronic case report form
EOS	End of study
FAS	Full Analysis Set
GGT	Gamma glutamyl transpeptidase
GLM	Generalized linear model

GSO-DM	Global Study Operations – Data Management
HDL	High density lipoprotein
IA	Interim analysis
IARSC	Interim Analysis Review Steering Committee
IBG	Independent Biostatistics Group
Ig	Immunoglobin
INR	International normalized ratio
IP	Investigational product
IPD	Important protocol deviation
IRT	Interactive response technology
LDL	Low density lipoprotein
LLDAS	Lupus Low Disease Activity State
LOCF	Last observation carried forward
MCV	Mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
NSAID	Non-steroidal anti-inflammatory drug
OCS	Oral corticosteroid
PD	Pharmacodynamic
PK	Pharmacokinetic
PGA	Physician Global Assessment
PtGA	Patient Global Assessment
PTT	Partial thromboplastin time
PRO	Patient-Reported Outcome
PROMIS-Fatigue SF7A	Patient-Reported Outcome Measurement Information System Fatigue Short Form 7a Instrument
QoL	Quality of Life
Q1	First quartile

Q3	Third quartile
Q2W	Every two weeks
Q4W	Every four weeks
RAR	Response adaptive randomization
RBC	Red blood cells
ROW	Rest of the world
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SE	Standard error
SFI	SLE Flare Index
SF36v2	Medical Outcomes Short Form-36 Questionnaire Version 2
SLE	Systemic Lupus Erythematosus
hSLEDAI	Hybrid Systemic Lupus Erythematosus Disease Activity Index
SGOT	Serum glutamic-oxaloacetic transaminase
SGPT	Serum glutamic-pyruvic transaminase
SOC	Standard of care
SRI	Systemic Lupus Erythematosus Responder Index
SSAP	Supplemental statistical analysis plan
WBC	White blood cells
WHO	World Health Organization
VAS	Visual analog scale

1. Introduction

The purpose of this Statistical Analysis Plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol amendment 3 for study 20170588, AMG 570 dated 24 May 2022**.** The scope of this plan is **reduced from those outlined in the protocol due to determination of futility at a preplanned interim analysis and subsequent study early termination. It** includes **the** final analysis that is planned and will be executed by the Amgen Global Biostatistical Science department unless otherwise specified.

2. Objectives, Endpoints and Hypotheses

2.1 Objectives and Endpoints/Estimands

Objectives	Endpoints
Primary	
• Evaluate the efficacy of rozibafusp alfa at week 52 as measured by the Systemic Lupus Erythematosus Responder Index (SRI-4)	• SRI-4 response at week 52
Primary Estimand The primary estimand is the difference in SRI-4 response rates between each rozibafusp alfa dose group and placebo at week 52 for all subjects with SLE with inadequate response to SOC therapy who are randomized, regardless of investigational product (IP) discontinuation; subjects will be considered non-responders for using more than protocol-allowed therapies as follows: • initiation or increase from baseline in oral corticosteroid (OCS) dose > 5 mg/day (prednisone or equivalent) or intravenous (IV) or intramuscular (IM) at any time point during the study • initiation or increase from baseline in OCS dose ≤ 5 mg/day (prednisone or equivalent) after week 10 • initiation or increase from baseline in OCS dose ≤ 5 mg/day (prednisone or equivalent) between week 0 and week 8 that was not returned to baseline dose within the following 2 weeks • initiation or increase from baseline of immunosuppressant/immunomodulator(s) at any time during the study	
Secondary	
• Evaluate the efficacy of rozibafusp alfa at week 24	• SRI-4 response at week 24 • British Isles Lupus Assessment Group (BILAG)-based Combined Lupus Assessment (BICLA) response at week 24
• Evaluate the efficacy of rozibafusp alfa at week 52	• Lupus Low Disease Activity State (LLDAS) at week 52

Objectives	Endpoints
	• BICLA response at week 52
• Evaluate the efficacy of rozibafusp alfa as measured by SRI-4 response with oral corticosteroid (OCS)-tapering	• SRI-4 response at week 52 with reduction of OCS to ≤ 7.5 mg/day by week 44 and sustained through week 52 in subjects with a baseline OCS dose ≥ 10 mg/day
• Evaluate the efficacy of rozibafusp alfa on disease flares	• Annualized moderate and severe flare rate (as measured by SELENA-SLEDAI Flare Index) over 52 weeks • Annualized severe flare rate (as measured by SELENA-SLEDAI Flare Index) over 52 weeks • Annualized flare rate (as measured by BILAG score designation of "worse" or "new" resulting in a B score in ≥ 2 organs or an A score in ≥ 1 organ) over 52 weeks
• Evaluate the efficacy of rozibafusp alfa on joints and skin	• Total tender and swollen joints count (limited to hands and wrists): ≥ 50% improvement from baseline at week 12, 24, 36, and 52 in subjects with ≥ 6 tender and swollen joints involving hands and wrists at baseline • Cutaneous Lupus Erythematosus Area and Severity Index (CLASI) activity score ≥ 50% improvement from baseline at week 12, 24, 36, and 52 in subjects with a CLASI activity score ≥ 8 at baseline
• Describe the efficacy of rozibafusp alfa using patient reported outcomes (PRO)	• Patient-Reported Outcome Measurement Information System Fatigue Short Form 7a Instrument (PROMIS-Fatigue SF7A) score and change from baseline at week 12, 24, 36, 44, and 52 • Medical Outcomes Short Form 36 version 2 Questionnaire (SF36v2) change from baseline in the physical component score, mental component score and individual domains of SF36v2 at week 12, 24, 36, 44, and 52

Objectives	Endpoints
	• Lupus Quality of Life questionnaire (LupusQoL) score and change from baseline at week 12, 24, 36, 44, and 52 • Patient Global Assessment (PtGA) score and change from baseline at week 12, 24, 36, 44, and 52
• Characterize the safety of rozibafusp alfa	• Treatment-emergent adverse events • Serious adverse events • Clinically significant changes in laboratory values and vital signs
• Characterize the pharmacokinetics (PK) of rozibafusp alfa	• Trough serum concentrations and terminal elimination half-life of rozibafusp alfa

Objectives	Endpoints
Exploratory	
[redacted]	

2.2 Hypotheses and/or Estimations

Rozibafusp alfa administered for 52 weeks will be well tolerated and will have greater efficacy than placebo as measured by the SRI-4 response at week 52 in subjects with active SLE with inadequate response to SOC therapy.

3. Study Overview

3.1 Study Design

This is a Bayesian adaptive phase 2b, multi-center, double-blind, randomized, placebo controlled, 52-week, dose-ranging study in subjects with active SLE and inadequate response to SOC therapies including OCS, immunosuppressants, and immunomodulators. Subjects will be randomized to receive either placebo or 1 of 3 doses of rozibafusp alfa with the last dose at week 50. Study duration for a single subject will be 52 weeks plus the screening period and safety follow-up period. Treatment will be administered every 2 weeks (Q2W). Subjects are required to attend as many follow-up visits as necessary to ensure a minimum of 16 weeks of safety follow-up after the last administration of the investigational product. Tapering of OCS from the baseline dose will be allowed from week 24 through week 44. The first interim analysis (IA) will be executed after the first 40 enrolled subjects have had the opportunity to complete the week 24 assessment. Additional IAs may be executed after approximately every 32 newly enrolled subjects have had the opportunity to complete the week 24 assessment. The last IA will occur when all 320 subjects are randomized and have had the opportunity to complete the week 24 assessment. This IA will be referred to as the 'all-subjects-week-24 IA'. The purpose and analyses planned at each IA are listed in Table 3-1.

Based on the Data Monitoring Committee's recommendation following review of the results from the 6th IA and subsequent decision by the Data Access Plan Team, the study was terminated due to futility and further enrollment in the study was stopped. Previously enrolled subjects were discontinued from IP and terminated the study after being followed for an additional 16 weeks following last dose of IP. The remaining planned IAs, including the last IA ('all-subjects-week-24-IA') and the primary analysis, were not conducted.

Table 3-1. Planned Analyses Schedule: Decisions at each Interim Analysis, Primary Analysis, and Final Analysis

Time point	Adaptive Decision	Number of subjects with opportunity to complete week 24
1st IA	RAR	40
2nd IA	Futility, RAR	72
3rd IA	Futility, RAR	104

4th IA	Futility, RAR	136
5th IA	Futility, RAR	168
6th IA	Futility, RAR	200
7th IA[ac]	Futility, RAR	232
Last IA[bc]	Administrative Success	320
Primary Analysis[c]	Not applicable	All subjects have had the opportunity to complete week 52
Final Analysis	Not applicable	All subjects reach EOS

IA = interim analysis; RAR = response adaptive randomization; EOS = end of study
[a] Interim analyses are planned after every 32 subjects are randomized and have had the opportunity to complete the week 24 assessment until full enrollment. Number of IAs will depend on the actual enrollment rate.
[b] Last IA is the all-subjects-week-24 IA. The administrative success analysis at the last IA will not result in any adaptive decision for the ongoing study.
[c] **Canceled due to study early termination.**

3.2 Sample Size

The approximate sample size of 320 subjects is chosen to provide 80% power to detect ≥ 25% absolute improvement for at least 1 rozibafusp alfa dose group relative to placebo in the primary endpoint of SRI-4 response rate at week 52 at a significance level of 0.025 (1-sided) using a Bayesian Hierarchical Model (BHM), assuming a 40% response rate in placebo group.

Because enrollment **was** stopped early, the actual sample size **is** smaller.

3.3 Adaptive Design

The prospectively defined adaptive features include RAR, early decision for futility and an early trigger for administrative success. The following adaptations are based on the planned IA results:

- Adaptive randomization begins after a fixed allocation period. After each IA (if before full enrolment), the randomization probability for each of the 3 active doses may be changed based on clinical efficacy after an IA, while the randomization allocation for placebo is kept constant at 25%.
- Efficacy is assessed against predefined early stopping rules for futility from the second IA until before the last IA. If futility is triggered, the study could be terminated, e.g., continued enrollment and IP dosing for previously enrolled subjects would be stopped, etc.
- Efficacy is assessed against predefined rules for early administrative success only at the last planned IA. If administrative success is determined, downstream activities

may be planned/initiated, but the execution of the trial would not be stopped or altered.

- All adaptive decisions will be based on the SRI-4 response at week 52 with longitudinal modeling of SRI-4 response data. For a subject who has not yet completed through week 52, their 52-week value will be imputed using a longitudinal model based on their week 16, 20, or 24 SRI-4 response data, whichever is the latest.

Modeling and simulations were used to design this Bayesian adaptive trial and refine study design features to achieve optimal operating characteristics, including but not limited to type I error, power and estimation bias.

The IAs will be conducted by an Independent Biostatistics Group (IBG) and will be reviewed by a Data Monitoring Committee (DMC), both external to Amgen. The study team, investigators and subjects will remain blinded to the results of the IAs.

4. Covariates and Subgroups

4.1 Planned Covariates

Not applicable.

4.2 Subgroups

The primary endpoint will be **summarized** within each of the subgroups listed below. If **a** subgroup sample size is less than 10% of the population, this subgroup may not be evaluated. The subgroups of interest include the following:

- baseline hSLEDAI (< 10, ≥ 10)
- region (North America + Western Europe, ROW)

5. Definitions

5.1 Basic Definition

Investigational Product (IP)

Rozibafusp alfa and placebo.

Enrollment/Randomization Date

The enrollment/randomization date is the date on which a subject is assigned to a treatment group.

First Dose Date

The first dose date is the date of administration of the first dose of IP; this may or may not be the same as randomization date.

Actual Treatment Received

The actual treatment received is the IP the subject actually received, regardless of what the subject was randomized to. In cases where a subject received both rozibafusp alfa and placebo, the actual treatment received will be the rozibafusp alfa dose. In cases where a subject received multiple rozibafusp alfa doses, the actual treatment received will be based on the highest rozibafusp alfa dose received.

Duration of IP exposure

The duration of IP exposure will be derived as the date of the last IP administration plus 13 days (14-day window - 1), end of study or cutoff date, whichever occurs first, minus the date of first IP administration plus 1 day.

Flare Exposure Time

The flare exposure time is the time from study day 1 up to the last flare assessment date.

Treatment-emergent adverse event

Treatment emergent adverse events are defined as events categorized as Adverse Events (AEs) starting on or after first dose of IP as determined by "Did event start before first dose of investigational product?" equal to "No" or missing on the Events eCRF and up to the end of study date.

Baseline medication

Baseline medication is defined as any medication with start date on or before study day 1 and ongoing while on study.

Concomitant medication

Concomitant medication is defined as any medication with start date prior to the study day 1 and ongoing while on study or any medication with start date on or after study day 1 and up to the end of study.

Prednisone-equivalent dose

Prednisone-equivalent dose will be computed based on the table below. Additionally, a multiplier will be applied depending on the dosing frequency: 2 for ‘twice daily’ or ‘every

12 hours' and 0.5 for every other day. For example, cortisone 25 mg is equivalent to prednisone 5 mg and a cortisone 25 mg twice daily is equivalent to prednisone 10 mg/day.

Corticosteroid	**Equivalent daily dose**
Prednisone	5 mg
Betamethasone	0.75 mg
Cortisone	25 mg
Deflazacort	6 mg
Dexamethasone	0.75 mg
Hydrocortisone	20 mg
Methylprednisolone/Meprednisone/Methylprednisolone Acetate	4 mg
Prednisolone	5 mg
Triamcinolone	4 mg

5.2 Study Points of Reference

Baseline

Baseline is defined as the closest recorded measurements prior to or on the first dose of study treatment for subjects receiving study treatment regardless of time collected. For subjects not receiving treatment, baseline is defined as the closest recorded measurements prior to or on the enrollment date.

Study Day 1

Study day 1 for each subject is the first day of IP administration or the day of randomization for subjects who did not receive IP.

Study Day

Study day for each subject is defined as (day of interest – study day 1) + 1 for dates on or after study day 1, or (day of interest – study day 1) for dates prior to study day 1.

5.3 Study Dates

End of 52-Week Treatment Period Date

End of 52-week treatment period date for each subject is defined as the date of the last assessment for the week 52 visit, the end of study date in case of **subject** early termination**,** study day 394, **or February 28, 2023 (end of study date due to early termination decision)**, whichever occurs first.

Start of Safety Follow-up Period Date

Start of safety follow-up period date for each subject is defined as the end of 52-week treatment period date plus one day.

End of Study Date

End of study for each subject is the date recorded on the End of Study CRF page.

5.4 Study Time Intervals

Treatment Period

The time period from study day 1 to the end of 52-week treatment period date inclusive.

Safety Follow-up Period

The time period from the end of treatment period date plus one day to the end of study date, inclusive.

5.5 Arithmetic Calculations

Change from baseline

The arithmetic difference between a post-baseline value and baseline for a given timepoint:

(post-baseline value – baseline value).

Percent change from baseline

The change from baseline divided by baseline value and multiplied by 100:

(change from baseline / baseline) * 100.

If the change from baseline is not equal to zero and the baseline value is zero then percent change from baseline is not defined. If the change from baseline is equal to zero and the baseline value is also zero then percent change from baseline is 100.

Fold change from baseline or ratio from baseline

Fold change from baseline equals the post-baseline value divided by the baseline value. If the change from baseline is not equal to 0 and the baseline value is 0 then fold change is not defined. If the change from baseline is equal to 0 and the baseline value is also 0 then fold change is 1.

5.6 Efficacy Assessments

Hybrid Systemic Lupus Erythematosus Disease Activity Index (hSLEDAI)

The hSLEDAI is a global index that evaluates disease activity and includes 24 items collecting specific manifestations in 9 organ systems: neurological, musculoskeletal, renal, mucocutaneous, general, heart, respiratory, vascular, and hematological. The maximum score is 105.

The hSLEDAI is identical to the SELENA-SLEDAI score except for the scoring of proteinuria, which uses the SLEDAI-2K definition (proteinuria is scored 4 points if proteinuria > 0.5 grams/24 hours [or equivalent urine protein/creatinine ratio]). The hSLEDAI includes scleritis and episcleritis for visual disturbances assessments and scores arthritis only if > 2 joints manifest signs of inflammation. In the hSLEDAI score, inflammation is strictly defined as the presence of tenderness plus one of the following: swelling, effusion, warmth or erythema. In this study, for hSLEDAI scoring purposes, arthritis is required to involve the small joints of hands and/or wrists.
Additional specifications were added in the protocol to reduce potential false positive scores and improve the instrument's ability to score inflammatory manifestations attributable to the disease activity, as described below.

1. Scleritis and episcleritis will be scored only if stability and SLE-relatedness are confirmed by an ophthalmologist.
2. For alopecia to be scored, subjects should have hair loss without scarring; should neither have alopecia areata nor androgenic alopecia; and should have a CLASI activity score for alopecia ≥ 2.
3. For oral ulcers to be scored, location and appearance must be documented by the clinician.
4. Pericarditis and pleurisy will be scored if symptoms of pericarditis and pleurisy are accompanied by objective findings (e.g. EKG, echocardiogram, and/or chest X-ray, among others).

hSLEDAI descriptors will be scored based on a review of medical history, physical examination, and clinical laboratory findings. Findings should reflect activity during the 30 calendar days prior to the current visit. Details of hSLEDAI descriptors and the scoring algorithm are described in Appendix B.

British-Isles Lupus Assessment Group (BILAG)

The BILAG index (BILAG 2004) evaluates disease activity in 9 separate organ systems and comprises a total of 97 items. Each item is measured qualitatively by review of medical history and physical examination (yes/no, improving/same/worse/new) or

quantitatively by measuring laboratory values. Based on these items, each of the 9 organ systems allocated an alphabetical score of A (most active), B (moderate activity), C (minor activity), D (no current activity) or E (never active).

BILAG descriptors will be scored based on a review of medical history, physical examination, and clinical laboratory findings. Findings should reflect activity during the 4-week period prior to the current visit and should be related to the subject's SLE. Details of BILAG descriptors and the scoring algorithm are described in Appendix B.

Physician Global Assessment (PGA)

The PGA is a visual analog scale (VAS) using 3 benchmarks for assessing disease activity over the last 4 weeks. When scoring the PGA, the previous visit score should be noted, and the current score should be relative to that previous visit. The score ranges from 0 to 3 with 3 indicating severe disease. This refers to the most severe possible disease and does not reflect the most severe ever seen in a particular subject, but the most severe disease ever seen in all SLE patients.

This is a global assessment, factoring in all aspects of the subject's lupus disease activity. It should not reflect non-lupus medical conditions. An increase of ≥ 0.3 points (scale 0 to 3) from baseline is considered clinically significant worsening of disease.

Systemic Lupus Erythematosus Responder Index (SRI-4)

A subject achieves SRI-4 response if all of the following criteria are met:

1. ≥ 4-point reduction from baseline in hSLEDAI score
2. no new BILAG 2004 A and no > 1 new BILAG 2004 B domain scores compared with baseline (e.g. no B, C, D or E scores at baseline becomes A or no more than 1 C, D or E score at baseline becomes B)
3. and < 0.3-point deterioration from baseline in PGA (scale 0 to 3).
4. no use of more than protocol-allowed therapies (i.e. initiation or increase from baseline in oral corticosteroid (OCS) dose > 5 mg/day (prednisone or equivalent) or intravenous (IV) or intramuscular (IM) at any time point during the study; initiation or increase in OCS dose ≤ 5 mg/day (prednisone or equivalent) after week 10; initiation or increase in OCS dose ≤ 5 mg/day (prednisone or equivalent) between week 0 and week 8 that was not returned to baseline dose within the following 2 weeks; initiation or increase from baseline of immunosuppressant/immunomodulator(s) at any time during the study.)

The latest visit date among hSLEDAI, BILAG and PGA will be considered the SRI-4 visit date. In the case when all three visit dates are missing, the upper limit of the analytic window of this visit will be the SRI-4 visit date.

BILAG-based Composite Lupus Assessment (BICLA)

A subject achieves BICLA response if all of the following criteria are met:

1. at least one gradation of improvement in baseline BILAG domain scores in all body systems with moderate or severe disease activity at entry (e.g., all A (severe disease) domain scores at baseline falling to B (moderate), C (mild), or D (no activity), and all B domain scores at baseline falling to C or D)
2. no new BILAG 2004 A domain score and no > 1 new BILAG 2004 B domain scores compared with baseline
3. no worsening of the hSLEDAI score from baseline
4. < 0.3-point deterioration from baseline in PGA (scale 0 to 3)
5. and no use of more than protocol-allowed therapies (i.e. initiation or increase from baseline in oral corticosteroid (OCS) dose > 5 mg/day (prednisone or equivalent) or intravenous (IV) or intramuscular (IM) at any time point during the study; initiation or increase in OCS dose ≤ 5 mg/day (prednisone or equivalent) after week 10; initiation or increase in OCS dose ≤ 5 mg/day (prednisone or equivalent) between week 0 and week 8 that was not returned to baseline dose within the following 2 weeks; initiation **or increase from baseline** of immunosuppressant/immunomodulator(s) at any time during the study.

Swollen and Tender Joint Count

Joints that have been replaced during the study, or have suffered trauma or received intra-articular injections, are considered non-evaluable.

Swollen Joint Count Assessments – A total 28 joints will be scored for presence or absence of swelling. A separated score for joints in the hands and wrists will be calculated.

Tender Joint Count Assessments – A total 28 joints will be scored for presence or absence of tenderness. A separated score for joints in the hands and wrists will be calculated.

Swollen and Tender Joint Count Assessments: joints in hands and wrists will be scored for the simultaneous presence of absence of swelling and tenderness.

Cutaneous Lupus Erythematosus Area and Severity Index (CLASI)

The CLASI consists of 2 scores, the first summarizes the activity of the disease while the second is a measure of the damage done by the disease. Activity is scored based on erythema, scale/hyperkeratosis, mucous membrane involvement, acute hair loss and non-scarring alopecia. Damage is scored in terms of dyspigmentation and scarring, including scarring alopecia. Subjects are asked whether dyspigmentation due to cutaneous lupus lesion usually remains visible for more than 12 months, which is taken to be permanent. If so, the dyspigmentation score is doubled. The scores are calculated by simple addition based on the extent of the symptoms. The CLASI is designed as a table where the rows denote anatomical areas, while the columns score major clinical symptoms. The extent of involvement for each of the skin symptoms is documented according to specific anatomic areas that are scored according to the worst affected lesion within that area for each symptom.

A CLASI response is defined as ≥ 50% improvement of CLASI activity score from baseline.

Lupus Low Disease Activity State (LLDAS)

A subject achieves LLDAS response if all of the following criteria are met:

1. hSLEDAI ≤ 4 with no activity in major organ systems (renal [proteinuria, haematuria, pyuria, urinary casts], central nervous system (CNS) [seizure, psychosis, organic brain syndrome, cranial nerve disorder, CVA], cardiopulmonary [pericarditis, pleurisy], vasculitis, fever) and no haemolytic anaemia or gastrointestinal activity in BILAG
2. no new lupus disease activity (i.e. no new descriptor scores in hSLEDAI, and no new activity in haemolytic anaemia or gastrointestinal domain in BILAG) compared with the previous assessment
3. PGA (scale 0-3) ≤1
4. Current prednisolone (or equivalent) dose ≤ 7.5 mg/day
5. Well tolerated standard maintenance doses of immunosuppressive drugs and approved biological agents (i.e. no increase or initiation of immunosuppressive drugs).

BILAG Flare Index

The BILAG flare index will be derived from BILAG 2004, as measured by BILAG score designation of “worse” or “new” resulting in a B score in ≥ 2 organs or an A score in ≥ 1 organ (Gordon et al, 2003) compared to previous available assessment.

SELENA-SLEDAI Flare Index

A mild or moderate flare is defined as satisfying 1 or more of the following compared with previous assessment:

a. change in SELENA-SLEDAI instrument score of 3 points or more points (but not more than 12)

b. new or worse discoid, photosensitive, lupus profundus, cutaneous vasculitis, or bullous lupus, nasopharyngeal ulcers, pleuritis, pericarditis, arthritis, or fever (SLE)

c. increase in prednisone, but not to ≥ 0.5 mg/kg/day

d. added nonsteroidal anti-inflammatory drugs (NSAIDs) or hydrochloroquine for SLE activity

e. ≥ 1 increase in PGA score but not to more than 2.5

A severe flare is defined as satisfying 1 or more of the following compared with previous visit:

a. change in SELENA-SLEDAI instrument score to greater than 12

b. New or worsening CNS-SLE, vasculitis, nephritis, myositis, platelet count < 60 X 10^9 cells/L, or hemolytic anemia (hemoglobin level < 70 g/L or decrease in hemoglobin level > 30 g/L requiring double prednisone or prednisone increase to > 0.5 mg/kg/day or hospitalization

c. increase in prednisone to >0.5 mg/kg/day

d. new cyclophosphamide, azathioprine, mycophenolate mofetil, or methotrexate

e. Hospitalization for SLE activity

f. increase of PGA to ≥ 2.5

To distinguish between mild and moderate flare an additional question, “Status”, was added to the classic instrument. Based on the investigator’s clinical judgement, the flare is to be classified as one of the following 4 options: No Flare,

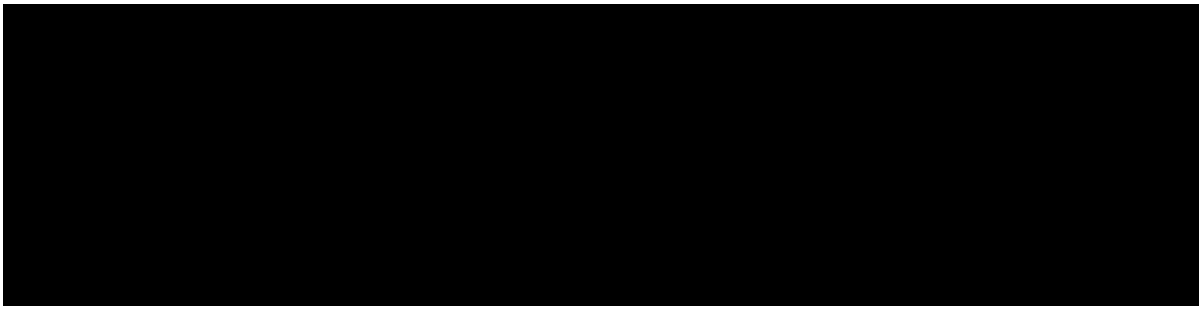
Mild Flare, Moderate Flare, Severe Flare. SELENA-SLEDAI flare data collected in ClinicalInk were transcribed to Rave and hence will be excluded from analyses to avoid duplicate records.

Annualized Flare Rate

The annualized flare rate will be calculated as the number of flares divided by the flare exposure time in days multiplied by 365.25.

Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI)

The SDI global score will be calculated as the sum of the damage scores (0-47) from 12 organ systems with the following possible scores: ocular (0-2), neuropsychiatric (0-6), renal (0-3), pulmonary (0-5), cardiovascular (0-6), peripheral vascular (0-5), gastrointestinal (0-6), musculoskeletal (0-7), skin (0-3), premature gonadal failure (0-1), diabetes (0-1) and malignancy (0-2).



5.7 Patient-reported Outcome (PRO) Assessments

Medical Outcome Short Form-36 Questionnaire Version 2 (SF36v2)

The SF-36v2 (acute version) Health Survey (Ware et al, 2000) contains 36 items and is a revised version of the SF-36 Health Survey. The SF-36v2 acute version is a patient-reported generic measure of health status. This survey yields assessments of 8 domains of health-related quality of life: 1) limitations in physical activities because of health problems; 2) limitations in social activities because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4) bodily pain; 5) general mental health (psychological distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions. The scores from the 8 domains will be evaluated independently and aggregated into 2 norm-based summary component measures of physical and mental

health. The recall period is the past 7 days. The scoring of SF-36v2 will be processed by QualityMetric's PRO CoRE Software.

Patient-Reported Outcome Measurement Information System Fatigue Short Form 7a Instrument (PROMIS Fatigue SF 7A v1.0)

The PROMIS Fatigue Short Form 7a is a 7-item instrument originally constructed by the PROMIS Fatigue team to represent the range of the fatigue trait (PROMIS Fatigue Scoring Manual). It assesses the experience of fatigue as well as its impact on physical, mental and social activities. Both psychometric properties and clinical input were used in the development of the short form from the PROMIS item bank. Estimates of responsiveness and minimally important differences have been reported for the 4 item PROMIS Fatigue instrument in SLE patients (Katz et al, 2019). The PROMIS Fatigue has also been able to differentiate disease activity in other rheumatologic diseases. (Wohlfahrt et al, 2019). Details of the scoring algorithm are described in Appendix C.

LupusQoL

The LupusQoL is a SLE-specific health-related quality of life instrument (Jolly et al, 2010; McElhone et al, 2007) for use in adults. The LupusQoL consists of 8 domains: physical health (8 items), pain (3 items), planning (3 items), intimate relationships (2 items), burden to others (3 items), emotional health (6 items), body image (5 items), and fatigue (4 items). The final instrument has demonstrated good internal reliability (Cronbach's 0.88 to 0.95), good test-retest reliability (r 0.72 to 0.93), good concurrent validity with the comparable domains of the SF-36 (r 0.71 to 0.79) and good discriminant validity for different levels of disease activity, measured by BILAG index, and damage (Systemic Lupus International Collaborating Clinics/ACR damage index) but not for all domains. The instrument also has acceptable ceiling effects and minimal floor effects.

Details of the scoring algorithm are described in Appendix C.

Patient Global Assessment (PtGA)

The patient global assessment of disease activity (PtGA) typically assesses disease activity on a 10 cm numeric rating scale (NRS; 0 to 10 cm) or 0-100 response. The scale for the assessment ranges from "very well" (0) to "very poor" (10) (Furie et al, 2009). The validity of the PtGA in SLE has been established (Khraishi et al, 2014; Liang et al, 1989). The PtGA scale has a 7-day recall.

6. Analysis Sets

6.1 Full Analysis Set

The Full Analysis Set (FAS) includes all randomized subjects. Data will be analyzed based on subjects' randomized treatment assignment.

6.2 Safety Analysis Set

The Safety Analysis Set includes all randomized subjects who received at least 1 dose of IP. Subjects in this set will be analyzed according to the actual treatment received.

6.3 Pharmacokinetic/Pharmacodynamic Analyses Set(s)

The PK concentration analysis set will contain all subjects who received at least one dose of IP and have at least one quantifiable PK sample collected. PK concentration data will be analyzed according to the actual treatment received.

The PK parameter analysis set will contain all subjects who received at least one dose of IP and for whom at least one PK parameter can be adequately estimated. PK parameter will be analyzed according to the actual treatment received.

6.4 Interim Analysis Set(s)

Each Interim Analys**i**s Set include all subjects randomized on or before the interim analys**i**s cutoff date and had the opportunity to complete week 16 (actual week 16 visit date ≤ the cutoff date, or randomization date + 126 - 1 < cutoff date). The interim data cutoff dates are **defined as the date the last enrolled subject among those that trigger an interim analysis (Table 3-1) has** the opportunity to complete **the** week 24 visit.

[redacted]

7. Planned Analyses

7.1 Interim Analysis and Early Stopping Guidelines

Interim analyses (IAs) will be conducted to allow adaptation of the randomization ratio to the 3 rozibafusp alfa treatment groups, holding the allocation to placebo constant at 25%, and to assess efficacy for early futility or administrative success decisions:

- Data will be subject to ongoing checks for integrity, completeness and accuracy in accordance with the Data Management Plan with the expectation that outstanding data issues are resolved ahead of the snapshot to the extent

possible. All available data up to and including the data cut-off date will be included in the analysis based on an "as-is" snapshot of the database without data locking.

- The study team, investigators, and patients will remain blinded to the study treatment, changes in randomization allocation probabilities, and results of the interim analyses unless futility is determined.
- The first IA will be executed after the first enrolled 40 subjects are randomized and have had the opportunity to complete the week 24 assessment. Subsequent IAs will be scheduled after every additional 32 subjects are randomized and have had the opportunity to complete the week 24 assessment until full enrollment. The last IA will occur when all 320 subjects are randomized and have had the opportunity to complete the week 24 assessment. This IA will be referred to as the 'all-subjects-week-24' IA.
- Efficacy analyses will be performed at the IAs to assess the likelihood of rozibafusp alfa being superior to placebo by a clinically meaningful difference. From the second IA until before the last IA, if this likelihood is unacceptably low for all dose levels, the trial is recommended to stop for futility.
- Additionally, at the 'all-subjects-week24 IA', if this likelihood is sufficiently high in at least 1 rozibafusp alfa dose level, IA triggers an administrative success signal. This would not alter ongoing or planned activities of this study, but downstream planning for subsequent trials (e.g. a phase 3 study) may commence. Analysis planned at each IA are listed in Table 3-1. **This interim analysis was cancelled due to the study being terminated for futility prior to this analysis.**

At IAs,

- For all IAs before full enrollment, the randomization ratio to each active treatment group is proportional to the posterior probability that each group has the highest response rate among the three active treatment groups with power scale of 2.

$$Allocation_d \propto \Pr\left(p_{d,52} = \max_c p_{c,52} \,\middle|\, \text{Interim Data} \right),\ c, d \in \{low, medium, high\}$$

where $p_{d,52}$ is the posterior probability of achieving SRI-4 response for group d, calculated using a Bayesian independent logistic model with non-informative priors.

- From the second IA until before the last IA, the interim futility analysis will be performed based on the BHM as described in the primary efficacy analysis. Enrollment to the study may be stopped for futility if the posterior probability of achieving a clinically meaningful difference in SRI-4 response rates of at least 15% between each active treatment group and placebo is below 2.5% for all 3 rozibafusp alfa doses.

  $$\max \Pr\left(p_{d,52} - p_{placebo,52} > \delta \mid \text{Interim Data } \right) < 0.025, d \in \{low, medium, high\}$$

  where $\delta = 0.15$ and $p_{d,52}$is the posterior probability of achieving SRI-4 response for group d, calculated using a BHM.

- The futility rule trigger **is** made using estimates of the response rate $p_{d,52}$ for each dose. However, due to the covariate adjustment, the Bayesian models provide estimates of $p_d \mid Z$, the response rate in each dose conditional on the stratification factor Z. The population level estimate for each dose is obtained using a weighted average of the response rate in each stratification group.

  $$p_d = \sum_{Z \in \mathcal{Z}} w_z \cdot p_d \mid Z \quad \text{for } d \in \{placebo, low, medium, high\}$$

  The set Z indicates the four possible stratification combinations (hSLEDAI stratification [≥10, <10] and region [North America + Western Europe vs ROW]). The weights w_z will be set equal to the observed proportion of patients in each stratification group at the time of the analysis.

- To address the challenge of the lag time between randomization and the week 52 outcome in the adaptive design, a longitudinal model is used to impute week 52 outcome for subjects not yet complete week 52 using their week 16, week 20, or week 24 response data.

7.2 Final Analysis

The final analysis will take place after study completion is reached **or study is early terminated**, i.e. after the last subject completes last visit or early terminates from the study, and all data are collected for the study.

Data will be subject to ongoing checks for integrity, completeness and accuracy in accordance with the Data Management Plan with the expectation that all outstanding data issues are resolved ahead of the final lock. The data supporting the final analysis will be locked to prevent further changes.

8. Data Screening and Acceptance

8.1 General Principles

The objective of the data screening is to assess the quantity, quality, and statistical characteristics of the data relative to the requirements of the planned analyses.

8.2 Data Handling and Electronic Transfer of Data

The Amgen Global Study Operations-Data Management (GSO-DM) department will provide all data to be used in the planned analyses. This study will use the RAVE database. All laboratory, PK, [redacted] and biomarker data from the central laboratory or vendor will be transferred to GSO-DM. All efficacy and PRO data from the central vendor will be transferred to GSO-DM. All other data will be captured on the eCRF.

8.3 Handling of Missing and Incomplete Data

Subjects may have missing specific data points for a variety of reasons. In general, data may be missing due to a subject's early withdrawal from study, a missed visit, or non-evaluability of a specific clinical or laboratory measurement at its planned clinical visit. **Additionally, as the study was terminated before the planned completion, subjects may have missed clinical visits including not completing the week 52 visit. For efficacy and PRO outcomes, subjects who terminated early due to sponsor decision will be excluded from analyses of data from visits for which they did not have the opportunity to participate in by the date of study termination decision being communicated to sites, i.e. February 28, 2023. Efficacy data collected after this date will be censored and excluded from analyses for that particular visit.** Unless specified, no imputation will be used. The general procedures outlined below describe what will be done when a data point is missing.

Hybrid SLEDAI

In case of completely missing hSLEDAI assessment (e.g. missed visit) or partially missing (e.g. laboratory measurements not available), the missing post-baseline hSLEDAI items will be imputed using last observation carried forward (LOCF) if the previous analysis visit is not missing. Baseline hSLEDAI will not be carried forward to post-baseline.

The hSLEDAI total score will then be computed based on the non-missing and imputed data. **If the hSLEDAI total score cannot be evaluated after imputation, the hSLEDAI score ≥ 4 points reduction from baseline will be considered not met.**

CLASI

In case of completely missing CLASI assessment (e.g. missed visit) or partially missing, the missing post-baseline CLASI items will be imputed using LOCF if the previous analysis visit is not missing. Baseline CLASI will not be carried forward to post-baseline.

The CLASI activity score will then be computed based on the non-missing and imputed data. If the CLASI activity score cannot be evaluated after imputation, the CLASI activity score ≥ 50% improvement endpoint will be considered not met.

BILAG

In case of completely missing BILAG assessment (e.g. missed visit), the missing post baseline BILAG item scores will be imputed using LOCF if the previous analysis visit is not missing.

In case of partially missing, missing laboratory items (e.g. UPCR, creatinine, etc.), **nephrotic syndrome or urinary sediment** will be imputed using LOCF if the previous visit is not missing; other missing items (e.g. active nephritis, Coombs' test positive, etc.) will be imputed as ‘No’ or ‘Not Present’.

Baseline BILAG will not be carried forward to missing post-baseline BILAG.

The BILAG organ domain scores will be based on the non-missing and imputed components. If the non-missing and imputed components are not sufficient to identify the organ domain score, worst scenario approach will be used to determine the BILAG organ domain scores.

PGA

Missing post-baseline PGA score will be imputed using LOCF if the previous analysis visit is not missing. Baseline PGA will not be carried forward to missing post-baseline PGA.

SRI-4, BICLA and LLDAS

The criteria for SRI-4, BICLA and LLDAS responses will be evaluated based on the non-missing and imputed PGA, hSLEDAI and BILAG data. If any criteria based on the non-

missing and imputed PGA, hSLEDAI and BILAG data cannot be evaluated, then the subject will be considered non-responder.

BILAG Flare index

BILAG flare index will be evaluated based on BILAG scores. In case of partially missing BILAG assessment, imputed BILAG scores will be used. In case of completely missing BILAG assessment, BILAG flare index will remain missing.

Tender and Swollen Joint Counts

If a joint is excluded (non-evaluable**,** missing or not done), the joint counts will be prorated based on the algorithm described below.

Prorated Joint Counts: If at least half but not all joints are evaluable (at least 14 joints for the total 28 joint count, at least 11 joints for the 22 hands and wrists joint count) then the observed prorated tender or swollen joint count will be calculated. The prorated scores will be adjusted based upon the number of evaluable joints: the counted score will be multiplied by 28 (for total) or 22 (for hands and wrists) as applicable and divided by the number of joints evaluated. If less than half of the joints are evaluable, the number of tender or swollen joints is missing. For example, if only 25 of the 28 joints are assessed at a visit and 10 of those 25 are tender and 8 of those 25 are swollen, the prorated total joint counts are:

Tender: (10/25)*28 = 11.20,

Swollen: (8/25)*28 = 8.96.

That is, the values of 11.20 and 8.96 will be used in calculating the % reductions in total joint counts, not the values of 10 and 8. Similar proration will be conducted for hands and wrists joint count.

If the tender and swollen joint counts cannot be evaluated after proration, the tender and swollen joint count ≥ 50% improvement from baseline endpoint will be considered not met.

[redacted]

Dates

Missing and incomplete dates will be imputed as outlined in Appendix E.

Laboratory Toxicity Grading

In cases when supplemental information for laboratory toxicity grading is missing or the result is ambiguous, the worse toxicity grade possible will be assigned.

Laboratory Measurements

Laboratory measurements that are below the lower quantification limits will be considered equal to the lower limit of quantification for all analyses unless explicitly noted otherwise.

8.4 Detection of Bias

Important protocol deviations and early withdrawal from treatment and from study may bias the results of the study. The incidence of these factors will be assessed and reason for early withdrawals will be tabulated.

8.5 Outliers

Scatter plots will be examined to identify potential outliers in any of the continuous variables and frequencies of the categorical data will be examined to identify questionable values. The validity of any questionable values will be verified, and observations found to be due to data entry errors will be queried. Potential outliers that are not due to data entry error will be included in the analysis.

8.6 Distributional Characteristics

Not applicable.

8.7 Validation of Statistical Analyses

Programs will be developed and maintained, and outputs will be verified in accordance with current risk-based quality control procedures.

Tables, figures, and listings will be produced with validated standard macro programs where standard macros can produce the specified outputs.

The production environment for statistical analyses consists of Amgen-supported versions of statistical analysis software; for example, the SAS System version 9.4 or later.

9. Statistical Methods of Analysis

9.1 General Considerations

All categorical variables will be summarized using the number and percent of subjects falling into each category and all continuous variables will be summarized using mean,

standard error (SE) or standard deviation (SD), median, first quartile (Q1), third quartile (Q3), minimum, maximum, and number of subjects with observations. Safety endpoints will be summarized descriptively, including the treatment-emergent adverse events and serious adverse events, clinically significant changes in laboratory values and vital signs, and incidence of [redacted]. All safety analyses will be performed using the Safety Analysis Set based on subject's actual treatment received.

9.2 Subject Accountability

The study dates for the first subject enrolled, last subject enrolled, last subject's end of IP **and** last subject last visit will be presented by randomized treatment group.

The number and percent of subjects who were enrolled, randomized, received investigational product, completed investigational product, discontinued investigational product and reasons for discontinuing, completed study, discontinued study and reasons for discontinuing study will be summarized by randomized treatment group. Summary of subjects who discontinue investigational product/study due to COVID-19 control measures will be included.

9.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject's initial visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, subcategory codes, and descriptions will be used during the course of the study. Eligibility deviations are defined in the protocol. The final IPD list will be used to produce the summary of IPDs table and the list of subjects with IPDs. IPDs related to COVID-19 control measures will be summarized separately.

9.4 Demographic and Baseline Characteristics

The following demographic, baseline characteristics, baseline disease characteristics and baseline therapies of interest will be summarized descriptively by randomized treatment group.

Demographics:

- sex (female, male)
- age (years)
- age groups (≥ 18 to < 50, ≥ 50 and 18 – 64, 65 – 74, 75 - 84)

- race (American Indian or Alaska Native, Asian, Black or African-American, Native Hawaiian or Other Pacific Islander, White, Other)
- ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- geographic region (North America [Canada and United States] + Western Europe [France, Germany, Sweden, Austria, Switzerland, and United Kingdom], ROW [Bulgaria, Czech Republic, Hungary, Italy, Poland, Russia, Turkey, Brazil, Australia, Japan, New Zealand, South Korea, Mexico, Hong Kong, Argentina, Spain, Portugal and Greece])

Baseline characteristics:

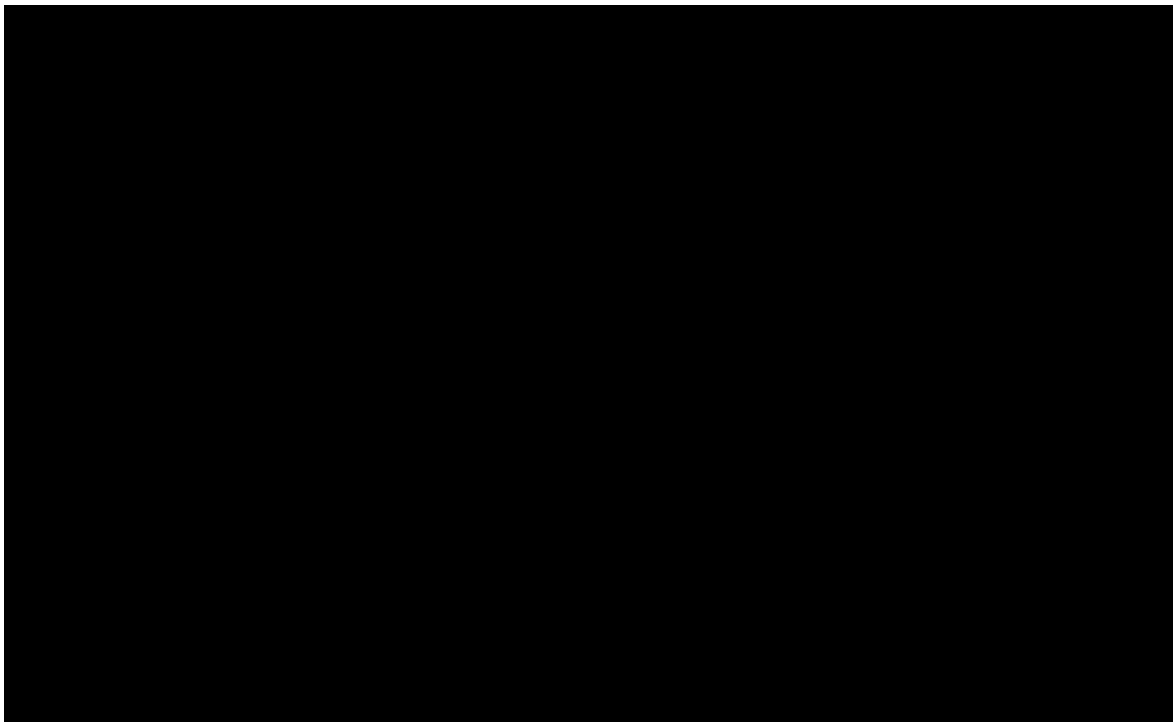
- height (cm)
- weight (kg)
- body mass index (BMI) (kg/m^2)

Baseline disease characteristics:

- years since SLE diagnosis
- hSLEDAI score (continuous, ≥ 10 or < 10)
- baseline hSLEDAI organ involvement (CNS [seizure, psychosis, organic brain syndrome, visual disturbance, cranial nerve disorder, lupus headache, cerebrovascular accident], vascular [vasculitis], renal [urinary casts, hematuria, proteinuria, pyuria], musculoskeletal [arthritis, myositis], serosal [pleurisy, pericarditis], dermal [rash, alopecia, mucosal ulcers], immunologic [low complement, increased DNA binding], constitutional [fever], hematologic [thrombocytopenia, leukopenia])
- BILAG domain score (A, B, C, D, E)
    - constitutional
    - mucocutaneous
    - neuropsychiatric
    - musculoskeletal
    - cardiorespiratory
    - gastrointestinal

- ophthalmic
- renal
- haematological

- CLASI score (continuous, 0, > 0 to < 8, ≥ 8)
- total swollen and tender joint counts
- hand and wrist swollen and tender joint counts
- PGA score
- baseline biomarkers



Baseline medications

- **background medications (detectable vs undetectable baseline SLE SOC drug levels)**
- immunosuppressant/immunomodulator (yes, no)
  - mycophenolate mofetil
  - azathioprine
  - methotrexate
  - hydroxychloroquine

- chloroquine
- dapsone
- other

- daily OCS dose (prednisone-equivalent mg/day; **continuous, 0 mg/day, > 0 to < 10 mg/day, ≥ 10 mg/day)**

9.5 Efficacy Analyses

Unless stated otherwise, the Full Analysis Set (FAS) will be used for analysis of efficacy endpoints. The efficacy analyses are summarized in Table 9-1, Table 9-2 and Table 9-3.

The SLE efficacy assessment data (BILAG, hSLEDAI, PGA, CLASI, swollen and tender joint counts, **and SELENA-SLEDAI flare index**) will be submitted to an external independent **clinician adjudicator** with expertise in managing and assessing SLE disease and clinical trials for adjudication. The committee will adjudicate the data and determine whether the efficacy assessments are scored correctly and consistently based on the instruments' specifications.

Unless stated otherwise, the analysis of the SLE efficacy assessment data and their derived endpoints (SRI-4, BICLA, LLDAS and BILAG flare index) will be based on the data resulting from the adjudication. Additionally**, to assess the concordance between the investigators and adjudicators, the SRI-4 responses at week 16, 20, 24 and 52 based on the two sources will be cross-tabulated and** the concordance rate will be computed as the frequency of cases where the data based on the investigator's assessment and the data resulting from adjudication agree on classification of a subject as an SRI-4 responder/non-responder. If the **overall** concordance rate at week 52 is < 0.95, the **primary endpoint may be summarized** using SRI-4 responses derived based on the investigator-reported data as a sensitivity analysis.

Table 9-1. Primary Efficacy Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity and Supplementary Analysis
SRI-4 response at week 52	**Number and percentage of subjects achieving SRI-4 response will be summarized by randomized treatment arm.**	To assess different data source, the analysis may be repeated using SRI-4 responses based on the investigator data, if the concordance rate is < 0.95.

Table 9-2. Secondary Efficacy Endpoints Summary Table

Endpoint	Primary Summary and Analysis Method
• SRI-4 response at week 24 • LLDAS response at week 52 • BICLA response at week 24 • BICLA response at week 52 • SRI-4 response at week 52 with reduction of OCS to ≤7.5 mg/day by week 44 and sustained through week 52 in subjects with a baseline OCS dose ≥10 mg/day	**Number and percentage of subjects achieving response of interest will be summarized by randomized treatment arm.**
• Annualized moderate and severe flare rate (as measured by SELENA-SLEDAI Flare Index) over 52 weeks • Annualized severe flare rate (as measured by SELENA-SLEDAI Flare Index) over 52 weeks • Annualized flare rate (as measured by BILAG score designation of “worse” or “new” resulting in a B score in ≥ 2 organs or an A score in ≥ 1 organ) over 52 weeks	**Descriptive statistics will be summarized by randomized treatment arm.**
• Total tender and swollen joint count **(**limited to hands and wrists) ≥ 50% improvement from baseline at week 12, 24, 36, and 52 in subjects with ≥ 6 tender and swollen joints involving the hands and wrists at baseline • CLASI activity score ≥ 50% improvement from baseline at week 12, 24, 36, and 52 in subjects with a CLASI activity score ≥ 8 at baseline	**Number and percentage of subjects achieving response of interest will be summarized by randomized treatment arm.**

• PROMIS-Fatigue score and change from baseline at week 12, 24, 36, 44, and 52 • SF36v2 score and change from baseline at week 12, 24, 36, 44, and 52 • LupusQoL score and change from baseline at week 12, 24, 36, 44, and 52 • PtGA score and change from baseline at week 12, 24, 36, 44, and 52	**Descriptive statistics will be summarized by randomized treatment arm.**

Table 9-3. Exploratory Efficacy Endpoints Summary Table

Endpoint	Primary Summary and Analysis Method
[redacted]	

9.6 Safety Analyses

In the safety analyses, data up to the end of study will be summarized in the final analysis. The safety analyses will be based on the safety analysis set.

9.6.1 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version **26.0** or later will be used to code all adverse events to a system organ class and a preferred term. The severity of each event will be graded using Common Terminology Criteria for Adverse Events (CTCAE) version 5.

The subject incidence of adverse events will be summarized for all treatment-emergent adverse events, serious adverse events, adverse events leading to withdrawal of IP, fatal adverse events, treatment-related adverse events, treatment-related serious adverse events, treatment-related adverse events leading to withdrawal of IP, treatment-related fatal adverse events **and adverse events of interest**. Subject incidence of treatment-emergent adverse events identified by COVID-19 standardized MedDRA queries and serious adverse events occurring on or after the COVID-19 infection will also be summarized.

Subject incidence of all treatment-emergent adverse events, serious adverse events, adverse events leading to withdrawal of IP, and fatal adverse events will be tabulated by system organ class in alphabetical order and preferred term in descending order of frequency.

Subject incidence of events of interest (standardized MedDRA queries [SMQ] and/or Amgen Medical Queries [AMQ]) will also be summarized according to their categories, preferred term and severity grade. The events of interests will include but not limited to hypersensitivity (SMQ), injection site reactions (AMQ), infections (infections and infestations system organ class) and malignancy (SMQ).

Summaries of treatment-emergent and serious adverse events will be tabulated by system organ class, preferred term and grade. An overall summary of adverse events by severity grade will be provided.

9.6.2 Laboratory Test Results

Selected clinical laboratory test results, change and percent change from baseline will be summarized over time by treatment arm. In addition, shift tables from baseline to the worst on-study laboratory toxicity based on the CTCAE version 5, will be presented.

Table 9-4 Selected Safety Laboratory Tests

Chemistry	Hematology	Urinalysis
Creatinine Total bilirubin AST (SGOT) ALT (SGPT) C-reactive Protein	RBC Hemoglobin Hematocrit MCV Reticulocytes Platelets WBC Differential • **Eosinophils** • **Basophils** • Lymphocytes	Protein/creatinine ratio WBC RBC

ALT = alanine aminotransferase; AST = aspartate aminotransferase; MCV = mean corpuscular volume; RBC = red blood cell count; SGOT = serum glutamic-oxaloacetic transaminase; SGPT = serum glutamic-pyruvic transaminase; WBC = white blood cell count

9.6.3 Vital Signs

The actual value, change and percent change from baseline in vital signs will be summarized over time by treatment arm for systolic/diastolic blood pressure, heart rate, respiratory rate, and temperature.

9.6.4 Physical Measurements

The actual value, change and percent change from baseline in weight will be summarized by treatment arm by scheduled visit.

9.6.5 Exposure to Investigational Product (IP)

Descriptive statistics will be produced to describe the exposure to IP by treatment arm. Summary statistics will be provided for the total dose received, and total duration of IP exposure by treatment arm.

9.6.6 Exposure to Other Protocol-required Therapy

Other protocol-required therapies for this study are immunomodulators and immunosuppressants, OCS, NSAIDs, topical corticosteroids and topical calcineurin inhibitors and anti-proteinuria agents.

The baseline exposure to OCS and immunomodulators/immunosuppressants will be summarized using descriptive statistics (number and percentages).

9.7 Other Analyses

9.7.1 Analyses of Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints

Amgen Clinical Pharmacology Modeling and Simulation (CPMS) group will conduct PK analysis. Serum concentration data will be summarized by treatment for each PK sampling time point. PK parameters may be summarized if adequately estimated.

[redacted]

10. Changes From Protocol-specified Analyses.

Based on the Data Monitoring Committee's recommendation after interim analysis 6 and subsequent decision by the Data Access Plan Team, enrollment in the study was stopped and the study terminated as the study met the pre-specified futility criteria. As such, some analyses that were initially planned in the protocol will not be performed. Key changes from protocol-specified analyses are as follows:

- **Removal of subsequent IAs following study early termination.**
- **Removal of some exploratory endpoints.**
- **Removal of statistical comparisons of treatment groups; descriptive statistics by treatment groups will be generated instead.**

11. Literature Citations / References

Agresti A. Categorical Data Analysis, New York: John Wiley & Sons, Inc, 1990.

Buyon JP, Petri MA, Kim MY, et. al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: A randomized trial. Ann Intern Med. 2005;142:953.

Franklyn K, Lau CS, Navarra SV, et al. Definition and initial validation of a Lupus Low Disease Activity State (LLDAS). *Ann Rheum Dis*. 2016 Sep;75(9):1615-1621.

Furie RA, Petri MA, Wallace DJ, et al. Novel evidence-based systemic lupus erythematosus responder index. Arthritis & Rheum. 2009;61(9):1143-1151.

Ge M, Durham LK, Meyer RD, Xie W, Thomas N. Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences. *Drug Inf J*. 2011;45:481-493.

Gladman D, Ginzler E, Goldsmith C, et al. The development and initial validation of the systemic lupus international collaborating clinics/American College of Rheumatology Damage Index for Systemic Lupus Erythematosus. Arthritis Rheum. 1996; 39(3):363-369.

Gordon C, Sutcliffe N, Skan J, et al. Definition and treatment of lupus flares measured by the BILAG index. *Rheumatology*. 2003;42(11):1372-9.

Jolly M, Pickard AS, Wilke C et al. Lupus-specific health outcome measure for US patients: the LupusQoL-US version. Ann Rheum Dis. 2010;69:29-33.

Katz P, Pedro S, Alemao E, et al. Estimates of responsiveness, minimally import differences and patient acceptable symptom state in five patient-reported outcomes measurement information system short forms in systemic lupus erythematosus. ACR Open. 2019: 1-8. Doi 10.1002/acr2.11100

Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997; 53: 983-997

Khraishi M, Aslanov R, Dixit S, et al 2014. The validity of patient and physician global disease activity assessments of systemic lupus erythematosus: results from the lupus activity coring tool (LAST) as compared to the Selena Sledai (SS) modification multicentre study. *ACR* 2014.

Liang MH, Socher SA, Larson MG, Schur PH. Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. *Arthritis Rheum*. 1989;32(9):1107-18.

Longford NT. Random Coefficient Models, New York: Oxford University Press, Inc, 1993

McElhone K, Abbott J, Shelmerdine J et al. Development and validation of a disease specific health-related quality of life measure, the LupusQol, for adults with systemic lupus erythematosus. Arthritis Rheum. 2007;57:972-979.

Navarra SV, Guzmán RM, Gallacher AE, et al. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *Lancet*. 2011 Feb 26;377(9767):721-731.

PROMIS Fatigue Scoring Manual. Accessed at http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Fatigue_Scoring_Manual.pdf on 29 Jan 2020

Ware JE. SF-36 Health Survey Update. SPINE. 2000; 25:3130-3139.

Wohlfahrt A, Bingham CO, Marder W, et al. Responsiveness of patient-reported outcomes measurement information system measures in rheumatoid arthritis patients starting or switching a disease-modifying antirheumatic drug. Arth Car Res 2019: (4): 521-529.

Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage 1997; 13: 63-74.

12. Appendices

Appendix A. Reference Values/Toxicity Grades

Assessment of severity for each adverse event and serious adverse event reported during the study will be based on:

The Common Terminology Criteria for Adverse Events (CTCAE), version 5 which is available at the following location:

http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm

Appendix B. SLE Assessment Forms/Instruments

HSLEDAI

The hSLEDAI score is computed as sum of the score across the descriptors.

Weight	Score	Descriptor	Definition
8		Seizure	Recent onset, exclude metabolic, infectious or drug causes
8		Psychosis	Altered ability to function in normal activity due to severe disturbance in the perception of reality. Include hallucinations, incoherence, marked loose associations, impoverished thought content, marked illogical thinking, bizarre, disorganized or catatonic behavior. Exclude uremia and drug causes
8		Organic brain syndrome	Altered mental function with impaired orientation, memory, or other intellectual function, with rapid onset and fluctuating clinical features, inability to sustain attention to environment, plus at least two of the following: perceptual disturbance, incoherent speech, insomnia or daytime drowsiness, or increased or decreased psychomotor activity. Exclude metabolic, infection, or drug causes.
8		Visual disturbance	Retinal changes of SLE. Include cystoid bodies, retinal hemorrhages, serous exudate or hemorrhages in the choroid, optic neuritis, scleritis or episcleritis. Exclude hypertension, infection, or drug causes.
8		Cranial nerve disorder	New onset of sensory or motor neuropathy involving cranial nerves.

8		Lupus headache	Severe, persistent headache; may be migrainous, but must be non-responsive to narcotic analgesia.
8		CVA	New onset of cerebrovascular accident(s). Exclude arteriosclerosis.
8		Vasculitis	Ulceration, gangrene, tender finder nodules, periungual infarction, splinter hemorrhages, or biopsy or angiogram proof of vasculitis.
4		Arthritis	>2 joints of hands and wrists with signs of inflammation (i.e., tenderness and swelling or effusion)
4		Myositis	Proximal muscle aching/weakness, associated with elevated creatinine phosphokinase/aldolase or electromyogram changes or a biopsy showing myositis.
4		Urinary casts	Heme-granular or red blood cell casts
4		Hematuria	> 5 red blood cells/high power field. Exclude stone, infection, or other cause.
4		Proteinuria	More than 0.5 gram/24 hours.
4		Pyuria	> 5 WBC/high power field. Exclude infection.
2		Rash	Inflammatory type rash.
2		Alopecia	Abnormal, patchy or diffuse loss of hair
2		Mucosal ulcers	Oral or nasal ulcerations
2		Pleurisy	Pleuritic chest pain or pleural rub with effusion or pleural thickening.
2		Pericarditis	Classic pericardial pain and/or rub effusion with ECG confirmation
2		Low complement	Decrease in CH50, C3, or C4 below the lower limit of normal for testing laboratory
2		Increased DNA binding	Increased DNA binding above normal range for testing laboratory
1		Fever	> 38 °C. Exclude infectious cause.
1		Thrombocytopenia	< 100 x 10^9 platelets / L, exclude drug causes.

1		Leukopenia	< 3 x 10^9 WBC/ L, exclude drug causes.

BILAG-2004 Index Form

BILAG2004 INDEX

Only record items due to SLE Disease Activity & assessment

Scoring: **ND Not Done**
1 Improving
2 Same
3 Worse
4 New
Yes/No OR Value (where indicated)
❑ indicate if not due to SLE activity
(default is 0 = not present)

CONSTITUTIONAL
1. Pyrexia - documented > 37.5°C ( )
2. Weight loss - unintentional > 5% ( )
3. Lymphadenopathy/splenomegaly ( )
4. Anorexia ( )

MUCOCUTANEOUS
5. Skin eruption - severe ( )
6. Skin eruption - mild ( )
7. Angio-oedema - severe ( )
8. Angio-oedema - mild ( )
9. Mucosal ulceration - severe ( )
10. Mucosal ulceration - mild ( )
11. Panniculitis/Bullous lupus - severe ( )
12. Panniculitis/Bullous lupus - mild ( )
13. Major cutaneous vasculitis/thrombosis ( )
14. Digital infarcts or nodular vasculitis ( )
15. Alopecia - severe ( )
16. Alopecia - mild ( )
17. Peri-ungual erythema/chilblains ( )
18. Splinter haemorrhages ( )

NEUROPSYCHIATRIC
19. Aseptic meningitis ( )
20. Cerebral vasculitis ( )
21. Demyelinating syndrome ( )
22. Myelopathy ( )
23. Acute confusional state ( )
24. Psychosis ( )
25. Acute inflammatory demyelinating polyradiculoneuropathy ( )
26. Mononeuropathy (single/multiplex) ( )
27. Cranial neuropathy ( )
28. Plexopathy ( )
29. Polyneuropathy ( )
30. Seizure disorder ( )
31. Status epilepticus ( )
32. Cerebrovascular disease (not due to vasculitis) ( )
33. Cognitive dysfunction ( )
34. Movement disorder ( )
35. Autonomic disorder ( )
36. Cerebellar ataxia (isolated) ( )
37. Lupus headache - severe unremitting ( )
38. Headache from IC hypertension ( )

MUSCULOSKELETAL
39. Myositis - severe ( )
40. Myositis - mild ( )
41. Arthritis (severe) ( )
42. Arthritis (moderate)/Tendonitis/Tenosynovitis ( )
43. Arthritis (mild)/Arthralgia/Myalgia ( )

Weight (kg):

CARDIORESPIRATORY
44. Myocarditis - mild ( )
45. Myocarditis/Endocarditis + Cardiac failure ( )
46. Arrhythmia ( )
47. New valvular dysfunction ( )
48. Pleurisy/Pericarditis ( )
49. Cardiac tamponade ( )
50. Pleural effusion with dyspnoea ( )
51. Pulmonary haemorrhage/vasculitis ( )
52. Interstitial alveolitis/pneumonitis ( )
53. Shrinking lung syndrome ( )
54. Aortitis ( )
55. Coronary vasculitis ( )

GASTROINTESTINAL
56. Lupus peritonitis ( )
57. Abdominal serositis or ascites ( )
58. Lupus enteritis/colitis ( )
59. Malabsorption ( )
60. Protein losing enteropathy ( )
61. Intestinal pseudo-obstruction ( )
62. Lupus hepatitis ( )
63. Acute lupus cholecystitis ( )
64. Acute lupus pancreatitis ( )

OPHTHALMIC
65. Orbital inflammation/myositis/proptosis ( )
66. Keratitis - severe ( )
67. Keratitis - mild ( )
68. Anterior uveitis ( )
69. Posterior uveitis/retinal vasculitis - severe ( )
70. Posterior uveitis/retinal vasculitis - mild ( )
71. Episcleritis ( )
72. Scleritis - severe ( )
73. Scleritis - mild ( )
74. Retinal/choroidal vaso-occlusive disease ( )
75. Isolated cotton-wool spots (cytoid bodies) ( )
76. Optic neuritis ( )
77. Anterior ischaemic optic neuropathy ( )

RENAL
78. Systolic blood pressure (mm Hg) value ( ) ❑
79. Diastolic blood pressure (mm Hg) value ( ) ❑
80. Accelerated hypertension Yes/No ( )
81. Urine dipstick protein (+=1, ++=2, +++=3) ( ) ❑
82. Urine albumin-creatinine ratio mg/mmol ( ) ❑
83. Urine protein-creatinine ratio mg/mmol ( ) ❑
84. 24 hour urine protein (g) value ( ) ❑
85. Nephrotic syndrome Yes/No ( )
86. Creatinine (plasma/serum) µmol/l ( ) ❑
87. GFR (calculated) ml/min/1.73 m^2 ( ) ❑
88. Active urinary sediment Yes/No ( )
89. Active nephritis Yes/No ( )

HAEMATOLOGICAL
90. Haemoglobin (g/dl) value ( ) ❑
91. Total white cell count (x 10^9/l) value ( ) ❑
92. Neutrophils (x 10^9/l) value ( ) ❑
93. Lymphocytes (x 10^9/l) value ( ) ❑
94. Platelets (x 10^9/l) value ( ) ❑
95. TTP ( )
96. Evidence of active haemolysis Yes/No ( )
97. Coombs' test positive (isolated) Yes/No ( )

Revision: 12/Jan/2007

Investigator initials: _____ Date of assessment: _____

BILAG-2004 INDEX SCORING

Scoring based on the principle of physician's intention to treat

Category	Definition
A	Severe disease activity requiring any of the following treatment: 1. systemic high dose oral glucocorticoids (equivalent to prednisolone > 20 mg/day) 2. intravenous pulse glucocorticoids (equivalent to pulse methylprednisolone ≥ 500 mg) 3. systemic immunomodulators (include biologicals, immunoglobulins and plasmapheresis) 4. therapeutic high dose anticoagulation in the presence of high dose steroids or immunomodulators e.g.: warfarin with target INR 3 - 4
B	Moderate disease activity requiring any of the following treatment: 1. systemic low dose oral glucocorticoids (equivalent to prednisolone ≤ 20 mg/day) 2. intramuscular or intra-articular or soft tissue glucocorticoids injection (equivalent to methylprednisolone < 500mg) 3. topical glucocorticoids 4. topical immunomodulators 5. antimalarials or thalidomide or prasterone or acitretin 6. symptomatic therapy e.g.: NSAIDs for inflammatory arthritis
C	Mild disease
D	Inactive disease but previously affected
E	System never involved

BILAG Constitutional Scoring Algorithm

Grade	Criteria
A	1. Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) AND 2. Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Weight loss b) Lymphadenopathy/splenomegaly c) Anorexia
B	1. Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) OR 2. Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Weight loss b) Lymphadenopathy/splenomegaly

	c) Anorexia 3. BUT do not fulfil criteria for Category A
C	1. Pyrexia recorded as 1 (improving) OR 2. One or more of the following recorded as > 0: a) Weight loss b) Lymphadenopathy/Splenomegaly c) Anorexia 3. BUT does not fulfil criteria for category A or B
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)
E	No previous involvement
Missing	

BILAG Mucocutaneous Scoring Algorithm

Grade	Criteria
A	Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Skin eruption - severe b) Angio-oedema - severe c) Mucosal ulceration - severe d) Panniculitis/Bullous lupus - severe e) Major cutaneous vasculitis/thrombosis
B	1. Any Category A features recorded as 1 (improving) OR 2. Any of the following recorded as 2 (same), 3 (worse) or 4 (new): 1. Skin eruption - mild 2. Panniculitis/Bullous lupus - mild 3. Digital infarcts or nodular vasculitis 4. Alopecia - severe
C	1. Any Category B features recorded as 1 (improving) OR 2. Any of the following recorded as > 0: a) Angio-oedema - mild b) Mucosal ulceration - mild c) Alopecia - mild d) Periungual erythema/chilblains e) Splinter haemorrhages
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)
E	No previous involvement
Missing	

BILAG Neuropsychiatric Scoring Algorithm

Grade	Criteria
A	Any of the following recorded as 2 (same), 3 (worse) or 4 (new): 1. Aseptic meningitis 2. Cerebral vasculitis 3. Demyelinating syndrome

	4. Myelopathy 5. Acute confusional state 6. Psychosis 7. Acute inflammatory demyelinating polyradiculoneuropathy 8. Mononeuropathy (single/multiplex) 9. Cranial neuropathy 10. Plexopathy 11. Polyneuropathy 12. Status epilepticus 13. Cerebellar ataxia
B	1. Any Category A features recorded as 1 (improving) OR 2. Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Seizure disorder b) Cerebrovascular disease (not due to vasculitis) c) Cognitive dysfunction d) Movement disorder e) Autonomic disorder f) Lupus headache - severe unremitting g) Headache due to raised intracranial hypertension
C	Any Category B features recorded as 1 (improving)
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)
E	No previous involvement
Missing	

BILAG Musculoskeletal Scoring Algorithm

Grade	Criteria
A	Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Severe Myositis b) Severe Arthritis
B	1. Any Category A features recorded as 1 (improving) OR 2. Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Mild Myositis b) Moderate Arthritis/Tendonitis/Tenosynovitis
C	1. Any Category B features recorded as 1 (improving) OR 2. Any of the following recorded as > 0: a) Mild Arthritis/Arthralgia/Myalgia
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)
E	No previous involvement
Missing	

BILAG Cardiorespiratory Scoring Algorithm

Grade	Criteria
A	Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Myocarditis/Endocarditis + Cardiac failure b) Arrhythmia c) New valvular dysfunction d) Cardiac tamponade e) Pleural effusion with dyspnoea f) Pulmonary haemorrhage/vasculitis g) Interstitial alveolitis/pneumonitis h) Shrinking lung syndrome i) Aortitis j) Coronary vasculitis
B	1. Any Category A features recorded as 1 (improving) OR 2. Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Pleurisy/Pericarditis b) Myocarditis - mild
C	Any Category B features recorded as 1 (improving)
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)
E	No previous involvement
Missing	

BILAG Gastrointestinal Scoring Algorithm

Grade	Criteria
A	Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Peritonitis b) Lupus enteritis/colitis c) Intestinal pseudo-obstruction d) Acute lupus cholecystitis e) Acute lupus pancreatitis
B	1. Any Category A feature recorded as 1 (improving) OR 2. Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Abdominal serositis and/or ascites b) Malabsorption c) Protein losing enteropathy d) Lupus hepatitis
C	Any Category B features recorded as 1 (improving)
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)

E	No previous involvement
Missing	

BILAG Ophthalmic Scoring Algorithm

Grade	Criteria
A	Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Orbital inflammation/myositis/proptosis b) Keratitis - severe c) Posterior uveitis/retinal vasculitis - severe d) Scleritis - severe e) Retinal/choroidal vaso-occlusive disease f) Optic neuritis g) Anterior ischaemic optic neuropathy
B	1. Any Category A features recorded as 1 (improving) OR 2. Any of the following recorded as 2 (same), 3 (worse) or 4 (new): a) Keratitis - mild b) Anterior uveitis c) Posterior uveitis/retinal vasculitis - mild d) Scleritis - mild
C	1. Any Category B features recorded as 1 (improving) OR 2. Any of the following recorded as > 0: a) Episcleritis b) Isolated cotton-wool spots (cystoid bodies)
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)
E	No previous involvement
Missing	

BILAG Renal Scoring Algorithm

Grade	Criteria
A	Two or more of the following providing 1, 4 or 5 is included: 1. Deteriorating proteinuria (severe) defined as a) urine dipstick increased by ≥ 2 levels (used only if other methods of urine protein estimation not available); or b) 24-hour urine protein > 1 g that has not decreased (improved) by ≥ 25% from previous visit; or c) urine protein-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ 25%; or d) urine albumin-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ 25% 2. Accelerated hypertension 3. Deteriorating renal function (severe) defined as

	a) plasma creatinine > 130 μmol/l and having risen to > 130% of previous value; or b) GFR < 80 ml/min per 1.73 m^2 and having fallen to < 67% of previous value; or c) GFR < 50 ml/min per 1.73 m^2, and last time was > 50 ml/min per 1.73 m^2 or was not measured. 4. Active urinary sediment 5. Histological evidence of active nephritis within last 3 months 6. Nephrotic syndrome
B	One of the following: 1. One of the Category A feature 2. Proteinuria (that has not fulfilled Category A criteria) a) urine dipstick which has risen by 1 level to at least 2+ (used only if other methods of urine protein estimation not available); or b) 24-hour urine protein ≥ 0.5 g that has not decreased (improved) by ≥ 25%; or c) urine protein-creatinine ratio ≥ 50 mg/mmol that has not decreased (improved) by ≥ 25%; or d) urine albumin-creatinine ratio ≥ 50 mg/mmol that has not decreased (improved) by ≥ 25% 3. Plasma creatinine > 130 μmol/l and having risen to ≥ 115% but ≤ 130% of previous value
C	One of the following: 1. Mild/Stable proteinuria defined as a) urine dipstick ≥ 1+ but has not fulfilled criteria for Category A & B (used only if other methods of urine protein estimation not available); or b) 24 hour urine protein > 0.25 g but has not fulfilled criteria for Category A & B; or c) urine protein-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Category A & B; or d) urine albumin-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Category A & B 2. Rising blood pressure (providing the recorded values are > 140/90 mm Hg) which has not fulfilled criteria for Category A & B, defined as a) systolic rise of ≥ 30 mm Hg; and b) diastolic rise of ≥ 15mm Hg
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)
E	No previous involvement
Missing	

Since the urine albumin-creatinine ratio and 24-hour urine protein are not collected in this study, the criteria for BILAG proteinuria will be based on urine dipstick and urine protein-creatinine ration only.

BILAG Hematological Scoring Algorithm

Grade	Criteria

A	1. TTP recorded as 2 (same), 3 (worse) or 4 (new) or 2. Any of the following: a) Evidence of haemolysis and Haemoglobin < 8 g/dl b) Platelet count < 25 x 10^9/l
B	1. TTP recorded as 1 (improving) or 2. Any of the following: a) Evidence of haemolysis and Haemoglobin 8 - 9.9 g/dl b) Haemoglobin < 8 g/dl (without haemolysis) c) White cell count < 1.0 x 10^9/l d) Neutrophil count < 0.5 x 10^9/l e) Platelet count 25 - 49 x 10^9/l
C	Any of the following: a) Evidence of haemolysis and Haemoglobin ≥ 10g/dl b) Haemoglobin 8 - 10.9 g/dl (without haemolysis) c) White cell count 1 - 3.9 x 10^9/l d) Neutrophil count 0.5 - 1.9 x 10^9/l e) Lymphocyte count < 1.0 x 10^9/L f) Platelet count 50 - 149 x 10^9/l g) Isolated Coombs' test positive
D	Previous involvement (i.e. A, B, C or D domain grade at any previous visits or domain previous involvement is yes)
E	No previous Involvement.
Missing	

LLDAS

Criteria	**Derivation**
HSLEDAI ≤ 4 with no activity in major organ systems (renal, CNS, cardiopulmonary, vasculitis, fever) and no activity in haemolytic anaemia or gastrointestinal in BILAG	HSLEDAI score ≤ 4. proteinuria, haematuria, pyuria, urinary casts, seizure, psychosis, organic brain syndrome, cranial nerve disorder, CVA, pericarditis, pleurisy, vasculitis and fever in hSLEDAI are marked 'not present.' BILAG items 56-64 are marked as 'not present' and 96 are marked as 'No.'
No new lupus activity compared with the previous assessment	All hSLEDAI items that are 'not present' in the previous assessment are also 'not present' in the current visit.

PGA ≤ 1	PGA ≤ 1
Current prednisolone (or equivalent) dose ≤ 7.5 mg/day	Prednisone equivalent is ≤ 7.5 mg/day at the visit
Well tolerated standard maintenance doses of immunosuppressive drugs and approved biological agents (i.e. no increase or initiation of immunosuppressive drugs.)	No new or increase in immunosuppressant at the visit

Appendix C. Patient-reported Outcome Forms/Instruments

C1. SF-36 Version 2

The scoring of SF-36v2 will be processed by QualityMetric's PRO CoRE Software.C2. Scoring algorithm for PROMIS Fatigue SF 7A.

PROMIS SF v1.0 – Fatigue 7A will be scored according to the PROMIS Fatigue Brief Guide including the scoring in the Instrument Section downloaded from

http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Fatigue_Scoring_Manual.pdf on 28 Feb 2019.

To calculate the score for PROMIS Fatigue SF 7A, sum the values of the responses to each of the 7 questions. The lowest and highest possible sum is 7 and 35, respectively. Then, use the table below to rescale the score into a standardize score with a mean of 50 and a standard deviation of 10.

Conversion Table	
Raw score	**T-score**
7	**29.4**
8	**33.4**
9	**36.9**
10	**39.6**
11	**41.9**
12	**43.9**
13	**45.8**
14	**47.6**

15	**49.2**
16	**50.8**
17	**52.2**
18	**53.7**
19	**55.1**
20	**56.4**
21	**57.8**
22	**59.2**
23	**60.6**
24	**62.0**
25	**63.4**
26	**64.8**
27	**66.3**
28	**67.8**
29	**69.4**
30	**71.1**
31	**72.9**
32	**74.8**
33	**77.1**
34	**79.8**
35	**83.2**

C3. Scoring algorithm for Modified LupusQoL

The Lupus QoL will be scored using the scoring algorithm from the RWS Life Science website accessed 10 April 2020. http://www.corptransinc.com/sites/lupusqol/instrument-information/instrument-scoring

To compute the LupusQOL raw domain score, sum the values of the responses to each question by domain based on the tables below and divide by the number of items. Then, transform the raw domain score by dividing by 4 and multiplying by 100.

Item Response	All the time	Most of the time	A good bit of the time	Occasionally	Never	Not applicable
Score	0	1	2	3	4	Do not score

Domain	Number of Items	Item Numbers
Physical Health	8	1-8
Pain	3	9-11
Planning	3	12-14
Intimate Relationship	2	15, 16
Burden to Others	3	17-19
Emotional Health	6	20-25
Body Image	5	36-30
Fatigue	4	31-34

C4. Scoring algorithm for Patient Global Assessment

The Patient Global Assessment (PtGA) is a 10 cm numeric rating scale with 0 being very well and 10 very poor.

Appendix D. Analytical Windows

The last measurement for the endpoint of interest taken prior to or on the first dose of IP in this study, unless stated otherwise, will be defined as a baseline visit and the analysis visit name will be 'Baseline'. For any visit up to Day 1 pre-dose which is not a baseline visit, the analysis visit will be 'Pre-analysis'.

Since the actual visit for a subject may not exactly coincide with their scheduled visit date, to allow for variations in scheduling, the following visit windows will be used to assign evaluations to a most appropriate nominal visit for analysis and summarization. Furthermore, there will be no gaps between visit windows in order to include as many data points as possible for summarization.

If more than one actual visit (including the unscheduled visits) falls within the same defined window, the visit closest to the target day with non-missing data will be considered for analysis. If two actual visit dates are at the same distance from the target day, the latest visit with non-missing data will be considered for analysis.

The rules above for selecting a visit from multiple ones within the same visit window are not applicable to retest values of lab data. If the lab measurement is a retest, the retest value will be used.

Efficacy assessment (tender/swollen joint count, hSLEDAI, BILAG, SELENA-SLEDAI Flare Index, PGA, CLASI), weight, and laboratory assessment [redacted], chemistry, urinalysis)

Study Visit	Target Day	Study Day
Baseline	1	Last evaluation prior to or on Study Day 1
Week 4	29	2 - 42
Week 8	57	43 - 70
Week 12	85	71 - 98
Week 16	113	99 - 126
Week 20	141	127 - 154
Week 24	169	155 - 182
Week 28	197	183 - 210
Week 32	225	211 - 238
Week 36	253	239 - 266
Week 40	281	267 - 294
Week 44	309	295 - 322
Week 48	337	323 - 350
Week 52	365	351 - 394

Hematology

Study Visit	**Target Day**	**Study Day**
Baseline	**1**	**Last evaluation prior to or on Study Day 1**
Week 4	**29**	**2 - 42**
Week 8	**57**	**43 - 70**
Week 12	**85**	**71 - 98**
Week 16	**113**	**99 - 126**
Week 20	**141**	**127 - 154**

Week 24	**169**	**155 - 182**
Week 28	**197**	**183 - 210**
Week 32	**225**	**211 - 238**
Week 36	**253**	**239 - 266**
Week 40	**281**	**267 - 294**
Week 44	**309**	**295 - 322**
Week 48	**337**	**323 - 350**
Week 52	**365**	**351 - 392**
Week 60	**421**	**393 - 448**
Week 68	**477**	**≥ 449**

Vital sign

Study Visit	Target Day	Study Day
Baseline	1	Last evaluation prior to or on Study Day 1
Week 2	15	2 - 21
Week 4	29	22 - 35
Week 6	43	36 - 49
Week 8	57	50 - 63
Week 10	71	64 - 77
Week 12	85	78 - 91
Week 14	99	92 - 105
Week 16	113	106 - 119
Week 18	127	120 - 133
Week 20	141	134 - 147
Week 22	155	148 - 161
Week 24	169	162 - 175

Week 26	183	176 - 189
Week 28	197	190 - 203
Week 30	211	204 - 217
Week 32	225	218 - 231
Week 34	239	232 - 245
Week 36	253	246 - 259
Week 38	267	260 - 273
Week 40	281	274 - 287
Week 42	295	288 - 301
Week 44	309	302 - 315
Week 46	323	316 - 329
Week 48	337	330 - 343
Week 50	351	344 - 357
Week 52	365	358 - 378
Week 56	**393**	**379 - 406**
Week 60	**421**	**407 - 434**
Week 64	**449**	**435 - 462**
Week 68	**477**	**≥ 463**

[redacted]**, SDI**

Study Visit	Target Day	Study Day
Baseline	1	Last evaluation prior to or on Study Day 1
Week 24	169	2 - 266
Week 52	365	267 - **394**

[redacted]

[redacted]

Pharmacokinetic assessment

Study Visit	Target Day	Study Day
Baseline	1	Last evaluation prior to or on Study Day 1
Week 4	29	2 - 42
Week 8	57	43 - 70
Week 12	85	71 - 98
Week 16	113	99 - 126
Week 20	141	127 - 154
Week 24	169	155 - 210
Week 36	253	211 - 280
Week 44	309	281 - 336
Week 52	365	337 - 378
Week 56	**393**	**379 - 406**
Week 60	**421**	**407 - 434**
Week 64	**449**	**435 - 462**
Week 68	**477**	**≥ 463**

PROs (SF-36v2, PROMIS Fatigue SF 7A, LupusQOL, PtGA)

Study Visit	Target Day	Study Day
Baseline	1	Last evaluation prior to or on Study Day 1
Week 4	29	2 - 56
Week 12	85	57 - 98
Week 24	169	99 - 210
Week 36	253	211 - 280
Week 44	309	281 - 336
Week 52	365	337 - 394

Appendix E. Handling of Dates, Incomplete Dates and Missing Dates

The reference date for the following rules is the date of first dose rozibafusp alfa.

Start Date		Stop Date						
		Complete: *yyyymmdd*		Partial: *yyyymm*		Partial: *yyyy*		Missing
		< 1st dose	≥ 1st dose	< 1st dose *yyyymm*	≥ 1st dose *yyyymm*	< 1st dose *yyyy*	≥ 1st dose *yyyy*	
Partial: *yyyymm*	= 1st dose *yyyymm*	2	1	n/a	1	n/a	1	1
	≠ 1st dose *yyyymm*		2	2	2	2	2	2
Partial: *yyyy*	= 1st dose *yyyy*	3	1	3	1	n/a	1	1
	≠ 1st dose *yyyy*		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1=Impute the date of first dose or the randomization date if not treated; 2=Impute the first of the month; 3=Impute January 1 of the year; 4=Impute January 1 of the stop year
Note: For subjects who were never treated (first dose date is missing), partial start dates will be set to the first day of the partial month or first day of year if month is also missing.

Imputation Rules for Partial or Missing Stop Dates

Initial imputation

- If the month and year are present, impute the last day of that month.
- If only the year is present, impute December 31 of that year.
- If the stop date is entirely missing, assume the event or medication is ongoing.

If the imputed stop date is before the start date, set stop date to missing.

If the imputed stop date is after the death date, impute as death date.