Integrated Analysis Plan

Clinical Trial Protocol Identification No.	MS200527-0018
Title:	A Phase II, Randomized, Double-Blind, Placebo-Controlled Dose-Ranging Study To Evaluate the Safety and Efficacy of M2951 in Subjects with Systemic Lupus Erythematosus (SLE)
Trial Phase	II
Investigational Medicinal Product(s)	Evobrutinib
Clinical Trial Protocol Version	22 May 2018 / Version 7.0

Integrated Analysis Plan Authors

Coordinating Author

PPD, Merck Healthcare KGaA /EMD Serono Research & Development Institute, Inc.

Function	Author(s) / Data Analyst(s)
PPD	PPD
PPD	PPD
PPD EMD-Serono	PPD

Integrated Analysis Plan Date and Version	11 June 2020 / Version 4.0

Integrated Analysis Plan Reviewers

Merck Healthcare KGaA/EMD Serono Research & Development Institute, Inc. Reviewers:

Function	Name
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD

PPD Reviewers:

Function	Name
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD

[redacted]
This document is the property of Merck KGaA, Darmstadt, Germany, or one of its affiliated companies.
It is intended for restricted use only and may not - in full or part - be passed on, reproduced, published or used without express permission of Merck KGaA, Darmstadt, Germany or its affiliate.
Copyright © 2020 by Merck KGaA, Darmstadt, Germany or its affiliate. All rights reserved.

Signature Page

Integrated Analysis Plan: MS200527-0018

A Phase II, Randomized, Double-Blind, Placebo-Controlled Dose-Ranging Study To Evaluate the Safety and Efficacy of M2951 in Subjects with Systemic Lupus Erythematosus (SLE).

Approval of the IAP by Merck Healthcare KGaA Responsible is documented within CARA. PPD Lead Statistician signature – for PPD use only.

Merck Healthcare KGaA responsible	**Date**	**Signature**
PPD		

PPD **responsible**	**Date**	**Signature**
PPD		

1 Table of Contents

Table of In-Text Tables

2 List of Abbreviations and Definition of Terms

ACR	American College of Rheumatology
ADaM	Analysis Data Model
AE	Adverse event
AESI	Adverse event of special interest
AFR	Annualized flare rate
ALT	Alanine amino-transferase
ANA	Antinuclear antibody(ies)
Anti-dsDNA	Anti-Double-Stranded Deoxyribonucleic Acid
Anti-La	An Antinuclear Antibody Associated with Autoimmune Diseases Including Sjögren Syndrome
Anti-Ro	An Antinuclear Antibody Associated with Autoimmune Diseases Including Sjögren Syndrome and Systemic Lupus Erythematosus
Anti-Sm	Anti-Smith Antibody, an Antinuclear Antibody Associated with Autoimmune Diseases Including Systemic Lupus Erythematosus
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic class
BICLA	BILAG-based composite lupus assessment
BID	Twice daily
BILAG	British Isles Lupus Assessment Group
BLQ	Below Limit of Quantification
BMI	Body mass index
BOA	Biostatistics outputs assembly
CCI [redacted]	
CDF	Cumulative distribution function
CFB	Change from baseline
CI	Confidence interval
CLASI	Cutaneous lupus erythematosus disease area and severity index
CLASI-A	CLASI activity
CLASI-D	CLASI damage
CS	Corticosteroid
CSR	Clinical Study Report
C-SSRS	Columbia-suicide severity rating scale

CTP	Clinical Trial Protocol
CV	Coefficient of variation
CXR	Chest X-ray
DILI	Drug-induced liver injury
DNA	Deoxyribonucleic acid
DRM	Data review meeting
EAC	Endpoint Adjudication Committee
EAIR	Exposure adjusted incidence rates
eCRF	Case Report Form, electronic
ECG	Electrocardiogram
EEA	European Economic Area
eGFR	estimated Glomerular Filtration Rate
EQ-5D-5L	EuroQoL 5 dimensions 5 levels
EOT	End of treatment
FACIT	Functional assessment of chronic illness therapy
FDA	Food and Drug Administration
FWER	Family-wise type 1 error rate
HBV	Hepatitis B Virus
HDA	High disease activity
HR	Hazard ratio
HRQoL	Health-related quality of life
HRU	Health resource utilization
IA	Interim analysis
IAP	Integrated Analysis Plan
IDMC	Independent Data Monitoring Committee
Ig	Immunoglobulin
IMP	Investigational Medicinal Product
IOV	Interoccasion variability
IQR	InterQuartile Range
ITT	Intent to Treat
IWRS	Interactive Web Response System
KM	Kaplan-Meier
LDA	Low disease activity
LFT	Liver function test
LOCF	Last observation carried forward

LLN	Lower limit of normal
LTE	Long term extension
MAR	Missing at random
MCP	Multiplicity control procedure
MCS	Mental component summary
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
mITT	Modified Intent-to-Treat
MMRM	Mixed model with repeated measures
MNAR	Missing not at random
mRNA	Messenger Ribonucleic Acid
MSM	Multi state Markov
M&S	Modeling & Simulation
N	Number of subjects
NB	Negative binomial
NCI-CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
NK	Natural killer
OCS	Oral Corticosteroids
OR	Odds ratio
PA	Primary analysis
PAC	Primary analysis cohort
PCS	Physical component summary
CCI [redacted]	
PGA	Physician's global assessment
PGIC	Patient global impression of change
CCI [redacted]	
PP	Per protocol
PRO	Patient-reported outcome
PT	Preferred term
Q1	25th percentile
Q3	75th percentile
QD	Once daily
QoL	Quality of life
RoW	Rest of the world

SAE	Serious adverse event
SAF	Safety analysis set
SAS	Statistical Analysis System
SD	Standard deviation
SDTM	Study Data Tabulation Model
SF-36	36-item short form health survey
SFI	SLEDAI flare index
SI	International System
SLE	Systemic lupus erythematosus
SLEDAI-2K	SLE disease activity index-2000
SLICC	Systemic lupus international collaborating clinics
SOC	System organ class
SoC	Standard of Care
SRI	SLE responder index
TEAE	Treatment emergent adverse event
TLF	Table/Listing/Figure
UPCR	Urine protein to creatinine ratio
ULN	Upper limit of normal
US	United States
VAS	Visual analog scale
WHO-DD	World Health Organization Drug Dictionary

3 Modification History

Unique Identifier for IAP Version	Date of IAP Version	Authors	Changes from the Previous Version
Version 1.0	26 July 2019	PPD	NA – first version
Version 2.0	04 February 2020	PPD	• Section 7: SRI definition updated (treatment discontinuation removed); EAC-treatment failure used in SRI, LLDAS, BICLA and LDA definition #2 CCI [redacted] • Section 9.4 added: Screening definition • Section 9.5: Baseline definition updated (time of dose is not accounted, assessments performed on the day of IMP administration are eligible for baseline). • Section 9.5: Start of LTE treatment is derived based on the end of 52-week treatment. • Section 9.6: Assessments collected more than 14 days after last exposure are excluded from analysis visit windows. CCI [redacted] Section 11.2.1: BILAG severity at Baseline updated, comparison with previous assessment removed. • Sections 13.1 & 13.2: update in compliance derivation (missing number of tablets is now imputed; last kit dispensed for 52-week treatment period is at Week 48; and correction of LTE compliance formula) • Section 14.1: o Update criterion #4 of SRI definition: treatment failure is confirmed by EAC, subjects with missing values at W52 will be considered as non-responder (including subjects who prematurely discontinued treatment) o Addition of intercurrent events o Sensitivity analysis: missing values at W52 will be imputed using MI, details added on MI procedures o New sensitivity analysis for post-PA DBL o CDF curves of SRI replaced with bar chart o New subgroup analysis for post-PA DBL • Section 14.2: o Clarification added for flare definition o Sensitivity analysis updated to use MI • Section 14.3: BILAG flare severity definition added; additional output for BILAG flare by visit and severity; additional analysis for BICLA (same logistic regression model as for primary endpoint); QoL: logistic regression removed for LupusQoL and added for PGIC with further details on MCID criteria; CS analysis: reference to Week 41 instead of W40 to be consistent through the document. • Section 14.6: Additional supportive analysis for post-PA DBL. • Appendices: clarification added for BILAG derivation (18.1.1); clarification added for SLEDAI derivation (18.1.3) and PGA

Unique Identifier for IAP Version	Date of IAP Version	Authors	Changes from the Previous Version
			(18.1.4); LLDAS definition updated to include EAC-treatment failure (18.1.7); BICLA definition clarified with new reference (18.1.8); clarification added for CS derivation rules (18.8); SAS codes updated; analysis visit for W52 extended (18.11); new appendices 18.16 & 18.17 for details on post-PA DBL analysis
Version 3.0	07 February 2020	PPD	Minor Clarification: All post-baseline efficacy response endpoints up to Week 52 (e.g. SRI responses, LDA, LDA def #2, LLDAS, BICLA, Remission) will use the similar imputation rules as applied for the Week 52 timepoint (e.g. primary endpoint, etc.)
Version 4.0	11 June 2020	PPD	Updates due to study termination: • Removal of Japanese cohort analysis and addition of safety summaries for APAC subjects • Changes in LTE analysis (changes in LTE baseline definition and efficacy endpoints) • Addition of COVID-19 impact analysis

4 Purpose of the Integrated Analysis Plan

The purpose of this Integrated Analysis Plan (IAP) is to document technical and detailed specifications for analyses of data collected for Clinical Trial Protocol (CTP) MS200527-0018 version 7.0 dated 22 May 2018. The IAP is based upon Section 8 (Statistics) of the trial protocol and is prepared in compliance with International Council for Harmonisation E9.

The previous version (version 3.0) of the IAP included details for all the planned analyses of this study as specified in the protocol but focusing on the Primary Analysis (PA) and analysis of Japanese cohort. This version (version 4.0) includes details of the analysis of Japanese cohort and Long-Term Extension (LTE) analysis considering the termination of the study, announced to the sites on March 23rd, 2020.

Results of the analyses described in this IAP will be included in the Clinical Study Report (CSR). Separate reports will be created to include the results of CCI modeling analysis, and additional biomarker analysis. Additionally, the planned analyses identified in this IAP will be included in regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective IAP will be clearly identified in the CSR.

The final clinical database cannot be locked until the final IAP has been approved and signed.

5 Objectives and Endpoints

Objectives and Endpoints are summarized in Table 1. Additional endpoints, not planned in the protocol, are summarized in Table 2.

Table 1: Objectives and Endpoints

	Objective	Endpoint	IAP Section
Primary Objective	To evaluate the efficacy and dose response of evobrutinib compared to placebo in reducing disease activity in adult subjects with active, autoantibody positive systemic lupus erythematosus (SLE) who are receiving standard of care (SoC) therapy based on SLE Responder Index (SRI)-4 response at Week 52 in all subjects, or on SRI-6 response at Week 52 in the High Disease Activity (HDA) subgroup, defined as SLE Disease Activity Index 2000 (SLEDAI-2K) ≥ 10	Primary Endpoint: The co-primary efficacy endpoints are: • SRI-4 response at Week 52 in all subjects • SRI-6 response at Week 52 in a HDA subgroup.	Section 14.1
	To evaluate the safety of evobrutinib in subjects with SLE on SoC therapy.	Primary Endpoint: • The primary safety endpoints are nature, severity, and incidence of adverse events (AEs), serious adverse events (SAEs); vital signs, electrocardiograms (ECGs); absolute values of and change from Baseline in serum total immunoglobulin (Ig) levels (IgG, IgA, IgM), total B cell counts, and clinical laboratory parameters.	Section 15
Key Secondary Objectives	To evaluate the efficacy and dose response of evobrutinib compared to placebo in delaying time to first severe flare during the Treatment Period, in subjects with SLE on SoC therapy, where a severe flare is defined as at least one British Isles Lupus Assessment Group (BILAG 2004) A in any organ system due to items that are new or worse, compared to the BILAG evaluation at the previous visit.	Key Secondary Endpoint: • Time to first severe flare, where a severe flare is defined as at least one BILAG A score in any organ system due to items that are new or worse, compared to the BILAG evaluation at the previous visit, during the Treatment Period.	Section 14.2
	To evaluate the efficacy and dose response of evobrutinib compared to placebo in reducing disease activity, based on the SRI-4 response at Week 52, in the serologically active subgroup, which is defined as subjects with positive anti-double-stranded deoxyribonucleic acid (anti-dsDNA) and/or low complement levels.	Key Secondary Endpoint: • SRI-4 response at Week 52, in the serologically active subgroup, which is defined as subjects with positive anti-dsDNA and/or low complement levels.	Section 14.2

Other secondary Objectives	• To evaluate the efficacy of evobrutinib compared to placebo on changes in disease activity over 52 weeks • To evaluate the efficacy of evobrutinib compared to placebo on changes in organ-specific disease activity over 52 weeks	Secondary Endpoints: • SRI-6 response at Week 52, in the serologically active subgroup, which is defined as subjects with positive anti-dsDNA and/or low complement levels. • SRI-4 Response at Week 52 with a Sustained Reduction of Oral Corticosteroids (OCS) Dose to 7.5 milligrams prednisone equivalent per day or less (≤ 7.5 mg/day) and less than or equal to (≤) Day 1 dose during Week 41 Through Week 52, in all subjects. • SRI-6 Response at Week 52 with a Sustained Reduction of OCS Dose to 7.5 milligrams prednisone equivalent per day or less (≤ 7.5 mg/day) and less than or equal to (≤) Day 1 dose during Week 41 Through Week 52, in the HDA subgroup, defined as SLEDAI-2K ≥ 10 at Screening. • SRI-4 Response at Week 52 with a Sustained Reduction of OCS Dose to 7.5 milligrams prednisone equivalent per day or less (≤ 7.5 mg/day) and less than or equal to (≤) Day 1 dose during Week 41 Through Week 52, in the serologically active subgroup, which is defined as subjects with positive anti-dsDNA and/or low complement levels. • Disease activity over time, during the Treatment Period, as measured by: ○ Low disease activity (LDA) status, defined by SLEDAI-2K ≤ 2, at Week 52 ○ LDA status, defined by clinical SLEDAI-2K (SLEDAI-2K excluding anti-dsDNA and low complement parameters) ≤ 2, at Week 52 ○ Lupus low disease activity state (LLDAS), defined as meeting all of the following (Franklyn 2016): ▪ SLEDAI-2K ≤ 4 ▪ No activity in any major organ systems (renal, central nervous system, cardiopulmonary, vasculitis, fever) ▪ No new features of disease activity compared with the previous assessment ▪ Physician's Global Assessment (PGA) ≤ 1 (on a 0-3 scale) ▪ Prednisone-equivalent ≤	Section 14.3

		7.5 mg/day ▪ Unchanged background immunosuppressive therapy ○ Change from Baseline in SLEDAI 2K score by visit ○ Change from Baseline in Cutaneous Lupus Erythematosus Disease Area and Severity Index-Activity (CLASI-A) by visit ○ BILAG-based Composite Lupus Assessment (BICLA) response by visit ○ Change from Baseline in BILAG 2004 by visit ○ Change from Baseline in PGA by visit.	
	To evaluate the effect of evobrutinib compared to placebo on the annualized flare rate	Secondary Endpoint: • Time to first flare, flare-free status at Week 52, and annualized flare rate (AFR), during the Treatment Period, will be analyzed separately, each assessed with flare defined as: ○ BILAG A Severe flare ○ BILAG A or 2B Moderate to Severe flare ○ SLEDAI Flare Index (SFI) Severe flare.	Section 14.3.2
	To evaluate the impact of evobrutinib treatment compared to placebo on subject reported health related quality of life (HRQoL) over 52 weeks	Secondary Endpoint: • HRQoL over time, during the Treatment Period, as measured by: ○ Change from Baseline in Medical Outcomes Study 36-Item Short Form Health Survey (SF-36v2®) Physical Component Summary (PCS) and Mental Component Summary (MCS) score (and their components) by visit ○ Change from Baseline in EuroQoL 5 Dimension 5 Levels (EQ-5D-5L) score by visit ○ Change from Baseline in Lupus Quality of Life (LupusQoL) scores by visit ○ Patient Global Impression of Change (PGIC) score by visit ○ Change from Baseline in Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue score by visit.	Section 14.3.4
	To evaluate the effect of evobrutinib on corticosteroid (CS) usage over 52	Secondary Endpoint: • Corticosteroid usage over time, during	Section 14.3.5

	weeks.	the Treatment Period, as measured by: ○ Reduction from Baseline in prednisone-equivalent CS dose by ≥ 25% to a dose of ≤ 7.5 mg/day, with no BILAG A or 2B flare in disease activity (at that visit) ○ Change from Baseline to Week 52 in prednisone-equivalent CS daily dose ○ Reduction from Baseline to Week 52 in prednisone equivalent CS daily dose of zero to < 25%, 25% to 50%, > 50%, or an increase ○ Cumulative prednisone-equivalent CS dose from Baseline until completion of the Treatment Period ○ Clinically meaningful reduction in CS dose from Baseline, defined by: ▪ A reduction of daily prednisone-equivalent CS dose ≥ 25% to a dose of ≤ 7.5 mg/day by Week 40 and sustained through Week 52 And ▪ No new BILAG A organ domain scores and no more than one new BILAG B organ domain score during Weeks 40 through 52.	

Exploratory Objectives	CCI [redacted]


CCI

Open-label Long-Term Extension (LTE) Period Objectives	To evaluate the long-term safety, efficacy, and HRQoL of evobrutinib at an initial dose of 50 mg twice daily or the eventual Phase III dose when decided for an additional two years.	Endpoints for LTE Period:
		
		<u>LTE Safety:</u>
		• Nature, severity, and incidence of AEs, SAEs, vital signs, ECGs, absolute values and change from Baseline in serum total Ig levels (IgG, IgA, IgM) and total B cell counts, and clinical laboratory parameters.
		
		<u>LTE Efficacy:</u>
		• The following LTE efficacy endpoints will be analyzed at Week 24, Week 52, and Week 104:
		  ○ Changes over time in SRI response
		  ○ Changes over time in Low Disease Activity status (LLDAS, SLEDAI-2K≤2, clinical SLEDAI-2K≤2).
		  ○ Changes over time in CLASI-A, CLASI-D, and SLICC/ACR Damage Index organ damage scores
		  ○ Changes over time in disease activity as measured by the BILAG, SLEDAI-2K, and PGA
		  ○ Change over time in prednisone-equivalent CS dose
		  ○ Changes over time in HRQoL
		  ○ Changes over time in autoantibodies and complement levels
		• Changes in HRU by visit, including but not limited to doctor/home/emergency visits, hospitalizations, paid assistance, and missed work
		• Time to first flare; flare-free status at Weeks 24, 52, and 104; and annualized flare rate, will be analyzed separately, each assessed with flare defined as:
		  ○ BILAG A Severe flare
		  ○ BILAG A or 2B Moderate to Severe flare
		  ○ SLEDAI Flare Index (SFI) Severe flare.
			Section 0

Table 2 : Additional Exploratory Endpoints not Planned in the Protocol

	Endpoint	IAP Section
Additional Exploratory Endpoints	CCI [redacted]	[redacted]

6 Overview of Planned Analyses

6.1 Independent Data Monitoring Committee review

An IDMC was established to assess safety and tolerability of IMP during the conduct of the study. Details of the statistical analyses of the IDMC are provided in a separate document (latest version of MS200527-0018_IDMC_SAP).

6.2 Primary Analysis

The Primary Analysis (PA) will occur only when:

- 100% of subjects in the primary analysis cohort (PAC):
  - Complete Week 52 of treatment and either enter LTE or complete Safety Follow-Up, or
  - Prematurely discontinue from treatment prior to Week 52 and complete Safety Follow-Up, or
  - Prematurely discontinue from study without Safety Follow-Up.

  The PAC is defined as all subjects who have been randomized till November 28, 2018 (see Section 8.2.2). Subjects not randomized (screen failure) will be included in the PAC if their screening date is on or before November 28, 2018.
- The protocol violations are determined,
- The database is locked.

Data collected during the 52 weeks treatment period as well as during the corresponding safety follow-up visit will be included in this analysis. Details will be provided in the PA Interim Database Lock Plan.

All endpoints as specified in the protocol will be analyzed for the PA. The unblinding process after PA will be detailed in the Unblinding Plan.

6.3 Analysis of Japanese Cohort

The plan was to perform two Japanese cohort analyses: 1) Partial Japanese cohort analysis at the time of primary analysis; and 2) Final Japanese cohort analysis at the time of final analysis. However, after review of PA results, the study was terminated. The data of the Japanese cohort will be included in data listings at the time of the LTE analysis when all subjects either completed or early terminated the study and the full database is locked.

The analysis of the partial Japanese cohort was performed with the Japanese subjects available in the database of primary analysis cohort who:

- Completed Week 52 of treatment and either entered LTE or completed Safety Follow-Up, or

- Prematurely discontinued from treatment prior to Week 52 and completed Safety Follow-Up, or
- Prematurely discontinued from study without Safety Follow-Up.

Data collected during the 52 weeks treatment period as well as during the corresponding safety follow-up visit was included in this analysis. Details were provided in the Interim Database Lock Plan for Analysis of Japanese Cohort.

Certain steps were taken to ensure protection of the blind of the Japanese cohort subjects who would be followed-up in a double-blind manner after the PA. Details can be found in the Unblinding Plan.

6.4 LTE Analysis

After review of PA results, the study was terminated. The LTE analysis will occur when all subjects either completed or early terminated the study and the full database is locked.

.

7 Changes to the Planned Analyses in the Clinical Trial Protocol

- The protocol states there may be an Interim Analysis (IA) for futility based on the highest dose of M2951, triggered when 100% of subjects enrolled in the PAC reach Week 24 of treatment, or prematurely discontinue from treatment prior to Week 24. It has been decided by Merck/EMD-Serono team not to perform this IA.

- The protocol refers to the co-primary endpoints as SRI-4 Response and SRI-6 Response for HDA subjects. Instead, they will be referred as two-primary endpoints in this IAP, according to the latest FDA guidance on multiplicity.

- The protocol mentions the McNemar's test to assess the change in a binary endpoint between two timepoints. For other secondary endpoints, the test will not be performed, only descriptive statistics will be provided.

- It is no longer planned that whole blood will be analyzed for phosphoproteomics or B cell activation and these endpoints will not be reported.

- The LLDAS status as defined in Section 18.1.7 must include the PGA ≤ 1 (0-3 scale) in the definition. This item was not mentioned in the protocol and has been added to this IAP.

- The protocol refers to the endpoint 'Changes over time in SRI response/in LDA status' for the LTE analysis but will be referred as 'SRI response/LDA status over time' in this IAP, for accuracy.

- The protocol states the PAC used for PA consists of the first 432 subjects randomized. However, due to a potentially higher than expected drop-out rate, the PAC will include all subjects screened/randomized as of November 28, 2018 (n > 432), to protect the power of the statistical analysis for which the sample size was calculated.

- The protocol states the definition of the subgroup severity of disease at screening (severe, mild/moderate), but a more detailed definition has been added to this IAP (see Section 8.2.3), for correctness.

- The protocol mentions a sensitivity analysis of the two-primary endpoints on completers. This analysis will not be performed as the primary analysis already considers subjects who discontinued treatment as non-responder.

CCI [redacted]

- The protocol specifies that flare-free status will be analyzed separately depending on severity, but in this IAP flare-free status is defined as not having any flares (defined with BILAG 2004 and SFI, separately) during the 52-week treatment period.

- The protocol mentions several endpoints for the analysis of corticosteroid usage over time. The first endpoint (i.e., reduction from baseline in prednisone-equivalent CS dose by ≥ 25% to a dose of ≤ 7.5 mg/day, with no BILAG A or 2B flare in disease activity at that visit) is removed from this IAP. Analysis of the clinically meaningful reduction in CS dose from Baseline (defined as a reduction of daily prednisone-equivalent CS dose ≥ 25% to a dose of ≤ 7.5 mg/day by Week 40 and sustained through Week 52 AND no new BILAG A organ domain scores and no more than one new BILAG B organ domain score during Weeks 40 through 52) will be sufficient.

- SRI definition specified in the protocol includes 'no discontinuation of investigational product', but this condition has been removed from this IAP (see Section 14.1.1). Subjects with missing response at Week 52 have been added as a condition of non-SRI responder, meaning that all subjects who prematurely discontinued treatment will be considered as non-SRI responder.

- For definition of SRI responses, LDA Definition #2, LLDAS and BICLA, EAC-treatment failure assessment will be used to consider subjects as non-responder in these 3 endpoints.

- Due to study termination, the analysis of Japanese cohort and LTE analysis will be performed simultaneously. The Japanese cohort will not be explored for consistency, the listings produced for the PA will be rerun including all Japanese subjects. A safety summary of Asia-Pacific subjects will be presented during the 52-Week treatment period (see Section 14.1.4). The LTE analysis will be reduced with a focus on safety, further details can be found in Sections 14.6 and 15.

8 Protocol Deviations and Analysis Sets

8.1 Definition of Protocol Deviations and Analysis Sets

Important protocol deviations are protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being. Important protocol deviations are defined in a separate document (latest version of MS200527-0018_List_of_IPDs).

All deviations will be identified and confirmed prior to or at the Data Review Meeting (DRM), which will occur before the database lock, including clinically important deviations leading to the exclusion of a subject from Per-Protocol (PP) analysis set (see Section 8.2).

The outcome of the DRM will document the important protocol deviations as well as the finalization of the analysis populations in a memo. Important protocol deviations will be documented in Clinical Data Interchange Standards Consortium Study Data Tabulation Model (SDTM) whether identified through sites monitoring, medical review and/or programming.

8.2 Definition of Analysis Sets and Subgroups

8.2.1 Sample Size

A sample size of 103 evaluable subjects per group provides 80% power at the $\alpha = 0.025$ one-sided significance level to detect

- an absolute improvement of 20% in Week 52 SRI-4 response proportion for M2951-treated subjects relative to placebo-exposed subjects, among subjects with SLEDAI-2K total score ≥ 6 at Screening, assuming a placebo response proportion of 40%,

or

- an absolute improvement of 25% in Week 52 SRI-6 response proportion for M2951-treated subjects relative to placebo-exposed subjects, among subjects with SLEDAI-2K total score ≥ 10 at Screening (i.e., HDA), assuming a placebo response proportion of 30%,

assuming that each two-primary endpoint is tested via a chi-squared test of the odds ratio (OR) for treatment effect at the 0.0125 one-sided level, and assuming that approximately 50% of randomized subjects are in the HDA group. If the absolute improvement in Week 52 SRI-6 response proportion in HDA subjects is only 20%, the power is 76%. Approximately 108 subjects will be randomized per treatment group to protect against a loss of information due to drop-out (for reasons other than efficacy/safety) of 5% over 52 weeks. Given the randomization ratio 1:1:1:1, the total sample size required for the PA is planned to be 432 subjects.

CCI [redacted]

CCI [redacted]

- a total evaluable sample size of 412 (i.e., 309:103 randomization for M2951:placebo) is involved in analysis of the two regions,

  and applying "Method 2" from the PMDA Guidance, 32 evaluable subjects in the Japan cohort are required so that both observed region-specific effect sizes exceed 0.04 with probability of 80%. Taking into account a loss of information due to 5% drop-out at Week 52, for reasons unrelated to efficacy or safety, the total number of Japan subjects to be randomized is 36, or 8.3% of the total planned enrollment of 432.

CCI [redacted]

8.2.2 Analysis Sets

Table 3 provides the definitions of the Analysis Sets and their purposes.

The analysis sets for PA will be drawn from all subjects randomized as of November 28, 2018 (n = 469), see Sections 6.2 and 7. Subjects not randomized (screen failure) will be included in PA analysis set if their screening date is on or before November 28, 2018. For the LTE analysis, the analysis sets will be drawn from all subjects who enter the LTE.

For analysis on HDA subpopulation, all subjects with SLEDAI-2K ≥ 10 at Screening will be included. The corresponding analysis sets will be drawn as specified above depending on the analysis (PA, LTE analysis).

Table 3: Analysis Sets

Analysis Set	Definition	Purpose
Enrolled	The Enrolled analysis set will include all subjects who sign the informed consent form.	The purpose of this analysis set is to count how many subjects signed the informed consent

ITT (Intent-To-Treat) Analysis Set	The ITT analysis set will include all randomized subjects. Subjects will be analyzed according to randomization treatment.	The ITT will be used for demographic/baseline characteristics.
mITT (modified Intent-To-Treat) Analysis Set	The mITT analysis set will include all randomized subjects who have received at least one dose of IMP (evobrutinib or placebo) and have at least one Baseline and one post-Baseline disease assessment (among the following: SFI, SLEDAI-2K, PGA, BILAG 2004, CLASI). Subjects will be analyzed according to randomized treatment. Subjects with Protocol Deviation (PD) code equal to PDEV80 are excluded from mITT.	The mITT will be used for all efficacy endpoints. The mITT will be used for demographic/baseline characteristics in case more than 10% of ITT are excluded from mITT.
Safety Analysis Set (SAF)	The Safety analysis set will include all subjects who have received at least one dose of IMP (evobrutinib or placebo). Subjects will be analyzed according to the actual treatment they receive.	The SAF will be used for the analysis of medications, exposure/compliance and safety endpoints.
QoL (Quality of Life) Analysis Set	The QoL analysis set will include all ITT subjects who have received at least one dose of IMP (evobrutinib or placebo) and have at least one Baseline and one post-Baseline QoL assessment (among the following: SF-36, EQ-5D-5L, LupusQoL, PGIC, FACIT). Subjects will be analyzed according to randomized treatment. Subjects with PD code equal to PDEV80 are excluded from QoL.	The QoL analysis set will be used for the analysis of QoL data.
PP (Per-Protocol) Analysis Set	The PP analysis set will include all subjects from mITT who do not have any clinically important protocol deviations (see Section 10.2). Subjects will be analyzed according to randomized treatment.	The PP will be used for the two-primary and key secondary efficacy endpoints. Efficacy analyses on PP will be performed if at least 10% of the mITT are excluded from PP. Demographic/baseline characteristics analyses will be repeated for the PP if more than 10% of mITT are excluded from PP.
CCI [redacted]		

CCI

CCI

CCI

It has been notified to the Sponsor that medical/scientific misconduct occurred at sites 653 and 107 during the study. CCI . As a result, subjects who have been enrolled in such sites will be included in the SAF and ITT analysis sets but excluded from mITT (and therefore PP), QoL, CCI CCI and CC analysis sets.

8.2.3 Subgroup Analysis

Table 4 below defines the subgroups to be used for subgroup analysis as described in Section 14 and Section 15.

Subgroup analysis on the stratification factors race (Black versus non-Black) and, disease activity prior to dosing on Day 1 (SLEDAI-2K total score < 10 versus ≥ 10 at Screening) will be defined using the eCRF data. The stratification factor region (US and Western Europe, Japan, and RoW) will be defined using Interactive Web Response System (IWRS) as only collected as part of randomization data.

Table 4: Subgroups and Strata

Subgroup	Content	Definition
A	Disease Activity at Screening (HDA) (stratification factor)	non-HDA: SLEDAI-2K total score < 10; HDA: SLEDAI-2K total score ≥ 10 **(from SLEDAI1 form of the eCRF)** *Some analysis will be presented on HDA subpopulation (i.e. including all HDA subjects) as detailed in Sections 14 and 15.*
B	Serological Activity status at Screening	Positive: positive anti-dsDNA (i.e. > 15 U/mL) OR complement levels (C3 or C4 below (<) LLN); Negative: otherwise, i.e. negative/indeterminate anti-dsDNA and C3/C4 complement levels above or equal (>=) LLN **(from Central Laboratory**: anti-dsDNA, serum C3 and serum C4**)**
C	Race (stratification factor)	Black or African American; other **(from DEM form of the eCRF)**
D	Region (stratification factor)	US and Western Europe; Japan; ROW **(from IWRS)**
E	Ethnicity 1	Hispanic/Latino VS Non-Hispanic/Latino (**from DEM form of the eCRF**)
F	Background Therapy of Immunosuppressant	Yes; No (derived from Concomitant Medication form of eCRF, using the L04A ATC code, see Section 14.1.2.2 for derivation detail)

IWRS = Interactive Web-Response System

CCI [redacted]

Table 5: Exploratory Subgroups and Strata

Subgroup	Content	Definition
CCI		
CCI		

9 General Specifications for Data Analyses

All statistical analyses will be performed by PPD Inc., except the modeling and simulation analysis and the gene expression, which will be performed by Merck/EMD Serono or delegate.

CCI [redacted]

9.1 Actual Treatment Assignment

A subject who received 2 different types of treatment regimen over the course of treatment should be tabulated according to the treatment regimen received most frequently. If there is a "tie", the highest dose will be chosen.

9.2 Presentation of Tables/Listings/Figures (TLFs)

Tables and figures will be sorted by treatment group (in the order stated in Table 6) and chronological time point (where applicable).

Table 6: Treatment Groups and Regimens

Treatment Group	Regimen	Presentation for the PA	Presentation for the LTE analysis (*)
1	Placebo	Placebo	Placebo/Evobrutinib 50 mg BID
2	Evobrutinib 25 mg QD (once daily)	Evobrutinib 25 mg QD	Evobrutinib 25 mg QD/Evobrutinib 50 mg BID
3	Evobrutinib 75 mg QD	Evobrutinib 75 mg QD	Evobrutinib 75 mg QD/Evobrutinib 50 mg BID
4	Evobrutinib 50 mg BID (twice daily)	Evobrutinib 50 mg BID	Evobrutinib 50 mg BID/Evobrutinib 50 mg BID
(*): Subjects who complete the 52-week Double-blind placebo-controlled period will be offered participation in an optional open label, LTE period for an additional 2 years. During this LTE period, subjects will be administered evobrutinib at an initial dose of 50 mg BID or the eventual Phase III dose, when decided.			

All data recorded during the trial will be presented in individual data listings (in the order stated in Table 6). All listings are sorted by treatment group, subject, and time point (where applicable), if not otherwise stated.

9.3 Data handling for the planned analyses

For the PA, data will be handled according to the corresponding Data Handling Report and Interim Database Lock Plan.

The stratification factors race (Black versus non-Black) and disease activity prior to dosing on Day 1 (SLEDAI-2K total score < 10 versus ≥ 10 at Screening) will be summarized using both the eCRF data and the IWRS data in demographics and baseline characteristics. However, for statistical analyses, eCRF data will be used. The stratification factor region (US and Western Europe, Japan, and RoW) will be defined using IWRS. (See Section 8.2.3, Table 4)

9.4 Presentation of continuous and qualitative variables

Continuous variables including CC but not CC will be summarized using descriptive statistics, i.e.,

- number of subjects (N)
- number and percentage of non-missing values
- number and percentage of missing values
- mean, standard deviation (SD)
- median, 25th Percentile - 75th Percentile (Q1-Q3)
- minimum and maximum

The number of digits for non-derived and derived data, presented in outputs or available in Analysis Data Model (ADaM) datasets, is specified in the Biostatistics Outputs Assembly (BOA) document. For efficacy endpoints from Section 14, median and SD will be presented with 1 more digit compared to the original data, whereas mean, min, max, Q1-Q3 will be presented with the same number of digits as the original data.

For both continuous and qualitative variables, percentages such as 0% or 100% should be reported with the same format used for the column, together with the count of observations. For example, if the count of observations is zero, then display '0 (0.0)'; if the count of observations represents 100%, then display 'xx (100.0)'.

Qualitative variables will be summarized by counts and percentages. Unless otherwise stated, the calculation of proportions will be based on the number of subjects in the analysis set of interest. Therefore, counts of missing observations will be included in the denominator and presented as a separate category.

In case the analysis refers only to certain visits, percentages will be based on the number of subjects in the analysis set of interest, unless otherwise specified. Total of missing and non-missing observations at each time-point will reflect the population still in the trial at that time. For example, if a subject is still in the trial at the time-point but with missing data, it should be counted in the number of missing observations.

CCI

9.5 Definition of Screening

Screening during the 52-week treatment period for the PA: last non-missing value before baseline.

9.6 Definition of Baseline

- Baseline during the 52-week treatment period for the PA: last non-missing value on the day of or prior to first administration of evobrutinib or placebo. Of note, time of first dosing administration and assessment of baseline (both done on the same baseline day) will not be used in the determination of baseline value, unless otherwise specified for specific analysis.
- Baseline during the LTE period for the LTE analysis: the baseline of the 52-week treatment period will be used.

9.7 Other Specifications

Derivation of Age (years):

Age will be derived according to the following formula:

$$Age\ (Years) = \frac{(date\ of\ given\ informed\ consent - date\ of\ birth + 1)}{365.25}$$

In case of missing day for the date of birth, but month and year available, the day of birth will be set to 1 in the formula above. In case of missing day and month, they will both be set to 1.

Definition of Change from Baseline (CFB):
CFB and percent CFB will be computed as follows:

- CFB = visit value – baseline value
- Percent CFB = 100 * (visit value – baseline value) / baseline value

Definition of duration:
Unless otherwise specified, duration will be calculated as the difference between start and stop date + 1 (eg, duration of AE (days) = AE stop date – AE start date + 1). Unless otherwise specified, missing dates will not be imputed.

Conversion factors:
The following conversion factors will be used to convert days into months or years: 1 month = 30.4375 days, 1 year = 365.25 days. For time windows calculation, 1 month is expressed as 30 days.

Definition of Body Mass Index (BMI) (kg/m²):
BMI is defined as weight / (height²), where weight is expressed in kg and height in m. The BMI is available in the vs1 form of the Case Report Form, electronic (eCRF) and will not be re-derived.

Handling of missing data:
For efficacy analysis, methods of missing data handling are detailed in Section 14.

Unless otherwise specified, missing data will not be replaced.

In all subject data listings, imputed values will be presented and will be flagged. Non-imputed partial dates will be presented in a format such as "_____YYYY". Where imputation is defined, an imputed date with flag (i.e., D for day, M for month) will be reported.

Missing statistics, e.g., when they cannot be calculated, should be presented as 'nd' (denoting "Not Done"). For example, if n=1, the measure of variability (SD) cannot be computed and should be presented as 'nd'. In case of zero records, an empty output with 0 occurrences, or a sentence mentioning that there are no data, will be presented. For tables of AEs and Deaths (outputs required for EudraCT and/or clinicaltrial.gov), if there are no observations, the output must contain the first line 'Subjects with…' or 'Subjects who died' displayed with 0 occurrence.

If the System Organ Class (SOC) or Anatomical Therapeutic Class (ATC) term is missing/not coded yet, then 'Uncoded SOC' (or 'Uncoded ATC') will be indicated at the TLF level. When a Preferred Term (PT) is missing, it will be set to 'Uncoded PT: verbatim text'.

Treatment day:
Treatment day is defined as relative day to the start date of treatment. Treatment day will be calculated as the number of days since the first administration of any IMP. The day before is defined as Treatment day –1 (no Treatment day 0 is defined). Treatment day will only be presented in subject safety data listings.
For the 52-week treatment period, start date of treatment is defined as the date of first administration of randomized study drug.
For the LTE treatment period, start date of treatment is defined as the date of first treatment administration on or after the date of end of 52-week treatment.

On-Treatment flag:
On-treatment values are results of assessments done from the first IMP administration on Day 1 till End of Study (completion or early termination).

Repeated and unscheduled measurements:
Repeated and unscheduled measurements are included in the listings. Data collected at both unscheduled and scheduled visits will be used for shift tables. For summary statistics, figures, or inferential analysis, data are re-allocated to an analysis visit using time window calculations (see below).

Gap between 52-Week treatment period and LTE treatment period:
Subjects with a gap greater than or equal to 2 weeks between 52-Week treatment period and LTE period will be analyzed the same way as subjects without a gap. The gap will be defined using the last dose received during the 52-week treatment period and the first dose received during the LTE treatment period. Specific data listings on safety data (AEs and laboratory data) collected during the gap period will be produced for these subjects for the LTE analysis. In addition, for the subgroup of subjects with a gap greater than or equal to 2 weeks, the following analyses will be provided:

- Summary statistics of the gap duration
- SLE disease baseline characteristics at LTE Day 1
- SRI-4 response and SRI-6 response among HDA subjects by LTE visit

Time windows:
For efficacy and safety analyses, each measurement will be assigned an analysis visit number first according to the prespecified time window. The analysis visit will then be used for missing data imputations, analysis variable derivations, statistical calculations and presentations. Safety follow-up will not be included in analysis visit calculations, only assessments collected under treatment visits and no more than 14 days after last exposure date will be used.
For by visit summary of safety endpoints (e.g. laboratory and vital sign assessments), data collected during the safety follow-up will be shown, presenting separately data from subjects who completed safety follow-up and who discontinued during safety follow-up.
For subject data listings by time point, the nominal visit (as collected in the database) as well as the analysis visit will be displayed.
For the calculation of analysis visit using the time windows, 1 month is expressed as 30 days.
As the schedule of assessments is different for each visit, specific time windows must be used for each endpoint. They are defined in Appendix 18.11.
If there are multiple assessments within a same visit window, then the one closest to the target day will be used. If two assessments have the same difference with target day, the earlier one will be used.

Unblinding:
Details regarding the unblinding process are available in the Unblinding Plan.

CCI [redacted]

10 Trial Subjects

The subsections in this section include specifications for reporting subject disposition and treatment/trial discontinuations. Additionally, procedures for reporting protocol deviations are provided.

10.1 Disposition of Subjects and Discontinuations

For each analysis (including analysis of HDA subpopulation), a table on screened subjects describing the number and percent of subjects in each disposition category will be produced by treatment group.

These categories are summarized in Table 7. In addition, the cumulative number and percentage of subjects who discontinue treatment during the 52-week treatment period will be provided by visit for each treatment group.

Table 7: Disposition

	PA	HDA subpopulation analysis	LTE analysis
• Total number of screened subjects, i.e., subjects that gave informed consent (overall summary only)	✓		
• Number of subjects who discontinued prior to randomization and reason (overall summary only)	✓		
• Number of randomized subjects	✓	✓	
• Number of randomized subjects who did not start treatment	✓	✓	
• Number of subjects who permanently discontinued treatment during 52-week treatment period and reason	✓	✓	
○ Number of subjects who completed safety follow-up for the 52-week treatment period	✓	✓	
○ Number of subjects who discontinued during safety follow-up for the 52-week treatment period and reason	✓	✓	
• Number of subjects who completed 52-week treatment period	✓	✓	
○ Number of subjects who agreed to participate in LTE	✓	✓	
○ Number of subjects who declined to participate in LTE	✓	✓	
▪ Number of subjects who completed safety follow-up for the 52-week treatment period	✓	✓	
▪ Number of subjects who discontinued during safety follow-up for the 52-week treatment period and reason	✓	✓	
• Number of subjects who entered the LTE*			✓
• Number of subjects who completed 52-week treatment period			
• Number of subjects who agreed to participate in LTE			✓
○ Number of subjects who completed safety follow-up for the 52-week treatment period			✓
○ Number of subjects who discontinued during safety follow-up for the 52-week treatment period and reason			✓
• Number of subjects who permanently discontinued treatment during the LTE period and reason			✓
○ Number of subjects who completed safety follow-up for the LTE period			✓
○ Number of subjects who discontinued during safety follow-up for the LTE period and reason			✓
* Subjects who entered the LTE are defined as subjects who signed the LTE inform consent.			

A table based on screened subjects, describing the number and percent of subjects in each analysis set by treatment group, will be produced, according to Table 8:

Table 8: Disposition - Analysis Sets

	PA	HDA subpopulation analysis	LTE analysis
Number of enrolled subjects	✓	✓	
Number of subjects included in the ITT Analysis Set (randomized subjects)	✓	✓	✓
Number of subjects included in the mITT Analysis Set	✓	✓	✓
Number of subjects included in the SAF	✓	✓	✓
Number of subjects included in the PP Analysis Set	✓		
CCI			
CCI			
Number of subjects included in the QoL Analysis Set	✓		
CCI			

The following summaries will be provided:

- Number and percentages of subjects in each analysis set (except for Enrolled subjects) by region, country and site
- Number and percentages of randomized subjects by randomization strata (as defined by groups A, C and D from Table 4) and country

Corresponding individual listings will be prepared, according to Table 9:

Table 9: Disposition Subject Data Listings

Description of the listing	PA	HDA subpopulation analysis	LTE analysis
Discontinued subjects (from treatment or study) will be listed with their reason for withdrawal (from treatment or study).	✓	✓	✓
A listing of the subjects screened but not randomized will be produced with the reason for not being randomized.	✓	✓	
A listing of randomized subjects with subject number, randomization date, and randomized treatment group, ordered by randomization number within randomization stratum, will be produced to assess whether randomization was conducted as planned.	✓	✓	
The list of re-screened subjects and corresponding subject identifiers will be provided. Only the second subject identifier will be used in statistical descriptions and analyses, while the first identifier will not be considered from the disposition counts.	✓		

10.2 Protocol Deviations

Table 10 summarizes how protocol deviations will be handled for the PA and LTE analysis.

Table 10: Protocol Deviations

Analysis	Period covered for protocol deviations reporting	Treatment Group	Observation Period	Analysis Set
PA	52-week treatment period	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	First dose of 52-week treatment period up to last observed dose during the 52-week treatment (safety follow-up included)	ITT
LTE analysis	LTE	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/ Evobrutinib 50 mg BID	First dose of LTE to last observed dose (safety follow-up included)	ITT

10.2.1 Important Protocol Deviations

The following summary tables and listings of important protocol deviations will be provided (separately for pre-/post inclusion deviations):

- Tables providing frequency for each type of important protocol deviation/clinically important protocol deviation
- Listing of important protocol deviations

10.2.2 Clinically Important Protocol Deviations

Clinically important protocol deviations, as defined in Section 8.1, will be summarized and listed. Applicable clinically important protocol deviations, based on selected deviations including protocol-prohibited medications, will be confirmed by an Endpoint Adjudication Committee (EAC) for the PA only. Other applicable clinically important protocol deviations will be confirmed by the Sponsor. Further details can be found in the EAC Charter.

For subjects excluded from the PP (based on mITT), the reasons for exclusion will be summarized and listed:

- Frequency table per reason of exclusion from the PP population
- Listing of reasons of exclusion from the PP population

For subjects excluded from the mITT, the reasons for exclusion will be listed.

11 Demographics and Other Baseline Characteristics

Demographics and other baseline characteristics will be summarized on ITT by treatment group and overall. For PA, it will be repeated on mITT if more than 10% of ITT subjects are excluded from mITT, and on PP if more than 10% of mITT subjects are excluded from PP.

11.1 Demographics

The demographic characteristics are summarized in Table 11.

Table 11: Demographic Characteristics

Demographic Characteristic	Modality
Gender	From DEM form of the eCRF: • Male • Female
Race	Modalities will be presented according to the DEM form of the eCRF, i.e.: • White, • Black or African American, • Asian, • American Indian or Alaska Native, • Native Hawaiian or Other Pacific Islander, • Not collected at this site, • Other
	From IWRS: • Black • Non-black
Ethnicity 1	From DEM form of the eCRF: • Hispanic or Latino • Not Hispanic or Latino
Ethnicity 2	From DEM form of the eCRF: • Japanese • Not Japanese
	From IWRS: • Japanese • Not Japanese
Age as a continuous variable	Summary statistics will be provided
Age in categories	• < 65 years • ≥ 65 years
Region	**From IWRS**: • US and Western Europe • Japan • RoW
European Economic Area (EEA)	• Yes • No

11.2 Other Baseline Characteristics

11.2.1 Disease History

SLE disease characteristics at Baseline will be presented in 4 parts according to Table 12, Table 13,

Table 14 and Table 15. Unless indicated otherwise, these characteristics will be presented according to the definition of baseline in Section 9.5.

For LTE analysis, the SLE disease characteristics will be produced using Baseline from 52-week treatment period and using Baseline from LTE Day 1. The CLASI total activity score (continuous and categorical) will also be produced using Baseline from 52-week treatment period and using Baseline from LTE Day 1.

The SLE disease characteristics summary at LTE Day 1 will be presented for the subjects with a gap greater than or equal to 2 weeks between 52-week and LTE treatment periods.

Table 12: SLE Disease Characteristics at Baseline

SLE Disease Characteristic at Baseline	Modality
Time (months) since confirmed diagnosis of SLE@	• Time will be computed as (Date of Informed Consent Signature – Date of confirmed diagnosis of SLE) / 30.4375. If the date of confirmed diagnosis of SLE is entirely missing, then time since confirmed diagnosis of SLE will not be computed. In case of partial missing date of confirmed

	diagnosis of SLE, this date will be imputed as follows: ○ If only the day is missing, then it will be replaced by the first day of the month ○ If both day and month are missing, then it will be replaced by the first of January Date of Informed Consent Signature will be obtained from DEM form of the eCRF. Date of confirmed diagnosis of SLE will be retrieved from DIAG form of the eCRF.
PGA Score (Baseline)	Converted value will be presented (from SLEDAI1/SLEDAI3 form of the eCRF)
Prednisone-equivalent CS daily dose (Baseline)	Continuous
Number of subjects with prednisone-equivalent CS dose categories 1 (Baseline)	• ≤7.5 mg/day, • >7.5 mg/day. • Missing
Number of subjects with prednisone-equivalent CS dose categories 2 (Baseline)	• 0 mg/day, • > 0 to 7.5 mg/day, • > 7.5 mg/day. • Missing
Number of subjects with prednisone-equivalent CS dose categories 3 (Baseline)	• 0 mg/day, • > 0 to 6 mg/day, • > 6 to 7.5 mg/day, • > 7.5 to 10 mg/day, • > 10 to 15 mg/day, • > 15 to 20 mg/day, • > 20 to 30 mg/day, • > 30 mg/day. • Missing
Number and percentage of subjects taking immunosuppressants (excluding anti-malarials) (Baseline)	• Yes • No
Number and percentage of subjects taking anti-malarials (Baseline)	• Yes • No
Number of subjects with urine protein to creatinine ratio (UPCR) (Baseline)	• <0.5 mg/mg, • ≥0.5-1 mg/mg, • ≥1-1.5 mg/mg, • ≥1.5-2 mg/mg, • ≥2-3 mg/mg,

	• ≥3-4 mg/mg, • ≥4 mg/mg, • Missing.
Number of subjects with negative, indeterminate, positive or missing anti-dsDNA antibodies (Screening)	• Negative defined as < 10.0 U/mL, • Indeterminate as ≥ 10.0 and ≤ 15.0 U/mL • Positive defined by value > 15 U/mL • Missing
Antinuclear antibodies (Screening)	• Negative (< 1:80), • Positive (≥ 1:80), • Missing
Antinuclear antibodies and/or anti-dsDNA antibodies (Screening)	• Negative antinuclear antibodies and negative anti-dsDNA antibodies, • Positive antinuclear antibodies or positive anti-dsDNA antibodies (at least one test result is needed), • Missing antinuclear antibodies and missing anti-dsDNA antibodies.
Anti-Smith (Screening)*	• Negative (< 1.0 AI), • Positive (≥ 1.0 AI), • Missing
anticardiolipin IgG (Baseline)	• Negative (≤ 14 G phospholipids [GPL] U/mL), • Intermediate (15-20 GPL U/mL), • Low to medium positive (21-80 GPL U/mL), • High positive (> 80 GPL U/mL) • Missing
anticardiolipin IgM (Baseline)	• Negative (≤ 12 M phospholipids [MPL] U/mL), • Intermediate (13-20 MPL U/mL), • Low to medium positive (21-80 MPL U/mL), • High positive (> 80 MPL U/mL), • Missing
Antinuclear antibody associated with autoimmune diseases including Sjögren syndrome and systemic lupus erythematosus (anti-Ro/SSA antibody) (Baseline)	• Negative (< 1.0 AI), • Positive (≥ 1.0 AI), • Missing
Antinuclear antibody associated with autoimmune diseases including Sjögren syndrome (anti-La/SSB antibody) (Baseline)	• Negative (< 1.0 AI), • Positive (≥ 1.0 AI), • Missing

Low C3 complement (Screening)	• Low (< LLN) • Normal (within normal ranges [NRs]) • High (>ULN) • Missing (NRs for C3: 90-180 mg/dL)
Low C4 complement (Screening)	• Low (< LLN) • Normal (within NRs) • High (>ULN) • Missing (NRs for C4: 10-40 mg/dL)
Serologically active subgroup (Screening) @	• Yes/No (serologically active is defined as anti-dsDNA positive OR low complement levels [C3 < LLN and/or C4 < LLN])
CCI	
* Anti-Smith test at Screening was added during a protocol amendment after the study start, therefore missing values are expected. @ Not to be repeated for LTE Day 1 summary.	

Table 13: ACR Classification Criteria at Screening

ACR Classification Criteria at Screening	Modality
Number of ACR classification criteria for SLE by categories:	From DIAG form of the eCRF: • < 4 • ≥ 4 and ≤ 5 • ≥ 6 and ≤ 7 • ≥ 8 and ≤ 9 • ≥ 10 and ≤ 11
ACR classification criteria for SLE	From DIAG form of the eCRF: • Discoid rash (Yes/No) • Malar rash (Yes/No) • Photosensitivity (Yes/No) • Oral ulcers (Yes/No) • Arthritis (Yes/No) • Serositis (Yes/No). If Yes, any of the following:

- Pleuritis
- Pericarditis

- Renal disorder (Yes/No). If Yes, any of the following:
  - Persistent proteinuria: > 0.5 g per day or > than 3+ if quantitation not performed
  - Cellular casts: may be red cell, hemoglobin, granular, tubular or mixed
- Neurologic disorder (Yes/No). If Yes, any of the following:
  - Seizures: in the absence of offending drugs or known metabolic derangements; e.g., uremia, ketoacidosis, or electrolyte imbalance
  - Psychosis: in the absence of offending drugs or known metabolic derangements; e.g., uremia, ketoacidosis, or electrolyte imbalance
- Hematologic disorder (Yes/No). If Yes, any of the following:
  - Hemolytic anemia with reticulocytosis
  - Leukopenia: < 4,000/mm3 on ≥ two occasions
  - Lymphopenia: < 1,500/mm3 on ≥ two occasions
  - Thrombocytopenia: < 100,000/mm3 in the absence of offending drugs
- Immunologic disorder (Yes/No). If Yes, any of the following:
  - Anti-DNA: antibody to native DNA in abnormal titer
  - Anti-Smith (anti-Sm): presence of antibody to Sm nuclear antigen
  - Positive finding of antiphospholipid antibodies based any of the following:
    1. an abnormal serum level of IgG or IgM anticardiolipin antibodies
    2. a positive test result for lupus anticoagulant using a standard method, or
    3. a false positive serologic test result for syphilis known to be positive for at least six months and confirmed by Treponema pallidum immobilization or fluorescent Treponemal antibody absorption test
- Antinuclear antibody (Yes/No)

Table 14: SLEDAI-2K Parameters and Scores at Screening

SLEDAI-2K Screening Characteristic	Modality

Parameter of the SLEDAI--2K score (Screening)	From SLEDAI1 form of the eCRF:
	• Central nervous system
	  ○ Cardiovascular accident (CVA)/Cranial nerve disorder/Lupus headache/Organic brain syndrome/Psychosis/Seizure/Visual disturbance
	• Musculoskeletal
	  ○ Arthritis/Myositis
	• Renal
	  ○ Hematuria/Proteinuria/Pyuria/Urinary casts
	• Skin
	  ○ Alopecia/Mucosal ulcers/Rash
	• Cardio-pulmonary
	  ○ Pericarditis/Pleurisy
	• Hematologic
	  ○ Leukopenia/Thrombocytopenia
	• Serologic
	  ○ Increase DNA binding/Low complement
	• Vascular
	  ○ Vasculitis
	• Constitutional
	  ○ Fever

Total SLEDAI-2K score (Screening)	• SLEDAI-2K < 10, SLEDAI-2K ≥ 10 (HDA subgroup [Yes/No]) from SLEDAI1 form of the eCRF • SLEDAI-2K < 10, SLEDAI-2K ≥ 10 (HDA subgroup [Yes/No]) **from IWRS** Note: Total SLEDAI-2K score and Clinical Total SLEDAI-2K score (excluding the components 'Increase DNA Binding' and 'Low Complement') at Screening will also be presented.
HDA Subject	• Yes/No (derived with Total SLEDAI-2K score at Screening from eCRF and from IWRS)

Table 15: Other Baseline Characteristics

Other Baseline Characteristic	Modality
SLICC/ACR damage	From SLICC form of the eCRF:

index score	• 0 • 1 • ≥ 2 and < 5 • ≥ 5
BILAG Severity	From BILAG form of the eCRF: • Severe = At least 1 BILAG A; vs. • Moderate = At least 2 BILAG B and no BILAG A; vs. • Mild = 1 BILAG B and no BILAG A and no more than 1 BILAG B. • No severity = Criteria for severe, moderate and mild not met
CLASI total activity score	From CLASI form of the eCRF: • <8 • ≥8 • Summary statistics for LTE analysis
SFI Flare	From SFI1 form of the eCRF: • No flares, • Mild or moderate flare, • Severe flare, • Missing

11.2.2 Other

Chest X-ray screening evaluations will be listed and tabulated by treatment group using the number and percentage of subjects for each interpretation category (Normal, Abnormal Not Clinically Significant, Clinically Significant, and Abnormal Overall).

Other screening evaluations like serum virology, tuberculosis assessment, t-spot, serum β-D-Glucan (Japan only), Quantiferon TB test, thyroid-stimulating hormone, follicle-stimulating hormone and prior surgeries will be listed only.

11.3 Medical History

The medical history will be summarized using Medical Dictionary for Regulatory Activities (MedDRA), current version, PT as event category and MedDRA SOC body term as Body System category. The MedDRA version used will be indicated in footnote. Medical history will be tabulated by SOC and PT. SOC and PT will be alphabetically sorted. Medical history will also be listed.

12 Previous or Concomitant Medications/Procedures

12.1 Previous or Concomitant Medications

For the PA, previous and concomitant medications will be summarized by treatment group on SAF. Data from the 52-week treatment period including safety follow up data will be included.

For the LTE analysis, concomitant medications will be summarized by treatment group using the LTE SAF, counting starts from the start of the LTE first dose of evobrutinib 50mg BID in LTE.

Previous medications will not be repeated for the LTE analysis.

Definition

Previous medications are medications, other than trial medications, which either:

1. started and stopped before first administration of any IMP (placebo or evobrutinib), or,
2. started prior to the first administration of IMP (placebo or evobrutinib) and are taken by subjects on or after the first administration of IMP.

Concomitant medications are medications, other than trial medications, which either:

1. started on or after the first administration of any IMP (placebo or evobrutinib), or,
2. started prior to the first administration of IMP (placebo or evobrutinib) and are taken by subjects on or after the first administration of IMP.

As result, a medication can be counted in both categories (previous/concomitant) at the same time.

Partial dates will be handled as follows:

- In case the dates do not allow a medication to be unequivocally allocated to previous medication, the medication will be considered as previous medication.
- In case the dates will not allow a medication to be unequivocally allocated to concomitant medication, the medication will be considered as concomitant medication.

Statistical analysis

The ATC-2nd level and PT will be tabulated as given from the World Health Organization Drug Dictionary (WHO-DD) current version. In case multiple ATCs are assigned to a drug, all ATC--2nd level will be used for reporting.

The number and proportion of subjects with previous or concomitant medications will be separately summarized by treatment group and will be presented by descending frequency of ATC-2nd level term and then by descending frequency of PT in total column. If multiple ATCs/PTs have the same frequency, they will be sorted alphabetically. The WHO-DD version used will be indicated in footnote.

Prior and concomitant Medications will be also listed.

12.2 Prior or Concurrent Procedures

Prior and concurrent procedures will be summarized the same way as medications (see Section 12.1).

Number of subjects with prior/concurrent procedures overall and by SOC and PT will be summarized by treatment group, using current version of MedDRA dictionary.

Concurrent procedures will be also listed.

13 Treatment Compliance and Exposure

For the PA, exposure, cumulative/calculated total dose and compliance will be summarized by treatment group on SAF. Data from the 52-week treatment period will be analyzed.

For the LTE analysis, exposure, cumulative/calculated total dose and compliance will be summarized by treatment group using the LTE SAF as follows:

- Data from the LTE period will be analyzed for all subjects who enter the LTE.
- Data from the entire study (52-week treatment period plus the LTE) will be analyzed for subjects who remain with evobrutinib 50mg BID throughout the study.

13.1 Exposure Calculation

Exposure for 52-week treatment period will be presented for PA, whereas exposure for LTE period and overall exposure will be presented for LTE analysis.

Dose interruptions/changes will not be considered in exposure calculation. Dose interruptions/changes with the associated reason will only be listed.

PA: 52-week treatment period

Details describing study therapy dosing and administration are provided in Table 16.

Table 16 : Dosage and Administration of Investigational Medicinal Product

Product Description	Dosage Form	Evobrutinib Dose	Number of Tablets (Morning Dose)		Number of Tablets (Evening Dose)	
			25 mg		25 mg	
			Evobrutinib	Placebo	Evobrutinib	Placebo
Evobrutinib	Oral	25 mg QD	1	2	0	2
Evobrutinib	Oral	75 mg QD	3	0	0	2
Evobrutinib	Oral	50 mg BID	2	1	2	0
Placebo	Oral	0 mg	0	3	0	2
BID – Twice Daily; QD – Once Daily.						

Treatment duration in weeks will be calculated according to the following formula:

$$Treatment\ duration\ (weeks) = \frac{(date\ of\ last\ dose\ -\ date\ of\ first\ dose\ +\ 1)}{7}$$

First dose refers to the first administration of any IMP in 52-week treatment period. Last dose refers to the last administration of any IMP in 52-week treatment period. Both dates of first and last dose will be retrieved from SDTM EC domain.

Treatment duration will be presented by summary statistics and according to the following categories:

- ≤ 1 week
- > 1 to 8 weeks
- > 8 to 16 weeks
- > 16 to 24 weeks
- > 24 to 32 weeks
- > 32 to 40 weeks
- > 40 to 48 weeks
- > 48 to 52 weeks
- > 52 weeks

The calculated total dose (mg) per subject for the 52-week treatment period will also be summarized for the active treatment groups, based on the actual treatment the subject receives. The calculated total dose is defined according to the following formula:

$$Calculated\ total\ dose\ (mg) = Treatment\ duration \times Daily\ dose$$

where the exposure time is expressed in days and the daily dose is provided in Table 16.

The cumulative actual dose (mg) per subject for the 52-week treatment period will also be summarized for the active treatment groups. The cumulative actual dose is defined according to the following formula:

$$Cumulative\ actual\ dose\ (mg) = Number\ of\ tablets\ ingested \times Fr \times 25mg$$

where the number of tablets ingested is defined below, and Fr is the fraction of tablets that contain evobrutinib 25 mg according to the actual treatment group:

$$Fr = 0\ for\ Placebo$$

$$Fr = \frac{1}{5} for\ Evobrutinib\ 25\ mg\ QD$$

$$Fr = \frac{3}{5} for\ Evobrutinib\ 75\ mg\ QD$$

$$Fr = \frac{4}{5} for\ Evobrutinib\ 50\ mg\ BID$$

The number of ingested tablets will be retrieved from the BATCH (Study Treatment Box Number) panel of the CRF, the number of tablets dispensed being equal to 170 for a given kit during the treatment period and 32 during the LTE period:

$$Number\ of\ tablets\ ingested = Number\ of\ tablets\ dispensed - number\ of\ tablets\ returned$$

Cumulative actual dose cannot be calculated for placebo treatment group.

For subjects with missing number of tablets returned (e.g. lost to follow-up), the cumulative actual dose will be calculated by imputing the missing returned amount with the number of tablets the subjects should have returned if 100% compliance was observed while on treatment for the associated dispensed kit.

*Example: Let's consider a subject that has been exposed to study drug for 90 days before withdrawal and assigned to Evobrutinib 50 mg BID group. The subject has received 3 kits of tablets (170 tablets each) and returned 10 tablets for the 1st kit, 12 for each of the two other kits. The number of tablets ingested is calculated as 170 * 3 – (10 + 12 + 12) = 476. The calculated total dose is equal to 90 * 100 = 9000 mg and the cumulative actual dose is equal to 476 * 4/5 * 25 = 9520 mg.*

LTE Analysis: LTE period

Treatment duration in months in the LTE period will be calculated according to the following formula:

$$Treatment\ duration\ (months) = \frac{(date\ of\ last\ dose\ -\ date\ of\ first\ dose\ +\ 1)}{30.4375}$$

First dose refers to the first administration of evobrutinib 50mg BID in LTE period.

Treatment duration will be presented by summary statistics and according to the following categories:

- ≤ 4 months
- > 4 to 8 months
- > 8 to 12 months
- > 12 to 16 months
- > 16 to 20 months
- > 20 to 24 months
- > 24 months

The calculated total dose and cumulative actual dose (mg) per subject for the LTE period will also be summarized, defined as for the 52-week treatment period. As all subjects will be receiving evobrutinib 50mg BID in the LTE period, cumulative actual dose will be calculated using:

$$Cumulative\ actual\ dose\ (mg) = Number\ of\ tablets\ ingested \times 25mg$$

***LTE analysis:* Overall period (i.e. 52-week treatment period + LTE)**

Overall exposure will be presented on LTE SAF only for subjects who received evobrutinib 50 mg BID during the 52-week treatment period and continue with the same dose of evobrutinib during LTE period.

Treatment duration in months will be calculated according to the following formula:

$$Treatment\ duration\ (months) = \frac{(date\ of\ last\ dose\ -\ date\ of\ first\ dose\ +\ 1)}{30.4375}$$

First dose refers to the first administration of evobrutinib 50mg BID in the study (i.e. first dose in the 52-week treatment period). For subjects with a gap greater than or equal to 2 weeks between the 52-week treatment period and LTE treatment period, the treatment duration will be derived by adding the 52-week treatment duration and the LTE treatment duration, both calculated separetely.

Treatment duration will be presented by summary statistics and according to the following categories:

- ≤ 6 months
- > 6 to 12 months
- > 12 to 18 months
- > 18 to 24 months
- > 24 to 30 months
- > 30 to 36 months
- > 36 months

The calculated total dose and cumulative actual dose (mg) per subject will be computed as well, defined as for the LTE period.

Subject data listings:

Study drug administrations will also be listed by treatment group, and subject, with start/end dates of administration and reason for dose change (if applicable).

13.2 Compliance Calculation

Compliance will be separately analyzed for the PA and LTE analysis, in the same manner as exposure was analyzed.

PA: 52-week treatment period

For the 52-week treatment period, compliance with treatment is defined as the number of tablets taken during a period divided by the number of tablets that should have been taken during that period, multiplied by 100 to yield a percentage, i.e.:

$$Compliance\ with\ \ treatment\ = 100 \times (\frac{N_1}{5 \times N_2})$$

where

- N_1 = number of tablets given minus number of tablets returned over N_2 days,
- N_2 = number of days between treatment start and treatment termination visit,

where N_1 will be computed using the BATCH form of the eCRF and N_2 corresponds to the exposure time expressed in days (see Section 13.1 for exposure time and number of tablets ingested formulas).

For the 52-week treatment period, the last kit dispensed is at Week 48 visit. The number of tablets returned for this last kit is entered during Week 52 visit.

Example: Let's consider the example from Section 13.1. The subject has been provided with 510 tablets, returned 34 tablets and has been exposed to study drug for 90 days. Therefore, N_1 is

*equal to 510 – 34 = 476 and N_2 is equal to 90, which gives a compliance equal to 100 * (476 / 5*90) = 105.8%.*

For subjects with missing number of tablets returned (e.g. lost to follow-up), the compliance will be calculated by imputing the missing returned amount with the number of tablets the subjects should have returned if 100% compliance was observed while on treatment for the associated dispensed kit.

Compliance with treatment will be tabulated by treatment group from first intake to last intake of 52-week treatment period.

Compliance with treatment will also be presented into categories as follows:

- < 60%
- ≥ 60% to < 80%
- ≥ 80% to ≤ 100%
- > 100% to ≤ 110%
- > 110%

***LTE analysis:* LTE period**

Compliance for the LTE period will be presented on subjects from LTE SAF.

For the LTE period, compliance with treatment is defined using the formula below (4 tablets taken per day, 2 in the morning and 2 in the afternoon):

$$Compliance\ with\ \ treatment\ = 100 \times (\frac{N_1}{4 \times N_2})$$

The compliance will be computed using the number of kits dispensed as well as the number of tablets returned. Compliance with treatment will be tabulated by treatment group from first intake in LTE to last intake during LTE period and also presented within categories, as for the 52-week treatment period.

The first kit dispensed in LTE treatment period is at Week 52 visit for subjects who rolled over directly from 52-week treatment to LTE.

***LTE analysis:* Overall period (i.e. 52-week treatment period + LTE)**

Overall compliance will be presented on LTE SAF only on subjects who received evobrutinib 50 mg BID during the 52-week treatment period and continue with the same dose of evobrutinib during LTE period.

Overall compliance with treatment is defined using the formula below:

$$Compliance\ with\ \ treatment\ = 100 \times [((\frac{N_1}{5 \times N_2}) + (\frac{N_3}{4 \times N_4}))/2]$$

where

- N_1 = number of tablets given minus number of tablets returned over N2 days for the 52-week treatment period
- N_2 = number of days between treatment start and treatment termination visit for the 52-week treatment period
- N_3 = number of tablets given minus number of tablets returned over N2 days for the LTE treatment period
- N_4 = number of days between treatment start and treatment termination visit for the LTE treatment period

Compliance with treatment will be tabulated from first intake to last intake of the study and also presented within categories, as for the 52-week treatment period.

Subject data listings:

For each analysis, the following listings will be provided:

- listing of kit numbers with date of dispense, and number of tablets returned (from BATCH).
- listing of start/end dates with number ingested tablets (from EXPOSUREDT).
- listing with exposure time, cumulative dose, and compliance.

14 Endpoint Evaluation

For all efficacy endpoints:

- Analysis visits will be used (as detailed in Section 9.7 and Appendix 18.11).
- Baseline definition from Section 9.5 should be applied.
- Statistical tests of the two-primary and key secondary efficacy endpoints entering the hierarchical testing procedure (see section 14.1.2.4) will be conducted at a one-sided α level of 0.025. The one-sided p-values and the two-sided 95% CIs will be presented as well as a note on how to interpret correctly the two-sided 95% CIs by considering appropriately the accurate boundary of the interval.

  For ALL other endpoints, statistical tests will be conducted at two-sided α level of 0.05, and two-sided p-values and 95% CI will be presented where applicable.

14.1 Two-primary Efficacy Endpoints

The two-primary efficacy endpoints will be analyzed for the PA using mITT Analysis Set.

14.1.1 Definitions

The two-primary efficacy endpoints are SRI response at Week 52:

- SRI-4 response in all subjects,
- SRI-6 response in the HDA subgroup.

The SRI-4 response, a measure of reduced SLE disease activity, is defined by meeting all of the following conditions compared to Baseline:

1. ≥ 4-point reduction in SLEDAI-2K total score.
2. No significant worsening in PGA score (< 0.3 increase assuming the PGA score is on a 0-3 scale).
3. No new BILAG A organ domain scores and ≤ 1 new BILAG B organ domain score compared to Day 1 using BILAG 2004.
4. No institution of EAC-determined protocol-prohibited medication/treatment (i.e., EAC-treatment failure)

Subjects with missing values at Week 52 after the implementation of rules defined in Appendices 18.1.1 and 18.1.3 will be considered as non-responder in the primary analysis.

Assessment of criterion #4 will be confirmed by an Endpoint Adjudication Committee (EAC). Further details can be found in the EAC Charter.

The SRI-6 response is defined similarly to SRI-4, based on a ≥ 6-point reduction in SLEDAI-2K total score.

Details of derivation are described in Appendix 18.1.

14.1.2 Primary Analysis of the Two-primary Endpoints

Table 17 outlines all planned statistical analyses of the two-primary endpoints. In the context of estimands, handling the intercurrent events is also explained in the table.

Table 17 : Analysis of the two-primary endpoints

Analysis	Analysis Population	Analysis Method	Data Handling Notes and handling of intercurrent events
Two Primary endpoints / MCP	mITT Analysis Set	• Presentation of adjusted Odds-Ratio (OR), 2-sided 95% CI of OR, nominal 1-sided p-value, adjusted 1-sided p-value (α=0.025) from Logistic Regression • Number and percentage of responders in each dose group, observed Δ in response proportions from Placebo for each dose group	A subject with a missing response at Week 52 will be considered a non-responder at Week 52. Subjects with endpoint adjudication committee (EAC)-determined treatment failure based on protocol-prohibited medication will be considered as non-responder. (Estimand: composite strategy)

Sensitivity 1 (Per Protocol Analysis)	PP Analysis Set	• Presentation of OR, 2-sided 95% CI of OR, nominal p-value from Logistic Regression • Number and percentage of responders in each dose group, observed Δ in response proportions from Placebo for each dose group	Same as Primary Analysis
Sensitivity 2 (Multiple Imputation Analysis)	mITT Analysis Set	• Presentation of OR, 2-sided 95% CI of OR, nominal p-value from Logistic Regression • Number and percentage of responders in each dose group, observed Δ in response proportions from Placebo for each dose group	For subjects with missing response at Week 52, the multiple imputation (MI) will be used to impute the missing data using the available observations. Subjects with endpoint adjudication committee (EAC)-determined treatment failure based on protocol-prohibited medication will be considered as non-responder. (Estimand: composite strategy)
Sensitivity 3 (Tipping Point Analysis)	mITT Analysis Set	• Presentation of study conclusion at each possible combinations of responders from each treatment group.	Tipping Point Analysis Subjects with endpoint adjudication committee (EAC)-determined treatment failure based on protocol-prohibited medication will be considered as non-responder. (Estimand: composite strategy)
Sensitivity 4 (Unadjusted Analysis)	mITT Analysis Set	• Presentation of unadjusted OR, 2-sided 95% CI of OR, nominal p-value from Logistic Regression • Number and percentage of responders in each dose group, observed Δ in response proportions from Placebo for each dose group	Same as Primary Analysis
Sensitivity 5 (Background therapy Adjusted Analysis)	mITT Analysis Set	• Presentation of adjusted OR, 2-sided 95% CI of OR, nominal p-value from Logistic Regression. The logistic model will include the background therapy of immunosuppressant use as covariate	Same as Primary Analysis
Supportive 1 (Trend test for Dose Response)	mITT Analysis Set	Test of linear trend in log odds of response with increasing dose, based on an appropriate contrast from the logistic model, ie, (3, 1, -1, -3).	Same as Primary Analysis

MCP = Multiplicity Control Procedure. See section 14.1.2.4 for details.

Treatment failures are determined by EAC based on their review of protocol-prohibited medications. Further details can be found in the EAC Charter.

14.1.2.1 Primary Analysis Method of SRI-4 and SRI-6

The null and alternative hypotheses corresponding to the two-primary endpoints are:

H_{0a}: $\log(OR_{4\_all}) \leq 0$ versus H_{1a}: $\log(OR_{4\_all}) > 0$

H_{0b}: $\log(OR_{6\_HDA}) \leq 0$ versus H_{1b}: $\log(OR_{6\_HDA}) > 0$

where $\log(OR_{4\_all})$ and $\log(OR_{6\_HDA})$ are the log of odds ratio for SRI-4 in all subjects and SRI-6 in the HDA subjects.

The study will be declared positive if the treatment effect is significant in at least one of the primary endpoints. The global hypothesis testing problem is stated as:

H_0: $H_{0a} \cap H_{0b}$

H_1: $H_{1a} \cup H_{1b}$.

The global null hypothesis is rejected if one or both individual null hypotheses are rejected.

To account for the multiplicity due to the two-primary endpoints and multiple treatment comparisons with placebo for each primary endpoint, the multiple testing procedure as described in Section 14.1.2.4 will be performed.

In the primary analysis of the two-primary endpoints, a subject could be considered as an SRI-4 or SRI-6 Non-Responder if the subject takes EAC-determined protocol-prohibited medications during the treatment period (i.e., EAC-treatment failure). The EAC will make the final determination on whether the subject should be classified as treatment failure. In addition, a subject with a missing response at Week 52 will be considered a non-responder at Week 52.

The primary analysis of the SRI-4 response at Week 52 among all subjects, and SRI-6 response at Week 52 among HDA subjects, will be based on the mITT analysis set, and analyzed using logistic regression model for the odds of a given SRI response adjusted for the stratification variables used in the randomization (race: black vs non-black, screening SLEDAI-2K total score: <10 vs ≥10 and region: US and Western Europe vs Japan vs ROW) and with treatment as a factor. The estimated OR, together with associated two-sided 95% CI of the estimated OR, nominal one-sided p-value and adjusted one-sided p-value (α=0.025) from the logistic regression model, comparing each evobrutinib treatment group to placebo will be presented.

If there is model convergence problem due to for example zero cells in covariates, appropriate method will be applied to resolve the issue.

An example of SAS code is available in Appendix 18.9.

In addition, the number and percentage of SRI-4/SRI-6 non-responders at Week 52 in each treatment group will be presented as follows:

- For the summary of SRI responders (Yes), the number and percentage of subjects from each treatment group who have met the responder criterion for each SRI component will be provided, including the delta from placebo.
- For the summary of SRI non-responders (No), the number and percentage of subjects from each treatment group who are imputed as non-responders for each reason will be provided. Reasons of non-response will be as follows:
  - Missing scoring at Week 52
    - Treatment discontinuation due to Safety (reasons = AEs, death)
    - Treatment discontinuation due to Lack of efficacy
    - Other treatment discontinuation (reasons = lost to follow-up, protocol non-compliance, withdrew consent and others)
    - Treatment completer and missing response at W52

In addition, a bar chart will be generated by treatment group to present SRI-4 through SRI-10 response rates at Week 52. The same figure will be repeated on HDA subgroup. A line plot representing the SRI response rate over time will be also provided for each primary endpoint, with all treatment groups on the same graph. The figure presenting data over time will use the similar imputation rules as applied for the Week 52 timepoint.

SRI-4/SRI-6 response rates over time will be also presented by visit, using similar imputation rules as applied for the Week 52 timepoint.

14.1.2.2 Sensitivity Analyses of SRI-4 and SRI-6

The following sensitivity analyses are planned to be performed using different approaches to the missing data handling from the primary analysis.

Sensitivity Analysis #1: Per Protocol Analysis

The primary analysis method including the missing data handling approach for the two-primary endpoints will be repeated on the PP Analysis Set if the size of this analysis set differs substantially (>10%) from mITT.

For this analysis, the estimated OR, together with associated two-sided 95% CI of the estimated OR and nominal p-value from the logistic regression model, comparing each evobrutinib treatment group to placebo will be presented.

In addition, the number and percentage of SRI-4/SRI-6 responders and non-responders at Week 52 in each treatment group will be presented the same way as for the primary analysis (see Section 14.1.2.1).

Sensitivity Analysis #2: MI

In this analysis, a subject will be considered as a non-responder if using EAC-determined protocol prohibited medications (i.e., EAC-treatment failure). For subjects with missing response at Week 52, then the MI will be used to impute the missing data using the available observations from previous visits and from components of SRI-4/SRI-6 response and also stratification factor used in primary analysis.

The MI analysis will include the components as described below:

1. SLEDAI-2K change from baseline total score will be imputed using SAS MI procedure (including 2 other components: PGA score and BILAG response) and then the occurrence of a ≥ 4-point reduction will be re-derived from the imputed values. For the SRI-6 the occurrence of a ≥ 6-point reduction will be also re-derived from the imputed values.

2. PGA score will be imputed using SAS MI procedure (including 2 other components: SLEDAI-2K change from baseline and BILAG response), assuming the distribution is approximately normal. The imputed values will be transformed to a 0-3 scale:

    i. Negative values will be imputed as 0,

    ii. Values > 3 will be imputed as 3.

    iii. The imputed score will then be used to calculate whether there was significant worsening in the PGA score.

3. Whether there was no new BILAG A organ domain scores and ≤ 1 new BILAG B organ domain score compared to Day 1 will be considered as a dichotomic response with value 1 when it is met and 0 otherwise. This score will be imputed using SAS MI procedure (including 2 other components: SLEDAI-2K change from baseline and PGA score) and the response will be deducted from the logic below:

    i. Values < 0.5 will be imputed as 0,

    ii. Values > 0.5 will be imputed as 1.

Components no. 1 to 3 will be imputed assuming a MAR pattern. The SAS MI procedure will be performed as follows:

- Monotone missing data structure will be created as follows: intermediate (non-monotone) missing data will be multiply imputed using the Markov chain Monte Carlo (MCMC) method and assuming MAR and multivariate normality. The SAS procedure PROC MI with the MCMC option will be used with seed number = 2005270018. The number of burn-in iterations will be set to 200, which is the default value. Nevertheless, if diagnosis plots show that the convergence has not yet occurred, this will be adjusted.
- Then, each component will be imputed with treatment group, race, region, SLEDAI-2K total score (<10;>=10) and data at previous visit from all SRI components as covariates.

- Imputation will be repeated 1000 times. The SRI response will then be calculated for each of the multiply imputed data sets from the imputed components. For this analysis, the estimated OR, together with associated two-sided 95% CI of the estimated OR and nominal p-value from the logistic regression model, comparing each evobrutinib treatment group to placebo will be presented. Results will be combined using MIANALYZE SAS procedure with the Rubin's rules as described in Appendix 18.9.

The SAS code to implement the MI and MIANALYZE procedures is in Appendix 18.9.

The same analysis method as the primary analysis of the two-primary endpoints will be performed using the mITT Analysis Set for SRI-4 and in the HDA subject subgroup for SRI-6.

In addition, the number and percentage of SRI-4/SRI-6 responders and non-responders at Week 52 in each treatment group will be presented the same way as for the primary analysis (see Section 14.1.2.10) using Rubin's rules to gather binomial proportion by each treatment (see Appendix 18.9).

Sensitivity Analysis #3: Tipping Point Analysis

For this analysis, a subject will be considered as an SRI-4 or SRI-6 Non-Responder if the subject takes EAC-determined protocol-prohibited medications during the treatment period (i.e. EAC-treatment failure).

Tipping point analysis will be performed as in Yan 2009, Section 2.3.2.

All possible combinations of success/failure may be posited in the 4 treatments groups. Results are obtained by a random single imputation of possible scenarios (responder or non-responder) to each arm. All possible combinations of imputations of response among subjects with missing response in the placebo group and evobrutinib 25 mg QD, 75 mg QD and 50 mg BID groups. For each combination, the imputed data were combined with the data from subjects with non-missing data and primary analysis.

Assumptions of missing data due to a missing response at Week 52 will be varied to see if there are any tipping points:

Let

$$n_1 = number\ of\ responders\ out\ of\ missing\ data\ due\ to\ various\ reasons\ for\ placebo$$

$$n_2 = number\ of\ responders\ out\ of\ missing\ data\ due\ to\ various\ reasons\ for\ evobrutinib\ 25\ mg\ QD$$

$$n_3 = number\ of\ responders\ out\ of\ missing\ data\ due\ to\ various\ reasons\ for\ evobrutinib\ 75\ mg\ QD$$

$$n_4 = number\ of\ responders\ out\ of\ missing\ data\ due\ to\ various\ reasons\ for\ evobrutinib\ 50\ mg\ BID$$

The same analysis method as the primary analysis of the two-primary endpoints will be performed using the mITT Analysis Set for each possible combination of n_1, n_2 , n_3 and n_4.

The values of n_1, n_2 , n_3 and n_4 where the study result will be tipped from statistically significant (not statistically significant) to not statistically significant (statistically significant) will be assessed for plausibility based on the corresponding response rate from each treatment. See Section 14.1.2.4 for the multiplicity adjustment procedure.

Sensitivity Analysis #4: Unadjusted Analysis

The analysis of the two-primary endpoints will be performed on the mITT Analysis Set.

In this analysis, the logistic regression model will be used with only treatment as a factor.

For this analysis, the estimated OR, together with associated two-sided 95% CI of the estimated OR and nominal p-value from the logistic regression model, comparing each evobrutinib treatment group to placebo will be presented.

In addition, the number and percentage of SRI-4/SRI-6 responders and non-responders at Week 52 in each treatment group will be presented the same way as for the primary analysis (see Section 14.1.2.1).

Sensitivity Analysis #5: Analysis including the background therapy of immunosuppressant

This sensitivity analysis of the SRI-4 response at Week 52 among all subjects, and SRI-6 response at Week 52 among HDA subjects, will be based on the mITT Analysis set, and analyzed using logistic regression model for the odds of a given SRI response adjusted for the stratification variables used in the randomization (race: black vs non-black, screening SLEDAI-2K total score: <10 vs ≥10 and region: US and Western Europe vs Japan vs ROW), background therapy of immunosuppressant (Yes/No) as covariate and with treatment as a factor.

Background therapy of immunosuppressant will be derived at Day 1 using concomitant medications data with class = “Immunosuppressant”, Anatomical Therapeutic Chemical (ATC) code level 3 = L04A.

For this analysis, the estimated OR, together with associated two-sided 95% CI of the estimated OR and nominal p-value from the logistic regression model, comparing each evobrutinib treatment group to placebo will be presented.

In addition, the number and percentage of SRI-4/SRI-6 responders and non-responders at Week 52 in each treatment group will be presented the same way as for the primary analysis (see Section 14.1.2.1).

14.1.2.3 Supportive Analyses of SRI-4 and SRI-6

The following supportive analysis will be performed for the two-primary endpoints:

Supportive analysis #1: linear trend test

For each of the two-primary endpoints, a test of linear trend in log odds of SRI response with increasing dose, will be based on an appropriate contrast from the logistic model (i.e., (3, 1, -1, -3)). The p-value reported tests the null hypothesis that the log odds of being SRI responder is the same across treatment groups, against the alternative that the log odds increases linearly with increasing treatment group order.

An example of SAS code is available in Appendix 18.9.

Both p-values based on Wald test and Likelihood ratio test will be presented. They will be added in the summary table of logistic regression.

14.1.2.4 Multiple testing

Based on the cumulative clinical and CCI data from Evobrutinib, it is expected that 50mg BID will be at least as effective as 75mg QD, or more effective than 75mg QD.

To maximize the probability of study success while controlling for the overall type 1 error, statistical testing of the hypotheses associated with the two primary endpoints and the two key secondary endpoints will be performed in a hierarchical manner such that the Evobrutinib dose groups will be compared to placebo from the highest dose to the lowest dose as detailed in the following:

1. The two-primary endpoints will be tested first for the highest dose using a Hochberg Procedure at a 1-sided α=0.025. If both null hypotheses are rejected, then go to step 2. If none or only one null hypothesis is rejected, then testing stops.

2. The 1st key secondary endpoint of SRI-4 at week 52 in serologically active subjects for the highest dose will be tested at a 1-sided α=0.025. If the null hypothesis is rejected, then go to Step 3. Otherwise testing stops.

3. The 2nd key secondary endpoint of time to first severe BILAG A flare for the highest dose will be tested at a 1-sided α=0.025. If the null hypothesis is rejected, then go to Step 4. Otherwise testing stops.

4. Repeat steps 1-3 for the medium dose, and the lowest dose sequentially.

Hochberg adjusted p-values will only be needed if the first step of the multiplicity testing is successful.

The corresponding hierarchical testing diagram using the approach described in Bretz 2009 can be found in Appendix 18.15.

14.1.3 Subgroup analysis of Two-primary Endpoints

To assess the consistency of the two-primary endpoints results among subgroups of subjects (See Table 4), the analysis will be performed on the mITT Analysis Set using the logistic regression model for the odds of a given SRI response with treatment, subgroup of interest, treatment by subgroup of interest interaction as main effects. The estimated OR, together with associated two-sided 95% CI of the estimated OR and nominal p-value from the model contrast will be presented for each treatment group. The same missing data imputation approach will be used as the primary analysis method. No multiple testing adjustments will be performed.

In addition, a forest plot will be presented for each of the two-primary endpoints for the overall treatment effect and for each subgroup of interest including the OR from logistic model and associated 2-sided 95% CI of the OR.

For subgroup D, the analysis performed on the Japan region will be reported as "Partial Japan" for the PA.

14.1.4 Analysis of SRI-4 Response at Week 52 in Japanese versus non-Japanese Subjects

Following the review of the PA, the Japanese cohort will not be explored. All the listings produced for the PA will be updated to include the data of all Japanese subjects. Safety summaries of AEs and Labs will be presented for the Asia-Pacific subgroup (defined as sites in the trial from Korea, Taiwan, Japan, Philippines and Malaysia) vs non-Asia-Pacific subgroup.

14.2 Key secondary endpoints

The key secondary endpoints will be analyzed for the PA on the mITT analysis set.

The key secondary endpoints are the following:

- SRI-4 response at Week 52 among serologically active subgroup.
- Time to first severe (BILAG A) flare

14.2.1 Definitions

SRI-4 response among serologically active subjects

The SRI-4 response will be analyzed at Week 52 among subjects who are serologically active at screening, i.e., positive anti-dsDNA or low complement levels.

Time to first severe (BILAG A) flare

A severe (BILAG A) flare is defined as at least one BILAG A score in any organ system due to items that are new or worse, compared to the BILAG evaluation at the previous visit, during the 52-week treatment period. "Due to items that were new or worse" means that if the assigned Grade A/B or C contains ANY ITEM which is WORSE or NEW, such grade will be included into the flare consideration. This is valid for clinical organ domains. For laboratory organ domains (i.e. renal and hematological) where items are not worse/new, there has to be an increase in score from B to A or C to B for flare to be considered. Appendix 18.1.1 provides further details on BILAG evaluation. EAC will review the dry-run BILAG analysis dataset to ensure appropriateness and correctness of programmed BILAG flare derivation. Further details can be found in the EAC Charter.

Since subjects are only to be assessed for BILAG flare at discrete planned visits, the corresponding date of visit will be used to derive the time to first severe (BILAG A) flare as follows:

$$\begin{aligned} &Time\ to\ first\ severe\ BILAG\ A\ flare\ (days) \\ &= Date\ of\ visit\ where\ the\ first\ severe\ BILAG\ A\ flare\ has\ been\ diagnosed \\ &- Date\ of\ randomization + 1 \end{aligned}$$

The following censoring rules will be used:

- A subject discontinuing the 52-week treatment period prior to Week 52 without severe flare will have his/her time to first severe flare right-censored at the last time point at which flare could be assessed (i.e. treatment discontinuation date).
- A subject completing the 52-week treatment period without severe flare will have his/her time to first severe flare right-censored at Week 52.

Additionally, a sensitivity analysis will be performed based on Multiple imputation approach of severe flare at each visit up to week 52:

- Monotone missing data structure will be created as follows: intermediate (non-monotone) missing data will be multiply imputed using the Markov chain Monte Carlo (MCMC) method and assuming MAR and multivariate log normality. The SAS procedure PROC MI with the MCMC option will be used with seed number = 2005270018. The number of burn-in iterations will be set to 200, which is the default value. Nevertheless, if diagnosis plots show that the convergence has not yet occurred, this will be adjusted.
- Then the response (Severe BILAG A: Yes/No) will be imputed with treatment group, race, region, SLEDAI-2K total score (<10;>=10) and response at previous visit covariates.
- Imputation will be repeated 1000 times. Then depending of the visit of censored observation or of response, the time to first severe BILAG A flare will be computed as described for the non- sensitivity analysis. Estimate and associated standard error of Cox regression analysis applied on each imputed sample will be combined using MIANALYZE SAS procedure with the Rubin's rules as described in Appendix 18.9.

14.2.2 Primary Analysis of Key Secondary Endpoints

The key secondary endpoints will be analyzed on mITT Analysis Set at Week 52, as indicated in

Table 18.

Table 18 : Analysis of the key secondary endpoints

Endpoint	Analysis	Analysis Population	Analysis Method	Data Handling Notes
SRI-4 response in serologically active subgroup	Primary Analysis	mITT Analysis Set	• Presentation of Odds-Ratio (OR), 2-sided 95% CI of OR, nominal 1-sided p-value (α=0.025) from Logistic Regression • Number and percentage of responders in each dose group, observed Δ in response proportions from Placebo for each dose group	A subject with a missing response at Week 52 will be considered a non-responder at Week 52.
	Sensitivity Analysis 1	PP Analysis Set	• Presentation of Odds-Ratio (OR), 2-sided 95% CI of OR, nominal p-value from Logistic Regression • Number and percentage of responders in each dose group, observed Δ in response proportions from Placebo for each dose group	As the Primary Analysis
	Supportive analysis 1	mITT Analysis Set	• Test of linear trend in log odds of response with increasing dose, based on an appropriate contrast from the logistic model, i.e., (3, 1, -1, -3).	As the Primary Analysis
Time to first severe (BILAG A) flare	Primary Analysis	mITT Analysis Set	• Stratified Log Rank Test for 1-sided p-values (α=0.025) for comparison between treatment groups vs Placebo • Cox Proportional Hazard model for Hazard Ratio (HR) and 95% CI of HR between each treatment group vs Placebo • Presentation of product-limit estimates • KM curves	See Section 14.2.1 for censoring rules.
	Sensitivity Analysis 1	PP Analysis Set	• Stratified Log Rank Test for p-values for comparison between treatment groups vs Placebo • Cox Proportional Hazard model for HR and 95% CI of HR between each treatment group vs Placebo • Presentation of product-limit	See Section 14.2.1 for censoring rules.

			estimates • KM curves	
	Sensitivity Analysis 2	mITT Analysis Set	• Stratified Log Rank Test for p-values for comparison between treatment groups vs Placebo • Cox Proportional Hazard model for HR and 95% CI of HR between each treatment group vs Placebo • Presentation of product-limit estimates • KM curves	For subjects with missing response at Week 52, the multiple imputation (MI) will be used to impute the missing data using the available observations.
	Supportive Analysis 1	mITT Analysis Set	• Test of dose response trend using the same Cox model as above	See Section 14.2.1 for censoring rules.

14.2.2.1 SRI-4 response among serologically active subjects

For this analysis, a subject will be considered as a SRI-4 Non-Responder if the subject takes any EAC-determined protocol-prohibited medications during the treatment period. It will be confirmed by the EAC. In addition, a subject with a missing response at Week 52 will be considered a non-responder at Week 52.

The SRI-4 response at Week 52 among serologically active subjects will be based on the mITT Analysis set, and analyzed using logistic regression model for the odds of a given SRI response adjusted for the stratification variables used in the randomization (race: black vs non-black, screening SLEDAI-2K total score: <10 vs ≥10 and region: US and Western Europe vs Japan vs ROW) and with treatment as a factor. The estimated OR, together with associated two-sided 95% CI of the estimated OR and nominal one-sided p-value (α=0.025) from the logistic regression model, comparing each evobrutinib treatment group to placebo will be presented. The multiplicity adjustment procedure as outlined in Section 14.1.2.4 will be applied to assess statistical significance.

The logistic model will be repeated on PP analysis set if the size of these analysis sets differs substantially (>10%) from mITT.

In addition, the number and percentage of SRI-4 responders and non-responders at Week 52 in each treatment group will be presented the same way as for the primary analysis (see Section 14.1.2.4).

SRI-4 response rates over time will be also presented by visit, using similar imputation rules as applied for the Week 52 timepoint.

14.2.2.2 Analysis of Time to first severe BILAG A flare

The time to first severe flare will be analyzed on the mITT Analysis Set using a stratified log rank test with treatment as main effect and stratification variables used in the randomization (race, screening SLEDAI-2K total score and region) as covariates. The one-sided p-values (α=0.025) from the stratified log rank test for treatment comparisons vs placebo will be reported. The multiplicity adjustment procedure as outlined in Section 14.1.2.4 will be applied to assess statistical significance.

In addition, a Cox proportional hazards regression model including treatment as factor and the stratification factor at randomization as covariates will be performed. Ties will be handled by replacing the proportional hazards model by the discrete logistic model. The HR together with its 95% CI will be presented as determined by the proportional hazards model. Placebo group will be used as reference group. The assumptions of proportional hazards will be verified graphically (the curves representing the relative risk will be plotted for each treatment group and should be parallel).

KM estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics (median, Q1 and Q3, minimum and maximum) including the corresponding two-sided 95% CI of the median. The KM 5%, 10%, 15% and 20% percentiles as well as the corresponding 95% CI will also be presented. In addition, the number of severe flares by treatment group will be presented.

The KM curves will also be provided for each treatment group.

As sensitivity analyses, stratified log-rank test, Cox model, KM estimates and KM curves will be repeated:

- Using MI as stated in Section 14.2.1,
- On PP analysis set if the size of these analysis sets differs substantially (>10%) from mITT.

In addition, a test for trend in dose-response, using the same Cox model, will be reported as a supportive analysis, using contrast statement and (-3 -1 1 3) coefficients.

An example of SAS code is available in Appendix 18.9.

If there is model convergence problem due to for example zero cells in covariates, appropriate method will be applied to resolve the issue.

14.2.2.3 Subgroup Analysis

Time to first severe (BILAG A) flare

To assess numerically the consistency among the subgroups (Table 4) of time to first severe flare, the analysis will be performed on the mITT Analysis Set using the Cox proportional regression model with treatment, subgroup of interest, treatment by subgroup of interest

interaction as main effects. The estimated HR for treatment comparison with placebo, together with associated two-sided 95% CI of the estimated HR, nominal p-value from the model contrast will be presented for each treatment group. The same censoring approach will be used as the main analysis. No multiple testing adjustments will be performed.

A forest plot of the HR and the 95% CI of the hazard ratio between each treatment group and placebo will be presented for each category of the subgroups.

SRI-4 response among serologically active subjects

Primary analysis of SRI-4 response among serologically active subjects will be repeated for subgroup A from Table 4 (HDA vs non-HDA). The estimated OR, together with associated two-sided 95% CI of the estimated OR, nominal p-values from the logistic regression model, comparing each evobrutinib treatment group to placebo will be presented. No multiple testing adjustments will be performed.

14.3 Other secondary endpoints

All other secondary endpoints will be analyzed for the PA using the mITT Analysis Set.

All statistical tests of the other secondary efficacy endpoints will be conducted at a two-sided α level of 0.05, with p-values considered nominal. P-values and the 95% CIs will be presented where applicable.

14.3.1 Disease Activity Endpoints

The following secondary endpoints to be analyzed on mITT analysis set are summarized in Table 19. LLDAS is defined in Appendix 18.1.7.

For SRI-4 and SRI-6 response endpoints as well as LDA status and LLDAS, the primary analysis method applied to the two-primary endpoints (Section 14.1.2.1) will be performed.

In addition, for LDA and LLDAS endpoints in Table 19, the number and percentage of subjects with response or non-response will be summarized over time (by visit), using similar imputation rules as applied for the Week 52 timepoint.

Table 19 : Disease Activity Endpoints

Endpoint	Analysis method	Data Handling Notes
• SRI-6 response at Week 52, in the serologically active subgroup • SRI-4 Response at Week 52 with a Sustained Reduction of OCS Dose to 7.5 milligrams prednisone-equivalent per day or less (≤ 7.5 mg/day) and less than or equal to (≤) Day 1 dose during Week 41 Through Week 52, in all subjects • SRI-6 Response at Week 52 with a Sustained Reduction of OCS Dose to 7.5 milligrams prednisone equivalent per day or less (≤ 7.5 mg/day) and less than or equal to (≤) Day 1 dose during Week 41 Through Week 52, in the HDA subgroup • SRI-4 Response at Week 52 with a Sustained Reduction of OCS Dose to 7.5 milligrams prednisone equivalent per day or less (≤ 7.5 mg/day) and less than or equal to (≤) Day 1 dose during Week 41 Through Week 52, in the serologically active subgroup	Same as the Primary Method in Table 17	Same as the Primary Method in Table 17
• LDA status, defined by SLEDAI-2K ≤ 2 at Week 52 • LDA, defined by clinical SLEDAI-2K (SLEDAI-2K excluding anti-dsDNA and low complement parameters) ≤ 2 at Week 52 • LLDAS at Week 52	Same as the Primary Method in Table 17	Same as the Primary Method in Table 17

In order to determine whether the subject had a sustained reduction of OCS Dose to ≤ 7.5 mg/day and ≤ Day 1 dose during Week 41 Through Week 52, the daily dose of CS will be calculated for each analysis visit from Week 41 to Week 52, according to Appendix 18.8. If the condition is met for all days during this period, the subject will be considered as having a sustained reduction of OCS Dose.

14.3.1.1 Subgroup Analysis

SRI-6 response among serologically active subjects

Analysis of SRI-6 response among serologically active subjects will be repeated for subgroup A from Table 4 (HDA vs non-HDA). The estimated OR, together with associated two-sided 95% CI of the estimated OR, nominal p-values from the logistic regression model, comparing each evobrutinib treatment group to placebo will be presented. No multiple testing adjustments will be performed.

14.3.2 Time to first flare, flare-free status and AFR

Time to first flare, flare-free status, and annualized flare rate, during the 52-week treatment period, will be analyzed separately, each assessed with flare defined as:

- Severe BILAG A flare (Key secondary endpoint, see Section 14.2.2.2)
- Moderate to Severe (BILAG A or 2B) flare
- Severe SFI flare

14.3.2.1 Definitions

BILAG flare severity is defined as follows:

- Severe flare: ≥ 1 BILAG A score in any organ system due to items that are new or worse, compared to the BILAG evaluation at the previous visit
- Moderate flare: at least 2 BILAG B scores in any organ systems due to items that were new or worse, and without meeting the criteria for the severe flare, compared to the BILAG evaluation at the previous visit and no BILAG A meeting the condition of severe flare
- Mild flare: 1 single BILAG B score in any organ systems due to items that were new or worse, or at least 3 BILAG C scores in any organ systems due to items that were new or worse, compared to the BILAG evaluation at the previous visit and no BILAG A and no more than one BILAG B meeting the condition of severe or moderate flare
- No flare: defined as not meeting any of the above criteria.

'Due to items that were new or worse' means that if the assigned Grade A/B or C contains ANY ITEM which is WORSE or NEW, such grade will be included into the flare consideration. This is valid for clinical organ domains. For laboratory organ domains (i.e. renal and hematological) where items are not worse/new, there must be an increase in score from B to A or C to B for flare to be considered.

For BILAG flares, the time to first flare is defined as follows:

$$Time\ to\ first\ flare\ (days) = \ Date\ of\ visit\ where\ the\ first\ flare\ has\ been\ diagnosed - Date\ of\ randomization + 1$$

For SFI severe flares, the time to first severe flare is defined as follows:

$$Time\ to\ first\ severe\ flare\ (days) = \ Date\ of\ first\ flare\ since\ last\ flare\ assessment\ from\ the\ SFI2\ form\ of\ the\ eCRF - Date\ of\ randomization + 1$$

A subject has a flare-free status at Week 52 if no flares have been reported during the 52-week treatment period. Subjects who discontinue study prior to Week 52, without having a flare are counted as not being flare-free at Week 52.

The unadjusted AFR is the total number of flares from all subjects in the analysis set, divided by the total exposure time from all subjects in the analysis set. At the subject level, the unadjusted AFR is the total number of flares experienced by the subject divided by the total exposure time of this subject.

14.3.2.2 Statistical analysis

Table 20, Table 21 and Table 22 provide details of the Flare endpoints.

Table 20 : Time to first flare

Endpoint	Analysis method	Data Handling Notes
• Time to first severe BILAG A flare (a) • Time to BILAG A or 2B Moderate to Severe flare (b) • Time to first severe SFI flare (c)	As described in Section 14.2.2.1 (no sensitivity/supportive analyses will be performed)	See Section 14.2.1 for censoring rules
(a) A Severe (BILAG A) flare is defined as at least one BILAG A score in any organ system due to items that are new or worse, compared to the BILAG evaluation at the previous visit, during the 52-week treatment period. Time to first severe BILAG A flare analysis is described in Section 14.2.2.1 as part of key secondary endpoint. (b) A Moderate to Severe (BILAG A or 2B) flare is defined as at least one BILAG A grade or two BILAG B grade in any organ system due to items that are new or worse, compared to the BILAG evaluation at the previous visit. (c) The severity of SFI flare is recorded in the eCRF (Mild/Moderate versus Severe). However, for statistical analysis purpose, the severity will be re-derived using the SLEDAI total score as well as each sign or symptom recorded in the SFI1, SFI2, SFI3 forms of the eCRF (see Appendix 18.1.4 for a definition of Mild/Moderate or Severe flare). As the onset of new flare date or stabilization/resolution date is not recorded in the eCRF, if an investigator records a flare in two consecutive visits, it will be assumed that the flares are distinct events (i.e., when a new SFI flare is recorded in the eCRF, it will be assumed the previous one is stabilized/resolved). Therefore, only unique records from the SFI forms of the eCRF will be counted, assuming that distinct records represent distinct flare events.		

Table 21 : Flare free status during the 52-week Treatment Period

Endpoint	Analysis method	Data Handling Notes
• BILAG 2004 Flare-free status (a) • SFI flare-free status (b)	• Logistic regression • Descriptive statistics	Subjects who discontinue study prior to Week 52, without having a flare are counted as not being flare-free
(a) BILAG 2004 flare-free is defined as having only BILAG scores equal to D or E during the 52-week treatment period. (b) SFI flare-free is defined as having no SFI flare during the 52-week treatment period.		

Logistic regression model:

For each definition of flare-free, the comparison of each evobrutinib group to the placebo group using proportion flare-free during the 52-week treatment period, will be based on a logistic model for the odds of flare-free status, with treatment group and stratifications of randomization as factors.

For each treatment comparison (evobrutinib versus placebo), the adjusted OR with 95% CI, and p-value will be reported.

An example of SAS code is available in Appendix 18.9.

Descriptive statistics:

The proportion of subjects having a BILAG flare will be presented by visit, severity and treatment group.

For each definition of flare, descriptive statistics (number and percentage of subjects) for flare-free status will be provided by treatment group. In particular, the number (proportion) of subjects with ≥ 1 flare, number (proportion) of subjects with 0 flares, number (proportion) of subjects discontinuing treatment prior to Week 52 among subjects with 0 flares will be reported.

Table 22 : AFR

Endpoint	Analysis method	Data Handling Notes
• Severe BILAG A AFR (a) • Moderate to severe (BILAG A or 2B) AFR (b) • Severe SFI AFR (c)	• Negative binomial (NB) model, with offset equal to log of exposure time, • Descriptive statistics	As observed
(a) A severe (BILAG A) flare is defined as at least one BILAG A score in any organ system due to items that are new or worse, compared to the BILAG evaluation at the previous visit, during the 52-week treatment period. (b) A Moderate to Severe (BILAG A or 2B) flare is defined as at least one BILAG A grade or two BILAG B grade in any organ system due to items that are new or worse, compared to the BILAG evaluation at the previous visit. (c) The severity of flare is recorded in the eCRF (Mild/Moderate versus Severe). However, for statistical analysis purpose, the severity will be re-derived using the SLEDAI total score as well as each sign or symptom recorded in the SFI1, SFI2, SFI3 forms of the eCRF (see Appendix 18.1.4 for a definition of Mild/Moderate or Severe flare). As the onset of new flare date or stabilization/resolution date is not recorded in the eCRF, if an investigator records a flare in two consecutive visits, it will be assumed that the flares are distinct events (i.e., when a new SFI flare is recorded in the eCRF, it will be assumed the previous one is stabilized/resolved). Therefore, only unique records from the SFI forms of the eCRF will be counted, assuming that distinct records represent distinct flare events.		

NB model:

The comparison of each evobrutinib treatment group to the placebo group using AFR at Week 52 will be based on a NB model for flare count, with offset equal to the log of exposure time. Subjects who prematurely discontinue from treatment will be analyzed according to the number of days on treatment and number of flares observed at the time of treatment discontinuation. Missing post-baseline flare assessment will be counted as no flare in AFR analysis.

The AFR analysis will be performed on the mITT analysis set using the NB model, with treatment group as main effect and randomization strata as covariates. The adjusted AFR ratio

for treatment comparison versus placebo, the 95% CI of the AFR ratio and the 2-sided nominal p-value from the NB model will be presented.

The NB model will be computed with the SAS GENMOD procedure. If the model does not converge, appropriate method will be applied to resolve the issue (e.g. removal of covariate). An example of SAS code is available in Appendix 18.9.

Descriptive statistics:

Summary statistics (mean, SD, median, Q1, Q3, min, and max) will be presented by treatment group for:

- Number of flares,
- Time on treatment (subject-years),
- Total number of flares per subject-year.

14.3.3 Disease activity over time

Table 23 provides details on disease activity endpoints.

Table 23 : Disease Activity Endpoints

Endpoint (*)	Analysis method	Data Handling Notes
CFB in SLEDAI-2K score	• Descriptive statistics for continuous variable, including absolute value, CFB and percent CFB	As observed
CFB in CLASI-A score	• Descriptive statistics for continuous variable, including absolute value, CFB and percent CFB	As observed
BICLA response	• Number and percentage of subject with response by each treatment group • Presentation of adjusted Odds-Ratio (OR), 2-sided 95% CI of OR, nominal p-value, adjusted p-value from Logistic Regression (same as primary analysis of primary endpoint)	As primary analysis .
CFB in BILAG-2004	• Descriptive statistics of total score, including absolute value at baseline and each post baseline visit, CFB and percent CFB at each post baseline visit	As observed
CFB in PGA	• Descriptive statistics, including CFB and percent CFB for the PGA score as a continuous variable on a 0-3 scale	As observed
(*) See Appendix 18.1 for definition of each endpoint.		

14.3.4 HRQoL

The following HRQoL endpoints are summarized in Table 24.

All endpoints are described in Appendices 18.3, 18.4, 18.5, 18.6 and 18.7. They will be analyzed on QoL Analysis Set.

Table 24 : HRQoL analysis

Questionnaire	Score to be analyzed	Analysis Set	Analysis
SF-36 v2	Change in PCS, MCS scores, and their components	• QoL Analysis Set	• Mixed model with repeated measures (MMRM) model • Descriptive statistics • Distribution %CFB curve • Logistic regression
LupusQoL	Change in each domain score	• QoL Analysis Set	• MMRM model • Descriptive statistics • Distribution %CFB curve
FACIT-Fatigue	Change in Total score	• QoL Analysis Set	• MMRM model • Descriptive statistics • Distribution %CFB curve • Logistic regression
EQ-5D-5L	Change in VAS score, EQ-5D-5L index	• QoL Analysis Set	• MMRM model • Descriptive statistics
PGIC	% of subjects with improvement, no change and worsening	• QoL Analysis Set	• Logistic regression • Descriptive statistics

MMRM model:

Missing data will be considered as MAR.

The QoL endpoints will be analyzed using the MMRM model, with treatment group, visit, and treatment by visit interaction as the main effects, baseline assessment and randomization strata as covariates. The unstructured covariance matrix will be considered. Denominator degree of freedom will be computed using Kenward and Roger's method (Kenward 1997).

The LSMEAN estimate of mean change from baseline score, difference of the LS Mean from placebo, 2-sided 95% CI of the difference, and nominal 2-sided p-value of the difference from the model contrast will be presented.

An example of SAS code is available in Appendix 18.9.

Descriptive statistics:

For each treatment group, descriptive statistics will be presented for absolute value, CFB and percent CFB value for each visit, from baseline to Week 52.

In addition, the following will be provided:

- Mean score CFB will be presented as a by-visit line plot for each treatment group, with all treatment groups included in a single figure, and horizontal axis extending to Week 52
- A bar-chart figure of Mean score CFB at Week 52 presenting each SF-36 sub-domain score type (PF, RP, BP, GH, VT, SF, RE, MH) for each treatment will be provided.
- For PGIC, the number (proportion) of subject of subjects with any improvement (i.e., PGIC = 1, 2 or 3), no change (i.e., PGIC = 4) and any worsening (i.e., PGIC = 5, 6 or 7) will be summarized by visit and treatment group.

Probability of achieving %CFB value or higher:

For each SF-36 endpoint (domain scores, MCS, PCS), a figure will be provided describing the distribution of % CFB at Week 52, one curve per treatment group. The curve will display proportion of subjects having a value for % CFB > x at Week 52, where the range of x depends on the data. The same will be applied for FACIT-Fatigue total score and each domain score of LupusQoL.

Logistic regression:

Response on the SF-36, FACIT-Fatigue total score and PGIC will be further analyzed based on differences in proportion of responders, comparing each dose of evobrutinib to placebo.

Responder definitions will be based on minimal clinically important difference (MCID) criteria for improvement or deterioration from baseline as follows (Strand 2005, Lai 2011):

- SF-36 – PCS & MCS:
    - Improvement is defined as an increase greater or equal to 2.5 compared to baseline
    - Deterioration is defined as a decrease greater or equal to 0.8 compared to baseline
- SF-36 – domain scores:
    - Improvement is defined as an increase greater or equal to 5.0 compared to baseline
    - Deterioration is defined as a decrease greater or equal to 2.5 compared to baseline
- FACIT-Fatigue Total Score:

- Improvement is defined as an increase greater or equal to 4.0 compared to baseline
- Deterioration is defined as a decrease greater or equal to 4.0 compared to baseline

- PGIC: improvement is defined as value = 1, 2 or 3

Logistic regression model for repeated measures will be applied, with adjustments for the randomization strata, baseline value and treatment as a factor. The estimated OR at Week 52, together with associated two-sided 95% CI of the estimated OR will be presented. For PGIC improvement, results will be presented at Week 12, 24, 40 and 52 visits. An example of SAS code is available in Appendix 18.9.

14.3.4.1 Subgroup analysis

MMRM, logistic regression, distribution %CFB curves and descriptive statistics will be repeated on QoL Analysis Set for each SF-36 endpoint, FACIT-Fatigue total score and each domain score of LupusQoL according to the following subgroups:

- Screening SLEDAI-2K scores:

a. <10 (non-HDA);

b. ≥10 (HDA)

- Baseline severity of disease:

a. Severe: defined as at least one BILAG A; vs.

b. Moderate: defined as at least 2 BILAG B and no BILAG A; vs.

c. Mild: defined as 1 BILAG B and no BILAG A and no more than 1 BILAG B

d. No severity: defined as criteria for severe, moderate and mild not met.

14.3.5 Corticosteroid usage over time

CS usage over time will be analyzed on the mITT set using data collected in "Relevant Previous Medications Details" and "Concomitant Medications" eCRF Forms.

Only CS taken at baseline or after will be taken into account. Any CS with an end date before screening date will not be taken into account in the following analysis.

General rules, missing data handling, conversions and cumulative dose is defined in Appendix 18.8.

The following endpoints will be described for statistical analysis:

- Clinically meaningful reduction in CS dose from Baseline, defined by:
  - A reduction of daily prednisone-equivalent CS dose ≥ 25% to a dose of ≤ 7.5 mg/day by Week 40 + 1 day (i.e. Week 41) and sustained through Week 52

AND

- No new BILAG A organ domain scores and no more than one new BILAG B organ domain score during Weeks 40 + 1 day (i.e. Week 41) through 52.

Subjects withdrawing from treatment before Week 40 or with missing data at Week 40 will be considered as non-responders. The analysis will be performed on the mITT subjects whose CS dose is greater than or equal to 7.5 mg at baseline.

- Change from Baseline to Week 52 in prednisone-equivalent CS daily dose.

Summary statistics for absolute value, CFB and percent CFB by visit during treatment period in prednisone-equivalent CS daily dose will be provided.

- Reduction by visit, from Baseline to Week 52, in prednisone-equivalent CS daily dose of zero to < 25%, 25% to 50%, > 50%, or an increase.

The same table will be produced with the following categories of prednisone-equivalent CS daily dose at Week 52: ≤ 6mg, > 6 - 7.5 mg, > 7.5 – 10 mg, > 10 – 20 mg, > 20 – 30 mg and > 30 mg. The proportions are based on the observed subjects as the denominator.

- Cumulative prednisone-equivalent CS dose from Baseline until completion of the Treatment Period (absolute value only) will be described using summary statistics.

Cumulative prednisone-equivalent CS dose will be compared between placebo arm and evobrutinib arms with a pairwise non-parametric Wilcoxon rank sum test.

All CS doses will be listed for each subject and flagged for any doses not taken into account in calculation of the scheduled-visit dose (e.g. dose taken between two scheduled visits or dose with missing information). Subjects non-compliant with protocol rules on CS usage will be listed based on protocol deviation data (i.e. all subjects with PD code = PDEV52e).

14.4 Exploratory Endpoints

Exploratory endpoints will be analyzed for the PA only, using the mITT Analysis Set.

CCI

[redacted]

CCI

[redacted]

14.5 Additional Exploratory Endpoints

CCI

[redacted]

CCI

[redacted]

14.6 LTE Efficacy Endpoints

LTE efficacy endpoints will be analyzed on mITT for the LTE Analysis only.

Same methods of imputation used for the PA will be applied to LTE analysis. However, missing efficacy assessments collected after the announcement to the sites of study early termination (i.e. 23 March 2020) will not be imputed.

Table 25 summarizes the analyses that will be performed for each LTE endpoint.

For all LTE efficacy endpoints, the baseline from the 52-weeks treatment period will be used (e.g., SRI responses, change from baseline, etc.).

Institution of EAC-determined protocol-prohibited medication/treatment (i.e., EAC-treatment failure) as well as time to first flare/flare-free/AFR will be censored at the time of announcement to the sites of study early termination (i.e., 23 March 2020).

Table 25: Analysis of LTE endpoints

Endpoint	Method
SRI response over time	SRI response rate will be summarized over time during LTE period. The baseline from the 52-week treatment period will be used (see Section 9.6). The same analysis will be repeated for subjects with a gap greater than or equal to 2 weeks between 52-week and LTE treatment periods. A line plot representing the SRI response rate over time will be also provided
Low disease activity status over time	Proportions of subjects with LLDAS, SLEDAI-2K≤2 or clinical SLEDAI-2K≤2 will be summarized over time.
Change over time in CLASI-A and SLICC/ACR Damage Index Score	Descriptive statistics will be presented over time for each endpoint.
Change over time in disease activity as measured by the BILAG, SLEDAI-2K, and PGA	Descriptive statistics will be presented over time for each endpoint.
BICLA response over time	Proportion of subjects with BICLA response will be summarized over time.
Change over time in prednisone-equivalent CS dose	Descriptive statistics will be presented as detailed in Section 14.3.5. This will be applied at LTE Week 24, Week 52 and Week 104.
CCI [redacted]	
Time to first flare, flare-free status and annualized flare rate	Time to first flare, flare-free status and annualized flare rate will be analyzed as in Section 14.3.2 at Week 24, Week 52 and Week 104 with descriptive statistics only. No statistical modeling will be performed. Time to first flare will be assessed starting from randomization (Day 1 of 52-week treatment period) and will be derived for the whole treatment period. Flare analyses will be censored at the time of announcements to the sites of study early termination (i.e. 23 March 2020)

15 Safety Evaluation

For all safety endpoints, data selection for PA and LTE analysis will be handled as indicated in the Table 26.

Table 26 : Data Handling for Safety Analysis

Data summarized	Analysis	Analysis sets	Period covered	Treatment groups
• AEs • Laboratory data • Vital signs • ECGs • Other safety evaluations	**PA**	SAF	All data from 52-week treatment period (safety follow-up included)	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID
	LTE analysis	LTE SAF	All data from LTE period (safety follow-up posterior to LTE included)	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID

15.1 Adverse Events

All analyses described in this section will be based on treatment-emergent adverse events (TEAEs) if not otherwise specified.

For the PA, TEAEs will be defined as:

- AEs starting on or after first treatment administration of any IMP (placebo or evobrutinib) until safety follow up (end of study) or first administration in LTE for subjects continuing in LTE,
- or if it was present prior to any IMP administration but exacerbated after.

Any AE which started before study first treatment administration of any IMP (placebo or evobrutinib) but improved during treatment period until safety follow up (end of study) or first administration in LTE for subjects continuing in LTE will not be counted as TEAE.

For the LTE analysis:

- For subjects who enter the LTE, TEAEs will be defined as:
    - AEs starting on or after first treatment administration of evobrutinib 50 mg BID during LTE until the end of study.
    - AEs starting prior first treatment administration of evobrutinib 50 mg BID during LTE, but exacerbated during the LTE

Only AEs from LTE period will be considered including AEs during the LTE safety follow-up period. Any AE which started before study first treatment administration of evobrutinib 50 mg BID but improved during LTE period will not be counted as TEAE.

Incomplete AE-related dates will be handled as follows:

- In case the onset date is missing completely or missing partially but the onset month and year, or the onset year are equal to the start of IMP then the onset date will be replaced by the minimum of start of IMP and AE resolution date.
- In all other cases the missing onset day or missing onset month will be replaced by 1.
- Incomplete stop dates will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of subject's death. In the latter case, the date of death will be used to impute the incomplete stop date.

Further information collected after the cut-off for an analysis (such as a fatal outcome) may be extracted from the Safety data base and presented separately in the CSR.

15.1.1 All Adverse Events

AEs will be coded according to the latest MedDRA version available at the time of analysis. The severity of AEs will be graded using NCI-CTCAE version 4.03 toxicity grades. AEs with missing classification concerning IMP relationship will be considered related to the IMP.

15.1.1.1 3-Tier Approach

The 3-tier approach is a systematic way to summarize and analyze adverse events (AEs) in clinical trials (Crowe 2009). AEs in different tiers are analyzed using different levels of statistical analyses.

3-Tier approach will be used for PA only.

The AEs identified by the Benefit Risk Action Team for Tier 1 reporting in this trial are in Table 27.

Table 27: Tier 1 AEs

Tier 1 AE (PT)	Within group summary	Between group comparison (Evobrutinib - Placebo) 95% CI
Transaminases elevations Alanine aminotransferase abnormal Alanine aminotransferase increased Aspartate aminotransferase abnormal Aspartate aminotransferase increased Hypertransaminasaemia Transaminases abnormal Transaminases increased	n (%)	Δ [xx%, xx%]

Tier 1 AE will be identified using sponsor-defined list of search term as defined in Table 27 (refer to Appendix 18.14).

All AEs will be further classified into Tier 2 or Tier 3 based on the Rule-of-3. If there are 3 or more subjects with the reported term in any treatment group, that term will be included in Tier 2. Otherwise, it will be included in Tier 3.

The Tier 1 and Tier 2 AEs will be assessed with a 95% CI for between-group comparisons. For the difference in crude rates, the CIs will be based on MN method (Miettinen 1985).

Exposure Adjusted Incidence Rate (EAIR) of TEAEs will also be separately presented for each Tier, by SOC and PT (see Section 15.1.1.3 for multiple events handling). For this trial, difference in EAIR will also be summarized. For Tier 1 and Tier 2 AEs, the CIs will be calculated using a Poisson binomial model. An example of SAS code is available in Appendix 18.9.

No multiplicity adjustment will be applied for Tier 1 and 2 AEs. The Tier 3 AEs will be assessed via summary statistics and risk differences.

For each comparison of evobrutinib group with placebo, forest trees for Tier 1 and Tier 2 AEs will be provided as well, displaying the incidence rate and associated 95% CI of difference.

15.1.1.2 Overview of TEAEs

Two summary tables of TEAEs will be provided as described in Table 28.

Table 28: Summary Tables of TEAEs

Summary	Modalities
Overview of TEAEs	• Any TEAE • IMP related TEAE • Serious TEAE • IMP related serious TEAE • TEAE with NCI-CTCAE grade 1 • IMP related TEAE with NCI-CTCAE grade 1 • TEAE with NCI-CTCAE grade 2 • IMP related TEAE with NCI-CTCAE grade 2 • TEAE with NCI-CTCAE grade 3 • IMP related TEAE with NCI-CTCAE grade 3 • TEAE with NCI-CTCAE grade 4 • IMP related TEAE with NCI-CTCAE grade 4 • TEAE leading to death • IMP related TEAE leading to death
Overview of TEAEs leading to Actions: • Change in dose, • Administration of medication, • Procedure, • Study termination	• TEAE with no change of dose • IMP related TEAE with no change of dose • TEAE leading to dose reduction • IMP related TEAE leading to dose reduction • TEAE leading to dose increase • IMP related TEAE leading to dose increase • TEAE leading to interruption of IMP • IMP related TEAE leading to interruption of IMP • TEAE leading to withdrawal of IMP • IMP related TEAE leading to withdrawal of IMP • TEAE leading to administration of concomitant medication • IMP related TEAE leading to administration of concomitant medication • TEAE leading to concomitant procedure • IMP related TEAE leading to concomitant procedure • TEAE leading to study termination • IMP related TEAE leading to study termination

15.1.1.3 Tabulation of Adverse Events by SOC and PT

The TEAE tables to be prepared are listed in Table 29.

Table 29: TEAE Tables to be produced

	Overall frequency	By primary SOC and PT	By PT only	By primary SOC, PT and worst grade	PA	LTE analysis	HDA subpopulation analysis for PA
TEAE overview summary	✓	*NA*	*NA*	*NA*	✓	✓	✓
TEAE leading to discontinuation of IMP/study/dose reduction of IMP overview summary	✓	*NA*	*NA*	*NA*	✓	✓	✓
All TEAEs	✓	✓	✓	✓	✓	✓	✓
Serious TEAEs	✓	✓			✓	✓	✓
Non-serious TEAEs*	✓	✓			✓	✓	✓
TEAEs leading to IMP withdrawal	✓	✓			✓	✓	✓
TEAEs leading to study termination	✓	✓			✓	✓	✓
TEAEs leading to death	✓	✓			✓	✓	✓
IMP-related TEAEs	✓	✓	✓	✓	✓	✓	✓
IMP-related serious TEAEs	✓	✓			✓	✓	✓
IMP-related TEAEs leading to IMP withdrawal	✓	✓			✓	✓	✓
IMP-related TEAE leading to study termination	✓	✓			✓	✓	✓
IMP-related TEAEs leading to death	✓	✓			✓	✓	✓
EAIR	✓	✓			✓	✓	✓

(*): A table with all non-serious TEAEs will be first provided and then only TEAEs exceeding a frequency of 5% in at least one of the treatment groups (> 5%), by SOC and PT will be provided.

Specific rules for SOC/PT tabulation

All AEs recorded during the trial (i.e., assessed from signature of informed consent until the end of the Follow-up/End of Trial visit) will be coded according to the MedDRA and assigned to a SOC and PT.

SOC terms will be sorted alphabetically.

- For the PA, each treatment group as well as evobrutinib combined (25 mg QD, 75 mg QD and 50 mg BID) will be computed and displayed in the table. PTs within each SOC will be sorted by descending frequency of this evobrutinib combined, and then alphabetically if multiple PTs have the same frequency.
- For the LTE analysis, PTs within each SOC will be sorted by descending frequency in the highest dose of evobrutinib (i.e., subjects who are administered evobrutinib 50 mg BID throughout the study), and then alphabetically if multiple PTs have the same frequency.

If a subject experiences more than one occurrence of the same TEAE (same SOC and same PT) during the trial, the subject will be counted only once for that treatment (the worst severity and the worst relationship to trial treatment will be tabulated).

In case a subject had events with missing and non-missing grades, the maximum of the non-missing grades will be displayed.

EAIR are calculated as number of subjects with AE divided by the sum of the individual times of all subjects in the safety population from start of treatment to first onset of AE. If a subject has multiple events, the exposure period of the first event is used. For a subject with no event, the exposure period is censored at the last follow-up time for the AE summarization period. The incidence rate multiplied with 1000 would give the number of AEs expected in 1000 subjects within 1-time unit (for example 1 year). EAIR of TEAEs will be presented by SOC and PT.

Subject data listings

TEAEs and non-TEAEs (i.e. pre-treatment AEs) will be listed separately by treatment group and subject. For subjects with a gap greater than or equal to 2 weeks between 52-Week treatment period and LTE treatment period, AEs collected during the gap period will be presented in specific data listings.

A listing of TEAEs leading to withdrawal of IMP, a listing of TEAEs leading to study termination, if any, will be provided as well.

15.1.1.4 Subgroup Analysis

TEAEs by SOC/PT/Worst grade will be summarized according to group F from Table 4: Subgroups and Strata. Only those with SOC = “Infections and infestations” will be presented.

15.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

15.2.1 Deaths

A summary of deaths will be provided including (clinicaltrials.gov requirement):

- Number and percentage of (all) deaths
- Number and percentage of the primary cause of death (categories: disease progression, adverse event, unknown, other)

The tabulation of TEAEs leading to death is described in Section 15.1.1.3. A listing of deaths, if any, will be provided.

In case there is no death in the trial, only the summary of death required by clinicaltrials.gov will be performed, neither tabulation of TEAE leading to death will be produced, nor the listing of death.

15.2.2 Serious Adverse Events

The tabulation of serious TEAEs is described in Section 15.1.1.3. A subject listing of serious TEAEs will be provided for PA.

15.2.3 TEAEs of Special Interest

The following events are defined as AEs of Special Interest (AESI):

- Liver AEs
    - Transaminase and bilirubin elevations, Hy's law cases
    - Hepatitis non-infectious
    - Hepatic failure, fibrosis and cirrhosis, and other liver damage-related conditions
    - Immune-medical hepatitis, alloimmune hepatitis, non-alcoholic fatty liver
- Severe Infections:
    - Serious AEs or Grade ≥ 3
    - Opportunistic infections Grade ≥ 3
- Amylase and Lipase elevations
    - Enzyme elevations Grade ≥ 3
    - Acute pancreatitis SMQ
- Seizures

AESI will be identified using Standardized MedDRA query (SMQ) if available or sponsor-defined list of search term (refer to Appendix 18.14).

An overview of AESI and related AESI will be presented by treatment group showing number and percentage of subjects experiencing AESI. AESI will be identified in listing of all TEAEs.

15.3 Clinical Laboratory Evaluation

The following laboratory parameters will be measured during the trial as part of the safety evaluation:

- Hematology,
- Biochemistry, including the following supplementary liver function tests (LFT):
    - Aspartate aminotransferase,
    - Alanine aminotransferase,
    - Alkaline phosphatase,
    - γ-Glutamyl-transferase,
    - Bilirubin (total),
- Urinalysis,
- Coagulation.

The clinical laboratory safety tests to be measured in this trial are provided in the protocol (refer to Section 7.4.3 Table 11 of the CTP). Parameters from the Section 7.4.3 of the CTP to be summarized and listed in the TLFs are provided in Appendix 18.12.

All laboratory data results will be presented using international system of units (SI).

Clinical laboratory findings (hematology, biochemistry, urinalysis, coagulation) will be summarized by treatment using descriptive statistics at baseline and each post baseline visit for absolute value and changes from baseline (see Section 8.2.3), excluding coagulation laboratory findings which will be summarized only at the screening visit.

Some laboratory results will be classified according to NCI-CTCAE Version 4.03 as provided by the central laboratory. In case a laboratory parameter has bi-directional toxicities (e.g., Potassium) both directions will be presented for the given parameter (i.e., Potassium Low and Potassium High).

On-treatment values are results of assessments done from the first IMP administration on Day 1 till End of Study (completion or early termination).

Laboratory results containing a modifier such as "<" or ">=" will be handled case by case for summary statistics (see Appendix 18.13) and will be reported as collected in the database and as imputed in subject data listings.

A shift table of baseline versus post-baseline based on the worst NCI-CTCAE grade will be presented by treatment group for hematology and biochemistry.

Subject data listings will be provided, with a flag for abnormal values, along with corresponding normal ranges:

- Laboratory gradable parameters part of NCI-CTCAE will be presented according to the categories based on normal ranges along with the grade. Abnormal values will be flagged according to the direction of toxicity as detailed in Appendix 18.12 (e.g., for a parameter such as Potassium Low, only values below the LLN will flagged).

- Laboratory parameters that are not part of NCI-CTCAE will be presented according to the categories based on normal ranges: below normal limits (Low), within normal limits (Normal), and above normal limits (High). Values that are either above ULN or below LLN will be flagged.

For subjects with a gap greater than or equal to 2 weeks between 52-Week treatment period and LTE treatment period, laboratory data collected during the gap period will be presented in specific data listings.

Boxplots of the laboratory values by treatment group and time point will be provided for the following parameters of Table 30:

Table 30: Laboratory Parameters

Category	Laboratory parameter	Conventional Unit	SI Units
Hematology	Hemoglobin	g/dL	g/L
	Red blood cell count	$10^6/\mu L$	$10^{12}/L$
	Reticulocyte count	$10^3/\mu L$	$10^9/L$
	White blood cell count	$10^3/\mu L$	$10^9/L$
	Neutrophil count	$10^3/\mu L$	$10^9/L$
	Lymphocyte count	$10^3/\mu L$	$10^9/L$
	Platelet count	$10^3/\mu L$	$10^9/L$
Biochemistry	Alanine aminotransferase (ALT)	U/L	U/L
	Albumin	g/dL	g/L
	Aspartate aminotransferase	U/L	U/L

	(AST)		
	Gamma-glutamyl transferase (GGT)	U/L	U/L
	Alkaline phosphatase	U/L	U/L
	Total bilirubin	mg/dL	µmol/L
	Amylase	U/L	U/L
	Lipase	U/L	U/L
	eGFR	mL/min/1.73m^2	mL/min/1.73m^2
	Blood urea nitrogen (BUN)	mg/dL	mmol/L

If consistent with BOA standards, the ULN and LLN will be added to the lab parameter boxplot, for any lab parameter where the normal range is the same for all subjects in the analysis set.

In addition, the following graphical displays will be provided:

- ALT values over time will be presented in a spaghetti plot for subjects with ALT grade 2 or higher elevation (i.e. grade ≥2) by treatment group
- Time in weeks from first dose to first ALT assessment of grade 2 or higher elevation (i.e. grade ≥2) for all subjects by treatment group (KM plot)
- Time in weeks from first dose to first eGFR assessment equal to less than 60 mL/min/1.73m^2 for subjects with a baseline assessment greater or equal than 60 mL/min/1.73m^2 by treatment group (KM plot)
- e-DISH (Evaluation of Drug-Induced Serious Hepatotoxicity) figure, as described in *Merz et al* (2014) will be presented including 5 panels: one for each evobrutinib dose group, one for all dose groups combined and one for placebo

For the 2 time-to-event figures, a subject reaching the end of study without experiencing the event will have his/her event time right-censored at the time of the last ALT/eGFR assessment or at day 398 (W56 SFU study day + 5) if last assessment is performed after. To be included in a time-to-event figure, a group must have at least one event. Below the horizontal axis of the time-to-event figure, # of subjects at risk will be displayed. The vertical axis may be restricted to 0.50 – 1.0, if none of the curves reach 50%. For LTE time-to-event figures, the starting point will be the Baseline of 52-week treatment period.

In this study, clinically significant lab abnormalities were recorded as adverse events. In lieu of a listing of clinically significant lab abnormalities for each domain, the following by-subject lab value listings will be provided:

- Listing of Grade ≥ 3 hematology values
- Listing of Grade ≥ 3 biochemistry values

- Listing of urinalysis with value ≥ 2 times ULN (excluding values for pH), or an increase of "++" for non-gradable parameters when applicable (for PA only).

- Listing of Grade ≥ 2 AST, ALT or Bilirubin. For each parameter, when a subject has an increase to Grade 2, all posterior values will be included in the listing.

15.4 Vital Signs

Vital signs (height (m), weight (kg), BMI (kg/m²), body temperature (°C), SBP (mmHg), DBP (mmHg), respiratory rate (breaths/min) and pulse rate (beats/min)) will be summarized by treatment group using descriptive statistics (see Section 8.2.3).

The descriptive statistics will be presented as follows:

- The baseline will be presented first

- Then each scheduled time point will be presented on subjects who have reached this time point: absolute values, CFB and percent CFB will be displayed (except for height and BMI as they are only collected at screening visit).

Body temperature, SBP, DBP, respiratory rate and pulse rate will be analyzed with shift tables of maximum CFB using the categories defined in Table 31 or PA; for FA, only SBP and DBP will be presented in shift tables:

Table 31: Vital Signs Categories

Parameter	Unit	Shift	Baseline categories	Post-baseline categories (absolute change)
Temperature	°C	Increase	<37 / ≥37 - <38 / ≥38 - <39 /≥39 - <40 / ≥ 40	≤0* / >0 - <1 / ≥1 - <2 /≥2 - <3 / ≥3
Pulse rate	bpm	Increase and decrease	<100 / ≥100	≤0* / >0 - ≤20 / >20 - ≤40 / >40
SBP	mmHg	Increase and decrease	<140 / ≥140	≤0* / >0 - ≤20 / >20 - ≤40 / >40
DBP	mmHg	Increase and decrease	<90 / ≥90	≤0* / >0 - ≤ 20 / >20 - ≤40 / >40
Respiratory rate	breaths/min	Increase and decrease	<20 / ≥20	≤0* / >0 - ≤5 / >5 - ≤10 / >10

* This category will include the subjects with no changes or decrease/increase in the increase/decrease part of the table respectively.

A listing of maximum CFB and a listing of all vital signs data will be provided.

15.5 12-Lead Electrocardiogram (ECG)

The 12-lead ECG data will be listed and summarized for observed values and CFB values by treatment group using descriptive statistics for PA:

- Rhythm (sinus rhythm, atrial fibrillation, other)
- Ventricular rate (beats/min)
- PR interval (ms)
- QRS duration (ms)
- QT (ms)
- Fridericia corrected QT (QTcF) (ms).

QTcF values will be presented into the following categories:

- ≤ 430 ms,
- > 430 – 450 ms,
- > 450 – 480 ms,
- > 480 – 500 ms,
- > 500 ms

and categorized according to their CFB into the categories

- ≤ 30 ms,
- > 30 – 60 ms,
- > 60 msec.

Listings of ECG quantitative values, morphological and rhythm results will be produced.

A shift table of rhythm results, from baseline to the end of 52-week treatment period (LTE period respectively for the LTE analysis), of the number and percentage of subjects for each category (Sinus rhythm, Atrial fibrillation, Other, Missing and Total) will be provided.

A shift table of morphological assessments, from baseline to the worst observation of the 52-week treatment period (LTE period respectively for the FA), of the number and percentage of subjects for each interpretation category (Normal, Abnormal-NCS, Abnormal-CS, Missing, and Total) will also be provided.

15.6 Physical Examination

No summary table will be provided since physical examination findings during screening will be recorded as medical history events and findings during the trial as AEs.

15.7 Pregnancy Test

Results of pregnancy test (urine or serum) will be listed.

15.8 Serum IgG, IgA, IgM Levels

Boxplots of Serum IgG, IgA and IgM levels (g/L) by treatment group and time point will be provided, as well as Line plots of median percent change from baseline by time point and treatment group.

Descriptive statistics (absolute value, change from baseline and percent change from baseline) by treatment group and time point will be performed as well.

Serum IgG, IgA, IgM and IgG subclass levels will be listed by treatment group, subject and time point (where applicable). IgG values < 3 g/L (severe hypogammaglobulinemia) and IgG values < 6 g/L (hypogammaglobulinemia) will be flagged in the listing.

15.9 Total B Cell Count

A boxplot of Total B cell count (cells/µL) by treatment group and time point will be provided, as well as Line plots of median percent change from baseline by time point and treatment group.

Descriptive statistics (absolute value, change from baseline and percent change from baseline) by treatment group and time point will be performed as well.

Total B cell counts will be listed by treatment group, subject and time point (where applicable).

15.10 Urinalysis Microscopic Evaluation

Urinalysis Microscopic Evaluation data will be listed by treatment group and time point (where applicable).

15.11 Columbia- Suicide Severity Rating Scale (C-SSRS)

The Columbia-Suicide Severity Rating Scale (C-SSRS) is a numerical score derived from 10 categories. The C-SSRS assesses the suicidal behavior and suicidal ideation in subjects.

Occurrence of suicidal behavior is defined as having answered “yes” to a least 1 of the 4 suicidal behavior subcategories (actual attempt, interrupted attempt, aborted attempt, and preparatory acts or behavior).

Occurrence of suicidal ideation is defined as having answered "yes" to at least one of the suicidal ideation sub-categories (wish to be dead, non-specific active suicidal thoughts, active suicidal ideation with any methods [not plan] without intent to act, active suicidal ideation with some intent to act [without specific plan], and active suicidal ideation with specific plan and intent).

Occurrence of suicidality is defined as having at least one occurrence of suicidal ideation or at least one occurrence of suicidal behavior.

For the PA and LTE analysis, the number and percentage of subjects with occurrence of suicidal behavior, occurrence of suicidal ideation and occurrence of suicidality will be summarized over time. Subject data listings will be also provided.

15.12 Urine Protein Creatinine Ratio

Absolute values, CFB and percentage CFB in UPCR at each time point during treatment period will be described.

A shift table will be provided to describe the shift from baseline to highest/lowest post-baseline value during treatment period by the following categories: <0.5 mg/mg, ≥0.5-1 mg/mg, ≥1-2 mg/mg, ≥2-3 mg/mg, ≥3-4mg/mg, ≥4 mg/mg, or missing.

UPCR values containing a modifier will be imputed as described in Appendix 18.13.

A subject data listing, by treatment group and time point (where applicable), will be provided as well. UPCR values containing a modifier such as "<" will be reported as collected in the database and as imputed in the subject data listing.

15.13 HBV DNA

Hepatitis B Virus (HBV) DNA polymerase chain reaction (PCR) testing results will be summarized and listed by treatment group and time point (where applicable). A shift table from baseline to highest post-baseline values including the categories < 20 IU/mL versus ≥ 20 IU/mL will be provided.

15.14 Local Laboratory Results

Local laboratory results that are not described in Section 15.3 will be listed by treatment group and time point (where applicable).

15.15 Drug-induced Liver Injury Kit Results

Drug-Induced Liver Injury (DILI) kit results will be listed by treatment group and time point (where applicable).

15.16 COVID-19 Impact

A listing of all subjects affected by the COVID-19 related study disruption (e.g. AEs/SAEs, missed visit/sample not done, treatment/study discontinuation) will be provided by treatment period and treatment group.

16 Analyses of Other Endpoints

CCI

CCI

CCI

CCI


CCI

CCI

[redacted]

CCI

[redacted]


CCI

17 References

Albrecht J, Taylor L, Berlin JA, et al. The CLASI (Cutaneous LE Disease Area and Severity Index): an outcome instrument for cutaneous lupus erythematosus. J Invest Dermatol. 2005;125: 889–94.

Amirfeyz R, Pentlow A, Foote J, et al. Assessing the clinical significance of change scores following carpal tunnel surgery. Int Orthop. 2009;33:181-5.

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009; 28(4):586-604.

Clowse MEB, Wallace DK, Furie RA, Petri MA, et al. Efficacy and safety of Epratuzumab in moderately to severely active systemic lupus erythematosus: results from two phase III randomized, double-blind, placebo-controlled trials. Arthritis & Rheumatology. 2017;69(2):362-375.

Crowe B, Xia A, Berlin J, et al. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. Clin Trials. 2009;6:430-40.

FACIT-Fatigue Scale: http://www.ser.es/wp-content/uploads/2015/03/FACIT-F_INDICE.pdf.

Franklyn K, Lau CS, Navarra SV, et al. Asia-Pacific Lupus Collaboration. Definition and initial validation of a Lupus Low Disease Activity State (LLDAS). Ann Rheum Dis. 2016;75(9):1615-21.

Furie R, Merrill J, Werth V, et al. Anifrolumab, an Anti-Interferon Alpha Receptor Monoclonal Antibody, in Moderate to Severe Systemic Lupus Erythematosus (SLE) [abstract]. Arthritis Rheum. 2015; 67 (suppl 10).

Furie R, Morand E, Bruce I, et al. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. The Lancet Rheumatology 2019;1:e208–19.Kalunian KC, Merrill JT, Maciuca R, et al. A Phase II study of the efficacy and safety of rontalizumab (rhuMAb interferon-α) in patients with systemic lupus erythematosus (ROSE). Ann Rheum Dis. 2016;75(1):196-202.

Khamashta M, Merrill JT, Werth VP, et al. Sifalimumab, an anti-interferon-α monoclonal antibody, in moderate to severe systemic lupus erythematosus: a randomised, double-blind, placebo-controlled study. Ann Rheum Dis. 2016;0(1-8).

Kennedy WP, Maciuca R, Wolslegel K, et al. Association of the interferon signature metric with serological disease manifestations but not global activity scores in multiple cohorts of patients with SLE. Lupus Sci Med. 2015;2(1):e0000080.

Kenward, M. and Roger, J. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics 53, 983-997.

Lai JS, Beaumont JL, Ogale S, et al. Validation of the Functional Assessment of Chronic Illness Therapy-Fatigue Scale in patients with moderately to severely active systemic lupus erythematosus, participating in a clinical trial. The Journal of Rheumatology. 2011;38:672-679.

Liu G.F., Wang J., Liu K. and Snavely D. B.; Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials; Stat Med. 2006 Apr 30; 25(8): 1275–1286.

Hochberg Y., Benjamini Y., (1990), "More Powerful Procedures for Multiple Significance Testing," *Statistics in Medicine*, 9, 811–818

Hurst H, Bolton J. Assessing the clinical significance of change scores recorded on subjective outcome measures. J Manipulative Physiol Ther. 2004;27:26–35.

Ibanez D, Gladman DD, Urowitz MB. Adjusted mean Systemic Lupus Erythematosus Disease Activity Index-2K is a predictor of outcome in SLE. J Rheumatol 2005;32:8247.

Kim M, Merrill JT, Kalunian KC, Hanrahan L, Izmirly PM. Estimating Duration of Response in Systemic Lupus Erythematosus (SLE) Trials [abstract]. Arthritis Rheumatol. 2017; 69 (suppl 10). http://acrabstracts.org/abstract/estimating-duration-of-response-in-systemic-lupus-erythematosus-sle-trials. Accessed May 8, 2018.

Manzi S, Sánchez-Guerrero J, Merrill J. Effects of belimumab, a B lymphocyte stimulator-specific inhibitor, on disease activity across multiple organ domains in patients with systemic lupus erythematosus: combined results from two phase III trials. Ann Rheum Dis 2012;71:1833–1838.

Maurer W, Bretz F. A note on testing families of hypotheses using graphical procedures. Statist Med. 2014; 33: 5340-5346. (related: Bretz, Correction. Statist Med. 2019; 38: 2504- 2504.)

Merz M, Lee KR, Kullak-Ublick GA, Brueckner A, Watkins PB. Methodology to assess clinical liver safety data. Drug Safety 2014; 37 (Suppl 1):S33-S45.

Miettinen O., Nurminen M.; Comparative analysis of two rates; Stat Med. 1985 Apr-Jun; 4(2): 213–226.

Petri M, Hellmann D, Hochberg M. Validity and reliability of lupus activity measures in the routine clinical setting. J Rheumatol. 1992;19:53-9.

Petri M, Orbai A-M, Alarcón GS, et al. Derivation and validation of the systemic Lupus international collaborating clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum. 2012;64(8):2677–86.

Strand V, Crawford B. Improvement in health-related quality of life in patients with SLE following sustained reductions in anti-dsDNA antibodies. Expert Review of Pharmacoeconomics and Outcomes Research. 2005;5(3):317-326.

US Department of Health and Human Services. Common Terminology Criteria for Adverse Events Version 4.0. 2009 (v4.03: June 14, 2010).

Van Vollenhoven R et al. A framework for remission in SLE: consensus findings from a large international task force on definition of remission in SLE (DORIS). Ann Rheum Dis. 2017;76(3):554-561.

Ware J, Sherbourne C. The MOS 36-ltem Short-Form Health Survey (SF-36). Med Care. 1992;30(6):473-83.

Ware JE Jr, Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME. User's manual for the SF-36 v2 Health Survey (2nd edition). Lincoln, RI: Quality Metric Incorporated, 2007.

Yan X, Shiowjen L, Li N. Missing data handling methods in medical device clinical trials. Journal of Biopharmaceutical Statistics, 19: 1085–1098, 2009.

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

Yee CS, Farewell V, Isenberg DA, Griffiths B, Teh LS, Bruce IN, Ahmad Y, Rahman A, Prabu A, Akil M, McHugh N, Edwards C, D'Cruz D, Khamashta MA, Maddison P, Gordon C. The BILAG-2004 index is sensitive to change for assessment of SLE disease activity. Rheumatology (Oxford). 2009 Jun;48(6):691-5.

18 Appendices

18.1 Definitions of Efficacy Endpoints

18.1.1 BILAG 2004 Disease Activity Index

The BILAG 2004 Disease Activity Index (Yee 2009, Yee 2010) evaluates SLE activity in a number of organ systems, based on the principle of "physician's intention to treat" (refer to Manual of Procedures). The primary purpose of the BILAG in this study is to assess possible worsening in specific organ systems. Additional analyses of improvements in disease activity as assessed by the BILAG 2004 will also be done.

A separate alphabetic score is assigned to each organ system, corresponding in general to the following definitions:

- BILAG A: Severe disease activity requiring any of the following treatments (e.g., systemic high dose oral CS, intravenous pulse CS, systemic immunosuppressants, or therapeutic high dose anticoagulation in the presence of high dose CS or ≥ 20 mg prednisone). Note that in the context of a CTP with medication restrictions and blinded IMP, the term "requiring" is not taken literally, but indicates that if all else were equal this would be the degree of treatment indicated. It is also understood that some subjects respond to different levels of medication than others, so that in assessing subjects with the BILAG "intent to treat" really means that most subjects with this degree of symptom would require this level of treatment, not necessarily the subject in question.
- BILAG B: Moderate disease activity requiring treatment with systemic low dose oral glucocorticoids, intramuscular or intra-articular or soft tissue CS injection, topical CS or immunosuppressants, or symptomatic therapy such as antimalarials or NSAIDs.
- BILAG C: Mild disease
- BILAG D: System previously affected but now inactive
- BILAG E: System never involved

The BILAG 2004 is evaluated by scoring each of a list of signs and symptom as: improving (1); same (2); worse (3); new (4); not present (0); not done (ND). For some items, appropriate responses may be: Y/N or numerical values where indicated, or Y/N confirm this is due to SLE activity.

All signs and symptoms scored must be due to SLE. Use of a glossary provided with the BILAG 2004 instrument and training of assessors in use of the instrument are essential to obtaining reliable and consistent results.

Use of the BILAG 2004 index for evaluating flares has been identified as a robust way of evaluating the efficacy of drugs; this judgment has been corroborated by external advisors and regulatory authorities.

BILAG assessments should be conducted by a trained evaluator.

18.1.1.1 Constitutional

Grade A:
Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) AND
Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Weight loss
- Lymphadenopathy/splenomegaly
- Anorexia

Grade B:
Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) **OR**
Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Weight loss
- Lymphadenopathy/splenomegaly
- Anorexia

BUT do not fulfil criteria for Grade A

Grade C
Pyrexia recorded as 1 (improving) **OR**
One or more of the following recorded as > 0:

- Weight loss
- Lymphadenopathy/Splenomegaly
- Anorexia

BUT does not fulfil criteria for Grade A or B

Grade D

If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

18.1.1.2 Mucocutaneous

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Skin eruption - severe
- Angio-oedema - severe
- Mucosal ulceration - severe
- Panniculitis/Bullous lupus - severe
- Major cutaneous vasculitis/thrombosis

Grade B
Any Grade A features recorded as 1 (improving) **OR**
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Skin eruption - mild
- Panniculitis/Bullous lupus - mild

- Digital infarcts or nodular vasculitis
- Alopecia - severe

Grade C
Any Grade B features recorded as 1 (improving) **OR**
Any of the following recorded as > 0:

- Angio-oedema - mild
- Mucosal ulceration - mild
- Alopecia - mild
- Periungual erythema/chilblains
- Splinter haemorrhages

Grade D
If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

18.1.1.3 Neuropsychiatric

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Aseptic meningitis
- Cerebral vasculitis
- Demyelinating syndrome
- Myelopathy
- Acute confusional state
- Psychosis
- Acute inflammatory demyelinating polyradiculoneuropathy
- Mononeuropathy (single/multiplex)
- Cranial neuropathy
- Plexopathy
- Polyneuropathy
- Status epilepticus
- Cerebellar ataxia

Grade B
Any Grade A features recorded as 1 (improving) **OR**
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Seizure disorder
- Cerebrovascular disease (not due to vasculitis)
- Cognitive dysfunction
- Movement disorder
- Autonomic disorder
- Lupus headache - severe unremitting
- Headache due to raised intracranial hypertension

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

Grade D
If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

18.1.1.4 Musculoskeletal

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

Grade B
Any Grade A features recorded as 1 (improving) **OR**
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):
- Mild Myositis
- Moderate Arthritis/Tendonitis/Tenosynovitis

Grade C
Any Grade B features recorded as 1 (improving) **OR**
Any of the following recorded as > 0:
- Mild Arthritis/Arthralgia/Myalgia

Grade D
If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

18.1.1.5 Cardiorespiratory

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):
- Myocarditis/Endocarditis + Cardiac failure
- Arrhythmia
- New valvular dysfunction
- Cardiac tamponade
- Pleural effusion with dyspnoea
- Pulmonary haemorrhage/vasculitis
- Interstitial alveolitis/pneumonitis
- Shrinking lung syndrome
- Aortitis
- Coronary vasculitis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Pleurisy/Pericarditis
- Myocarditis - mild

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

Grade D
If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

18.1.1.6 Gastrointestinal

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Peritonitis
- Lupus enteritis/colitis
- Intestinal pseudo-obstruction
- Acute lupus cholecystitis
- Acute lupus pancreatitis

Grade B
Any Grade A feature recorded as 1 (improving) **OR**
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Abdominal serositis and/or ascites
- Malabsorption
- Protein losing enteropathy
- Lupus hepatitis

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

Grade D
If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

18.1.1.7 Ophthalmic

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Orbital inflammation/myositis/proptosis
- Keratitis - severe
- Posterior uveitis/retinal vasculitis - severe
- Scleritis - severe
- Retinal/choroidal vaso-occlusive disease
- Optic neuritis
- Anterior ischaemic optic neuropathy

Grade B
Any Grade A features recorded as 1 (improving) **OR**
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

- Keratitis - mild
- Anterior uveitis
- Posterior uveitis/retinal vasculitis - mild
- Scleritis - mild

Grade C
Any Grade B features recorded as 1 (improving) **OR**
Any of the following recorded as > 0:

- Episcleritis
- Isolated cotton-wool spots (cytoid bodies)

Grade D
If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

18.1.1.8 Renal

Items recorded as "Not SLE related" won't be considered for the scoring of the RENAL Grade. When an item is ticked "Not SLE related" this means that the item is considered as Normal, no grading. This item does not have to be considered as missing.

Grade 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; **or**
(b) 24 hour urine protein > 1 g that has not decreased (improved) by ≥ 25%; **or**
(c) urine protein-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥25% **or**
(d) urine albumin-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ 25%

(b) and (d) are not measured in this study and will be disregarded from grade derivation.

(a) will not be used in grade derivation neither, see the paragraph 'Notes' at the end of the Renal grade section.

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 m2 and having fallen to < 67% of previous value; **or**
   (c) GFR < 50 ml/min per 1.73 m2, and last time was > 50 ml/min per 1.73 m2 or was not measured.

4. Active urinary sediment
5. Histological evidence of active nephritis within last 3 months
6. Nephrotic syndrome

Grade B
One of the following:
1. One of the Grade A feature
2. Proteinuria (that has not fulfilled Grade A criteria)

   (a) urine dipstick which has risen by 1 level to at least 2+; **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%

   (b) and (d) are not measured in this study and will be disregarded from grade derivation.
   (a) will not be used in grade derivation neither, see the paragraph 'Notes' at the end of the Renal grade section.

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

Grade C
One of the following:

1. Mild/Stable proteinuria defined as
   (a) urine dipstick ≥ 1+ but has not fulfilled criteria for Grade A & B (used only if UPCR result is not available); **or**
   (b) 24 hour urine protein > 0.25 g but has not fulfilled criteria for Grade A & B ; **or**
   (c) urine protein-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grade A & B; **or**
   (d) urine albumin-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grade A & B

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

   (b) and (d) are not measured in this study and will be disregarded from grade derivation.
   (a) will not be used in grade derivation neither, see the paragraph 'Notes' at the end of the Renal grade section.

Grade D
If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

Notes:
Although albumin-creatinine ratio and protein-creatinine ratio are different, the same cut-off values are used for both in this index. However, albumin-creatinine ratio is not collected and therefore not used.
Urine dipstick results will not be used for BILAG scoring due to the unreliable results of urine dipstick for urine protein and lack of correlation with UPCR values (e.g., normal, ++ and +++ for urine protein by urine dipstick were reported in subjects with normal or high UPCR values).
When the criterion is evaluated by comparison with a previous assessment (i.e. deteriorating proteinuria, deteriorating renal function, evolution of plasma), the value at the previous visit will be taken. As a consequence, the RENAL grade cannot be derived at screening visit.

18.1.1.9 Haematological

Items recorded as "Not SLE related" won't be considered for the scoring of the HAEMATOLOGICAL Grade.

Grade A
TTP recorded as 2 (same), 3 (worse) or 4 (new) **OR**
Any of the following:
- Evidence of haemolysis and Hemoglobin < 8 g/dl
- Platelet count < 25 x 10^9/l

Grade B
TTP recorded as 1 (improving) **OR**
Any of the following:
- Evidence of haemolysis and Hemoglobin 8 - 9.9 g/dl
- Hemoglobin < 8 g/dl (without haemolysis)
- White cell count < 1.0 x 10^9/l
- Neutrophil count < 0.5 x 10^9/l
- Platelet count 25 - 49 x 10^9/l

Grade C
Any of the following:
- Evidence of haemolysis and Hemoglobin ≥ 10 g/dl
- Hemoglobin 8 – 10.9 g/dl (without haemolysis)
- White cell count 1 - 3.9 x 10^9/l
- Neutrophil count 0.5 - 1.9 x 10^9/l
- Lymphocyte count < 1.0 x 10^9/L
- Platelet count 50 - 149 x 10^9/l
- Isolated Coombs' test positive

Isolated Coombs's test will not be considered missing if not done.

Grade D
If criteria for Grade A or B or C are not fulfilled, then subject is by default in Grade D.

Handling of missing items for organ specific scores:

It is anticipated that there are no missing items for the following clinical organ systems: constitutional, mucocutaneous, neuropsychiatric, musculoskeletal, cardiorespiratory, gastrointestinal and ophthalmic.

For renal and haematological organ systems, "Not SLE related" items will not be considered as missing values and will therefore not be considered for the scoring of the organ domain grade. Missing items (including 'Not done'') will be imputed using the value carried forward from previous analysis visit (only value assessed post-Day 1 dose [i.e. after baseline] can be used for imputation). For renal grade, UPCR will be the only parameter to consider assessing proteinuria. The following logic will be applied: if UPCR is missing (including 'Not done'), then missing UPCR (and attribution to SLE) will be imputed using the value carried forward from previous analysis visit.
If there are missing items that lead to a missing grade for one or two of the laboratory organ domains, the missing domain grade will be imputed using the value carried forward from previous analysis visit. In such case, BILAG flare will be assessed based on domain grades that were not carried over from previous visit (i.e. excluding the renal and/or haematological domain grade).

If a subject completely missed a visit, the BILAG grades from previous analysis visit will be used and BILAG flare will be derived as missing.

18.1.1.10 Total and organ specific numeric scores

Attribution of numeric values to each grade (derived above) is as follow:

Table 32: BILAG scoring

Grade	Numeric value
A	12
B	8
C	1
D	0
E	0

The total score is calculated as the sum of the organ domain scores.

If at least one organ domain score is missing, then the Total score is missing.

18.1.2 Physician Global Assessment

The PGA is used to quantify disease activity and is measured using an anchored VAS (see Figure 1). The PGA will be determined on a continuous VAS that asks the Investigator to assess the subject's current disease activity from a score of 0 (none) to 3 (severe), with the assessment made relative not to the subject's most severe state but the most severe state of SLE per the Investigator's assessment. As per its validation method, the PGA is recommended to be completed prior to laboratory results from the individual study visit being available (Petri 1992).

This version of the PGA is not scored blindly. The assessor is instructed to look at the previous month's PGA, decide whether the overall condition of the subject is same, better or worse and move the line accordingly.

A score change, where score is moved to the right from 2.5 or below in the previous month to > 2.5 this month, denotes an arbitrary threshold for severe flare to be considered when determining if enough change has occurred to justify assessing the criterion as severe flare. A score change ≥ 1 unit to the right denotes a designation of mild/moderate flare.

Figure 1: Physician Global Assessment Visual Analog Scale with Anchors

If a subject completely missed a visit, the PGA score from previous analysis visit will be used.

18.1.3 SLEDAI-2K

The Total score of the SLEDAI-2K is calculated as the sum of the weights of the descriptors marked as present by the investigator. The weights corresponding to each of the 24 descriptors are described in the table below:

Table 33: SLEDAI scoring

Descriptor	Weight
Seizure	8
Psychosis	8
Organic brain syndrome	8
Visual disturbance	8
Cranial nerve disorder	8
Lupus headache	8
Cerebrovascular accident	8
Vasculitis	8
Arthritis	4
Myositis	4
Urinary casts[#]	4
Hematuria[#]	4
Proteinuria[#]	4
Pyuria[#]	4
Rash	2
Alopecia	2
Mucosal ulcers	2
Pleurisy	2
Pericarditis	2
Low complement*[#]	2
Increased DNA binding*[#]	2
Fever	1
Thrombocytopenia[#]	1
Leukopenia[#]	1

*blinded descriptor
[#]laboratory components

Handling with blinded descriptor: As "Increased DNA binding" and "Low complement" are blinded items, the SLEDAI-2K total score will have to be derived for primary analysis with the laboratory unblinded values. "Increased DNA binding" item is defined as ">25% binding by Farr assay or above normal range for testing laboratory". The parameter to take into account in laboratory data is "anti-Double Stranded DNA". Increased DNA binding will be derived as 'present' if the anti-dsDNA value is positive (i.e. > 15 U/mL). "Low complement" item is defined as "decrease in C3 or C4 below the lower limit of normal for testing laboratory". The parameters to take into account in laboratory data are "Complement C3" and "Complement C4".

Total score used for statistical analysis will be re-calculated for all timepoints except screening and Day 1 with the unblinded values of Increased DNA binding and Low complement. The total score for the next visits as stated by the investigator in the eCRF will not be used in any statistical analysis.

Handling of missing items in SLEDAI-2K scores and change from baseline scores:

Actual scores

If the missing clinical component has weight > 4 in the overall score (see Table 33), then no imputation is performed. The SLEDAI-2K score will be missing.

If the missing clinical component has weight ≤ 4 in the overall score (see Table 33) or laboratory components missing, then the missing values will be imputed with the value of previous analysis visit. Only value assessed post-Day 1 dose (i.e. after baseline) can be used for imputation. If at least one item is still missing, then SLEDAI-2K score will be calculated based on available data.

If a subject completely missed a visit, the SLEDAI-2K total score from previous analysis visit will be used.

Change from baseline and SRI response

In general, SLEDAI clinical components are not expected to be missing.

In unlikely case of missing clinical value, change from baseline and thus SRI response (see Section 14.1.1) will be derived as follows:

- If the missing clinical component has weight > 4 in the overall score (see Table 33), then no imputation is performed. The SLEDAI-2K score will be missing and thus subject will be considered as non-SRI responder.

- If the missing clinical component has weight ≤ 4 in the overall score (see Table 33) or laboratory components missing, then the first step of imputation will be to carry forward the value of previous analysis visit (only value assessed while subject is on-treatment can be used for imputation). If that corresponding previous analysis visit value is also missing, then the component value will remain missing and the steps below will be followed:

    - Baseline SLEDAI-2K score will be re-derived with corresponding missing values at scheduled visit set to missing at Baseline visit

    - Change from baseline to analysis visit will be calculated and use as follows for SRI response derivation:

        - If there is a change from baseline ≥ 4 points reduction based on available non-missing components, then criterion #1 for SRI response is met

        - If there is an increase (i.e. worsening) or no change, or 1 to 3 points reduction based on available non-missing component, then the subject will be considered as non-SRI responder for that analysis visit

18.1.4 SLEDAI Flare Index

The SFI can be used with any version of the SLEDAI and will be used with the SLEDAI-2K for the purposes of this trial.

A mild/moderate flare is defined as any of the following

- Increase in SLEDAI-2K total score of 3 points or more (but total score not to more than 12).
- New or worse discoid, photosensitive, profundus, cutaneous vasculitis, bullous lupus; or nasopharyngeal ulcers; or pleuritic; or pericarditis; or arthritis; or fever due to SLE.
- Increase in prednisone, but not to > 0.5 mg/kg/day.
- Added NSAID or hydroxychloroquine (or chloroquine) for SLE activity.
- ≥ 1.0 increase in PGA score, but score not to exceed 2.5 (assuming the PGA score has been transformed to a 0-3-point scale).

A severe flare is defined as any of the following:

- Increase in SLEDAI-2K total score leading to total score > 12.
- New or worse central nervous system SLE; or vasculitis; or nephritis; or myositis; or platelets < 60,000/mm3, or hemolytic anemia with hemoglobin < 70 g/L or decrease in hemoglobin > 30 g/L AND requiring: double prednisone, or prednisone increase to > 0.5 mg/kg/day, or hospitalization due to SLE.
- Increase in prednisone to > 0.5 mg/kg/day.
- New cyclophosphamide, azathioprine, methotrexate, or mycophenolate for SLE activity.
- Hospitalization for SLE activity.
- Increase in PGA score leading to total score > 2.5 (assuming the PGA score has been transformed to a 0-3-point scale).

SFI used for statistical analysis will be re-calculated with the SLEDAI-2K total score derived with the unblinded value of Increased DNA binding and Low complement.

Subjects with missing SFI assessment for a visit will be counted as missing in the SFI summary statistic table for that visit.

18.1.5 CLASI

Cutaneous Lupus Erythematosus Disease Area and Severity Index is a validated measurement instrument for lupus erythematosus developed for use in clinical studies that consists of separate scores for the activity of the disease (CLASI-A) and the damage done by the disease (CLASI-D).

The CLASI activity score is calculated on the basis of erythema, scale/hyperkeratosis, mucous membrane involvement, acute hair loss and non-scarring alopecia. The CLASI damage score is calculated on the basis of dyspigmentation and scarring, including scarring alopecia. Dyspigmentation due to lesions is defined as lesions that remain visible for more than 12 months, which are considered to then be permanent (Albrecht 2005).

18.1.6 SLICC/ACR Damage Index

The SLICC/ACR Damage Index evaluates damage occurring in subjects with SLE regardless of its cause. The Total Score is automatically calculated in the eCRF as the sum of the individual items which can be grouped into 12 domains. As shown in the table below, a weight of 1 is attributed to each damage occurring since onset of disease and present for at least 6 months. A weight of 2 can be attributed to some damages which occurred more than once. A weight of 3 is attributed to the occurrence of end-stage renal disease.

Table 34: SLICC scoring

Domain	Item	Score
Ocular Score	Any Cataract Ever	1
	Retinal Change or Optic Atrophy	1
Neuropsychiatric Score	Cognitive Impairment	1
	Seizures Requiring Therapy For 6 Months	1
	Cerebrovascular Accident Ever	1 (2)
	Cranial or Peripheral Neuropathy	1
	Transverse Myelitis	1
Renal Score	Estimated or Measured Glomerular filtration rate	1
	Proteinuria	1
	End-Stage Renal Disease	3
Pulmonary Score	Pulmonary Hypertension	1
	Pulmonary Fibrosis	1
	Shrinking Lung	1
	Pleural Fibrosis	1
	Pulmonary Infarction	1
Cardiovascular Score	Angina or Coronary Artery Bypass	1
	Myocardial Infarction Ever	1 (2)
	Cardiomyopathy	1
	Valvular Disease	1
	Pericarditis For 6 Months	1
Peripheral Score	Claudication For 6 Months	1
	Minor Tissue Loss	1
	Significant Tissue Loss Ever	1 (2)
	Venous Thrombosis	1
Gastrointestinal Score	Infarction or Resection of Bowel, Spleen, Liver or Gall Blader	1 (2)
	Mesenteric Insufficiency	1
	Chronic Peritonitis	1
	Upper Gastrointestinal Tract Surgery	1
Musculoskeletal Score	Muscle Atrophy or Weakness	1
	Deforming or Erosive Arthritis	1
	Osteoporosis With Fracture	1
	Avascular Necrosis	1 (2)
	Osteomyelitis	1
Skin Score	Scarring Chronic Alopecia	1
	Extensive Scarring	1
	Skin Ulceration	1
Gonadal Score	Premature Gonadal Failure	1
Endocrine Score	Diabetes	1
Malignancy Score	Malignancy	1 (2)
Total Score		Sum of individual items

18.1.7 LLDAS

LLDAS status is defined as meeting all the following (Franklyn 2016):

- SLEDAI-2K ≤ 4 and no activity in any major organ systems (renal, central nervous system, cardiopulmonary, vasculitis, fever):
    - Derived from CRF form 'SLEDAI1', 'SLEDAI2', 'SLEDAI3' depending on the visit. In order to meet this criterion, the descriptors "seizure", "psychosis", "organic brain syndrome", "visual disturbance", "cranial nerve disorder", "lupus headache", "CVA", "vasculitis", "Urinary casts", "hematuria", "proteinuria", "pyuria", "pleurisy", "pericarditis" & "fever" must be absent (score=0) with the exception of the following descriptors: "arthritis", "myositis", "rash", "alopecia", "mucosal ulcers", "low complement", "increase DNA binding", "thrombocytopenia", "leukopenia". The above items can be present, however, the total score of them (if present) cannot be larger than 4, assuming all zeros being observed (i.e., absence) for the other descriptors.
- PGA ≤ 1 (scale 0-3)
- No new features of disease activity compared with the assessment of previous analysis visitt
    - this criterion is derived from the SLEDAI-2K questionnaire: no new features (absent) compared with previous visit (absent = score 0 / present = score>0) for each items
- Prednisone-equivalent ≤ 7.5 mg/day at each visit (see Table 49 and Table 50)
- No institution of EAC-determined protocol-prohibited medication/treatment (i.e., EAC-treatment failure)
    - This must be a case-by-case review and defined via Medical Review, and then confirmed by the EAC.

18.1.8 BICLA Response

Requirements for BILAG-based Composite Lupus Assessment (BICLA) response are (Clowse 2017), to be assessed in the subgroup of subjects with at least one A and/or two B at Baseline:

1. BILAG 2004 improvement (all A scores at Baseline improved to B/C/D, and all B scores improved to C or D);

2. No worsening in disease activity (no new BILAG 2004 A scores and ≤ 1 new B score) from Baseline;

3. No worsening of total SLEDAI-2K score from Baseline;

4. No significant deterioration (< 10% worsening) in visual analogue PGA from Baseline

5. No institution of EAC-determined protocol-prohibited medication/treatment (i.e., EAC-treatment failure).

Assessment of treatment failure will be confirmed by the EAC.

Handling of missing criteria
Missing values in BILAG organ domain scores and SLEDAI-2K total scores will be handled as detailed in appendices 18.1.1 and 18.1.3.

18.2 Revised American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus

For the purpose of identifying subjects in clinical studies, a person shall be said to have systemic lupus erythematosus if any four or more of the 11 criteria are present, serially or simultaneously, during any interval of observation (Hochberg 1997, Tan 1982).

Table 35: ACR Criteria for Classification of SLE

Criterion	Definition
1. Malar Rash	Fixed erythema, flat or raised, over the malar eminences, tending to spare the nasolabial folds
2. Discoid Rash	Erythematous raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring occur in older lesions
3. Photosensitivity	Skin rash as a result of unusual reaction to sunlight, by patient history or physician observation
4. Oral Ulcers	Oral or nasopharyngeal ulceration, usually painless, observed by physician
5. Nonerosive Arthritis	Involving two or more peripheral joints, characterized by tenderness, swelling, or effusion
6. Pleuritis or Pericarditis	1. Pleuritis: convincing history of pleuritic pain or rubbing heard by a physician or evidence of pleural effusion or 2. Pericarditis: documented by electrocardiogram or rub or evidence of pericardial effusion
7. Renal Disorder	1. Persistent proteinuria: > 0.5 g per day or > than 3+ if quantitation not performed or 2. Cellular casts: may be red cell, hemoglobin, granular, tubular, or mixed
8. Neurologic Disorder	1. Seizures: in the absence of offending drugs or known metabolic derangements; e.g., uremia, ketoacidosis, or electrolyte imbalance or 2. Psychosis: in the absence of offending drugs or known metabolic derangements; e.g., uremia, ketoacidosis, or electrolyte imbalance
9. Hematologic Disorder	1. Hemolytic anemia with reticulocytosis or 2. Leukopenia: < 4,000/mm^3 on ≥ two occasions or 3. Lymphopenia: < 1,500/mm^3 on ≥ two occasions or 4. Thrombocytopenia: < 100,000/mm^3 in the absence of offending drugs
10. Immunologic Disorder	1. Anti-DNA: antibody to native DNA in abnormal titer or 2. Anti-Sm: presence of antibody to Sm nuclear antigen or 3. Positive finding of antiphospholipid antibodies based <u>any</u> of the following: 1. an abnormal serum level of IgG or IgM anticardiolipin antibodies 2. a positive test result for SLE anticoagulant using a standard method, or a false positive serologic test result for syphilis known to be positive for at least six months and confirmed by Treponema pallidum immobilization or fluorescent

	Treponemal antibody absorption test
11. Positive Antinuclear Antibody	An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time and in the absence of drugs known to be associated with "drug-induced lupus" syndrome

18.3 36-Item Short Form Survey (SF-36) Scoring Instructions

The SF-36 is a validated 36-item, subject-reported indication of overall health status not specific to any age, disease, or treatment group (Ware 1992).

The SF-36 includes multi-item scales measuring each of the following 8 health concepts: (1) physical functioning; (2) role limitations because of physical health problems; (3) bodily pain; (4) social functioning; (5) general mental health (psychological distress and psychological wellbeing); (6) role limitations because of emotional problems; (7) vitality (energy/fatigue); and (8) general health perceptions (see Table 36). These are summarized in two summary measures of physical and mental health: the PCS and MCS.

Questions in the standard version of the SF-36 refer to a 4-week time period. Scales are scored according to the Likert method. Lower scores equate to higher disability and higher scores equate to lower disability.

The SF-36v2 multi-item scales yield a health profile (8 scores) or can be aggregated into two summary scores, the PCS score and MCS score obtained through a linear combination of weighted transformed scores from the 8 subscales. PCS and MCS are standardized, with an average of 50 and a standard deviation of 10 in the general American population. PCS and MCS are computed only if all of the 8 scale scores are available. This Appendix details how these two scores should be calculated, as described in Ware 2007.

Table 36: SF-36 – Abbreviated Item Content for the SF-36v2 Health Domain Scales

Scale	Original item#	Item# in the MS200527-0018 study	Abbreviated Item Content
Physical Functioning (PF)	3a	3	Vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports
	3b	4	Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
	3c	5	Lifting or carrying groceries
	3d	6	Climbing several flights of stairs
	3e	7	Climbing one flight of stairs
	3f	8	Bending, kneeling, or stooping
	3g	9	Walking more than a mile
	3h	10	Walking several hundred yards
	3i	11	Walking one hundred yards
	3j	12	Bathing or dressing oneself
Role-Physical (RP)	4a	13	Cut down the amount of time spent on work or other activities
	4b	14	Accomplished less than you would like
	4c	15	Limited in kind of work or other activities
	4d	16	Had difficulty performing work or other activities (eg, it took extra effort)
Bodily Pain (BP)	7	21	Intensity of bodily pain
	8	22	Extent pain interfered with normal work
General Health (GH)	1	1	Is your health: excellent, very good, good, fair, poor
	11a	33	Seem to get sick a little easier than other people
	11b	34	As healthy as anybody I know
	11c	35	Expect my health to get worse
	11d	36	Health is excellent
Vitality (VT)	9a	23	Feel full of life
	9e	27	Have a lot of energy
	9g	29	Feel worn out
	9i	31	Feel tired
Social Functioning (SF)	6	20	Extent health problems interfered with normal social activities
	10	32	Frequency health problems interfered with social activities
Role-Emotional (RE)	5a	17	Cut down the amount of time spent on work or other activities
	5b	18	Accomplished less than you would like
	5c	19	Did work or other activities less carefully than usual
Mental Health (MH)	9b	24	Been very nervous
	9c	25	Felt so down in the dumps that nothing could cheer you up
	9d	26	Felt calm and peaceful
	9f	28	Felt downhearted and depressed
	9h	30	Been happy
Self-Evaluated Transition (SET)	2	2	How health is now compared to 1 year ago

Step 1: Recoding Item Response Values
Some of the SF-36v2 items will be re-coded so that across all questions, a higher score will indicate a better health state. Questions 2, 3a-3j, 4a-4d, 5a-5c, 9b, 9c, 9f, 9g, 9i, 10, 11a, 11c will be scored as recorded; the other questions will have the scores transformed as shown in Table 37. If multiple answers are given to the same item, then the item score will be left as missing.

Table 37: SF-36 – Recoding

Question	Original code and re-code response					
Question number: 1						
Original response	1	2	3	4	5	
Re-coded response	5	4.4	3.4	2	1	
Questions numbers: 6, 11b, 11d						
Original response	1	2	3	4	5	
Re-coded response	5	4	3	2	1	
Question number: 7						
Original response	1	2	3	4	5	6
Re-coded response	6	5.4	4.2	3.1	2.2	1
Question number: 8 (if question number 7 is answered)						
Original response to #8	1	1	2	3	4	5
Original response to #7	1	2-6	1-6	1-6	1-6	1-6
Re-coded response	6	5	4	3	2	1
Question number: 8 (if question number 7 is NOT answered)						
Original response	1	2	3	4	5	
Re-coded response	6	4.75	3.5	2.25	1	
Questions number: 9a, 9d, 9e, 9h						
Original response	1	2	3	4	5	
Re-coded response	5	4	3	2	1	

Step 2: Determining Health Domain Scale Scores (0-100 Scores)
After item recoding, a total raw score is computed for each health domain scale. The total raw score is the simple algebraic sum of the final response values for all items in a given scale, as shown in Table 38. The total raw score for each scale is transformed to a 0-100 scale score using the following formula:

$$Total\ raw\ score = 100 \times \frac{(\text{Raw score} - \text{Lowest possible raw score})}{\text{Possible raw score range}}$$

Table 38: SF-36 – Values used in transforming SF-36v2 Health Survey Health Domain Scale Total Raw Scores on the 0-100 Scale

Scale	Sum of Final Response Values	Lowest and highest possible total raw scores	Possible total raw score range
PF	3a+3b+3c+3d+3e+3f+3g+3h+3i+3j	10, 30	20
RP	4a+4b+4c+4d	4, 20	16
BP	7+8	2, 12	10
GH	1+11a+11b+11c+11d	5, 25	20
VT	9a+9e+9g+9i	4, 20	16
SF	6+10	2, 10	8
RE	5a+5b+5c	3, 15	12
MH	9b+9c+9d+9f+9h	5, 25	20

Raw and transformed scale scores are not calculated for the Reported Health Transition (HT) item.

As recommended by the developers of the questionnaire, missing item responses will be treated using the "Half-scale rule", which states that a score can be calculated if the respondent answers at least 50% of the items in a multi-item scale. In such cases, the missing item data will be replaced by the mean of the answered items of its scale. If more than 50% of the items are missing within a scale, the scale score will be missing.

Step 3: Calculating Normalized Health Domain Scores
The normalized scale scores will then be calculated using the following formulas:

$$Health\ Domain\ Z_{\text{score}} = {(Health\ Domain_{0-100\ score} - a)}/{b}$$

$$Normalized\ Health\ Domain\ Score = 50 + (Health\ Domain\ Z_{\text{score}} \times 10)$$

where a and b are the Mean and Standard Deviation of the Health Domain scale in the 1998 U.S. general population as shown in Table 39.

Table 39: SF-36 – 1998 General US Population Means and Standard Deviations used to Calculate Normalized Health Domain Scores

Health Domain Scales	Mean	Standard Deviation
PF	83.29094	23.75883
RP	82.50964	25.52028
BP	71.32527	23.66224
GH	70.84570	20.97821
VT	58.31411	20.01923
SF	84.30250	22.91921
RE	87.39733	21.43778
MH	74.98685	17.75604

The advantages of the normalization of the eight health domain scales are that results for one health domain scale can be meaningfully compared with those from the other scales and that domain scores have a direct interpretation in relation to the distribution of scores in the 1998 U.S. general population.

Step 4: Scoring the Physical and Mental Component Summary Measures
The Physical Component Summary (PCS) and Mental Component Summary (MCS) measures are scored using a three-step procedure:

1. First, the 8 health domain scale scores are standardized using means and standards deviations from the 1998 U.S. general population (see Table 39).

2. Second, these Z-scores are aggregated using weights (factor score coefficient) from the 1990 U.S general population.

3. Third, aggregate PCS and MCS scores are standardized by multiplying the standardized scale by 10 and adding 50.

U.S. general population statistics used in the standardization and in the aggregation of SF-36v2 Health Survey health domain scale scores are presented in Table 40.

Table 40: SF-36 – Factor Score Coefficients used to Calculate PCS and MCS Scores for the SF-36v2

	Summary component measure factor score coefficients	
Scales	**PCS**	**MCS**
PF	0.42402	-0.22999
RP	0.35119	-0.12329
BP	0.31754	-0.09731
GH	0.24954	-0.01571
VT	0.028877	0.23534
SF	-0.00753	0.26876
RE	-0.19206	0.43407
MH	-0.22069	0.48581

Example:

Let's consider the following answers from a subject:

Table 41: SF-36 – Numerical Example – Raw Data

#	Item description	Answer
1	In general, would you say your health is:	1 – Excellent
2	Compared to one year ago, how would you rate your health in general now?	2 – Somewhat better now than one year ago
3	Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	2 – Yes, limited a little
4	Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	2 – Yes, limited a little
5	Lifting or carrying groceries	2 – Yes, limited a little
6	Climbing several flights of stairs	2 – Yes, limited a little
7	Climbing one flight of stairs	2 – Yes, limited a little
8	Bending, kneeling or stooping	1 – Yes, limited a lot
9	Walking more than a mile	2 – Yes, limited a little
10	Walking several hundred yards	2 – Yes, limited a little
11	Walking one hundred yards	2 – Yes, limited a little
12	Bathing or dressing yourself	2 – Yes, limited a little
13	Cut down on the amount of time you spent on work or other activities	2 – Most of the time
14	Accomplished less than you would like	3 – Some of the time
15	Were limited in the kind of work or other activities	2 – Most of the time
16	Had difficulty performing the work or other activities (for example, it took extra effort)	3 – Some of the time

17	Cut down on the amount of time you spent on work or other activities	2 – Most of the time
18	Accomplished less than you would like	3 – Some of the time
19	Don't do work or other activities as carefully as usual	2 – Most of the time
20	During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?	2 – Slightly
21	How much bodily pain have you had during the past 4 weeks?	3 – Mild
22	During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?	2 – A little bit
23	Did you feel full of life?	4 – A little of the time
24	Have you been a very nervous person?	3 – Some of the time
25	Have you felt so down in the dumps that nothing could cheer you up?	2 – Most of the time
26	Have you felt calm and peaceful?	4 – A little of the time
27	Did you have a lot of energy?	3 – Some of the time
28	Have you felt downhearted and low?	2 – Most of the time
29	Did you feel worn out?	4 – A little of the time
30	Have you been a happy person?	3 – Some of the time
31	Did you feel tired?	2 – Most of the time
32	During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?	4 – A little of the time
33	I seem to get ill more easily than other people	3 – Don't know
34	I am as healthy as anybody I know	2 – Mostly true
35	I expect my health to get worse	3 – Don't know
36	My health is excellent	2 – Mostly true

After having recoded the answers, the 8 domain scores are equal to:

Table 42: SF-36 – Numerical Example – Domain Scores

	Raw Score	Raw Score$_{0\text{-}100}$	Health Domain Z-Score	Normalized Health Domain Z-Score
PF	19	100 x (19-10) /20 = 45	45 - 83.29094 / 23.75883 = -1.61	-1.61 x 10 + 50 = 33.9
RP	10	100 x (10-4) /16 = 37.5	37.5 - 82.50964 / 25.52028 = -1.76	-1.76 x 10 + 50 = 32.4
BP	8.2	100 x (8.2-2) /10 = 62	62 - 71.32527 / 23.66224 = -0.39	-0.39 x 10 + 50 = 46.1
GH	19	100 x (19-5) /20 = 70	70 - 70.84570 / 20.97821 = -0.04	-0.04 x 10 + 50 = 49.6
VT	11	100 x (11-4) /16 = 43.75	43.75 - 58.31411 / 20.01923 = -0.73	-0.73 x 10 + 50 = 42.7
SF	8	100 x (8-2) / 8 = 75	75 - 84.30250 / 22.91921 = -0.41	-0.41 x 10 + 50 = 45.9
RE	7	100 x (7-3) / 15 = 27	27 - 87.39733 / 21.43778 = -2.82	-2.82 x 10 + 50 = 21.8
MH	12	100 x (12-5) / 20 = 35	35 - 74.98685 / 17.75604 = -2.25	-2.25 x 10 + 50 = 27.5

Finally, PCS and MCS are provided below using Table 40:

- PCS = -1.61 x 0.42402 + … + -2.25 x -0.22069 = -0.41
- Normalized PCS = 10 x PCS + 50 = 45.9

- MCS = -1.61 x -0.22999 + … + -2.25 x 0.48581 = -2.62
- Normalized MCS = 10 x MCS + 50 = 23.8

18.4 FACIT-Fatigue Scoring Instructions

FACIT-Fatigue is a 13-item questionnaire that assesses self-reported fatigue and its impact upon daily activities and function (Wolfe 1996). It uses a 5-point Likert-type scale (0 = not at all; 1 = a little bit; 2 = somewhat; 3 = quite a bit; 4 = very much). As each of the 13 items of the FACIT-Fatigue scale ranges from 0–4, the range of possible scores is 0–52, with 0 being the worst possible score and 52 the best. To obtain the 0–52 score, each negatively worded item response is recoded so that 0 is a bad response and 4 is good response. All responses are added with equal weight to obtain the total score.

Table 43 provides the list of items from the FACIT-Fatigue questionnaire, whereas Table 44 indicates how to derive the final score, taking into account potential missing items.

Table 43: FACIT-Fatigue: Description of 13-item Questionnaire

Item #	Description	Not at all	A little bit	Somewhat	Quit a bit	Very much
1	I feel fatigued	0	1	2	3	4
2	I feel weak all over	0	1	2	3	4
3	I feel listless (“washed out”)	0	1	2	3	4
4	I feel tired	0	1	2	3	4
5	I have trouble starting things because I am tired	0	1	2	3	4
6	I have trouble finishing things because I am tired	0	1	2	3	4
7	I have energy	0	1	2	3	4
8	I am able to do my usual activities	0	1	2	3	4
9	I need to sleep during the day	0	1	2	3	4
10	I am too tired to eat	0	1	2	3	4
11	I need help doing my usual activities	0	1	2	3	4
12	I am frustrated by being too tired to do the things I want to do	0	1	2	3	4
13	I have to limit my social activity because I am tired	0	1	2	3	4
Scoring: Items are scored as follows: 4=Not At All; 3=A Little Bit; 2=Somewhat; 1=Quite A Bit; 0=Very Much, EXCEPT items #7 and #8 which are reversed scored. Score range 0-52. A score of less than 30 indicates severe fatigue. The higher the score, the better the quality of life.						

Table 44: FACIT-Fatigue: Guideline to Compute Total Score

Item #	Reversed Item?		Item Response	Item Score
1	4	-		=
2	4	-		=
3	4	-		=
4	4	-		=
5	4	-		=
6	4	-		=
7	0	+		=
8	0	+		=
9	4	-		=
10	4	-		=
11	4	-		=
12	4	-		=
13	4	-		=
Sum individual item scores: _____ **Multiply by 13: _____** **Divide by number of items answered: _____ = Fatigue Subscale score**				

Example:

Let's consider a subject who answered the questionnaire according to Table 45.

Table 45: FACIT-Fatigue : Example

Item #	Description	Item Response	Item Score
1	I feel fatigued	2	2
2	I feel weak all over	2	2
3	I feel listless ("washed out")	3	1
4	I feel tired	3	1
5	I have trouble starting things because I am tired	3	1
6	I have trouble finishing things because I am tired	3	1
7	I have energy	1	1
8	I am able to do my usual activities	1	1
9	I need to sleep during the day	2	2
10	I am too tired to eat	2	2
11	I need help doing my usual activities	3	1
12	I am frustrated by being too tired to do the things I want to do	2	2
13	I have to limit my social activity because I am tired	2	2

The Fatigue Score is equal to the sum of 13 item scores, i.e. 19.

18.5 Lupus QoL

The Lupus QoL is a lupus-specific health related QoL questionnaire consisting of 34 items grouped in 8 domains as shown below:

Table 46: Lupus Scoring

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	26-30
Fatigue	4	31-34

Subjects indicate their responses on a 5-point Likert response format, where 4=never, 3=occasionally, 2= a good bit of the time, 1=most of the time, and 0=all of the time (data received from the external vendor will be transformed to match this scale). Summary scores can be calculated for all 8 domains. First, the mean raw domain score must be calculated by adding up the item response scores for each domain and dividing this total by the number of items in that domain. Then, the transformed domain scores are calculated by dividing by 4 and multiplying by 100 each mean raw domain score. Transformed scores range from 0 (worst HRQoL) to 100 (best HRQoL).

Transformed domain scores can be obtained when at least 50% of the items are answered. The mean raw domain score is then calculated by totaling the item response scores of the answered items and dividing by the number of answered items. A non-applicable response is treated as unanswered and the domain score calculated as above.

Example:

Let's consider a subject who has answered the questionnaire according to Table 47.

Table 47: Lupus – Example

Domain	Item #	Item Description	Response
Physical health	1	I need help to do heavy physical jobs such as digging the garden, painting and/or decorating, moving furniture	3
	2	I need help to do moderate physical jobs, such as vacuuming, ironing, shopping, cleaning the bathroom	2
	3	I need help to do light physical jobs, such as cooking/preparing meals, opening jars, dusting, combing my hair or attending to personal hygiene	3
	4	I am unable to perform everyday tasks, such as my job, childcare, housework as well as I would like to	3
	5	I have difficulty in climbing stairs	2
	6	I have lost some independence and am reliant on others	1
	7	I have to do things at a slower pace	1
	8	My sleep pattern is disturbed	4
Pain	9	I am prevented from performing activities the way I would like to because of pain due to lupus	3
	10	The pain I experience interferes with the quality of my sleep	3
	11	The pain due to my lupus is so severe that it limits my mobility	2
Planning	12	I avoid planning to attend events in the future	1
	13	Because of the unpredictability of my lupus, I am unable to organize my life efficiently	2
	14	My lupus varies from day to day, which makes it difficult for me to commit myself to social arrangements	2
Intimate relationship	15	Because of the pain I experience due to lupus, I am less interested in a sexual relationship	1
	16	I am not interested in sex	0
Burden to others	17	I am concerned that my lupus is stressful for those who are close to me	3
	18	I am concerned that I cause worry to those who are close to me	3
	19	I feel that I am a burden to my friends and/or family	3
Emotional health	20	I have found my lupus makes me resentful	3
	21	I have found my lupus makes me so fed up nothing can cheer me up	3
	22	I have found my lupus makes me sad	4
	23	I have found my lupus makes me anxious	2
	24	I have found my lupus makes me worried	2
	25	I have found my lupus makes me lacking in self-confidence	2
Body image	26	My physical appearance due to lupus interferes with my enjoyment of life	2
	27	My appearance (e.g. rash, weight gain/loss) makes me avoid social situations	2
	28	Lupus-related skin rashes make me feel less attractive	2
	29	The hair loss I have experienced makes me feel less attractive	2
	30	The weight gain I have experienced during treatment makes me feel less attractive	2
Fatigue	31	I cannot concentrate for long periods of time	2
	32	I feel worn out and sluggish	1
	33	I need to have early nights	3
	34	I am often exhausted in the morning	2

Domain Scores are provided in Table 48.

Table 48: LupusQoL – Example – Domain Scores

Item	Mean Score	Transformed Mean Score
Physical health	19/8 = 2.375	100 x 2.375 / 4 = 59.4
Pain	8/3 = 2.667	100 x 2.667 / 4 = 66.7
Planning	5/3 = 1.667	100 x 1.667 / 4 = 41.7
Intimate relationship	1/2 = 0.5	100 x 0.5 / 4 = 12.5
Burden to others	9/3 = 3	100 x 3 / 4 = 75
Emotional health	16/6 = 2.667	100 x 2.667 / 4 = 66.7
Body image	10/5 = 2	100 x 2 / 4 = 50
Fatigue	8/4 = 2	100 x 2 / 4 = 50

18.6 PGIC

The PGIC is a self-rated scale that asks the subject to describe the change in activity limitations, symptoms, emotions, and overall quality of life (QoL) related to the subject's painful condition on the following scale: 1 (very much improved), 2 (much improved), 3 (minimally improved), 4 (no change), 5 (minimally worse), 6 (much worse) and 7 (very much worse). The subject will select the number that matches the subject's degree of change since beginning the treatment with evobrutinib (Hurst 2004).

The PGIC can be used as an anchor based method to assess clinically important change in which the judgment of meaningful change is made by the subject (Amirfeyz 2009).

18.7 EQ-5D-5L

The EQ-5D questionnaire comprises 5 questions (items) relating to current problems in the dimensions mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Responses in each dimension are divided into 5 ordinal levels coded (1) no problems, (2) slight problems, (3) moderate problems, (4) severe problems, (5) extreme problems. This part, called the EQ-5D descriptive system, provides a 5-dimensional description of health status.

A unique health state is defined by combining one level from each of the 5 dimensions. A total of 3125 possible health states is defined in this way. Each state is referred to in terms of a 5-digits code. For example, state 1111 indicates no problems on any of the 5 dimensions, while state 12345 indicates no problems with mobility, slight problems with self-care, moderate problems with doing usual activities, severe pain or discomfort, and extreme anxiety or depression.

The EQ-5D-5L descriptive system is followed by a VAS (EQ VAS), similar to a thermometer, ranging from 0 (worst imaginable health state) to 100 (best imaginable health state). The EQ VAS records the respondent's self-rated valuation of health state, i.e., a value which is based on the respondent's preferences (EQ VAS score). With respect to psychometric properties, the EQ VAS showed acceptable validity and reliability in subject with SLE.

Based on the EQ-5D-5L, the numeric EQ-5D-5L index can be derived, using validated translations provided by EuroQoL presented in the following XLS file. The EQ-5D-5L index is frequently used in economic evaluations: it represents societal preference values for the full set of EQ-5D-5L health states with the best state (perfect health) and "death" being assigned values of 1 and 0, respectively.

The UK country specific value set will be used in deriving the index value for all subjects.

This questionnaire will be used in the countries for which validated translations are available.

EQ-5D-5L_Crosswalk_
Index_Value_Calculato

18.8 Corticosteroid usage over time

18.8.1.1 General rules

To perform derivations with regards to CS endpoints, the following rules will be applied:

- CS will be medications with an ATC code which begins with "H02A" (CORTICOSTEROID FOR SYSTEMIC USE, PLAIN).
- All routes are considered except ophthalmic, topical, nasal and respiratory (inhalation).
- CS medications that started on or after end of treatment will not be taken into account in the analysis.
- When dose per administration or dosing frequency is missing or uninterpretable, the corresponding CS will not be included in endpoint derivation but will be listed instead. The reason why this CS is not considered in endpoint derivation will be provided in this listing. An exception is when the start and stop dates occur the same day, no dosing frequency is applicable: these administrations will be included in calculations (daily dosing frequency will be assumed as equal to 1).

18.8.1.2 Missing dates handling

The following rules will apply to ensure the completion of start/end dates for all CS:

- For start date:
  - If the start date of the medication/procedure is unknown (i.e. complete missing date), the start date will be imputed with the date of 1st of January 2015 (one year before the year the trial started).
  - If both day and month are missing, but the year is available, the date will be imputed to the 1st of January.
  - If the day is missing only, it will be imputed to the first day of the month.

If the stop date is complete or partially missing, and the imputed start date is after the stop date then the start date will be imputed using the stop date

- For end date:
  - If the end date of the CS is unknown (i.e. complete missing date and end relative to reference time point is not equal to ONGOING), the end date will be imputed to the 31st of December 2099.
  - If both day and month are missing, but the year is available, it will be imputed to the 31st of December.
  - If the day is missing only, it will be imputed to the last day of the month.

If the imputed stop date is before the start date then the imputed stop date will be equal to the start date.

18.8.1.3 Conversions

Table 49 and Table 50 provide conversions to be performed in order to assess the CS dosage:

Table 49: Conversion factors in prednisone equivalent

Who-drug Term	Conversion factor in prednisone equivalent
METHYLPREDNISOLONE	1,25
METHYLPREDNISOLONE SODIUM / METHYLPREDNISOLONE SODIUM SUCCINATE	1,25
METHYLPREDNISOLONE ACETATE	1,25
HYDROCORTISONE / HYDROCORTISONE SODIUM SUCCINATE	0,25
BETAMETHASONE	8.333
BETAMETHASONE DIPROPIONATE	8.333
BETAMETHASONE VALERATE	8.333
BETAMETHASONE PHOSPHATE	8.333
DEXAMETHASONE	6,667
TRIAMCINOLONE / TRIAMCINOLONE ACETONIDE	1,25
CORTISONE	0,2
PREDNISOLONE	1
PREDNISONE	1
DEFLAZACORT	0,83333
MEPREDNISONE	1,25
PARAMETHASONE	2,5
CRONOLEVEL	8,333
BETROSPAM	8,333

Table 50: Frequency conversions for CS

Frequency	Conversion factor	Numerical Conversion factor
OAM	1/30	0,0333
QOD	1/2	0,5000
QW	1/7	0,1429
QWK	1/7	0,1429
BID/Q12	2	2,0000
TID/Q8	3	3,0000
QD – including QD with AM or PM (e.g. QD(AM), QD(PM), QD at AM, QD at PM)	1	1,0000
ONCE	1	1,0000
QAM	1	1,0000
Q2H	12	12,0000
Q3H	8	8,0000
Q4H	6	6,0000
Q6H	4	4,0000
Q8H	3	3,0000
Q12H	2	2,0000
QHS	1	1,0000
QID/Q6	4	4,0000
QPM	1	1,0000
QH	24	24,0000
Q3W	1/21	0,0476
TIW or Three times a week	3/7	0,4286
Four times a week	4/7	0,5714
Q4W	1/28	0,0357
BIW or Twice a week	2/7	0,2857
EVERY 3 HOURS	8	8,0000
X1	1	1,0000
EVERY OTHER DAY	1/2	0,5000
EVERY OTHER DAY ALTERNTATLY WITH30 MG Q0D	1/2	0,5000
EVERY OTHER DAY ALTERNTATLY WITH 20 MG QOD	1/2	0,5000
ONCE A WEEK	1/7	0,1429
QD (EVERY MORNING)	1	1,0000
SINGLE DOSES	1	1,0000
ALTERNATE DAYS	1/2	0,5000

18.8.1.4 Corticosteroid usage, cumulative and average daily dose

The study protocol defines rules of CS usage during the study. The non-compliance with these rules will be described using data collected in “Concomitant Medications” eCRF form and

subjects' compliance with CS usage rules checked during the medical review and protocol deviation process (i.e. subjects with PD code = PDEV52e).

According to protocol secondary criterion the following criteria will be summarized:

- Cumulative prednisone-equivalent dose (mg) from baseline visit until end of 52-week treatment period. If one subject does not receive any CS between baseline and end of treatment period then his cumulative dose will be set to 0 mg.

Cumulative dose [mg] = sum of (dose per administration * daily dosing frequency * Duration),

with:

- Sum of () = sum of all CS recorded in Concomitant Medication eCRF form, from baseline visit until the end of 52-week treatment period
- Duration = End CS dose date during treatment period – Start CS dose date + 1
- Daily dosing frequency using frequency conversions (See Table 49 and Table 50).

As CS usage has to be restricted to the period between screening visit and end of treatment period, the following specific rules will be applied only for the calculation of cumulative prednisone-equivalent dose:

- Start CS dose date:
  - If start dose date is before screening date, baseline visit date will be considered as the start date for duration calculation.
  - If start dose date is after end of treatment period, CS will not be considered for duration calculation.
- End CS dose date:
  - If end is missing or after date of last IMP dose + 14 days or CS ongoing, the end CS dose date will be imputed by the last IMP dose date + 14 days for duration calculation. If last IMP dose is missing then end dose date will not be imputed and duration will not be calculated (patient not treated not taken into account in the analysis).
  - If end CS dose date is before baseline date, CS will not be considered
- Average daily steroid dose (mg) from baseline visit until end of treatment period. If one subject does not receive any CS between screening and end of treatment period then his average daily dose will be set to 0 mg.
  - Average daily dose [mg] = $\left(\frac{\text{Cumulative dose [mg]}}{\text{Duration}}\right)$
  - For completers, Duration = Week 52 date – baseline visit date + 1
  - For discontinuers, Duration = Last IMP dose date + 14 – baseline visit date + 1

- If one subject does not receive any CS between baseline and end of treatment period then his daily dose will be set to 0 mg and change from screening value will be equal to 0 mg.

- Prednisone-equivalent daily dose [mg] by visit:

- The prednisone-equivalent daily dose at visit X is calculated only if visit X is performed during treatment period (date of visit X ≤ last treatment intake + 14 days).

- At a scheduled visit, prednisone-equivalent CS daily dose is equal to the dose the subject is on at the scheduled visit. If several CS are taken at the same scheduled visit then all CS doses should be added. If a prednisone-equivalent CS dose stopped/changed at the scheduled visit date (Week X visit date), then the prednisone-equivalent CS dose started before (Week X visit date - 1 day) will be considered as the CS daily dose for that visit.

- CS dose taken between two scheduled visits will not be taken into account in calculation of the scheduled-visit dose.

- For subjects who completed the treatment period, CS dose that started on or after end of treatment will be excluded from the calculation.

- All comparisons will be done compared to CS dosing at baseline. CS dosing at baseline is defined at CS dose taken the day of baseline visit (day 1). If patient did not receive any CS at baseline then CS dose is set to 0 mg.

- For defining CS sparing effect endpoint (i.e. sustained reduction by W41 to Week 52), the derivation of prednisone-equivalent CS daily dose will be repeated at Week 40 scheduled visit using a different rule. If a prednisone-equivalent CS dose stopped/changed at the Week 40 scheduled visit date (Week 40 visit date), then the prednisone-equivalent CS dose started after (Week 40 visit date + 1 day) will be considered as the CS daily dose for that visit. Sustained reduction by Week 41 will be then programmed by checking that CS reduction is maintained between this new post-Week 40 (=Week 41) visit and Week 52 visit (i.e. CS dose at Week 52 is less or equal than CS dose at Week 41). The associated conditions related to CS sustained reduction (i.e. CS dose ≤ 7.5 mg/day and ≤ Day 1 CS dose, CS dose reduction from baseline ≥ 25% and CS dose ≤ 7.5 mg/day) will be checked for Week 41, 44, 48 and 52 visits.

18.9 SAS Sample Code

This Appendix details how the statistical methods will be implemented with SAS software application.

18.9.1 SRI Response

Logistic regression: PROC LOGISTIC with SAS

The logistic regression can be implemented using the following SAS code:

```
proc logistic data=<dataset> descending;
      class TreatmentGroup(ref='Placebo') Region Race SLEDAI_2K_screening;
      model AVALC(event="Y") = TreatmentGroup Region Race
                               SLEDAI_2K_screening / orpvalue clodds=wald;
      ods output CLoddsWald=CLoddsWald;
run;
```

For primary analysis of the 2-primary endpoints and key secondary endpoint *SRI-4 response at Week 52 among serologically active subgroup*, output the 1-sided p-value (α=0.025) as below:

- When the OR favors the active treatment (i.e. OR >1), one-sided p-value = two-sided p-value / 2
- When the OR does not favor the active treatment (i.e. OR ≤ 1), one-sided p-value = 1 – (two-sided p-value / 2)

The interaction can be implemented using the "*" symbol. For example, the following statement takes into consideration the interaction between treatment and region:

```
proc logistic data=<dataset> descending;
      class TreatmentGroup(ref='Placebo') Region Race SLEDAI_2K_screening;
      model AVALC(event="Y") = TreatmentGroup Region Race SLEDAI_2K_screening
                               TreatmentGroup * Region/ orpvalue clodds=wald;
      ods output ParameterEstimates=ParameterEstimates CLoddsWald=CLoddsWald;
run;
```

Linear trend in log odds of SRI response for two-primary and key secondary endpoints: PROC LOGISTIC with SAS

In addition to the standard logistic model, the linear trend test can be implemented using the ESTIMATE and CONTRAST statements:

```
proc logistic data=<dataset>;
      class TreatmentGroup Region Race SLEDAI_2K_screening / param=glm;
      model AVALC(event="Y") = TreatmentGroup Region Race
                               SLEDAI_2K_screening;
      ESTIMATE "SRI response rate trend linear in treatment" TreatmentGroup 3 1 -1 -
3;
      CONTRAST "SRI response rate trend linear in treatment" TreatmentGroup 3 1 -1 -3
      / e;
      ods output
      Estimates=Estimates ContrastTest=ContrastTest;
run;
```

Logistic regression subgroup analysis: PROC GLIMMIX with SAS

In addition to the standard logistic model, the subgroup analysis can be implemented using the following code:

```
proc glimmix data=<dataset>;
       class TreatmentGroup (ref='Placebo') SLEDAI_2K_screening;
       model  AVALC(event="Y")  =  TreatmentGroup  SLEDAI_2K_screening  TreatmentGroup*
       SLEDAI_2K_screening / alpha=0.05 dist=binary;
       lsmeans  TreatmentGroup*  SLEDAI_2K_screening  /  slicediff=  SLEDAI_2K_screening
       oddsratio ilink cl;
       ods output Glimmix.LSMeans.SliceDiffs;
run;
```

Adjusted P-value:

The 12 raw p-values generated (for the two-primary endpoints and two key secondary endpoints with 3 treatment comparisons with placebo) are adjusted based on the hierarchical procedure by applying the methodology described in Section 14.1.2.4 (and further explained in Appendix 18.15). SAS PROC IML presented in Maurer 2014 will be used to generate the final adjusted p-values (i.e. adjusted for multiplicity).

MI process – MAR pattern: PROC MI with SAS

The following SAS code can be used to implement the MI with MAR pattern. The order of the SRI components must follow the one defined in the Section 14.1.2.4.

```
/*************************************************************
      STEP 1 : creation of monotone missing data structure
*************************************************************/

/*** _burn-in_NBITER_ will have to be adjusted by a convergence plot,
if convergence is not yet diagnosed after 200 iterations ***/
title "Step 1 : Create monotone structure" ;
proc MI data=<Dataset> out=outMAR1 seed=&seed. nimpute=&nimpute. noprint;
       var
       TreatmentGroup
       black (0=non-black 1=black)
       SLEDAI_2K_screening
       US (0=not US 1=US region)
       West (0=not Western Europe, 1=Western Europe region)
       Jap (0=not Japan region, (1=Japan region)
       SRIComp1_base SRIComp3_base
       SRIComp1_Day14 SRIComp2_Day14 SRIComp3_Day14

       /* continue to the last endpoint */

       ;
       mcmc chain=single initial=EM nbITER=200 NITER=100 impute=monotone ;
       EM MAXITER = 5000 ;
run ;

/*****************************************
```

```
     STEP 2 : Apply multiple imputation
*****************************************/
title "Step 2 : Apply MI procedure (MAR pattern)" ;
proc mi data=outMAR1 nimpute=1 seed=&seed. out=outMAR2 noprint;
       by _Imputation_;
       class TreatmentGroup race region SLEDAI_2K_screening;
       var
              TreatmentGroup race region SLEDAI_2K_screening
              SRIComp1change_Day14 SRIComp2_Day14 SRIComp3_Day14

       /* continue to the last endpoint */

       ;
       monotone regression;
run;
```

Then after applying the needed derivation of SRI score and the logistic analysis of interest for SRI-4, the next step consists in:

```
/*********************************************************
     STEP 4a : Pooling ODDS RATIOS and CI using Log Transformation
*********************************************************/
*** Log-transform odds ratio estimates and obtain standard error from confidence
intervals ***;
DATA lgsodds_t;
       SET lgsodds(WHERE=(INDEX(EFFECT,"TRT")));
       log_or_lr_value=LOG(ODDSRATIOEST);
       qn=quantile('NORMAL',.975);
       log_or_lr_se=(LOG(UPPERCL)-LOG(LOWERCL))/(2*qn);
RUN;
*** Combine transformed estimates;
PROC MIANALYZE DATA=lgsodds_t;
       MODELEFFECTS log_or_lr_value;
       STDERR log_or_lr_se;
       ODS OUTPUT PARAMETERESTIMATES=mian_lgsodds_t;
QUIT;
*** Back-transform combined values;
DATA mian_lgsodds_bt;
       SET mian_lgsodds_t;
       estimate_back = EXP(ESTIMATE); *Pooled odds ratio;
       qn=quantile('NORMAL',.975);
       LCL_back=Estimate_back*EXP(-qn*STDERR); *Pooled lower limit;
       UCL_back=Estimate_back*EXP(+qn*STDERR); *Pooled upper limit;
RUN;

/*********************************************************
     STEP 4b : Pooling p-value
*********************************************************/
PROC MIANLYZE PARMS(CLASSVAR=CLASSVAL)=ParameterEstimatesMI
       CLASS TreatmentGroup;
       MODELEFFECTS TreatmentGroup;
       ODS OUTPUT PARAMETERESTIMATES=mian_logres
QUIT;
```

Then Step 4a and 4b are to be repeated for the analysis of SRI-6.

Pooling binomial proportion in each treatment arm after MI

```
*** Estimate proportions of responders in each treatment arm;
PROC FREQ DATA=datain_mi;
 TABLES resp / cl binomial(level=2);
```

```
 Exact binomial;
 BY _Imputation_ TreatmentGroup;
 ODS OUTPUT BINOMIALPROP=prop;
RUN;
*** From ODS output dataset BINOMIALPROP, create a dataset
 containing estimated proportion of responders in each
 treatment arm and their standard errors;
DATA prop_trt;
 MERGE
 prop(WHERE=(Label1="Proportion")
 KEEP=_Imputation_ TreatmentGroup nValue1 Label1
 RENAME=(nValue1=prop))
 prop(WHERE=(Label1="95% Lower Conf Limit")
 RENAME=(nValue1=LOWERCL)
 prop(WHERE=(Label1="95% Upper Conf Limit")
 RENAME=(nValue1=UPPERCL)
 qn = Quantile('NORMAL',.975)
 prop_se = (UPPERCL - LOWERCL) / (2 * qn)
 BY _Imputation_ TreatmentGroup;
RUN;
*** Combine proportion estimates;
PROC SORT DATA=prop_trt; BY TreatmentGroup _Imputation_; RUN;
PROC MIANALYZE DATA=prop_trt;
 MODELEFFECTS prop;
 STDERR prop_se;
 BY TreatmentGroup;
 ODS OUTPUT PARAMETERESTIMATES=mian_prop_trt;
RUN;
```

Tipping point sensitivity analysis

*1- Start by computing the number of missing response in each treatment at the willing visit in macrovariables: `nmisspbo nmissd1 nmissd2 nmissd3`.

*2- Then complete the tipping point analysis:

```
***TIPPING POINT ANALYSIS FOR BINARY DATA;
data resptip
  set resp;
  if TreatmentGroup='Pbo' then trtn=1;
  else if TreatmentGroup='Dose 1' then trtn=2;
  else if TreatmentGroup='Dose 2' then trtn=3;
  else if TreatmentGroup='Dose 3' then trtn=4;
  randnum=ranuni(200527);
  misspbo=&nmisspbo;
  missd1=&nmissd1;
  missd2=&nmissd2;
  missd3=&nmissd3;
run;
***Sort by random number allow cumulative imputation of failure /
success to subjects in random order;
PROC SORT data=resptip; BY randnum; QUIT;
DATA resptip2
  SET resptip;
  WHERE trtn in (1 2 3 4);
  ARRAY randcount(4);
  ARRAY t(4);
  *Each missing data will be identified by its number starting from 1, by each
treatment;
  RETAIN randcount 0;
```

```
  newresp=resp;
  IF resp=. THEN randcount(trtn)+1;
*Output misspbo*missd1 + misspbo*missd2 + misspbo*missd3 scenarios, ie all possible
combination of pbo with the dose;
  DO t1=1 to misspbo
    DO t2=1 TO missd1;
     DO t3=1 TO missd2;
      DO t4=1 TO missd3;
       IF resp=. THEN DO;
        IF randcount(trtn)<=t(trtn) THEN newresp=1;
        ELSE newresp=0;
        OUTPUT;
       END;
       ELSE OUTPUT;
      END;
     END;
    END;
  END;
```

*3- Then apply primary analysis to the new dataset according to each tipping point scenarios (ie by t1 t2 t3 t4).
From that we could plot p-value of each dose versus pbo.
Looking at t1 vs t2 for the p-value of dose 1.
Looking at t1 vs t3 for the p-value of dose 2.
Looking at t1 vs t4 for the p-value of dose 4.

```
*Label variable in result dataset that will be displayed in plot;
data p;
 set pvalue;
 label pvalue="P-value"
       t1='N of responders among missing - Placebo'
       t2='N of responders among missing - Dose 1;
       t3='N of responders among missing - Dose 2;
       t4='N of responders among missing - Dose 3;
run;
*Select scenarios and treatment to plot;
%macro selscenartrt (scenarios=, trtnum=,out=)
  data &out.;
   set p;
   where bygroup in (&scenarios) and trtn in (&trtnum.)
%mend selscenartrt;
%selscenartrt (scenarios=%str("t1" "t2"), trtnum=%str(1 2),out=pt1t2)
%selscenartrt (scenarios=%str("t1" "t3"), trtnum=%str(1 3),out=pt1t3)
%selscenartrt (scenarios=%str("t1" "t4"), trtnum=%str(1 4),out=pt1t4)

*Plot the wanted comparison;
proc sort data=pt1t2;
  by t1 t2;
quit;
 proc sgplot data=p;
  title1 "Adjusted p-value 1-side : Placebo vs Dose 1";
   heatmapparm x=t2 y=t1 colorresponse=pvalue  / colormodel=TwoColorRamp discretex
discretey transparency=0.2 OUTLINE outlineattrs=(thickness=0.5pt color=gray);
   text x=j y=i text=nvalue2 / textattrs=(size=5pt) strip;
   gradlegend;
```

```
   footnote1 justify=left "Placebo: N (ITTm) = xxx1, Observed Resp = xxresp1, Observed
Non-Resp = xxnresp1 and Missing = &nmisspbo.";
   footnote2 justify=left "Dose 1:  N (ITTm) = xxx2, Observed Resp =  xxresp2,
Observed Non-Resp = xxnresp2, Missing = &nmissd1.";
run;
```

18.9.2 Time to first flare, flare-free status, AFR

KM estimates for the time to first flare: proc LIFESTEST with SAS

The KM estimates and KM curve will be presented using the LIFETEST procedure as follows:

```
ODS listing gpath="<gpath>";
ods graphics on;
proc   lifetest   data=<dataset>   outsurv=<OutDataset>   plots=(survival(atrisk=0   to
&LastPatientAtRisk. by <step>)) method=km;
       time aval*cnsr(1);
       strata TreatmentGroup;
run;
ods graphics off;
```

In order to get the KM percentiles and their corresponding 95% CI, the following steps will be performed:

1. Run PROC LIFETEST 'standard': need outsurv and ProductLimitEstimates
2. From ProductLimitEstimates, create one dataset per group, keeping only events and non-missing survival estimates
3. In each group dataset, compute LOGLOG estimate of each survival estimate: formula from SAS support (1): $\frac{g(S(t))-g(1-0.25)}{g'(S(t))\sigma(S(t))}$ where S is survival, sigma is stderr, g=LOG(-LOG(..)) and g' is first derivative of the function. Replace 0.25 with expected percentile – 0.05, 0.10, 0.15, 0.20 - (example in SAS is for first quartile)
4. Then compute 95% CI: lowest/highest values from (1) that lie between $\pm z_{1-\frac{0.05}{2}} = \pm 1.965$
5. Retrieve the p percentile value from outsurv
6. Combine with the datasets containing CI

In addition to the KM estimates, the subgroup analysis can be implemented using the following code:

```
proc lifetest data=<dataset> outsurv=<OutDataset>;
     time AVAL*CNSR(1);
     strata &TRT.N;
     where SLEDAI_2K_screening = <value_of_interest>;
run;
```

Log-rank p-value:

```
proc lifetest data=<dataset> notable;
     time aval*cnsr(1);
      strata  Region  Race  SLEDAI_2K_screening  /  group=  TreatmentGroup  test=logrank
diff=control('1');
     ods output SurvDiff=SurvDiff;
run;
```

For primary analysis of key secondary endpoint *Time to first severe (BILAG A) flare* output the 1-sided p-value (α=0.025) as below:

- When the HR favors the active treatment (i.e. HR ≤ 1), one-sided p-value = two-sided p-value / 2
- When the HR does not favor the active treatment (i.e. HR >1), one-sided p-value = 1 – (two-sided p-value / 2).

Cox model for the time to first flare: proc PHREG with SAS

The Cox regression model will be performed using the PHREG procedure from SAS – get HR and 95%CI:

```
proc phreg data=<dataset>;
       class TreatmentGroup(ref='Placebo') Region Race SLEDAI_2K_screening;
       model AVAL*CNSR(1) = TreatmentGroup Region Race
                            SLEDAI_2K_screening / RL TIES=DISCRETE;
       Hazardratio TreatmentGroup / DIFF=REF;
run;
```

In addition to the Cox regression model, the subgroup analysis can be implemented using the following code:

```
proc phreg data=<dataset>;
       class TreatmentGroup(ref='Placebo') SLEDAI_2K_screening;
       model AVAL*CNSR(1) = TreatmentGroup SLEDAI_2K_screening
       TreatmentGroup*SLEDAI_2K_screening / TIES=DISCRETE;
       Hazardratio TreatmentGroup / DIFF=REF;
run;
```

For trend test for key secondary endpoint time to first severe (BILAG A) flare:

```
proc phreg data=<dataset>;
       class TreatmentGroup Region Race SLEDAI_2K_screening / param=GLM;
       model AVAL*CNSR(1) = TreatmentGroup Region Race
                            SLEDAI_2K_screening / TIES=DISCRETE;
        ESTIMATE "Trend linear in treatment" TreatmentGroup -3 -1 1 3;
       CONTRAST "Trend linear in treatment" TreatmentGroup -3 -1 1 3 / e;
run;
```

Sensitivity analysis of first severe BILAG A flare:

```
/*************************************************************
      STEP 1 : creation of monotone missing data structure
*************************************************************/

/*** _burn-in_NBITER_ will have to be adjusted by a convergence plot,
if convergence is not yet diagnosed after 200 iterations ***/
title "Step 1 : Create monotone structure" ;
proc MI data=<Dataset> out=outMAR1 seed=&seed. nimpute=&nimpute. noprint;
       var
       TreatmentGroup
       black (0=non-black 1=black)
       SLEDAI_2K_screening
       US (0=not US 1=US region)
```

```
       West (0=not Western Europe, 1=Western Europe region)
       Jap (0=not Japan region, (1=Japan region)

       Resp_base Resp_Day14-Resp_Day52;


       /* continue to the last endpoint */

       ;
       mcmc chain=single initial=EM nbITER=200 NITER=100 impute=monotone ;
       EM MAXITER = 5000 ;
run ;

/*******************************************
     STEP 2 : Apply multiple imputation
*******************************************/
title "Step 2 : Apply MI procedure (MAR pattern)" ;
proc mi data=outMAR1 nimpute=1 seed=&seed. out=outMAR2 noprint;
       by _Imputation_;
       class  TreatmentGroup race region SLEDAI_2K_screening
              Resp_base Resp_Day14-Resp_Day52;
       var    TreatmentGroup race region SLEDAI_2K_screening
              Resp_base Resp_Day14-Resp_Day52;
       monotone logistic;
run;

/*******************************************
     STEP 3 : Create dataset with Time to
*******************************************/
This should be define in specification of ADAM and same rule should be applied to
outMAR2 dataset by imputation.

/*******************************************
     STEP 4 : Create estimate on imputed dataset
*******************************************/
proc phreg data=<dataset>;
       by _Imputation_;
       class TreatmentGroup(ref='Placebo') Region Race SLEDAI_2K_screening;
       model AVAL*CNSR(1) = TreatmentGroup Region Race
                            SLEDAI_2K_screening / RL TIES=DISCRETE;
       ods output ParameterEstimates=_es; run;
run;


/*******************************************
     STEP 5 : Combined estimates
*******************************************/
/* Combine model coefficients. */
proc sort data=_es;
by Parameter CLassVal0 _Imputation_;
 run;
proc mianalyze data=_es;
       by  Parameter ClassVal0;
       modeleffects Estimate;
       stderr stderr;
       ods output ParameterEstimates = _es_mianal;
run;

/*Exponentiate in order to obtain hazard ratio estimates and confidence intervals */
data resprefix_hr;
       set _es_mianal;
       Log_HR_comb=Estimate;
```

```
      HR_comb=exp(Estimate);
      HR_LCL_comb=exp(LCLMean);
      HR_UCL_comb=exp(UCLMean);
      keep Parameter ClassVal0 Log_HR_comb HR_comb HR_LCL_comb HR_UCL_comb Probt;
      rename Probt=HR_pval_comb;
run;
```

Logistic regression for the flare-free status

The same procedure as in Appendix 18.9.1 must be applied.

NB model for BILAG A Severe flare / BILAG A or 2B Moderate to Severe flare/ SFI Severe flare

The NB model will be performed using the GENMOD procedure from SAS:

```
proc genmod data=<Dataset>;
      class TreatmentGroup Region Race SLEDAI_2K_screening;
      model AVAL = TreatmentGroup Region Race SLEDAI_2K_screening/ alpha=0.05
offset=LOGTIME
                                          DIST=NB MAXITER=<MaxIter>;
      ESTIMATE "log flare RR evobrutinib 25mg QD vs placebo" TreatmentGroup -1 1 0 0
      / EXP;
      ESTIMATE "log flare RR evobrutinib 75mg QD vs placebo" TreatmentGroup -1 0 1 0
      / EXP;
      ESTIMATE "log flare RR evobrutinib 50mg BID vs placebo" TreatmentGroup -1 0 0 1
      / EXP;
      ods output estimates=estimates1;
run;

where <maxIter> is the number of iteration (must be sufficiently high to run the
model). The offset is equal to the logarithm of years on treatment.
```

18.9.3 Other Secondary Endpoints

MMRM model: proc MIXED with SAS

For each continuous endpoint, the MMRM model will be implemented using the MIXED procedure.

A mixed model can be performed with SAS using the following code below. This sample code assumes the treatment group is assigned in the following order:

- 1 = Placebo
- 2 = Evobrutinib 25 mg QD
- 3 = Evobrutinib 75 mg QD
- 4 = Evobrutinib 50 mg BID

The interaction between treatment group and visit must be tested.

```
proc mixed data=<dataset>;
              where avisitn>1;
              class SubjectID Visit TreatmentGroup Region Race SLEDAI_2K_screening;
              model CHG = BASE TreatmentGroup Visit Region Race SLEDAI_2K_screening
TreatmentGroup*Visit / ddfm=KR s;
              repeated Visit / sub=USUBJID type=UN ;
              lsmeans TreatmentGroup*Visit / pdiff /*e*/ cl;

              %*use estimates or results of PDIFF (DIFFS dataset) but inverse as pdiff
gives trt=1 minus Trt=2/3/4;

              estimate 'Evobrutinib 25 mg QD - Placebo at W52' TreatmentGroup -1 1 0 0
                     TreatmentGroup*Visit 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 -1 1 0 0 / cl;
              estimate 'Evobrutinib 75 mg QD - Placebo at W52' TreatmentGroup -1 0 1 0
                     TreatmentGroup*Visit 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 -1 0 1 0 / cl;
              estimate 'Evobrutinib 50 mg BID - Placebo at W52' TreatmentGroup -1 0 0 1
                     TreatmentGroup*Visit 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 -1 0 0 1 / cl;   run ;

              ods output lsmeans=lsmeans diffs=diffs Tests3=Tests3;
```

where <TypeMatrix> represents the matrix of covariance (either type=CS or type=UN, see corresponding Section of the iAP for each endpoint). The 0 and 1 for visit and interaction treatment*visit must be adapted, depending on the number of assessments.

Logistic regression with repeated measures: GENMOD procedure from SAS

A logistic model with repeated measures can be implemented using the GENMOD procedure from SAS ® as follows:

```
ods listing close;
       proc genmod data=<Dataset> descending;
       class USUBJID Visit TreatmentGroup Region;
       model AVAL = TreatmentGroup Visit Region TreatmentGroup*Visit / dist=bin
       link=logit covb;
       repeated subject=USUBJID / type=exch within=Visit;
       estimate 'Evobrutinib 25 mg QD - Placebo at W52'
              TreatmentGroup -1 1 0 0
              TreatmentGroup*Visit 0 0 0 0 0 0 0 0 -1 1 0 0 / exp;
       estimate 'Evobrutinib 75 mg QD - Placebo at W52'
              TreatmentGroup -1 0 1 0
              TreatmentGroup*Visit 0 0 0 0 0 0 0 0 -1 0 1 0 / exp;
       estimate 'Evobrutinib 50 mg BID - Placebo at W52'
              TreatmentGroup -1 0 0 1
              TreatmentGroup*Visit 0 0 0 0 0 0 0 0 -1 0 0 1 / exp;

       lsmeans TreatmentGroup / ILINK CL;
       ods output Estimates=Estimates LSmeans=LSmeans Contrasts=Contrasts;
run;
quit;
ods listing;
```

18.9.4 EAIR of TEAEs: computation of Tier 1 and Tier 2 95% CIs

95%CI of the incidence rate will be computed using the GENMOD procedure from SAS ® with lsmeans instruction using a Poisson binomial model as follows:

```
ods listing close;
       proc genmod data=<Dataset>;
       where rate > 0;
       by <SOC/PT>;
       class <TreatmentGroup>;
       model rate = <TreatmentGroup> / dist=poisson link=log
       offset=log(PatientYears/1000);
       lsmeans <TreatmentGroup> / ilink cl;
       ods output lsmeans=OutLSmeans;
run;
ods listing;

where
       <TreatmentGroup> represents the treatment group variable (usually TRT01AN)
       <SOC/PT> represents the sorting variable (label for Overall, by SOC and by PT
       within SOC).
```

The 95%CI of difference with Placebo will be computed using MN method as follows:

```
proc freq data=<Dataset>;
       tables TreatmentGroup *AVALC / out=<outDataset> RISKDIFF (CL=MN) alpha=0.05;
       by AESOC AEDECOD;
       ods output PdiffCLs=<outPdiffCLs >;
run;
```

CCI

[redacted]

CCI

[redacted]

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

18.11 Time Windows

Table 51: Time Windows for SFI, SLEDAI-2K, PGA, BILAG 2004

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA • Japanese cohort analysis	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 126)	Week 16 – Day 113
			[126 ; 154)	Week 20 – Day 141
			[154 ; 182)	Week 24 – Day 169
			[182 ; 210)	Week 28 – Day 197
			[210 ; 238)	Week 32 – Day 225
			[238 ; 266)	Week 36 – Day 253
			[266 ; 294)	Week 40 – Day 281
			[294 ; 322)	Week 44 – Day 309
			[322 ; 350)	Week 48 – Day 337
			≥ 350*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 52: Time Windows for CLASI

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 21)	Week 2 – Day 15
			[21 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 126)	Week 16 – Day 113
			[126 ; 154)	Week 20 – Day 141
			[154 ; 182)	Week 24 – Day 169
			[182 ; 210)	Week 28 – Day 197
			[210 ; 238)	Week 32 – Day 225
			[238 ; 266)	Week 36 – Day 253
			[266 ; 294)	Week 40 – Day 281
			[294 ; 322)	Week 44 – Day 309
			[322 ; 350)	Week 48 – Day 337
			≥ 350*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 53: Time Windows for Urinalysis, Hematology, Chemistry

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 21)	Week 2 – Day 15
			[21 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 126)	Week 16 – Day 113
			[126 ; 154)	Week 20 – Day 141
			[154 ; 182)	Week 24 – Day 169
			[182 ; 210)	Week 28 – Day 197
			[210 ; 238)	Week 32 – Day 225
			[238 ; 266)	Week 36 – Day 253
			[266 ; 294)	Week 40 – Day 281
			[294 ; 322)	Week 44 – Day 309
			[322 ; 350)	Week 48 – Day 337
			≥ 350*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 42)	Week 4 - Day 29 – LTE
			[42 ; 70)	Week 8 - Day 57 – LTE
			[70 ; 126)	Week 12 - Day 85 – LTE
			[126 ; 224)	Week 24 - Day 169 – LTE
			[224 ; 322)	Week 40 - Day 281 – LTE
			[322 ; 454)	Week 52 - Day 365 – LTE
			[454 ; 644)	Week 76 - Day 545 – LTE
			[644 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 54: Time Windows for Urine Microscopy, UPCR

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 21)	Week 2 – Day 15
			[21 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 126)	Week 16 – Day 113
			[126 ; 154)	Week 20 – Day 141
			[154 ; 182)	Week 24 – Day 169
			[182 ; 210)	Week 28 – Day 197
			[210 ; 238)	Week 32 – Day 225
			[238 ; 266)	Week 36 – Day 253
			[266 ; 294)	Week 40 – Day 281
			[294 ; 322)	Week 44 – Day 309
			[322 ; 350)	Week 48 – Day 337
			≥ 350*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 55: Time Windows for Supplementary Liver Function Tests (LFTs)

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 21)	Week 2 – Day 15
			[21 ; 35)	Week 4 – Day 29
			[35 ; 49)	Week 6 – Day 43
			[49 ; 63)	Week 8 – Day 57
			[63 ; 77)	Week 10 – Day 71
			[77 ; 91)	Week 12 – Day 85
			[91 ; 105)	Week 14 – Day 99
			[105 ; 126)	Week 16 – Day 113
			[126 ; 154)	Week 20 – Day 141
			[154 ; 182)	Week 24 – Day 169
			[182 ; 210)	Week 28 – Day 197
			[210 ; 238)	Week 32 – Day 225
			[238 ; 266)	Week 36 – Day 253
			[266 ; 294)	Week 40 – Day 281
			[294 ; 322)	Week 44 – Day 309
			[322 ; 350)	Week 48 – Day 337
			≥ 350*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 21)	Week 2 - Day 15 – LTE
			[21 ; 35)	Week 4 - Day 29 – LTE
			[35 ; 49)	Week 6 - Day 43 – LTE
			[49 ; 63)	Week 8 - Day 57 – LTE
			[63 ; 77)	Week 10 - Day 71 – LTE
			[77 ; 91)	Week 12 - Day 85 – LTE
			[91 ; 105)	Week 14 - Day 99 – LTE
			[105 ; 140)	Week 16 - Day 113 – LTE
			[140 ; 224)	Week 24 - Day 169 – LTE
			[224 ; 322)	Week 40 - Day 281 – LTE
			[322 ; 454)	Week 52 - Day 365 – LTE
			[454 ; 644)	Week 76 - Day 545 – LTE
			[644 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 56: Time Windows for Total Ig Levels (IgG, IgA, IgM) and IgG CCI [redacted]

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 21)	Week 2 – Day 15
			[21 ; 42)	Week 4 – Day 29
			[42 ; 126)	Week 12 – Day 85
			[126 ; 210)	Week 24 – Day 169
			[210 ; 308)	Week 36 – Day 253
			≥ 308*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 42)	Week 4 - Day 29 – LTE
			[42 ; 70)	Week 8 - Day 57 – LTE
			[70 ; 126)	Week 12 - Day 85 – LTE
			[126 ; 224)	Week 24 - Day 169 – LTE
			[224 ; 322)	Week 40 - Day 281 – LTE
			[322 ; 454)	Week 52 - Day 365 – LTE
			[454 ; 644)	Week 76 - Day 545 – LTE
			[644 ; 751)	Week 104 - Day 745 – LTE

CCI [redacted]

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 57: Time Windows for SF-36v2, LupusQoL, FACIT-Fatigue, EQ‑5D-5L

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 140)	Week 16 – Day 113
			[140 ; 196)	Week 24 – Day 169
			[196 ; 252)	Week 32 – Day 225
			[252 ; 322)	Week 40 – Day 281
			≥ 322*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 58: Time Windows for HRU

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	[1 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 126)	Week 16 – Day 113
			[126 ; 154)	Week 20 – Day 141
			[154 ; 182)	Week 24 – Day 169
			[182 ; 210)	Week 28 – Day 197
			[210 ; 238)	Week 32 – Day 225
			[238 ; 266)	Week 36 – Day 253
			[266 ; 294)	Week 40 – Day 281
			[294 ; 322)	Week 44 – Day 309
			[322 ; 350)	Week 48 – Day 337
			≥ 350*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 42)	Week 4 - Day 29 – LTE
			[42 ; 70)	Week 8 - Day 57 – LTE
			[70 ; 98)	Week 12 - Day 85 – LTE
			[98 ; 140)	Week 16 - Day 113 – LTE
			[140 ; 224)	Week 24 - Day 169 – LTE
			[224 ; 322)	Week 40 - Day 281 – LTE
			[322 ; 454)	Week 52 - Day 365 – LTE
			[454 ; 644)	Week 76 - Day 545 – LTE
			[644 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 59: Time Windows for Vital Signs

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 21)	Week 2 – Day 15
			[21 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 126)	Week 16 – Day 113
			[126 ; 154)	Week 20 – Day 141
			[154 ; 182)	Week 24 – Day 169
			[182 ; 210)	Week 28 – Day 197
			[210 ; 238)	Week 32 – Day 225
			[238 ; 266)	Week 36 – Day 253
			[266 ; 294)	Week 40 – Day 281
			[294 ; 322)	Week 44 – Day 309
			[322 ; 350)	Week 48 – Day 337
			≥ 350*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 42)	Week 4 - Day 29 – LTE
			[42 ; 70)	Week 8 - Day 57 – LTE
			[70 ; 98)	Week 12 - Day 85 – LTE
			[98 ; 140)	Week 16 - Day 113 – LTE
			[140 ; 224)	Week 24 - Day 169 – LTE
			[224 ; 322)	Week 40 - Day 281 – LTE
			[322 ; 454)	Week 52 - Day 365 – LTE
			[454 ; 644)	Week 76 - Day 545 – LTE
			[644 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 60: Time Windows for PGIC

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	[1 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 140)	Week 16 – Day 113
			[140 ; 196)	Week 24 – Day 169
			[196 ; 252)	Week 32 – Day 225
			[252 ; 322)	Week 40 – Day 281
			≥ 322*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 61: Time Windows for ECG

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 126)	Week 12 – Day 85
			[126 ; 224)	Week 24 – Day 169
			[224 ; 322)	Week 40 – Day 281
			≥ 322*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 126)	Week 12 - Day 85 – LTE
			[126 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 454)	Week 52 - Day 365 – LTE
			[454 ; 644)	Week 76 - Day 545 – LTE
			[644 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 62: Time Windows for C-SSRS

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	[1 ; 56)	Week 4 – Day 29
			[56 ; 112)	Week 12 – Day 85
			[112 ; 168)	Week 20 – Day 141
			[168 ; 224)	Week 28 – Day 197
			[224 ; 280)	Week 36 – Day 253
			[280 ; 336)	Week 44 – Day 309
			≥ 336*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 126)	Week 12 - Day 85 – LTE
			[126 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 454)	Week 52 - Day 365 – LTE
			[454 ; 644)	Week 76 - Day 545 – LTE
			[644 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 63: Time Windows for SLICC/ACR Damage Index

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			≥ 2*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 64: Time Windows for Total B Cell Count

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 84)	Week 4 – Day 29
			[84 ; 266)	Week 24 – Day 169
			≥ 266*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 84)	Week 4 - Day 29 – LTE
			[84 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

CCI

[redacted]

Table 66: Time Windows for Autoantibodies, anti-Sm

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
			[2 ; 266)	Week 24 – Day 169
			≥ 266*	Week 52 – Day 365
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.

Table 67: Time Windows for ANA

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	1	Day 1
• LTE analysis	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	52-week treatment period	NA	NA
		LTE period	1	Day 1 – LTE
			[2 ; 266)	Week 24 - Day 169 – LTE
			[266 ; 554)	Week 52 - Day 365 – LTE
			[554 ; 751)	Week 104 - Day 745 – LTE

CCI

[redacted]

Table 69: Time Windows for HBV DNA assessments

Analysis	Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• PA	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	52-week treatment period	[2 ; 42)	Week 4 – Day 29
			[42 ; 70)	Week 8 – Day 57
			[70 ; 98)	Week 12 – Day 85
			[98 ; 126)	Week 16 – Day 113
			[126 ; 154)	Week 20 – Day 141
			[154 ; 182)	Week 24 – Day 169
			[182 ; 266)	Week 36 – Day 253
			[266 ; 350)	Week 48 – Day 337
			≥ 350*	Week 52 – Day 365
• LTE	• Placebo/Evobrutinib 50	52-week treatment period	NA	NA

analysis	mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	LTE period	1	Day 1 – LTE
			[2 ; 35)	Week 4 – Day 29 - LTE
			[35 ; 63)	Week 8 – Day 57 - LTE
			[63 ; 91)	Week 12 – Day 85 - LTE
			[91 ; 140)	Week 16 - Day 113 – LTE
			[140 ; 224)	Week 24 - Day 169 – LTE
			[224 ; 322)	Week 40 - Day 281 – LTE
			[322 ; 454)	Week 52 - Day 365 – LTE
			[454 ; 644)	Week 76 - Day 545 – LTE
			[644 ; 751)	Week 104 - Day 745 – LTE
* Assessments performed more than 14 days after last exposure date will not be included as well as assessments collected during safety follow-up visit.				

18.12 Laboratory Parameters to be summarized in the TLFs

Table 70: Laboratory Parameters to be Summarized in the TLFs

	Names of Clinical Safety Laboratory Evaluations in Protocol version 3.0	If gradable parameter, corresponding evaluation names in NCI-CTCAE 4.03	Worst on treatment value based on normal range
Biochemistry	Albumin	Hypoalbuminemia	LOW
	Aspartate aminotransferase	Aspartate aminotransferase increased	HIGH
	Alanine aminotransferase	Alanine aminotransferase increased	HIGH
	Alkaline phosphatase	Alkaline phosphatase increased	HIGH
	γ-Glutamyl-transferase	GGT increased	HIGH
	Lactate dehydrogenase		HIGH
	Bilirubin (total)	Blood bilirubin increased	HIGH
	Protein (total)		LOW
	Creatinine	Creatinine increased	HIGH
	eGFR	Chronic kidney disease	LOW
	Amylase	Serum amylase increased	HIGH
	Lipase	Lipase increased	HIGH
	Bicarbonate		LOW
	Blood urea nitrogen		HIGH
	Glucose	Hyperglycemia	HIGH
	Glucose	Hypoglycemia	LOW
	Sodium	Hypernatremia	HIGH
	Sodium	Hyponatremia	LOW
	Potassium	Hyperkalemia	HIGH
	Potassium	Hypokalemia	LOW
	Chloride		NA
	Calcium	Hypercalcemia	HIGH
	Calcium	Hypocalcemia	LOW
	Magnesium	Hypermagnesemia	HIGH
	Magnesium	Hypomagnesemia	LOW
	Phosphate	Hypophosphatemia	LOW
	Uric Acid	Hyperuricemia	HIGH
Hematology	Hematocrit		LOW/HIGH
	Hemoglobin	Hemoglobin increased	HIGH
	Hemoglobin	Anemia	LOW
	Red blood cell count (Erytrocytes)		NA
	Mean corpuscular volume		NA
	Mean corpuscular hemoglobin		NA

	Mean corpuscular hemoglobin concentration		NA
	Reticulocyte count		NA
	Platelet count	Platelet count decreased	LOW
	White blood cell count (Leucocytes)	Leukocytosis	HIGH
	White blood cell count (Leucocytes)	White blood cell decreased	LOW
	White blood cell differentials and absolute counts: Basophils		NA
	White blood cell differentials and absolute counts: Eosinophils		NA
	White blood cell differentials and absolute counts: Lymphocytes	Lymphocyte count increased	HIGH
	White blood cell differentials and absolute counts: Lymphocytes	Lymphocyte count decreased	LOW
	White blood cell differentials and absolute counts: Monocytes		NA
	White blood cell differentials and absolute counts: Neutrophils	Neutrophil count decreased	LOW
Urinalysis	pH		NA
	Nitrite		NA
	Protein	Proteinuria	NA
	Creatinine	Creatinine increased	NA
	Blood		NA
	Glucose		NA
	Ketones		NA
	Urobilinogen		NA
	Bilirubin		NA
	UPCR		NA
Coagulation	Prothrombin International normalized ratio	INR increased	NA
	Activated Partial thromboplastin time	Activated partial thromboplastin time prolonged	NA
	Prothrombin Time		NA
	Fibrinogen		NA
Urine Microscopy	White blood cells		HIGH
	Red blood cells		HIGH
	Casts		NA
	Crystals		NA

18.13 Laboratory Modifiers

Laboratory results containing a modifier such as "<" or ">" will be reported both as collected in the database and imputed in subject data listings as per the rules defined below.

In general (for exceptions see below at end of section), laboratory modifiers will be handled as follows:

- When the modifier is equal to "<" the derived laboratory value will be equal to the original value minus a quantity. The quantity will be equal to:
  - 1 if the original value is an integer
  - 0.1 if the original value has one significant digit after the decimal place
  - 0.01 if the original value has two significant digits after the decimal place
  - Etc…

  As the derived value must always be positive, there is an exception for original values equal to 1, 0.1, 0.01, and so on. For such cases, the quantity will be equal to the quantity defined above divided by 10.

  Examples:

  1. If the original value is equal to "<15", then the corresponding numeric value for summary statistics will be equal to 14
  2. If the original value is equal to "<0.5", then the corresponding numeric value for summary statistics will be equal to 0.4
  3. If the original value is equal to "<0.001", then the corresponding numeric value for summary statistics will be equal to 0.0009

- When the modifier is equal to ">" the derived laboratory value will be equal to the original value plus a quantity. The quantity will be equal to:
  - 1 if the original value is an integer
  - 0.1 if the original value has one significant digit after the decimal place
  - 0.01 if the original value has two significant digits after the decimal place
  - Etc…

  Examples:

  1. If the original value is equal to ">20", then the corresponding numeric value for summary statistics will be equal to 21
  2. If the original value is equal to ">8.5", then the corresponding numeric value for summary statistics will be equal to 8.6

3. If the original value is equal to ">1.045", then the corresponding numeric value for summary statistics will be equal to 1.046

CCI [redacted] plicable to all parameters/values listed in the enclosed Excel File.

For the parameters Cytomegalovirus IgG/IgM Antibody, as well as for Epstein-Barr parameters, when the original value is equal to "<x", the corresponding numeric value used for summary statistics will be equal to x/2 (see Excel File for examples).

Special case for the UPCR:

The rule to be followed for the UPCR is the following: when urine protein is < 5 (BLQ), then impute a numeric value of 2.5 to urine protein and derive UPCR accordingly.

For laboratory parameters other than UPCR and not listed in the enclosed Excel File (or if there is any change in the listed parameters), results containing a modifier will be handled case by case for summary statistics. Decision for each laboratory parameter will be documented in a Note-to-File prior to the database lock.

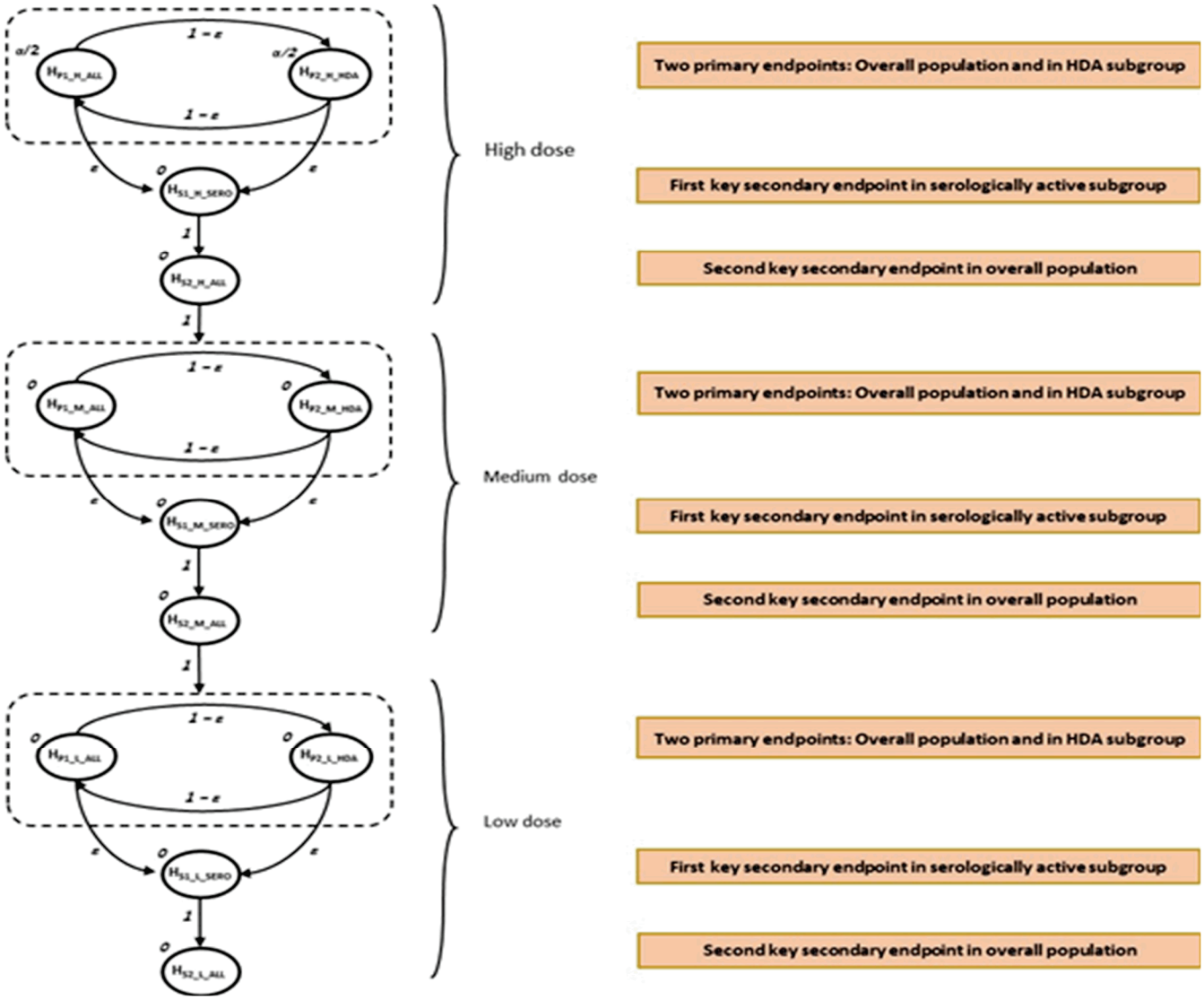
18.14 Tier 1 AEs and AEs of Special Interest

The spreadsheet enclosed contains the list of AE terms to programmatically flag subjects with Tier 1 AEs and AEs of special interest:

Evobrutinib_AESIs_
and_Tier1_AEs_MedI

This is the latest version at the time of IAP signoff, as of 25 November 2019.
The spreadsheet will be regularly updated by the Safety Team according to MedDRA version update. For the actual analysis, the latest current version received at the time of the analysis will be used.

18.15 Hierarchical Testing Diagram



The figure above uses the notation and approach described by Bretz 2009. This figure is constructed by using vertices, associated weights as well as directed edges. As explained in the paper in 'section 3.3. *Shifting significance levels between families of hypotheses*', the use of ε in the figure has a unique meaning that is different to the overall general notation used in that paper.

ELECTRONIC SIGNATURES

Document: ctp-ms200527-0018-iap-v4

Signed By	Event Name	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'GMT'Z)
PPD	Task Completed (Approval eSign): Approved	Technical Approval	16-Jun-2020 07:33