STATISTICAL ANALYSIS PLAN

Study Code	D3468C00003
Edition Number	2.0
Date	16-May-2025

A Multicentre, Randomized, Double-blind, Placebo-controlled, Phase III Study Evaluating the Efficacy and Safety of Anifrolumab in Asian Participants with Active Systemic Lupus Erythematosus

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

List abbreviations and definitions of specialized or unusual terms, measurements, or units.

Abbreviation or Specialized Term	Definition
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse event of special interest
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANA	Antinuclear antibody
AR(1)	1st order autoregressive
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
BICLA	British Isles Lupus Assessment Group-based Composite Lupus Assessment
BILAG	British Isles Lupus Assessment Group
BLQ	Below the limit of quantification
BMI	Body mass index
BP	Blood pressure
C3	Third component of complement
C4	Fourth component of complement
CDL	Clinical Database Lock
CH50	Total hemolytic complement
CI	Confidence interval
CLASI	Cutaneous Lupus Erythematosus Disease Area and Severity Index
CLE	Cutaneous Lupus Erythematosus
cm	Centimeter
CMH	Cochran-Mantel-Haenszel
CNS	Central nervous system
COVID-19	Coronavirus disease-19
CRF	Case report form (electronic/paper)
CS	Compound symmetry
C-SSRS	Columbia suicide severity rating scale
CSP	Clinical study protocol
CSR	Clinical study report
CTMS	Clinical trial management system

Abbreviation or Specialized Term	Definition
CV	Coefficient variation
DACRT	Disease Activity Central Review Team
DBP	Diastolic blood pressure
DNA	Deoxyribonucleic acid
dsDNA	Double stranded deoxyribonucleic acid
EAIR	Exposure-adjusted incident rate
ECG	Electrocardiogram
EDV	Early discontinuation visit
EQ-5D-5L	EuroQoL 5 dimensions
FAS	Full analysis set
GGT	Gamma glutamyl transferase
ICF	Informed consent form
ICU	Intensive care unit
IFN	Interferon
IM	Intramuscular
INR	International normalized ratio
IP	Investigational product
IPD	Important protocol deviations
IV	Intravenous
kg	Kilogram
LLDAS	Lupus Low Disease Activity State
LLOQ	Lower limit of quantitation
LOCF	Last observation carried forward
m	Meter
MACE	Major adverse cardiovascular event
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified intention-to-treat
mmHg	Millimeters of mercury
nAb	Neutralizing antibodies
NR	Non-response or non-responder
NSAIDs	Nonsteroidal anti-inflammatory drugs
OCS	Oral corticosteroids
PD	Pharmacodynamic

Abbreviation or Specialized Term	Definition
PD Plan	Protocol deviation plan
PEOT	Premature end of treatment
PGA	Physician's Global Assessment
PK	Pharmacokinetic(s)
PT	Preferred Term
Q4W	Every 4 weeks
QoL	Quality of life
RM	Restricted medication use
SAE	Serious adverse event
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SD	Standard deviation
SDI	Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index
SI	International System of units
SLE	Systemic lupus erythematosus
SLEDAI-2K	Systemic Lupus Erythematosus Disease Activity Index 2000
SOC	System organ class
SRI(X)	Systemic Lupus Erythematosus Responder Index of ≥X
TELVC	Treatment emergent laboratory/vital signs changes
ULN	Upper limit of the normal
VAS	Visual analogue scale
WDS	Withdraw from study
WHO-DD	WHO Drug Dictionary
WPAI	Work productivity and activity impairment

AMENDMENT HISTORY

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
N/A	06-Nov-2021	Initial draft SAP	Yes, version 1.0	N/A
Derivation of secondary endpoint(s)	05-Jan-2024	Added text “compared to baseline” in section 4.2.13.1.	Yes, version 3.0	To clarify the definition of participants who achieve Major Clinical Response.
Data presentation	05-Jan-2024	Updated that Week 56 is 8 weeks after last dose and Week 60 is 12 weeks after last dose in section 3.3.1.2.	Yes, version 3.0	To clarify the study periods.
Primary endpoint(s)	05-Jan-2024	Clarified the LOCF logic for the primary endpoints, relevant secondary and exploratory endpoints, and that LOCF records are not used in the repeated measures models.	Yes, version 3.0	To clarify the LOCF logic and the method of imputing missing data using LOCF in this study.
Primary endpoint(s)	05-Jan-2024	Added information for SLEDAI-2K in section 4.1.6.1.	Yes, version 3.0	To clarify the definition of SLEDAI-2K scores.
Data presentation	05-Jan-2024	Remove reference Appendix U from the CSP.	Yes, version 3.0	Removed as this Appendix is not in the CSP.
Data presentation	05-Jan-2024	Updated period for C-SSRS to use “treatment and follow-up period” in section 4.6.6.1.	Yes, version 3.0	To clarify the reporting period for C-SSRS.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Data presentation	05-Jan-2024	Add reference range for U-Protein test in 7.2. Add INR and total Cholesterol ranges to 7.2.	Yes, version 3.0	Information added for the analysis.
Data presentation	05-Jan-2024	Added text to define the number of total expected dose in the calculation of Drug compliance % in section 4.1.9.1.	Yes, version 3.0	To clarify the definition of total expected dose.
Data presentation	05-Jan-2024	Updated Table 4 to Table 3 in text in section 4.2.16.	Yes, version 3.0	To correct the typo.
Secondary endpoint(s)	05-Jan-2024	Removed supplementary analysis targeting the impact of PEOT and/or RM for CLASI. Moved flares endpoint to objective 4 as key secondary endpoint. Moved CLASI to objective 5 as other secondary endpoint.	Yes, version 3.0	To follow the updated CSP v3.0.
Secondary endpoint(s)	05-Jan-2024	Added text “Individual components of BICLA and SRI(4) at weeks 24 and 52 will be presented in a table by treatment group.” To section 4.2.10.4.	Yes, version 3.0	To justify the analysis for individual component of BICLA and SRI(4).
Primary endpoint(s)	05-Jan-2024	Updated formula in part D and F of section 4.2.1.5	Yes, version 3.0	To clarify the formula for the primary analysis.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Data presentation	05-Jan-2024	Updated the rounding rules for decimal places in section 3.3.	Yes, version 3.0	To follow AZSOL guiding principle.
Data presentation	05-Jan-2024	Added criteria for treatment emergent changes for urinalysis for Glucose, Ketone Bodies, and qualitative Erythrocytes in section 4.6.3.2.	Yes, version 3.0	To provide information for the analysis.
Secondary endpoint(s)	05-Jan-2024	Added PEOT and Death in the calculation of exposure time and number of flares in section 4.2.4.2.	Yes, version 3.0	To clarify the calculation of flare endpoint.
Data presentation	05-Jan-2024	Added text to define "screening" analysis window in section 3.3.1.2.	Yes, version 3.0	To clarify the analysis window.
Secondary endpoint(s)	05-Jan-2024	Updated to "Not applicable" in section 4.2.16.3.	Yes, version 3.0	To update the missing data imputation rule for Lupus QoL.
Secondary endpoint(s)	15-Jan-2024	Updated LLDAS definition in section 4.2.14.1.	Yes, version 3.0	To follow the definition in protocol v3.0.
Data presentation	15-Jan-2024	Updated "AEs occurring during treatment period including follow-up period" to "AEs occurring during treatment period and follow-up period" in section 4.6.2.1.	Yes, version 3.0	To align with previous anifrolumab studies at program level

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Data presentation	15-Jan-2024	Updated region group "Others" to "not mainland China" in section 3.2, 4.1.1.2, 4.2, 4.2.1.5 4.2.1.8, 4.2.2.6, 4.2.3.4, 4.2.3.6, 4.2.4.8, 4.2.9.4.	Yes, version 3.0	To follow protocol v3.0.
Primary endpoint(s)	15-Jan-2024	For primary/key secondary/other secondary endpoint, no PEOT, no RM are removed from the response criteria. Instead, these are handled through ICE strategies. Changes were made to section 4.2.1.1, 4.2.1.3, 4.2.2.1, 4.2.2.3, 4.2.3.1, 4.2.3.3, 4.2.5.1, 4.2.5.3, 4.2.6.1, 4.2.6.3, 4.2.14.1,4.2.14.3.	Yes, version 3.0	To follow protocol v3.0.
Data presentation	13-Feb-2024	Changed the number of randomized patients from 328 to 260 in section 3.2.	Yes, version 3.0	To follow protocol v3.0.
Data presentation	13-Feb-2024	Added abbreviation for ANA, ICU, MAR to the List of Abbreviations.	Yes, version 3.0	To add the required information.
Data presentation	08-Apr-2024	Update the missing day and missing day and month imputation rule in section 3.3.1.3	Yes, version 3.0	To correct the imputation rule.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Primary endpoint(s)	10-Apr-2024	For the primary endpoint, the stratified Cochran-Mantel-Haenszel (CMH) analysis is now a supportive analysis. The primary analysis will use the Bayesian robustified mixture prior approach with borrowing from the pooled global phase III studies (study 05 and study 04). This change applies to section 4.2.1.4 and 4.2.1.5.	Yes, version 3.0	To follow protocol v3.0.
Primary endpoint(s)	16-Jul-2024	For the time to BICLA analysis in section 4.2.1.7.7, updated the title to include "Supplementary Analysis of the Primary Endpoint". Clarified that intercurrent events are considered as non-responders per the primary BICLA endpoint, with time at risk based on the latest assessment during the double blind treatment period.	Yes, version 3.0	To clarify the handling of intercurrent events for the analysis.
Secondary endpoint(s)	16-Jul-2024	Removed flares versus baseline analysis in section 4.2.4.7.	Yes, version 3.0	Removed the analysis as robust results were observed between flare versus the previous visit and flare versus baseline.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Secondary endpoint(s)	16-Jul-2024	Replace "BICLA" with "BILAG" in the title of section 4.2.1.7.6.	Yes, version 3.0	To correct the text.
Primary endpoint(s)	26-Jul-2024	Edited Table 2 to reflect the analyses performed.	Yes, version 3.0	To provide the correct list of analyses.
Primary endpoint(s)	26-Jul-2024	Added text to reference BICLA response definition in section 4.2.1.7.7.	Yes, version 3.0	To provide more information.
Primary endpoint(s)	29-Jul-2024	Added new section 4.2.1.7.1 to add the supplementary analysis targeting the impact of PEOT due to different reasons. Table 2 was updated.	Yes, version 3.0	To evaluate the impact of PEOT due to different reasons.
Secondary endpoint(s)	29-Jul-2024	Updated the rule for dropouts and missing data handling in section 4.2.4.3.	Yes, version 3.0	To clarify how to handle dropouts and missing data for flare related analyses.
Data presentation	30-Jul-2024	Removed changes under "Changes to protocol Planned analyses" section as all the planned analyses are now following the latest CSP v3.0.	Yes, version 3.0	To update according to the latest CSP v3.0.
Data presentation	04-Sept-2024	Update text "Week 0 (Day 1) OCS dose" to "Randomization (day 1) OCS dose" throughout the SAP.	Yes, version 3.0	To update according to the latest CSP v3.0.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Secondary endpoint(s)	04-Sept-2024	Removed section 4.2.4.5.2 Missing data imputation analysis for flares.	Yes, version 3.0	Robust results observed based on global studies. Remove to avoid repeated analysis.
Secondary endpoint(s)	04-Sept-2024	Removed section 4.2.5.5.1 sensitivity analysis targeting the impact of PEOT and/or RM, tipping point analysis and section 4.2.5.6 subgroup analyses for skin lesions.	Yes, version 3.0	Reduce the number of sensitivity analysis as CLASI has been downgraded to other secondary endpoint per CSP 3.0.
Secondary endpoint(s)	16-Sept-2024	Updated the table "Conversion factors to oral prednisone equivalent dose" in Appendix 7.4.	Yes, version 3.0	To include all the OCS potentially used in the study.
Secondary endpoint(s)	17-Sept-2024	Updated that Pharmacokinetics is analyzed by IFN gene status not by treatment group in section 4.2.7.4.	Yes, version 3.0	PK parameters are only tested for active group.
Data presentation	18-Oct-2024	Removed Appendix D – ECG Potentially clinically significant post-baseline values.	Yes, version 3.0	To reflect no numerical ECG data is collected in this study.
Data presentation	18-Oct-2024	Updated the Herpes zoster classification in section 4.6.2.1.	Yes, version 3.0	To reflect the data collection approach.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Data presentation	23-Oct-2024	Added new section 3.3.7 - Identifying Cases of PEOT Due to Lack of Therapeutic Response and section 3.3.8 - Identifying Cases of PEOT due to COVID Lockdown.	Yes, version 3.0	To reflect the data collection approach.
Data presentation	23-Oct-2024	Removed Appendix A - Derivation of Prednisone Equivalent Daily Dose. Moved the content from this Appendix A to the current Appendix D - Restricted and prohibited medications – Programmatic Rules. Update the rules in Appendix D.	Yes, version 3.0	To clarify the programmatic rules for restricted and prohibited medications, keep alignment with anifrolumab program level.
Data presentation	24-Mar-2025	Capitalize Lupus-QoL Domains in Table 3 in section 4.2.16.1.	Yes, version 3.0	To follow industry standards.
Data presentation	24-Mar-2025	Pharmacokinetics by ADA status figure description updated in section 4.2.7.4.	Yes, version 3.0.	To align with previous anifrolumab studies.
Secondary endpoint(s)	25-Mar-2025	Section 3.3.9 added to describe EQ-5D-5L data excluded from analysis based on Clinical Data Standards and Dictionary Waiver Form (AZDoc0165290).	Yes, version 3.0.	To clarify that only data without errors is included for analysis.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Data presentation	25-Mar-2025	Section 3.3.1.2 change phrasing "primary and secondary endpoints" to "efficacy endpoints and pharmacodynamics endpoints including complements and anti-dsDNA" are assessed during the double-blind treatment period.	Yes, version 3.0.	To clarify the data included for efficacy analyses.
Secondary endpoint(s)	25-Mar-2025	Sections 4.2.4.2, 4.2.4.3 and 4.2.4.6 updated to clarify that LOCF records are included in analyses, and clarify the analysis periods for the primary and on-treatment analyses.	Yes, version 3.0.	To clarify the data included for efficacy analyses.
Derivation of secondary endpoint(s)	25-Mar-2025	Section 4.6.2.1 update the definition for opportunistic and non-opportunistic infections to align to AESIs.	Yes, version 3.0.	To clarify the derivation rules.
Data presentation	25-Mar-2025	Sections 4.6.2.1 and 4.6.2.2 Added definitions, derivations and analyses for AE exposure adjusted incidence rates.	Yes, version 3.0.	To clarify the derivation rules and types of AEs to be summarized by EAIR.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Data presentation	25-Mar-2025	Section 4.7 update to use China subpopulation (instead of China cohort and China Analysis sets).	Yes, version 3.0.	Per CSR requirement.
Data presentation	25-Mar-2025	Appendix B Added Urine Glucose, Ketone Bodies and qualitative Erythrocytes to tests.	Yes, version 3.0.	To clarify all the tests to be presented.
Data presentation	07-Apr-2025	Section 4.6 clarify that safety presents for the treatment and follow-up period, unless otherwise specified. Section 4.6.7.2 clarify that analysis reports the treatment period.	Yes, version 3.0	To clarify the data presentation.
Data presentation	07-Apr-2025	Section 4.1.8 updated to present restricted medication ICE, instead of On IP/Off IP restricted medications.	Yes, version 3.0	To clarify the data presentation. To align to restricted medication process.
Data presentation	09-Apr-2025	Section 3.3 added R version used for primary analysis of the primary endpoints.	Yes, version 3.0	To clarify software version ensuring reproducibility of results.
Secondary endpoint(s)	29-Apr-2025	Section 4.2.4 clarify that stratification factor can be exclude from the covariates of model when there are limited events in the corresponding stratum	Yes, version 3.0	To maintain model stability and robustness in case of limited events.

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Data presentation	29-Apr-2025	Section 4.2.8.1 clarify the definitions and derivation for ADA categories	Yes, version 3.0	To clarify the definitions and derivation for ADA categories
Primary endpoint(s)	29-Apr-2025	Section 4.2.1.6 and Table 2 Added a supportive analysis of the Primary Estimand: Bayesian Borrowing based on odds ratio.	Yes, version 3.0	To evaluate the robustness of Bayesian dynamic borrowing approach
Primary endpoint(s)	12-May-2025	Section 4.2.1.4 clarify the prior weights and thresholds used in primary analysis.	Yes, version 4.0	To align with the latest CSP

1 INTRODUCTION

The purpose of this document is to give details for the statistical analysis of study D3468C00003, a multicenter, multinational, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of an intravenous (IV) treatment regimen of 300 mg anifrolumab versus placebo in adult Asian participants with moderate to severe active, autoantibody-positive systemic lupus erythematosus (SLE) despite receiving standard of care treatment.

This statistical analysis plan (SAP) provides details of the summaries and analyses to be performed to support the clinical study report (CSR) of the study. The reader is referred to the latest version of the clinical study protocol (CSP) of the study D3468C00003.

2 CHANGES TO PROTOCOL PLANNED ANALYSES

No change to protocol planned analyses listed in CSP v3.0.

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

There will be 2 clinical database locks (CDLs) in the study. The primary analysis CDL will occur after all randomized participants have completed Week 52 (or have withdrawn from the study). The primary analysis CDL includes the assessments of the primary, key secondary, and secondary objectives, including safety. Exploratory objectives available at the time of the primary CDL may also be included. No changes to the clinical database up to Week 52 will be made once the primary CDL has occurred. The final analysis CDL will be performed after all participants have completed their last visits/assessments.

3.2 Analysis Populations

Approximately 260 eligible participants receiving standard of care treatment are randomized in a 1:1 ratio to receive either a fixed IV dose of 300 mg anifrolumab or placebo every 4 weeks (Q4W) for a total of 13 doses (Week 0 to Week 48), with the primary endpoint evaluated at the Week 52 visit.

Randomization will be stratified using the following factors: Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K) score at screening (<10 points versus ≥10 points); randomization (Day 1) oral corticosteroids (OCS) dose (<10 mg/day versus ≥10 mg/day prednisone or equivalent); and region (mainland China versus not mainland China).

3.2.1 All Participants Analysis Set

All participants who signed the informed consent form (ICF). The all participants analysis set is used to summarize disposition and screening failures.

3.2.2 Full Analysis Set

The full analysis set (FAS) is used as the primary population for reporting efficacy. This comprises all participants randomized into the study who receive at least 1 dose of investigational product (IP) and is analyzed according to randomized treatment (modified Intention-To-Treat [mITT]).

3.2.3 Safety Analysis Set

The safety analysis set consists of all participants who have received at least 1 dose of IP. Erroneously treated participants, who are randomized to receive anifrolumab but are actually given placebo for the entirety of their time in the study, are accounted for in the placebo arm. A participant who has on one or several occasions received anifrolumab is classified in the respective anifrolumab group. This population is used in the analysis of safety endpoints.

3.2.4 Pharmacokinetic Analysis Set

The pharmacokinetic (PK) analysis set consists of all participants in the FAS who received anifrolumab and who have at least 1 quantifiable serum PK observation post first dose, with no protocol deviations thought to significantly impact on the analysis of the PK data.

The exclusion of any participants or time points from the calculation of the PK parameters are documented, including the reasons for exclusion. The available concentration data and PK parameter data for any participants excluded from the PK analysis set is listed only.

The PK analysis set is used to summarize the PK data.

3.3 General Considerations

- Summary tables are produced by treatment group (Anifrolumab 300mg and Placebo) unless otherwise stated. Total columns are produced only for tables of study population data, unless otherwise stated.

- Continuous variables are summarized by the number of non-missing observations, mean, standard deviation (SD), median, and the minimum and maximum as appropriate.

- Log-transformed continuous variables are summarized in a similar manner. The geometric mean and coefficient of variation (CV%), median, minimum and maximum are

presented as appropriate. In addition, the relative change from baseline is calculated by back-transforming the change from baseline of the logarithmic variable as necessary.

- Categorical variables are summarized as counts (n) and percentages (%). Unless otherwise stated, percentages are calculated using the relevant analysis set population total by treatment group as the denominator. Percentages are rounded to one decimal place; except 100% which are displayed without any decimal places. Percentages are not to be presented for zero counts.

- If a model is used to estimate the treatment difference, the corresponding confidence interval (CI) according to the model is presented. Otherwise, the unadjusted CI is used.

- Mean, SD, medians, 1st quartile and 3rd quartile are rounded to one additional decimal place relative to the original data, the maximum, minimum will be displayed with the same accuracy as the original data and the CV% is presented with 1 decimal place. The statistics are displayed in the order of n, mean, minimum, first quartile (optional), median, third quartile (optional), maximum. 95% CIs are presented for treatment comparison to one more decimal place than the raw data. For the confirmatory treatment comparisons, p-values will be reported to 3 decimal places. Percentages (proportion) will be rounded to one decimal place.

- Summaries are provided by time point of assessment where appropriate. Where summaries are by time point, study day is calculated in relation to the date of first dose of study treatment. For assignment of data to time points using the visit windows, study day is calculated in relation to the date of first dose of investigational product.

· Study Day 1 is defined as the day of first study drug administration.

· Study Days prior to the date of first study drug administration, are defined as: (Date of assessment – date of first administration of IP)

· Study Days after the first dose of investigational product, are defined as: (Date of assessment – date of first administration of IP) +1. Using this definition, the day of first dose of IP is Day 1 and the scheduled visit date of Week 4 is study day 29 (=28+1) for example.

- Corresponding listings are provided for certain results unless stated otherwise. Any additional listings that are required will be described in the appropriate sections of the SAP. The listings are sorted by treatment group and subject number.

• SAS® version 9.4 or higher will be used for all data analyses, except for Bayesian Dynamic Borrowing approach (Section 4.2.1.4 and 4.2.1.6), which was implemented using the RBesT package in R (version 3.6.3 or higher).

3.3.1 General Study Level Definitions

3.3.1.1 Definition of Baseline

Baseline is defined as the last measurement prior to the first dose administration of anifrolumab on Day 1. If the Day 1 value is missing or is invalid or is collected after administration of IP, the latest assessment prior to dose administration on Day 1 serves as baseline.

3.3.1.2 Study Periods

The allocation to the study periods is performed after any imputation of missing dates (for AEs only) as described in Section 3.3.1.3. The study periods are defined as follows:

- Screening Period: Up to 35 days.
- Double Blind Treatment Period: A 52-week double-blind treatment period from first dose of IP through Week 52 visit or date of study withdrawal/lost to follow-up (FU)/death/ Early discontinuation visit (EDV), whichever comes first; efficacy endpoints and pharmacodynamics endpoints including complements and anti-dsDNA are assessed during the double-blind treatment period.
- Safety Follow-up: After Week 52 or the date of EDV (the completion of the double-blind treatment period), participants may continue in the study for further 8 weeks to complete the safety follow-up period, which is comprised of two visits, at Week 56 (or 8 weeks after the last dose) and Week 60 (or 12 weeks after the last dose).

Measurements on Study Day 1 are with a time prior to the first IV administration of IP, use the study period label as “Screening”. If any of the measurement occurs after the start time of IP administration, the study period label is as “Treatment”. In case of the time of measurement at Study Day 1 is not available, it is assumed to be prior to first administration of IP and study period label is “Screening”. If the measurement is after Week 52 or Early Withdrawal and participants will continue in the study for further 8 weeks to complete the safety follow-up period, then the study period label is as “Safety Follow-up”.

Study periods for the analyses of adverse events (AE) are defined separately in Section 4.6.2.1.

3.3.1.3 Handling of Missing Data

If in case at any site, the device to collect participant reported outcomes are not synchronized with the real time, then baseline for the following participant reported outcomes are derived based on the date only, and not time of the assessment.

- Lupus quality of life scale
- EuroQol 5 dimensions
- Work Productivity and Activity Impairment (WPAI) – Lupus
- Medical Resource Use
- Columbia-Suicide Severity Rating Scale (C-SSRS)

In general, missing data is not imputed and is treated as missing unless specifically described in an analysis section. The following considerations are made for missing safety data, AE dates, and concomitant medication/diseases dates:

- Safety assessment values of the form '<x' or '>x' (i.e., above or below the limits of quantifications) are imputed as 'x' in the calculation of summary statistics but displayed as '<x' or '>x' in the listings.
- Adverse events that have missing causality after data querying are assumed to be related to study drug.
- For missing AE/concomitant medication/disease start dates, the following are applied:
    - Missing day: Impute the 1st of the month unless the month is the same as the month of the first dose of study drug and the end date is on or after the first dose of study drug or ongoing then impute first dose date.
    - Missing day and month: Impute 1st January unless year is the same as first dose date and the end date is on or after the first dose of study drug or ongoing then impute first dose date.
    - Missing year: Impute the year of dosing.
    - Completely missing: Impute first dose date unless the end date suggests it could have started prior to this in which case impute 1st January of the same year as the end date.

- For missing AE/concomitant medication/disease end dates, the following are applied:
    - Missing day: Impute the last day of the month unless the month is the same as that of the last dose of the study drug in which case impute the date of last dose.
    - Missing day and month: Impute 31st December unless the year is the same as that of the last dose of study drug in which case impute date of last dose.
    - Missing year: Impute the year of dosing.
    - Completely missing: Do not impute a date i.e., assume that AE/concomitant medication/disease is ongoing.

If a participant is known to have died where only a partial death date is available, then the date of death is imputed as the latest of

- the last known date to be alive + 1 from the database
- the death date using the available information provided:
    - Missing day only: Use the 1st of the month.
    - Missing day and month: Use 1st January.
    - Missing year: Impute the year of dosing.

For all missing start/end dates, flags are retained in the analysis datasets indicating where any programmatic imputation has been applied, and in such cases, any durations would not be calculated.

3.3.2 Visit Window

For visit based analyses, the variables will be summarized based on the scheduled days with adjusted analysis-defined visit windows. The adjusted analysis-defined windows are summarized below:

Table 1 Visit windows

Adjusted Defined Windows Visit	Scheduled Study Day	Maximum Windows
Baseline / Day 1	1	Study Day ≤1
Week 4	29	2≤ Study Day ≤42
Week 8	57	43≤ Study Day ≤70

Week 12	85	71≤ Study Days ≤98
Week 16	113	99≤ Study Day ≤126
Week 20	141	127≤ Study Day ≤154
Week 24	169	155≤ Study Day ≤182
Week 28	197	183≤ Study Day ≤ 210
Week 32	225	211≤ Study Day ≤238
Week 36	253	239≤ Study Day ≤266
Week 40	281	267≤ Study Day ≤294
Week 44	309	295≤ Study Day ≤322
Week 48	337	323≤ Study Day ≤350
Week 52	365	351≤ Study Day ≤378
Week 56	393	379≤ Study Day ≤406
Week 60	421	407≤ Study Day

If there is more than one value per participant within an analysis visit window then the closest value to the scheduled visit date is summarized, or the earlier value if the values are equidistant from the nominal visit date. If two observations are equidistant from the scheduled visit on the same day, the non-missing one with the earlier collection time will be included in the analysis. Listings display all values contributing to a time point for a participant and highlight the value that contributed to the summary table where feasible.

If a visit window does not contain any observations, then the data will remain missing.

For the Baseline / Day 1 visit window, the visit label "Baseline" is used for baseline record as defined in Section 3.3.1.1. For measurements at Study Day 1 with a time of measurement after the start of IP administration, the visit label "Day 1" is used (indicating a post-baseline measurement). If the time of measurement at Study Day 1 is not available, it is assumed to be prior to the first administration of IP. If Study Day ≤1 and records not selected as "Baseline", the visit label "Screening" will be used.

For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment is used, regardless of where it falls in an interval.

Listings display all values contributing to a time point for a participant and highlight the value that contributed to the summary table where feasible.

3.3.3 Handling of Unscheduled Visits

Visits that fall outside the visit windows defined in the CSP will be classified as Unscheduled Visits. If an unscheduled visit is the closest visit to the scheduled visit study day or the earlier non-missing one when the values are equidistant from the nominal visit date (described in Section 3.3.2), they are included in the by visits summary tables. All

visits are considered for summaries of treatment emergent abnormalities, minimum/maximum on treatment values etc. and will be presented in the listings.

3.3.4 Multiplicity/Multiple Comparisons

The primary analysis for the primary endpoint is performed using Bayesian dynamic borrowing approach with supportive analysis performed using frequentist hypothesis testing approach. Multiplicity adjustments are not performed when testing the primary and key secondary endpoints. Nominal p-values are presented for the primary and key secondary endpoints.

3.3.5 Handling of Protocol Deviations in Study Analysis

All protocol deviations identified during monitoring of the study are recorded in the clinical trial management system (CTMS). Important protocol deviations (IPDs) are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being.

Protocol deviations are classified as important or not important by different methods, either programmatically or manually identified, as defined in the Protocol Deviation Plan based on periodic reviews of CTMS deviation reports and Dry Runs. The Sponsor reviews the classification and provides the final determination. The list of IPDs is finalized and documented prior to CDL and unblinding. Only IPDs are listed and tabulated in the CSR.

A detailed description on how IPDs are collected and identified is referenced in the Protocol Deviation plan.

The following general categories are considered IPDs and are listed and discussed in the CSR as appropriate for the study:

· Deviation 1: Inclusion Criteria Deviations

· Deviation 2: Exclusion Criteria Deviations

· Deviation 3: Discontinuation Criteria for study product met but subjects not withdrawn from study treatment

· Deviation 4: Discontinuation Criteria for overall study withdrawal met but patient not withdrawn from study

· Deviation 5: IP Deviation

- Deviation 6: Excluded Medications Taken
- Deviation 7: Deviations related to study procedure
- Deviation 8: Other Important Deviations

3.3.6 Identifying Cases of Restricted Medication Use

Restricted medications are concomitant medications that have restrictions applied to their use. Restricted medications that are used outside of what is permitted, that are deemed to affect endpoints of interest, are considered intercurrent events.

These cases are confirmed by the medication review team who reviews potential restricted medication use (RM) cases that have been identified via programmatic rules. See the restricted medication review charter for details on the review process. See Appendix 7.4 for the programmatic rules.

Note that prohibited medications are considered a subset of restricted medications that are not permitted during the study and result in immediate premature end of treatment if taken during the study treatment period.

3.3.7 Identifying Cases of PEOT due to Lack of Therapeutic Response

The reason for early IP discontinuation is classified as "Lack of therapeutic response related" if persistent or increased disease activity at time of study intervention discontinuation compared with baseline activity was evident upon review of specific endpoints. Lack of therapeutic response related reasons include discontinuation reasons of "Condition under investigation worsened" and "Condition under investigation not improved", and reasons that require blinded review to determine the relationship with lack of therapeutic response. These reasons are "Subject decision", "Adverse event", "Development of study specific discontinuation criteria", and "Other".

The manual review is conducted blindly by a sponsor-led Central Review Team prior to the primary CDL. PEOT charter details the data review process and documentation processes.

3.3.8 Identifying Cases of PEOT due to COVID Lockdown

The reason for early IP discontinuation is classified as "COVID lockdown" if the reason for discontinuation is related to COVID lockdown. This is identified if main reason for investigational product discontinuation = 'Other' with 'COVID'COVID specified in the text on the eCRF form DOSDISC.

3.3.9 Data excluded from analysis

During the study, an error was identified in the electronic patient-reported outcome (ePRO) software for the simplified Chinese and traditional Chinese versions, impacting sites in mainland China and Taiwan (See more details in QE-273452 and QE-386615). This error affects the last question of the EQ-5D-5L questionnaire in ePRO software version 06.01 and earlier. The correct instruction above the visual analogue scale indicates “100 represents the best health you can imagine” and the instruction below indicates “0 represents the worst health you can imagine”. In the erroneous translated screen patients are seeing, the instruction texts upside down with the above instruction saying 100 for worst and the below instruction saying 0 for best. This discrepancy makes it extremely difficult to retrospectively verify each patient's response for each visit and ensure accuracy. After consulting with a PRO expert, the study team has decided to only use correct data for the EQ-5D-5L analysis, which will be treated as exploratory.

On 18-Dec-2024, the decision was made to exclude data collected from affected sites (sites in China mainland and Taiwan) before the software was updated to version 7.0 from the EQ-5D-5L analysis.

Therefore, the EQ-5D-5L analysis only included all data from unaffected sites (sites not in China mainland and Taiwan) and data collected from affected sites (sites in China mainland and Taiwan) after the software was updated to version 7.0.

4 STATISTICAL ANALYSIS

This section provides information on definitions, derivation and analysis/data presentation per domain.

4.1 Study Population

The domain study population covers disposition, analysis sets, protocol deviations, demographics, baseline characteristics, disease characteristics, medical history and concomitant disease, prior and concomitant medication and study drug compliance.

4.1.1 Disposition and Completion Status

4.1.1.1 Definitions and Derivations

Not applicable.

4.1.1.2 Presentation

Disposition is summarized based on the all-participant analysis set and includes the following information:

- the number of participants screened for the eligibility criteria
- the number of participants who are randomized
- the number of participants who are randomized and who received at least one dose of study drug
- the number and percentage of participants who completed treatment/discontinued treatment (including reasons for IP discontinuation)
- the number and percentage of participants who completed or discontinued study (including reasons for study discontinuation)

A randomization listing is presented and includes the following information: randomization number, full enrolment number, date of randomization and randomized treatment group.

The impact of COVID-19 (i.e., participants discontinued/withdraw from study, protocol deviations) is collected in the electronic case report form (eCRF) as a variable related to global/country situation.

Tables on disposition due to global/country situation, global/country situation study disruptions (i.e., subjects with at least one disruption due to global/country situation) are provided.

A summary of number and percentage of participants in each region, each country/area and each site by treatment group and overall are provided for the FAS. Additionally, the stratification factors at randomization (as calculated from the data for SLEDAI-2K score at screening and Randomization [Day 1] OCS dose and region (mainland China versus not mainland China)) are presented with count and percentage of the participants by treatment group and overall, for the FAS.

4.1.2 Analysis Sets

4.1.2.1 Definitions and Derivations

Analysis sets are defined in Section 3.2.

4.1.2.2 Presentation

Analysis sets are summarized based on all the participants analysis set. The number of participants included and excluded (including reason for exclusion) in each analysis population is presented by treatment group and overall.

Any participants excluded from any analysis set are listed including the reasons for exclusion.

4.1.3 Protocol Deviations

4.1.3.1 Definitions and Derivations

IPDs are defined in the PD Plan and are considered as those deviations from the protocol likely to have a significant impact the completeness, accuracy, and/or reliability of the study data or that may significantly participant's rights, safety, or well-being of study treatment. The final list of IPDs is determined before CDL.

4.1.3.2 Presentation

IPDs are summarized based on the FAS. The number and percentage of participants meeting each IPD criterion are summarized by treatment group and overall. Participants who deviate from a given criterion more than once are counted once for that criterion. Any participants with more than one IPD are counted once in the overall summary.

4.1.4 Demographics

4.1.4.1 Definitions and Derivations

Demographic characteristics include geographic region, age, age group reported as [>=18 to < 65, ≥ 65 years], sex, ethnicity and race.

Age is measured in years and calculated as:

Age (years) = (Date of randomization – date of birth + 1) / 365.25. Age variable as collected at screening in eCRF (case report form) is considered in case date of birth is missing or not available due to local/country regulations.

4.1.4.2 Presentation

Demographic characteristics are summarized as per Section 3.3, based on the FAS. Summaries are presented by treatment group and overall.

4.1.5 Baseline Characteristics

4.1.5.1 Definitions and Derivations

Baseline characteristics include height (cm), weight (kg), body mass index [BMI] (kg/m^2) and BMI group [≤28 and 28 kg/m^2].

BMI is calculated as:

$$BMI\ (kg/m^2) = weight\ (kg) / (height\ (m))^2.$$

4.1.5.2 Presentation

Baseline characteristics as well as the cardiovascular risk are summarized as per Section 3.3, based on the FAS. Summaries are presented by treatment group and overall.

4.1.6 Disease Characteristics

4.1.6.1 Definitions and Derivations

SLE disease characteristics are assessed and represented by the following variables:

- SLEDAI-2K score/ Clinical SLEDAI-2K score at screening and at baseline
- Adjudicated British Isles Lupus Assessment Group 2004 (BILAG-2004) score at screening and at baseline
- BILAG-2004 global score at baseline
- Physician Global Assessment (PGA) score at baseline
- Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI) activity and damage scores at baseline
- Joints including swollen joints, tender joints and active joints at baseline
- Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI) at baseline
- Time from initial SLE diagnosis to randomization (months)
- Interferon (IFN) 4-gene test status evaluated at screening
- anti-double stranded deoxyribonucleic acid (anti-dsDNA) levels, antinuclear antibody (ANA), complement levels, third component of complement (C3), fourth component of complement (C4) and total hemolytic complement (CH50) evaluated at baseline.

SLEDAI-2K

The SLEDAI-2K disease activity index consists of a list of organ manifestations, each with a definition. The SLEDAI-2K organ systems are defined as follows:

- Central nervous system (6 items): seizure, psychosis, organic brain syndrome, visual disturbance, cranial nerve disorder, and lupus headache
- Vascular (2 items): CVA (cerebrovascular accident) and vasculitis
- Musculoskeletal (2 items): arthritis and myositis
- Renal (4 items): urinary casts, haematuria, proteinuria, and pyuria
- Mucocutaneous (3 items): rash, alopecia, and mucosal ulcer
- Cardiovascular system and respiratory (2 items): pleurisy and pericarditis
- Immunology (2 items): low complement and increased DNA binding

- Haematological and fever (3 items): fever, thrombocytopenia, and leukopenia.

These 24 lupus-related items in which signs, symptoms, and laboratory tests for each organ systems are given a weighted score. Items present in the last 28 days are scored and summed to create score for each organ system, and a total score.

The clinical SLEDAI-2K score is derived as the sum of the scores for the following clinical components: arthritis, myositis, rash, alopecia, mucosal ulcers, pleurisy, pericarditis and vasculitis. The clinical SLEDAI-2K score is the SLEDAI-2K score excluding items attributable to any urine or laboratory results including immunologic measures.

Scores for the SLEDAI organ systems will be derived in the way as SLEDAI-2K but using the scores for the respective items only.

BILAG-2004

The BILAG-2004 contains 97 (from Appendix F of CSP) clinical and laboratory parameters which are divided into a translational index with 9 organ systems (Constitutional, Mucocutaneous, Neuropsychiatric, Musculoskeletal, Cardiorespiratory, Gastrointestinal, Ophthalmic, Renal and Hematology). The first 7 organ systems (Constitutional, Mucocutaneous, Neuropsychiatric, Musculoskeletal, Cardiorespiratory, Gastrointestinal, Ophthalmic) contain clinical parameters which are assessed by the treating physician as new (4), worse (3), same (2), improving (1) and not present (0). The Renal and Hematologic scoring is based on laboratory values. BILAG organ system scores are assigned into 5 different levels as A, B, C, D or E scores to all study visits by strictly following this Index Scoring. Results from the original scores are used for calculation of the primary efficacy endpoint, BICLA response at Week 52, defined in Section 4.2.1.1. The treatment has no bearing on the scoring index. Only the presence of active manifestations influences the BILAG-2004 scoring.

BILAG-2004 grades will be presented by organ system. Furthermore, a BILAG-2004 global score is derived by summing-up the numerical score equivalents for each organ system with the numerical score equivalents given as: A = 12, B = 8, C = 1, D = 0, and E = 0.

Unless otherwise defined, scores that are derived from summing up items will be considered missing if any of the items are still missing after applying the LOCF (last observation carried forward) rules. This should apply to both SLEDAI-2K and BILAG-2004.

Physician's Global Assessment (PGA)

The PGA is a modification of the classic analogue scale in that it is anchored with numbers from 0 to 3 demarcating no, mild, moderate, and severe disease. The number 3 indicates severe disease and is at the end of the scale. Any disease rated greater than 2.5 is very severe. The range of moderate disease covers approximately 1.5 to 2.4. Mild disease falls below 1.5. The instrument is similar to a logarithmic scale, with greater distances or demarcations possible among more mild-moderate symptoms. This is a global assessment, factoring in all aspects of the participant's lupus disease activity. It should not reflect non-lupus medical conditions.

Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI) inflammatory disease activity

The CLASI is a validated index used for assessing the cutaneous lesions of SLE and consists of 2 separate scores: the first summarizes the inflammatory activity of the disease; the second is a measure of the damage done by the disease. The activity score takes into account erythema, scale/hypertrophy, mucous membrane lesions, recent hair loss, and non-scarring alopecia. The damage score represents dyspigmentation, scarring/atrophy/panniculitis, and scarring of the scalp. Participants are asked if their dyspigmentation lasted 12 months or longer, in which case the dyspigmentation score is doubled. Each of the above parameters is measured in 13 different anatomical locations, included specifically because they are most often involved in cutaneous lupus erythematosus (CLE). The most severe lesion in each area is measured.

Joints

The joint count consists of swollen and tender joints. An active joint is defined as a joint with both swelling and tenderness.

SDI

The SDI is defined for 12 organ systems (possible scores): peripheral vascular (0 to 5), ocular (0 to 2), neuropsychiatric (0 to 6), renal (0 to 3), pulmonary (0 to 5), cardiovascular (0 to 6), gastrointestinal (0 to 6), musculoskeletal (0 to 7), skin (0 to 3), endocrine (diabetes) (0 to 1), gonadal (0 to 1) and malignancies (0 to 2). The SDI global score is the sum of the damage scores for all 12 organ systems.

Unless otherwise defined, scores that are derived from summing up items will be considered missing if any of the items are still missing.

Interferon (IFN)

The IFN 4-gene test status (high/low) is used to monitor SLE activity and severity. High mRNA expression of these IFN genes is associated with greater SLE disease activity. The IFN 4-gene test status is assessed at baseline.

ANA

ANA is used to evaluate the immunogenicity of anifrolumab using a titer approach. ANA positivity, or presence of ANA, is defined as ANA titres greater than or equal to limit of detection. ANA titres for ANA positive participants are summarized at baseline.

Anti-dsDNA and complements

Anti-dsDNA is the SLE autoantibodies and complement levels include C3, C4 and CH50.

Complements C3, C4 and CH50 levels are categorised as normal or abnormal.

Abnormal complement level is defined as complement level below lower limit of normal (LLN).

Time from initial SLE diagnosis to randomization is calculated in months as (date of randomization – date of diagnosis) converted to months.

4.1.6.2 Presentation

The above variables are summarized as descriptive statistics (for both continuous and categorical variables as applicable) defined in Section 3.3 on the FAS.

SLEDAI-2K (derived as the sum of scores for all items) will be evaluated using the difference in mean change from baseline longitudinally over time to Week 52.

4.1.7 Medical History and Concomitant Disease

4.1.7.1 Definitions and Derivations

Medical/surgical history and concomitant diseases are classified according to the terminology of the Medical Dictionary for Regulatory Activities (MedDRA), version 24.0 (March 2021) or higher. The imputation method described in Section 3.3.1.3 is used in the instance that partial or missing dates are recorded. After imputation of dates, concomitant diseases will be classified as prior or concomitant. A prior disease is defined as any disease with a start and end date prior to the first dose date (exclusive). A concomitant disease is defined as any disease with an end date on or after the first dose date. A disease with a completely missing end date is considered as concomitant.

4.1.7.2 Presentation

Medical/surgical history and concomitant diseases are summarized based on the FAS. The number and percentage of participants with relevant medical/surgical history/concomitant diseases are presented by treatment group and summarized by system organ class (SOC) and Preferred Term (PT). Participants with histories in more than one SOC/PT are counted

only once in that SOC/PT. Tables are sorted by international order for SOC and in alphabetical order for PT.

4.1.8 Prior and Concomitant Medications

4.1.8.1 Definitions and Derivations

Prior and concomitant medications are coded using the WHO Drug Dictionary (WHO-DD). The imputation method described in Section 3.3.1.3 is used in the instance that partial or missing dates are recorded.

Any medication taken by the participant prior to the first dose date of IP is considered prior medication. Any medication taken by the subject at any time between the date of the first dose (including the date of the first dose) of IP up to Week 52 (Visit 14/EDV), inclusive, is considered concomitant medication. Any medication started prior to the first dose of IP and did not end before first dose or was ongoing is considered as both prior and concomitant medication.

Disease related treatments at baseline are defined as all medications with therapy reason containing "disease under study" with an intake at the date of first dose of IP (i.e., start date on or before the date of first dose and end date on or after date of first dose or was ongoing). The medications will be presented in the following categories in a separate table:

- OCS
- Anti-malarial
- Immunosuppressant, defined for medications of azathioprine, mycophenolate mofetil/mycophenolic acid, methotrexate, cyclosporine, tacrolimus, or mizoribine.
- NSAIDs (nonsteroidal anti-inflammatory drugs)
- Other SLE medication, defined as SLE medications not covered within the above categories.

Restricted and prohibited medications are detailed in completeness in Appendix D. There are some special cases of restricted medication use intercurrent events where subjects are eligible to respond after some duration of time. Steroid bursts during the first 12 weeks of the study might be needed for some subjects to allow for a reasonable amount of time for the anifrolumab drug response to be observed. Given the biologic properties of steroids, an early burst is not thought to confound efficacy assessments for the primary endpoint at week 52.

These are the following rules for intermittent ICE of RM use:

- If a subject has greater than 1 steroid burst before Week 12, the subject will be considered to have an ICE of RM use from the burst breaking the rule until 12 weeks after the end date of the burst i.e. until the end date of the burst + 84 days.

- If a subject has OCS or equivalent total daily dose above their Day 1 dose for a period >14 days, ending before or on Week 12, the subject will be considered to have an ICE of RM use from the start date of the rule break until 12 weeks after the end date of the burst i.e. until the end date of the burst + 84 days.

12 weeks takes into consideration the biologic half-life of the treatment (18-36 hours) and the clinical effect of the increased steroids.

These intermittent ICEs of RM use will be documented and captured in the clinical database.

For non-SLE steroid use, details are provided in the 7.4, rule # RM 20.1-20.3. As stated in the CSP, Section 7.7.2, the participants are allowed during the DB treatment period to receive one additional course of burst and taper corticosteroids for increased non-SLE disease activity between Week 12 and Week 40. To not restrict the manual review, the time point of when a potential burst occurred and the number of bursts has not been implemented in the programming rule. This to ensure that all concomitant therapy for this reason are reviewed manually by the MRT.

4.1.8.2 Presentation

Prior and concomitant medications are summarized based on the FAS. The number and percentage of participants is presented by treatment group and overall and summarized by ATC class. Participants with medications in more than one ATC class will be counted only once in that ATC class. Tables are sorted alphabetically by ATC class.

The number and percentage of subjects with restricted and prohibited medication use during double-blind phase of the study (i.e., up until Week 52 or EDV) will be summarized by the latest available WHO-Drug Dictionary Preferred Term. Only the first occurrence per subject will be presented.

4.1.9 Study Drug Compliance

4.1.9.1 Definitions and Derivations

Study drug compliance is calculated for each participant and expressed as a percentage. The percent study drug compliance is calculated as the number of dosing occasions relative to the total number of dosing occasions expected up until the date of permanent discontinuation of IP or end of the treatment period, whichever is earlier.

$$\text{Study drug compliance (\%)} = 100 \times \frac{\text{total } n\text{umber of dosing occasions administered}}{\text{total number of dosing occasions expected}}$$

Total number of dosing occasions expected = sum of expected anifrolumab based on treatment exposure. It is calculated over the entire treatment period excluding the period of permanent discontinuation of IP. For subjects completing the study, the total number of dosing occasions expected is 13. For subjects who discontinue from the study early, the total number of dosing occasions expected will be based on treatment exposure, i.e. exposure/28 (rounded to the nearest integer). The duration of treatment exposure is calculated per Section 4.6.1.1.

4.1.9.2 Presentation

Study drug compliance is summarized based on the FAS and by treatment group and overall, using the principles outlined in Section 3.3 for continuous variables.

4.2 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, key secondary and other endpoints including sensitivity and supportive analyses.

If not stated otherwise, change from baseline will be calculated as value at the respective time point minus value at baseline. The percent change from baseline is defined as change from baseline divided by baseline value multiplied with 100.

Intercurrent events are defined as events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. In this study the intercurrent events are:

- Restricted Medication use
- Premature End of Treatment
- Death.

If a patient observes an intermittent ICE of RM use (see Section 4.1.8.1), the ICE strategy of RM will be applied to the period defined.

Hereafter the intercurrent events of Restricted Medication use is referred to as RM, and Premature End of Treatment is referred to as PEOT. Additionally, Non-Response or Non-Responder are referenced as NR. PEOT is identified by the eCRF form DOSDISC "Discontinuation of Investigational Product"). Subjects with IP being prematurely and permanently discontinued are considered as PEOT status from the date of last IP dose administered + 28 days.

RM are identified from the start date of the concomitant medication use reported in the CM form meeting the criteria to be classified as RM. The start date for RM status is identified during the restricted medication review – see Appendix 7.4.

Where specified, endpoint analyses will take into account the stratification factors. The stratification factors are derived as follows:

- SLEDAI-2K score at screening (<10 points versus ≥10 points)
  If different measurements are available for re-screened subjects, the value at re-screening is used

- Randomization (Day 1) OCS dose (<10 mg/day versus ≥10 mg/day prednisone or equivalent)
  For the classification, the derived OCS dose rounded to 1 decimal place is used

- Region (mainland China versus not mainland China)

SLEDAI-2K score at Screening and Randomization (Day 1) OCS strata is derived programmatically from the data recorded in the eCRF.

Where necessary (e.g., for subgroup analysis based on stratification factors), the model factors will be reduced. If not stated otherwise, the subgroup analysis will be suppressed if any of the sub-populations in any treatment group will consist of less than 25 participants.

Table 2 Summary of Analyses

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Objective 1: To evaluate the effect of anifrolumab compared to placebo on disease activity as measured by the difference in the proportion of participants achieving BICLA response at Week 52.					
Primary Analysis of the Primary Endpoint	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (Bayesian analysis), Week 52	4.2.1.4
Supportive Analysis of the Primary Estimand	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH), Week 52	4.2.1.5
Supportive Analysis of the Primary Estimand	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/PEOT/Death=NR	Odds ratio between interventions (Bayesian analysis), Week 52	4.2.1.6
Supplementary Analysis of the Primary Endpoint – targeting impact of PEOT due to different reasons	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM, Deaths, PEOT due to lack of therapeutic response = NR Hypothetical strategy where PEOT due to COVID lockdown = MAR Treatment policy where PEOT not due to lack of therapeutic response or COVID lockdown = ignored	Difference in proportions between interventions (CMH; multiple imputation), Week 52	4.2.1.7.1
Supplementary Analysis of the Primary Endpoint - targeting the impact of PEOT	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/Death=NR Treatment policy where PEOT = ignored	Difference in proportions between interventions (CMH), Week 52	4.2.1.7.2

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Supplementary Analysis of the Primary Endpoint - targeting impact of different discontinuation rates	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/Death=NR PEOT by tipping point analysis	Difference in proportions between interventions (Tipping point: Pearson's chi-squared test), Week 52	4.2.1.7.3
Supplementary Analysis of the Primary Endpoint - targeting the impact of RM	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where PEOT/Death=NR Treatment policy where RM = ignored	Difference in proportions between interventions (CMH), Week 52	4.2.1.7.4
Supplementary Analysis of the Primary Endpoint - targeting the impact of PEOT and RM	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where Death=NR Treatment policy where RM/PEOT = ignored	Difference in proportions between interventions (CMH), Week 52	4.2.1.7.5
Supplementary Analysis of the Primary Endpoint - targeting the impact of being automatically classified as responder due to their baseline visual analogue scale (VAS) and/or BILAG scores	BICLA response at week 52	Asian adults with moderate-to-severe SLE (Excluding subjects with no baseline BILAG A or B, and subjects with baseline PGA VAS >2.7)	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH), Week 52	4.2.1.7.6
Supplementary Analysis – time to BICLA response	Time to BICLA response through Week 52	Asian adults with	Composite strategy where RM/PEOT/Death=NR	Estimated hazard ratio between interventions by Cox PH model	4.2.1.7.7

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
		moderate-to-severe SLE			
Subgroup Analyses of the Primary Endpoint	BICLA response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH), Week 52 by subgroups	4.2.1.8
Objective 2: The proportion of participants who achieve SRI(4) response at week 52					
Primary Analysis of the Key Secondary Endpoint – SRI(4)	SRI(4) response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH), Week 52	4.2.2.4
Supplementary Analysis of the Key Secondary Endpoint – targeting the impact of PEOT due to different reasons	SRI(4) response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM, Deaths, PEOT due to lack of therapeutic response = NR Hypothetical strategy where PEOT due to COVID lockdown = MAR Treatment policy where PEOT not due to lack of therapeutic response or COVID lockdown = ignore	Difference in proportions between interventions (CMH; multiple imputation), Week 52	4.2.2.5
Supplementary Analysis of the Key Secondary Endpoint – targeting the impact of PEOT	SRI(4) response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/Death=NR Treatment policy where PEOT = ignored	Difference in proportions between interventions (CMH), Week 52	4.2.2.5
Supplementary Analysis of the Key Secondary Endpoint – targeting impact of different discontinuation rates	SRI(4) response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/Death=NR PEOT by tipping point analysis	Difference in proportions between interventions (Tipping point: Pearson's chi-squared test), Week 52	4.2.2.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Supplementary Analysis of the Key Secondary Endpoint – targeting the impact of RM	SRI(4) response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where PEOT/Death=NR Treatment policy where RM = ignored	Difference in proportions between interventions (CMH), Week 52	4.2.2.5
Supplementary Analysis of the Key Secondary Endpoint – targeting the impact of PEOT and RM	SRI(4) response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where Death=NR Treatment policy where RM/PEOT = ignored	Difference in proportions between interventions (CMH), Week 52	4.2.2.5
Subgroup Analyses of the Key Secondary Endpoint – SRI(4)	SRI(4) response at week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH), Week 52 by subgroups	4.2.2.6
Objective 3: The proportion of participants who achieve an OCS dose ≤7.5 mg/day at Week 40, which is maintained through Week 52 in the sub-group of those with baseline OCS ≥10 mg/day.					
Primary Analysis of the Key Secondary Endpoint – OCS	Reduction in OCS dose to ≤7.5 mg/day at Week 40, and maintained through Week 52	Asian adults with moderate-to-severe SLE with baseline OCS >= 10mg/day	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH)	4.2.3.4
Supplementary Analyses of the Key Secondary Endpoint – targeting the impact of PEOT	Reduction in OCS dose to ≤7.5 mg/day at Week 40, and maintained through Week 52	Asian adults with moderate-to-severe SLE with baseline	Composite strategy where RM/Death=NR Treatment policy where PEOT = ignored	Difference in proportions between interventions (CMH)	4.2.3.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
		OCS >= 10mg/day			
Supplementary Analysis of the Key Secondary Endpoint – targeting the impact of different discontinuation rates	Reduction in OCS dose to ≤7.5 mg/day at Week 40, and maintained through Week 52	Asian adults with moderate-to-severe SLE with baseline OCS >= 10mg/day	Composite strategy where RM/Death=NR PEOT by tipping point analysis	Difference in proportions between interventions (Tipping point: Pearson's chi-squared test)	4.2.3.5
Supplementary Analysis of the Key Secondary Endpoint – targeting the impact of RM	Reduction in OCS dose to ≤7.5 mg/day at Week 40, and maintained through Week 52	Asian adults with moderate-to-severe SLE with baseline OCS >= 10mg/day	Composite strategy where PEOT/Death=NR Treatment policy where RM=ignored	Difference in proportions between interventions (CMH)	4.2.3.5
Supplementary Analysis of the Key Secondary Endpoint – targeting the impact of PEOT and RM	Reduction in OCS dose to ≤7.5 mg/day at Week 40, and maintained through Week 52	Asian adults with moderate-to-severe SLE with baseline OCS >= 10mg/day	Composite strategy where Death=NR Treatment policy where PEOT/RM=ignored	Difference in proportions between interventions (CMH)	4.2.3.5
Subgroup Analyses of the Key Secondary Endpoint – OCS	Reduction in OCS dose to ≤7.5 mg/day at Week 40, and	Asian adults with moderate-to-	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH) by subgroups	4.2.3.6

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
	maintained through Week 52	severe SLE with baseline OCS >= 10mg/day			
Objective 4: The annualized flare rate through 52 weeks.					
Primary Analysis of the Key Secondary Endpoint – Flares	Annualized flare rate through Week 52	Asian adults with moderate-to-severe SLE	Treatment policy where RM/PEOT/Death = ignored	Ratio of anifrolumab group compared to placebo group in annualized flare rates (Negative binomial regression model) over 52-weeks	4.2.4.4
Sensitivity Analysis of Key Secondary Endpoint – tipping point analysis	Annualized flare rate through Week 52	Asian adults with moderate-to-severe SLE	Treatment policy where RM/PEOT/Death = ignored	Ratio of anifrolumab group compared to placebo group in annualized flare rates (Negative binomial regression model) over 52-weeks	4.2.4.5
Supplementary Analysis of Key Secondary Endpoint	Annualized flare rate through Week 52	Asian adults with moderate-to-severe SLE	While on treatment for PEOT Treatment policy where RM/Death = ignored.	Ratio of anifrolumab group compared to placebo group in annualized flare rates (Negative binomial regression model) over 52-weeks	4.2.4.6
Supportive Analysis of Key Secondary Endpoint	Time to first flare through Week 52	Asian adults with moderate-to-severe SLE	Treatment policy where RM/PEOT/Death = ignored	Estimated Hazard ratio between interventions by Cox PH model	4.2.4.7
Subgroup Analyses of the Secondary Endpoint	Annualized flare rate through Week 52	Asian adults with	Treatment policy where RM/PEOT/Death = ignored	Ratio of anifrolumab group compared to placebo group in annualized flare rates (Negative	4.2.4.8

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
		moderate-to-severe SLE		binomial regression model) over 52-weeks by subgroups	
Objective 5: The proportion of participants with a ≥50% reduction in Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI) activity score at Week 12 in the sub-group of those with baseline CLASI activity score ≥10.					
Primary Analysis of the Secondary Endpoint – Skin Lesions	≥50% reduction in CLASI score	Asian adults with moderate-to-severe SLE with baseline CLASI activity score >= 10	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH), Week 12	4.2.5.4
Supplementary Analysis of the Secondary Endpoint – targeting the impact of burst and taper.	≥50% reduction in CLASI score	Asian adults with moderate-to-severe SLE with baseline CLASI activity score >= 10, excluding subjects administered a burst and taper of OCS or Intramuscular (IM)	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH), Week 12	4.2.5.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
		steroids during the first 12 weeks of treatment			
Supportive Analysis of the Secondary Endpoint - change from baseline in CLASI activity and damage score	Change from baseline in CLASI activity and damage score.	Asian adults with moderate-to-severe SLE	Treatment policy where PRM/PEOT/Death=ignored	Difference in mean change from baseline between interventions (repeated measures model), over time up to Week 52	4.2.5.6
Objective 6: To assess the difference between anifrolumab and placebo on the proportions of participants with reduction in the number of swollen and tender joints at Week 52 in the sub-group of those with ≥ 6 swollen and ≥ 6 tender joints at baseline					
Primary Analysis of the Secondary Endpoint – joint count	50% improvement in joint counts at Week 52	Asian adults with moderate-to-severe SLE with ≥ 6 swollen and ≥ 6 tender joints at baseline	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH)	4.2.6.4
Supportive Analysis of the Secondary Endpoint – Change from baseline in the number of joints	Change from baseline in number of swollen, tender and active joints	Asian adults with moderate-to-severe SLE	Treatment policy where PRM/PEOT/Death=ignored	Difference in mean change from baseline between interventions (repeated measures model), over time up to Week 52	4.2.6.5
Supportive analysis of the Secondary Endpoint – joint count	20% improvement in joint counts at Week 52	Asian adults with moderate-to-severe SLE	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH)	4.2.6.6

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
		with ≥ 6 swollen and ≥ 6 tender joints at baseline			
Supportive analysis of the Secondary Endpoint – joint count	20% and 50% improvement in joint counts at Week 52	Asian adults with moderate-to-severe SLE with ≥ 8 swollen and ≥ 8 tender joints at baseline	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH)	4.2.6.6
Objective 7: To evaluate the population pharmacokinetics (PK), immunogenicity, and pharmacodynamics (PD) of anifrolumab.					
Primary analysis of other Secondary Endpoint - pharmacokinetics, immunogenicity, and pharmacodynamics	Anifrolumab concentration and PK parameters, ADA, 21-gene type I IFN gene signature, dsDNA antibodies, C3, C4, and CH50 levels	Asian adults with moderate-to-severe SLE	Included all data captured during double blind treatment period and follow-up period in analyses regardless of treatment discontinuation, delayed and/or irrespective of protocol adherence	Descriptive statistics	4.2.7 4.2.8 4.2.9
Objective 8: To explore the difference between anifrolumab and placebo on measures of disease activity including levels of SRI response other than 4, BICLA and SRI(4) dual response, the individual components of BICLA and SRI, LLDAS, as well as BICLA and SRI over time					
Primary analysis of exploratory Endpoint - individual components of BICLA and SRI	The individual components of BICLA and SRI	Asian adults with moderate-to-severe SLE	Treatment policy where PRM/PEOT/Death=ignored	Difference in mean change from baseline between interventions; Descriptive statistics	4.2.10

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
response, BILAG-2004, SLEDAI-2K and PGA					
Primary analysis of Exploratory Endpoints - SRI(5), SRI(6), SRI(7), SRI(8), BICLA and SRI(4) dual response, Major Clinical Response, Partial Clinical Response, LLDAS.	Binary endpoints: SRI(5), SRI(6), SRI(7), SRI(8), BICLA and SRI(4) dual response, Major Clinical Response, Partial Clinical Response, LLDAS	Asian adults with moderate-to-severe SLE	Composite strategy where RM/PEOT/Death=NR	Difference in proportions between interventions (CMH), over time up to Week 52	4.2.11 4.2.12 4.2.13 4.2.14
Objective 9: To assess the difference between anifrolumab and placebo on measures of organ damage, i.e., Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI) at Week 52.					
Primary analysis of Exploratory Endpoint – Change in SDI	Change in SDI at Week 52	Asian adults with moderate-to-severe SLE	Treatment policy where PRM/PEOT/Death=ignored	Descriptive statistics	4.2.15
Objective 10: To assess the difference between anifrolumab and placebo on participant-reported health status, and other participant reported outcome measures at Week 52.					
Primary analysis of Exploratory Endpoints – Lupus QoL, EuroQoL 5 dimensions (EQ-5D-5L), Work productivity and activity impairment (WPAI)-Lupus, and Medical Resource Use Questionnaire	Lupus Quality of life (QoL), EuroQoL 5 dimensions (EQ-5D-5L), Work productivity and activity impairment (WPAI)-Lupus, and Medical Resource Use Questionnaire at Week 52	Asian adults with moderate-to-severe SLE	Treatment policy where PRM/PEOT/Death=ignored	Repeated Mean difference between interventions by repeated measures model for Lupus QoL, EQ-5D-5L and WPAI-Lupus; Descriptive statistics for EQ-5D-5L, WPAI-Lupus and medical resource use questionnaire	4.2.16 4.2.17 4.2.18 4.2.19

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Objective 11: To evaluate the safety and tolerability of anifrolumab.					
Primary analysis of Safety Endpoints – Adverse Events, Clinical Laboratory, Electrocardiogram (ECGs), Vital signs, C-SSRS, Cushingoid features	Adverse Events (including adverse event of special interest (AESI))	Asian adults with moderate-to-severe SLE	Includes all data captured during the double-blind treatment period in analyses regardless of treatment discontinuation, delayed and/or irrespective of protocol adherence	Descriptive statistics	4.6.2
	Clinical laboratory parameters (clinical chemistry, hematology and urinalysis)				4.6.3
	Vital signs, and ECG				4.6.4 4.6.5
	C-SSRS				4.6.6
	Cushingoid features				4.6.7

4.2.1 Primary Endpoint

4.2.1.1 Definition

The primary endpoint used to evaluate the effect of anifrolumab compared to placebo on disease activity is the difference in the proportion of participants achieving BICLA response at Week 52, where a participant is a BICLA responder if all the following criteria are met:

- Reduction of all baseline BILAG-2004 A to B/C/D and baseline BILAG-2004 B to C/D, and no BILAG-2004 worsening in other organ systems, as defined by ≥1 new BILAG-2004 A or ≥2 new BILAG-2004 B.
- No worsening from baseline in SLEDAI-2K, where worsening is as defined as an increase from baseline of >0 points in SLEDAI-2K
- No worsening from baseline in participants' lupus disease activity, where worsening is defined by an increase ≥0.30 points on a 3-point PGA VAS

In addition, time to onset of a BICLA response that is sustained up to Week 52 is assessed. A participant is considered to have achieved BICLA response sustained up to Week 52 if a response is achieved at Week 52 with "time to" defined as the first timepoint from Study Day 1 where a BICLA response is achieved and maintained through Week 52.

4.2.1.2 Derivations

The following participants are considered as having met the criteria for BILAG and PGA VAS, respectively.

- Participants with no BILAG A or B at baseline and no worsening in any organ systems where worsening is defined as: ≥1 new BILAG-2004 A or ≥ 2 new BILAG-2004 B,
- Participants with a baseline PGA VAS score > 2.7.

4.2.1.3 Handling of Dropouts and Missing Data

If any of the criteria cannot be evaluated at any visit up to Week 52 (e.g., due to missing values), the data for that criterion is imputed using LOCF and BICLA response is derived based on the complete data. This applies only once. Missing data after LOCF will be imputed as non-responder. For example, if any of the Week 52 criteria data is missing and Week 48 data is not missing, the missing Week 52 criteria data will be imputed using the corresponding Week 48 data. Otherwise, the participant is defined as a BICLA non-responder at Week 52.

4.2.1.4 Primary Analysis of Primary Endpoint

The intercurrent events: PEOT, RM and death are unfavourable outcomes. Therefore, participants with these intercurrent events will be non-responders from the time such events occur.

For the primary analysis, Bayesian inference will be performed for the estimation of the treatment difference on BICLA response at Week 52 in the Asian population. A robustified mixture prior approach will be adopted to partially extrapolate the response rates of anifrolumab 300mg and placebo groups in the pooled study 05 and study 04, to this study. This robustified mixture prior approach closely follows the MAP prior approach as described by Schmidli et al (2014).

Informative prior

First, an informative prior is constructed by pooling results from study 05 and study 04. We assume the number of responders in the placebo and anifrolumab 300mg groups of study 05 and study 04, denoted as $Y_{hk}, h = 1, 2;\ k = 0, 1$, follow binomial distributions,

$$Y_{hk} \sim Bin(N_{hk}, \pi_{hk}),$$

where $h = 1, 2$ represent study 05 and study 04, k=0, 1 represent placebo and anifrolumab 300mg groups, $N_{hk}, k = 0, 1$ represent the number of participants in the placebo and anifrolulmab 300mg groups respectively in study 05 and study 04 respectively, and π_{hk}, k=0, 1 represent the probability of being a responder in placebo and anifrolumab 300mg groups respectively. Let $Y_k, k = 0,1$ be the number of responders and $N_k, k = 0, 1$ be the number of participants in the placebo and anifrolumab 300mg groups in the current study. Therefore Y_k follows a binominal distribution with parameter $\pi_k, k = 0, 1$. The informative prior for π_k, is chosen as beta distribution with parameters $(a_k, b_k), i.e.$

$$\hat{p}_{Hk}(\pi_k | Y_{hk}, h = 1,2) = Beta(a_k, b_k), k = 0, 1$$

where $Beta(a, b)$ is the beta function

$$Beta(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt,$$

and (a_k, b_k) are to be estimated from the results of study 05 and study 04. Given the homogeneous response rates from study 05 and study 04 as shown in [redacted] 8 of the CSP v3.0, we take $(a_0, b_0) = (112, 254)$ and $(a_1, b_1) = (171, 189)$ from the pooled results from the placebo and anifrolumab 300mg groups respectively. The choice of (a_k, b_k) can be interpreted in a way where a_k represents the number of responders and $a_k + b_k$ represents the number of participants in that group.

Robustified mixture prior

We construct a robustified mixture prior for π_k, as

$$\hat{p}_{HRk}(\pi_k) = w * \hat{p}_{Hk}(\pi_k) + (1-w) * Beta(1,1), k = 0,1$$

where $\hat{p}_{Hk}$ is the estimated informative prior, and the second component $Beta(1,1)$ is the noninformative, flat prior. The weight term, w is the prior (initial) weight to place on the source data in the mixture distribution above. It can range between 0-1 and represents the *a priori* degree of relevance the source data has to the target population.

At CDL, the mixture prior distribution $\hat{p}_{HRk}$ is then updated with observed data on the target population, i.e. our current trial data results. Importantly, the prior weight w, is also updated to posterior weight. This accounts for any conflict between the informative prior based on the source data and the observed data. Therefore, in the posterior estimation, the source data can be further discounted if results in target population are observed to be very different than results of source data.

This produces posterior distributions of the response rates in Asian population that borrows from what is already known about the response rates in study 05 and study 04. By the conjugate properties of beta-binomial distribution, the posterior distribution can be estimated as

$$\hat{p}_{HRk}(\pi_k|Y_k) = \tilde{w} * Beta(a_k + Y_k, b_k + N_k - Y_k) + (1-\tilde{w}) * Beta(1 + Y_k, 1 + N_k - Y_k), k = 0,1$$

where $\tilde{w}$ is the updated mixture weight, N_k and Y_k are the numbers of responders after applying the intercurrent event strategy as mentioned in Section 4.2.1.4 The MCMC analysis will be used to estimate the posterior distribution of the parameters in the model.

Analyses

[redacted] prior weight [redacted] will be used for the primary analysis for BICLA response at Week 52. Posterior median estimates for the treatment difference and the corresponding [redacted] credible intervals will be presented. Posterior probabilities of efficacy (treatment difference greater than 0), i.e. $P(\pi_1 - \pi_0 > 0|data)$, will also be estimated and compared with a number of thresholds including [redacted] Results across a range of prior weights will be also presented as for the primary endpoint.

4.2.1.5 Supportive Analysis of the Primary Estimand - Stratified Cochran-Mantel-Haenszel (CMH) approach

This estimand answers a clinically relevant question comparing the proportion of participants able to both complete the study treatment and to achieve adequate response without further medication being required. The summary measure is the difference in the proportion of subjects achieving BICLA response at Week 52 using the FAS, comparing the anifrolumab to the placebo groups.

The null hypothesis is that the proportion of participants achieving BICLA response on anifrolumab is equal to that on placebo. The alternative hypothesis is that the proportion of participants achieving BICLA response on anifrolumab is not equal to that on placebo, i.e.,

H_0: difference in proportion achieving BICLA response (anifrolumab vs Placebo) = 0
H_a: difference in proportion achieving BICLA response (anifrolumab vs Placebo) ≠ 0

The proportion of participants achieving BICLA response in the anifrolumab treatment group will be compared to that in the placebo group using a CMH (Clowse et al, 2017) stratified by:

- SLEDAI-2K score at screening (<10 points versus ≥10 points).
  If different measurements are available for re-screened participants, the value at rescreening will be used.
- Randomization (Day 1) OCS dose (<10 mg/day versus ≥10 mg/day prednisone or equivalent). For the classification, the derived OCS dose rounded to 1 decimal place will be used.
- Region (mainland China versus not mainland China)

SLEDAI-2K score at Screening and Randomization (Day 1) OCS strata will be derived programmatically from the data recorded in the database. Region (mainland China versus non-mainland China) stratum will be taken as recorded at randomization by the Interactive voice/web response system (IXRS). Strata with low counts will be pooled with adjacent strata prior to the analysis. If a substratum within the non-mainland China region stratum has less than 20 participants (in the pooled treatment group), then all non-mainland China region strata will be pooled together. If a sub-stratum within the mainland China region stratum has less than 20 participants (in the pooled treatment group), then within the mainland China region stratum the SLEDAI-2K score <10 points sub-strata will be pooled together and the SLEDAI-2K score ≥10 points sub-strata will be pooled together. If any of these two sub-strata within the mainland China region stratum still has less than 20 participants (in the pooled treatment group), then all China region sub-strata will be pooled together. If all not mainland China region sub-strata are pooled together and all mainland

China region sub-strata are pooled together and any of these strata has less than 20 participants (in the pooled treatment group) then all strata are pooled together.

The analysis can be described as follows:

A. There are n_{ij} participants in each stratum, where i is the stratum, and j is the treatment group. The number of participants achieving BICLA response is x_{ij}. The proportion of participants achieving BICLA response is denoted as $p_{ij} = x_{ij} / n_{ij}$.
B. For each stratum, the difference in proportion of participants achieving BICLA response is calculated as $d_i = p_{iA} – p_{iP}$, where *A* and *P* denote the different treatment groups (anifrolumab and placebo, respectively).
C. Weights for each stratum, w_i, are calculated as $n_{iP} * n_{iA} / (n_{iA} + n_{iP})$. The weighted difference is calculated as:

$$\text{WD} = \frac{\sum w_i d_i}{\sum w_i}$$

D. The standard error (SE) of the weighted difference under the null hypothesis is given by:

$$\text{SE} = \sqrt{\frac{\sum[w_i^2 \, Var\,(d_i)]}{(\sum w_i)^2}}$$

E. For deriving the CI for the weighted difference in proportions, a correction will be applied to the variance, providing a CI with more accurate coverage. This will be applied to all strata and is derived as follows.

$$Var\,(d_i) = \frac{p^*_{iA}\,(1 - p^*_{iA})}{n_{iA}} + \frac{p^*_{ip}\,(1 - p^*_{ip})}{n_{ip}} \text{ with}$$

$$p^*_{ij} = \frac{x_{ij}+2}{n_{ij}+4}$$

The 95% CI can be generated using the weighted difference ± $z_{0.975}$ * SE.

The value of the test statistic is calculated as $\frac{WD}{SE}$. The p-value from the two-sided test of no difference in treatment groups is calculated as $2(1 - \text{Prob}(|\frac{WD}{SE}|))$, where Prob() is the distribution function of the standard normal distribution.

F. The 95% CI for the weighted proportion $\frac{\sum_i w_i p_{ij}}{w}$ in a treatment group j can be generated using a normal approximation and assuming independence between strata, where p_{ij}^* are used as above.

$$S_{ij}^2 = \text{Var}\,(p_{ij}) = \frac{p^*_{ij}\,(1 - p^*_{ij})}{n_{ij}}$$

$$SE_j^2 = \frac{\sum_i w_i^2 \, S_{ij}^2}{w^2}, \; w = \sum_i w_i$$

The 95% CI can be generated using the weighted proportion ± $z_{0.975}$ *SEj.

If the resulting lower or upper limit is <0% or >100%, it will be set to 0% or 100%, respectively.

The estimated treatment effect (i.e., the difference in response rate for anifrolumab versus placebo), corresponding 95% CI, and 2-sided p-value for the difference at Week 52 will be presented. In addition, the estimated response rate (weighted proportion) and the corresponding 95% CI within each treatment group will be presented.

Longitudinal presentations of results over time (i.e., for each post-baseline visit up to Week 52) based on the same CMH analysis, with the corresponding 95% CI, will be created. Line plot will be created accordingly.

4.2.1.6 Supportive Analysis of the Primary Estimand – Bayesian Borrowing based on odds ratio

The proportion of participants achieving BICLA response at Week 52 in the anifrolumab treatment group will be compared to that in the placebo group using a logistic regression model with treatment arm and the stratification factors as covariates. The results of the analyses will be presented using odds ratio (OR), together with associated 95% CI.

Bayesian inference will be performed for the estimation of the OR on BICLA response at Week 52 in the Asian population. A robust mixture prior approach will be applied to partially extrapolate the treatment effect observed in the global study to this study. This approach was reviewed and evaluated for belimumab intravenous infusion in children 5 to 17 years of age with SLE (FDA: BLA 125370/s-064 and BLA 761043/s-007).

The pooled result of Study 04 and Study 05 is used as an informative prior distribution on the OR of BICLA response rate in this study. The log odds ratio log(OR) for BICLA response is assumed normally distributed. Let θ_p be the log(OR) for this study and let $p_a(\theta_p)$ be the informative prior for θ_p. Therefore,

$$p_a(\theta_p) \sim N(\log(2.03), 0.15^2)$$

where 2.03 is the observed OR for BICLA at Week 52 from the pooled global data for anifrolumab versus placebo and 0.15^2 is the variance of the OR represented as the standard error squared.

The robust mixture prior θ_p is constructed, as

$w * p_a(\theta_p) + (1 - w) * N(0, \sigma^2_{NonInf})$ where $N(0, \sigma^2_{NonInf})$ represents the non-informative prior assuming an OR = 1 representing no difference in the treatments resulting

in a mean of 0 [i.e., log(1) = 0] and a standard deviation consistent with information from one participant. The term w reflects the prior amount of weight given to the informative component of the prior and $1 - w$ is the remaining weight that is given to the non-informative component.

Once this data collection is complete, this mixture prior distribution is updated with the observed study results, and inference is conducted on the resulting posterior distribution. Importantly, the prior weight, w, is updated based on the observed data in order to account for any conflict between the informative component of the prior based on the global data and the observed data. Therefore, in the posterior estimation, additional weight can be given to informative component of the prior if Asian are observed to be similar, or the informative component can be discounted if results in Asian are observed to be very different. This method produces a posterior distribution of the treatment effect in Asian that dynamically borrows from what is already known about the treatment effect in the pooled result of Study 04 and Study 05.

Prior weight ranging from $w = 0$ to $w = 100\%$, with a step size of 5%, will be used for the analysis for BICLA response at Week 52. Posterior median estimates for the OR, the corresponding [redacted] credible intervals and posterior probabilities of efficacy (treatment difference greater than 0), i.e. $P(\mathrm{OR} > 1|data)$, will be presented.

4.2.1.7 Supplementary Analyses of the Primary Endpoint

4.2.1.7.1 Supplementary Analysis of the Primary Endpoint – targeting the impact of PEOT due to different reasons

A supplementary analysis for BICLA response is to assess the impact of PEOT due to different reasons by combining composite, treatment policy and hypothetical strategies to handle PEOT due to different reasons. That is, intercurrent events of PEOT due to lack of therapeutic response (after blinded review of the discontinuation reasons, see Section 3.3.7 for details) will result in a non-responder (NR) status per composite strategy. PEOT due to COVID lockdown will be handled per hypothetical strategy. PEOT not due to lack of therapeutic response or COVID lockdown will be ignored per treatment policy. RM and death will still be handled using composite strategy and will be imputed as non-responders from the time such ICEs occur up to week 52. More details on the derivation rules are specified as below.

For subjects with intercurrent events of RM, death and PEOT due to lack of therapeutic response, they will be imputed as non-responders from the time such ICEs occur up to Week 52. For subjects with PEOT due to COVID lockdown, data after PEOT will not be used and will be imputed. Intermediate and terminal missing values for BILAG-2004, SLEDAI-2K and PGA will be imputed separately for each BICLA component using multiple imputation under the assumption of missing at random. Missing values of BICLA

are imputed based on the imputed values of the BILAG-2004, PGA and SLEDAI-2K components.

For each outcome and visit, 100 imputations are generated by randomized treatment group. The procedure is initiated with a seed of 12345. For analysis, each imputed dataset is analyzed separately using the CMH approach described in Section [redacted] 4.2.1.5, and the single estimates is combined using PROC MIANALYZE. Each component is imputed as follows:

- BILAG-2004 is imputed as a binary variable reflecting the BICLA criterion, i.e., "Reduction of all baseline BILAG-2004 A to B/C/D and baseline BILAG-2004 B to C/D, and no BILAG-2004 worsening in other organ systems, as defined by ≥1 new BILAG-2004 A or ≥2 new BILAG-2004 B". This binary BILAG variable for participant i at visit t is modelled as $BILAG_{i(t)}$~Binomial (1, $\pi_{i(t)}$) where $\pi_{i(t)}$= $\theta_1 BILAG_{i(t-1)} + \theta_0(1- BILAG_{i(t-1)})$ for t>1 and $\pi_{i(t)}= \theta_2$ for t = 1 (Week 4). Independent *Beta*(1,1) priors are specified for θ_0, θ_1 and θ_2. 10,000 burn-in iterations are used, followed by 10,000 main iterations. Imputations are then taken from every 100th iteration of the main chain (i.e., after burn-in).

- SLEDAI-2K total score is imputed with PROC MI using the MCMC IMPUTE=FULL specification, and a VAR statement specifying the variables in order of visit. Specify MINIMUM=0 to ensure imputed values are non-negative. Specify MAXIMUM=105 to avoid imputation of values above the maximum possible score. 10,000 burn-in iterations to be used, with 100 iterations between each imputation.

- PGA will be imputed in the same way as SLEDAI-2K total score but using MAXIMUM=3 (instead of MAXIMUM=105).

The estimated treatment effect, corresponding 95% CI, and 2-sided p-value for the difference at Week 52 will be provided using the same method as presented in Section **错误!未找到引用源。**.

4.2.1.7.2 Supplementary Analysis of the Primary Endpoint – targeting the impact of PEOT

A supplementary analysis for BICLA response at week 52, using the CMH approach described in Section 4.2.1.5, is repeated removing the criterion of PEOT. Therefore, PEOT does not result in a NR status, RM is retained as part of the composite endpoint definition and thus results in a NR status. Death is imputed as NR from the time such events occur up to Week 52.

The estimated treatment effect, corresponding 95% CI, and 2-sided p-value for the difference at Week 52 will be provided using the same method as presented in Section 4.2.1.5.

4.2.1.7.3 Supplementary Analysis of the Primary Endpoint – targeting the impact of different discontinuation rates; based on tipping point analysis

In order to examine the impact of different discontinuation rates between treatment groups, a tipping point analysis will be performed if p-value for the primary analysis is <= 0.05. This analysis will vary the assumptions about outcomes among the subsets of participants on the treatment arms who prematurely discontinue IP and can be described as:

- The proportions of participants achieving BICLA response will be analyzed using a Pearson's chi-squared test, thus the stratification factors that are used in the main analysis (CMH) will be disregarded. Since the strata will be balanced in respect of treatment assignment by virtue of the stratified randomization scheme used, the only impact of this simplification should be that the inferences from the unstratified analysis will be somewhat conservative.

- For the primary analysis, participants who prematurely discontinue IP having not received restricted medications prior to discontinuation of IP are by definition, imputed as non-responders. For this analysis, these participants will be altered from non-responder to responder in an iterative manner.

- At each step of the analysis one of these participants switches from not achieving BICLA response to achieving BICLA response, and the Pearson's chi-squared test is re-run. The results (statistical significance) are presented in a grid where the x-axis and the y-axis represent the added number of participants assumed to be achieving BICLA response for placebo and anifrolumab 300 mg, respectively. The region where the conclusion changes, will be considered as the tipping point.

4.2.1.7.4 Supplementary Analysis of the Primary Endpoint - targeting the impact of RM

A supplementary analysis for BICLA response, using the CMH approach described in Section 4.2.1.5, is repeated removing the criterion of RM. Therefore, this analysis does not consider participants as NR if they have used restricted medications. The PEOT is considered as NR and Death is imputed as NR from the time such events occur up to Week 52.

The estimated treatment effect, corresponding 95% CI, and 2-sided p-value for the difference at Week 52 will be provided using the same method as presented in Section 4.2.1.5.

4.2.1.7.5 Supplementary Analysis of the Primary Endpoint - targeting the impact of PEOT and RM

A supplementary analysis for BICLA response, using the CMH approach described in Section 4.2.1.5, is repeated removing both the criterion PEOT and RM. Therefore, PEOT and/or RM do not result in a NR status. Death is imputed as NR from the time such events occur up to Week 52.

The estimated treatment effect, corresponding 95% CI, and 2-sided p-value for the difference at Week 52 will be provided using the same method as presented in Section 4.2.1.5.

4.2.1.7.6 Supplementary Analysis of the Primary Endpoint - targeting the impact of being automatically classified as responder due to their baseline VAS and/or BILAG scores

A supplementary analysis using the CMH approach described in Section 4.2.1.5 is performed excluding from the analysis those subjects with no BILAG A or B at baseline, as well as participants with a baseline PGA VAS score > 2.7.

The estimated treatment effect, corresponding 95% CI, and 2-sided p-value for the difference at Week 52 will be provided using the same method as presented in Section 4.2.1.5

4.2.1.7.7 Supplementary Analysis of the Primary Endpoint – Time to BICLA response

Time to BICLA response sustained up to Week 52 is defined as the visit of first BICLA response which is sustained up to, and including, Week 52. Subjects without a BICLA response sustained up to Week 52 are censored at the date of last available BICLA assessment up to and including Week 52 during the double-blind treatment period. BICLA response is defined in Section 4.2.1.1.

The PEOT, RM and death are intercurrent events. Subjects with PEOT, RM and/or death will be considered NR (not achieving a sustained BICLA response) from the date they prematurely end treatment, take restricted medications or death respectively, through to and including Week 52. Subjects who are WDS status are handled by censoring the subject from the date of the WDS.

The supportive outcome of time to BICLA response sustained up to Week 52 is analyzed using a Cox proportional hazard models (using a profile likelihood approach with ties=Efron) including the covariates of treatment and the stratification factors. The estimated hazard ratios and corresponding CIs will be presented for the effect of the treatment group. Furthermore, the time to BICLA response sustained up to, and including, Week 52 will be presented as Kaplan-Meier plot including the number of participants at risk at each visit.

4.2.1.8 Subgroup Analyses

To explore the uniformity of the detected overall treatment effect on the primary endpoint (i.e., only for the primary estimand excluding the supplementary and sensitivity analyses), subgroup analyses may be performed for the following factors:

- Region (mainland China versus not mainland China)
- SLEDAI-2K score at screening (<10 points versus ≥10 points)
- OCS dose at baseline (<10 mg/day versus ≥10 mg/day prednisone or equivalent)
- Sex (male versus female)
- Age (≥18 to <65 versus ≥65 years)
- Onset of disease (adult versus pediatric onset)
- BMI (⩽28 versus >28 kg/m^2)
- IFN 4-gene test (high versus low)
- ADA Result (positive at any time, negative, persistently positive, positive with titer >median of maximum titer)
- Baseline anti-dsDNA, C3 and C4 (at least one positive/abnormal, all negative/normal)

Treatment effect is estimated by the stratified CMH model. A forest plot reporting response rates, estimated differences with its 95% CI for treatment derived by the stratified CMH model is presented overall and by subgroups.

4.2.2 Key Secondary Endpoint - SRI (4)

4.2.2.1 Definition

The key secondary efficacy endpoint, the assessment of disease activity will be evaluated based on the difference in proportions between anifrolumab and placebo in SRI(4) response at Week 52. SRI(4) is defined as a participant meeting all the following criteria:

- Reduction from baseline of ≥4 points in the SLEDAI-2K will be derived as the sum of the scores for all items. A ≥4 points reduction is reached if the change from baseline is ≤-4.
- No new organ systems affected as defined by 1 or more new BILAG-2004 A or 2 or more new BILAG-2004 B items compared to baseline
- No worsening from baseline in participants' lupus disease activity, where worsening is defined by an increase ≥0.30 points on a 3-point PGA VAS

4.2.2.2 Derivations

Not applicable.

4.2.2.3 Handling of Dropouts and Missing Data

If any of the criteria mentioned in the Section 4.2.2.1 cannot be evaluated at any visit up to Week 52 (e.g., due to missing values), the data for that criterion will be imputed using LOCF and SRI(4) response is derived based on the complete data. This applies only once. Missing data after LOCF will be imputed as non-responder. For example, if any of the Week 52 criteria data is missing and Week 48 data is not missing, the missing Week 52 criteria data will be imputed using the corresponding Week 48 data. Otherwise, the participant is defined as a SRI(4) non-responder at Week 52.

4.2.2.4 Primary Analysis of Key Secondary Endpoint - SRI (4)

The same CMH approach as described in Section 4.2.1.5 for the primary endpoint is used to estimate the treatment difference between anifrolumab and placebo as well as the response rates. Nominal p-values will be presented for Week 52 only. Using the same CMH approach, longitudinal presentations of the results over time (i.e., for each post-baseline visit up to Week 52 during double blind treatment period) is created.

The estimated treatment effect, corresponding 95% CI, and 2-sided p-value for the difference at Week 52 will be provided using the same method as presented in Section 4.2.1.5.

4.2.2.5 Supplementary Analyses of Key Secondary Endpoint

A supplementary analysis to assess the impact of PEOT due to different reasons by combining composite, treatment policy and hypothetical strategies to handle PEOT due to different reasons is performed using the similar method as described in Section 4.2.1.7.1, where component of BILAG-2004 is imputed as a binary variable reflecting the SRI (4) criterion, i.e., “No new organ systems affected as defined by 1 or more new BILAG-2004 A or 2 or more new BILAG-2004 B items compared to baseline”.

Supplementary analyses targeting the impact of intercurrent invents: PEOT and/or RM are performed following the strategy discussed in Section 4.2.1.7.2, 4.2.1.7.4 and 4.2.1.7.5.

An analysis to examine the impact of different discontinuation rates based on tipping point analysis is performed following the strategy discussed in Section 4.2.1.7.3 by replacing BICLA response by SRI(4) response, if the nominal p-value is <= 0.05 for the primary analysis of SRI(4).

4.2.2.6 Subgroup Analyses

The subgroup analyses will be performed for the following stratification factors, i.e.

- Region (mainland China versus not mainland China)
- SLEDAI-2K score at screening (<10 points versus ≥10 points)
- OCS dose at baseline (<10 mg/day versus ≥10 mg/day prednisone or equivalent).

The analysis methods are the same as described above for SRI (4) endpoint in Section 4.2.2.4.

4.2.3 Key Secondary Endpoint - OCS

4.2.3.1 Definition

The effect of anifrolumab versus placebo on the ability to reduce the OCS dose in participants with baseline OCS ≥10 mg/day prednisone or equivalent will be evaluated. A maintained OCS reduction is defined as meeting all the following criteria:

- Achieve an OCS dose of ≤7.5 mg/day prednisone or equivalent by Week 40.
- Maintain an OCS dose ≤7.5 mg/day prednisone or equivalent from Week 40 to Week 52

  A maintained OCS dose is defined as no dose increase (i.e., no dose greater than the dose at Week 40 + 1 day) between Week 40 + 2 days and Week 52, inclusive. The date used for Week 52 is the date of last assessment used for efficacy analysis (SLEDAI-2K, PGA and BILAG) in the time window of Week 52 (as described in Section 3.3.2). If no such assessment falls into the respective time window, then the target date for the timepoint will be used instead.

The derivation of the prednisone equivalent daily dose is given in Appendix 7.4.

4.2.3.2 Derivations

The dose of OCS at Week 40 used in the evaluation of this endpoint is the dose on the date of Week 40 + 1 day.

The date used for Week 40 is the date of last assessment used for efficacy analysis (SLEDAI-2K, PGA and BILAG) in the time window of Week 40 (as described in Section 3.3.2).

4.2.3.3 Handling of Dropouts and Missing Data

If any of the conditions cannot be evaluated at Week 52 (e.g., due to missing values) the participant is considered as not reaching a maintained OCS reduction.

If there is no assessment in Week 40 window, then the target date for the timepoint will be used instead.

4.2.3.4 Primary Analysis of Key Secondary Endpoint - OCS

The same CMH approach as described in Section 4.2.1.5 for the primary endpoint will be used to estimate the treatment difference between anifrolumab and placebo in subjects with baseline OCS ≥10 mg/day prednisone or equivalent. Nominal p-value is presented for Week 52 only. For the stratified CMH analyses, the stratification factors will be reduced to SLEDAI-2K score at screening (<10 points versus ≥10 points) and region (mainland China versus not mainland China).

4.2.3.5 Supplementary Analyses of the Key Secondary Endpoint

Supplementary analyses targeting the impact of intercurrent invents: PEOT and/or RM are performed following the strategy discussed in Section 4.2.1.7.2, 4.2.1.7.4 and 4.2.1.7.5.

Another supplementary analysis to examine the impact of different discontinuation rates based on tipping point analysis is performed following the strategy discussed in Section 4.2.1.7.3 by replacing BICLA response by maintained OCS reduction, if the nominal p-value is <= 0.05 for primary analysis of maintained OCS reduction.

For subjects with baseline OCS ≥10 mg/day prednisone or equivalent, a shift table of reaching a maintained OCS reduction at Week 52 (as defined in Section 4.2.3.1) versus achieving BICLA response at Week 52 (as defined in Section 4.2.1.1) will be provided.

4.2.3.6 Subgroup Analyses

The subgroup analyses will be performed for the following stratification factors excluding the subgroup OCS dose at baseline (< 10 mg/day versus ≥ 10 mg/day prednisone or equivalent), i.e.

- Region (mainland China versus not mainland China)
- SLEDAI-2K score at screening (<10 points versus ≥10 points).

The analysis methods are the same as described above for the maintained OCS reduction endpoint in Section 4.2.3.4.

4.2.4 Key Secondary Endpoint - Flares

4.2.4.1 Definition

The key secondary endpoint used to evaluate the effect of anifrolumab versus placebo on flares is the difference in annualized flare rate through Week 52.

A flare is defined as either 1 or more new BILAG-2004 A or 2 or more new BILAG-2004 B items compared to the previous visit (i.e., a worsening from an E, D, or C score to a B score in at least 2 organ systems or a worsening from an E, D, C, or B score to an A score in any one organ system compared to the previous visit).

4.2.4.2 Derivations

The occurrence of a new flare will be checked for each available visit versus the previously available visit up to Week 52. If no flare occurred, the number of flares will be set to 0. Otherwise, all flares will be counted leading to a maximum number of flares of 13.

The annualized flare rate is calculated as the number of flares divided by the flare exposure time in days multiplied with 365.25. The flare exposure time is the time up to the date of last available BILAG-2004 assessment up to and including Week 52 during the double-blind treatment period. It is derived as the date of last BILAG-2004 assessment minus date of first administration of IP + 1.

For the supplementary analyses of flares while on treatment, the flare exposure time is the time up to the date of last dose of IP + 28 days. It is derived as date of last dose of IP + 28 days - date of first dose of IP + 1. All flares occurring within the double-blind treatment period and the flare exposure time are considered for this analysis.

4.2.4.3 Handling of Dropouts and Missing Data

Participants who prematurely discontinue from treatment are analyzed according to the date of last available BILAG-2004 assessment up to and including Week 52, except for Section 4.2.4.6 on-treatment analysis. Missing post-baseline BILAG-2004 assessments are imputed using LOCF. This applies only once. Missing post-baseline assessments after LOCF are counted as no flare in annualized flare rate analysis.

4.2.4.4 Primary Analysis of Key Secondary Endpoint – Flares

The flare rate in the anifrolumab treatment group is compared to the flare rate in the placebo group using a negative binomial regression model. The response variable in the model is the number of flares over the 52-week treatment period. The model includes covariates of treatment group, and the stratification factors. The logarithm (to base e) of the follow-up time (flare exposure time as defined in Section 4.2.4.2) is used as an offset variable in the model to adjust for participants having different exposure times. If there are less than 10 flare events in total for a subgroup based on stratification factor or less than 2 events in either treatment group for that subgroup, then the corresponding stratification factor will be removed in the model. The estimated treatment effect and the corresponding 95% CI, as well as the 2-sided nominal p-value are presented.

4.2.4.5 Sensitivity Analysis of Key Secondary Endpoint – tipping point analysis

Firstly, a missing at random (MAR) analysis will be performed where for each participant the rate after withdrawal λ_1 is assumed to be the same as their rate before withdrawal λ_2, which itself is calculated based on their randomized treatment group and baseline covariates included in the negative binomial regression model as section 4.2.4.4.

A tipping point analysis will then be performed if the p-value for the primary analysis is <= 0.05, where the rate after withdrawal will be modified to $\delta\lambda_2$. A series of analyses will be performed with a range of increasing deltas for the two arms (δ_P and δ_A for placebo and anifrolumab 300 mg groups, respectively) so that one could assess at which point the study conclusions would change from favorable to unfavorable, i.e., to identify a tipping point.

In this assessment, the placebo group is assumed to improve after withdrawal and the anifrolumab group is assumed to worsen after withdrawal. Therefore, $\log(\delta_P)$ will be varied from -1.5 to 0 in increments of 0.5 and $\log(\delta_A)$ will be varied from 0 to 1.5 in increments of 0.5. This corresponds to deltas between 0.22 and 1 for the placebo group and deltas between 1 and 4.5 for the anifrolumab 300 mg group. If statistical significance is maintained among the matrix of possible δ combinations, the comparison is deemed robust to missing data. For a given comparison, if a tipping point is observed with analysis at 0.5 increments, the δ values will be further refined down to 0.25 increments for the relevant interval. For example, if a tipping point is identified when increasing $\log(\delta_A)$ from 1 to 1.5, the matrix will be expanded to include also the value $\log(\delta_A) = 1.25$. The values for δ (and the corresponding increments) will be checked during the dry run and adapted as necessary.

4.2.4.6 Supplementary Analysis of the Key Secondary Endpoint - Flares while on treatment – targeting the impact of PEOT

The flare rate in the anifrolumab treatment group will be compared to the flare rate in the placebo group using a negative binomial regression model as section 4.2.4.4. The response variable in the model is the number of flares while on treatment (i.e., up to last administration of IP + 28 days during the double-blind treatment period). The model includes covariates of treatment group, and the stratification factors. The logarithm (to base e) of the follow-up time (flare exposure time as defined in Section 4.2.4.2) is used as an offset variable in the model to adjust for participants having different exposure times. The estimated treatment effect and the corresponding 95% CI are presented.

A summary of the annualized flare rate by descriptive statistics as well as a summary of the number and percentage of participants with no flares, at least one flare, 1 flare, 2 flares, and 3 or more flares, respectively, are presented by treatment group.

4.2.4.7 Supportive Analysis of the Key Secondary Endpoint

Time to first flare is assessed for flares as defined in Section 4.2.4.1 . The time to first flare is derived as date of first flare minus date of first administration of IP. If the participant did not have a flare, the time to flare is censored at the end of the flare exposure time (as defined in Section 4.2.4.2).

Cox proportional hazards models, as reported in Section 4.2.1.7.7, (using a profile likelihood approach with ties=Efron) including the covariates of treatment and the

stratification factors are used to estimate the treatment effect. The estimated HRs and corresponding 95% CIs are presented for the effect of the treatment group.

Time to first flare distribution is compared between anifrolumab and placebo using a stratified log-rank test adjusting for the same randomization stratification factors used in the stratified Cox model.

4.2.4.8 Subgroup Analyses

The subgroup analyses will be performed for the following stratification factors, i.e.

- Region (mainland China versus not mainland China)
- SLEDAI-2K score at screening (<10 points versus ≥10 points)
- OCS dose at baseline (<10 mg/day versus ≥10 mg/day prednisone or equivalent).

The analysis methods are the same as described above for the flare rate endpoint in Section 4.2.4.4.

4.2.5 Other Secondary Endpoint - Skin Lesions

4.2.5.1 Definition

The secondary endpoint used to evaluate the effect of anifrolumab versus placebo on inflammatory cutaneous lupus lesions in participants with baseline CLASI activity score ≥10 is the difference in the proportion of participants with ≥50% reduction in CLASI activity score at Week 12. An at least 50% reduction in CLASI activity score is defined as meeting all of the following criteria:

- Achieve ≥50% reduction of CLASI activity score at Week 12 compared to baseline.

4.2.5.2 Derivations

The CLASI activity score is defined in Section 4.1.6.1. A ≥ 50% reduction is reached if the percentage change is ≤ -50%.

4.2.5.3 Handling of Dropouts and Missing Data

If any of the criteria cannot be evaluated at any visit up to Week 12 (e.g., due to missing values), the data for that criterion will be imputed using LOCF and at least 50% reduction in CLASI activity score derived based on the complete data. This applies only once. Missing data after the imputed LOCF value will be imputed as non-responder. For example, if any of the Week 12 criteria data is missing and Week 8 data is not missing, the missing Week 12 criteria data will be imputed using the corresponding Week 8 data. Otherwise, the participant will be defined as not reaching a ≥50% reduction in CLASI activity score at Week 12.

LOCF records are not used in the repeated measures models.

4.2.5.4 Primary Analysis of the Secondary Endpoint – Skin Lesions

The analysis includes only subjects with baseline CLASI activity score ≥ 10.

The same CMH approach as described in Section 4.2.1.5 for the primary endpoint will be used to estimate the treatment difference between anifrolumab and placebo in subjects with baseline CLASI activity score ≥10. Longitudinal presentations of the results over time (i.e., for each post-baseline visit up to Week 52 during double blind treatment period) based on this analysis will be created.

4.2.5.5 Supplementary Analysis of the Secondary Endpoint – targeting the impact of burst and taper

A supplementary analysis is provided if at least 10 participants in a treatment arm have a burst and taper of OCS or IM (intramuscular) steroids during the first 12 weeks of treatment. The CMH analysis will be repeated for the at least 50% reduction in CLASI activity score (including all criteria) at Week 12 excluding subjects administered a burst and taper of OCS or IM steroids during the first 12 weeks of treatment. A burst and taper of OCS or IM steroids is defined as an OCS increase above the daily dose at Day 1 or any IM steroid dose.

4.2.5.6 Supportive Analysis of the Secondary Endpoint- change from baseline in CLASI activity and damage score

Change from baseline in CLASI activity score and CLASI damage score, respectively, are analyzed using the repeated measures models with fixed effects for baseline value, treatment group, visit, treatment*visit interaction and stratification factors, and participant as random effect. Covariance parameters will be estimated using restricted Maximum Likelihood method and Kenward Rogers denominator degrees of freedom will be used for the tests of fixed effects. An unstructured covariance matrix will be used. In case of convergence issues, the following alternative structures are used for fitting (in that order): Heterogeneous Toeplitz, heterogeneous AR(1) (1st order autoregressive), Heterogeneous CS (compound symmetry), Homogeneous CS. Results will be presented in terms of the adjusted means for each treatment group, estimates of treatment differences, and associated 2-sided CIs for each visit up to Week 52 during double blind treatment period.

4.2.6 Other Secondary Endpoint – Joint count

4.2.6.1 Definition

The endpoints used to evaluate the effect of anifrolumab versus placebo on joints are:

- Difference in the proportion of participants with at least 6 swollen and at least 6 tender joints at baseline who achieve at least a 50% reduction from baseline in both the number of swollen and tender joints at Week 52;

A reduction of at least 50% is reached if all the following criteria are met:

- The percentage reduction from baseline in both the number of swollen joints and the number of tender joints, separately, is ≥ 50%;

An active joint is defined as a joint with both swelling and tenderness.

4.2.6.2 Derivations

Not applicable.

4.2.6.3 Handling of Dropouts and Missing Data

If the change from baseline in the number of swollen and tender joints cannot be evaluated at any visit up to Week 52 (e.g., due to missing values), the data for the change from baseline is imputed using LOCF. This applies only once. Missing data after LOCF will be imputed as non-responder. For example, if any of the Week 52 criteria data is missing and Week 48 data is not missing, the missing Week 52 criteria data will be imputed using the corresponding Week 48 data. Otherwise, the participant will be defined as non-responder.

LOCF records are not included in the repeated measures models.

4.2.6.4 Primary Analysis of the Secondary Endpoint – Joint count

For participants with at least 6 swollen and at least 6 tender joints at baseline, the same CMH approach as described in Section 4.2.1.5 for the primary endpoint is used to estimate the treatment difference between anifrolumab and placebo in the proportions of participants achieving an at least 50% reduction in both swollen and tender joints, respectively. Longitudinal presentations of the results over time (i.e., for each post-baseline visit up to Week 52 during double blind treatment period) based on this analysis are created. 95% CIs will be presented.

4.2.6.5 Supportive Analysis of the Secondary Endpoint – Change from baseline in the number of joints

The change from baseline in the number of active, swollen and tender joints, respectively, is analyzed using the repeated measures models with fixed effects for baseline value, treatment group, visit, and treatment*visit interaction stratification factors, and participant as random effect. Covariance parameters will be estimated using restricted Maximum Likelihood method and Kenward Rogers denominator degrees of freedom will be used for the tests of fixed effects. Thereby, an unstructured covariance matrix is used. In case of

convergence issues, the following alternative structures are used for fitting (in that order): Heterogeneous Toeplitz, heterogeneous AR(1), Heterogeneous CS, Homogeneous CS.

4.2.6.6 Supportive Analysis of the Secondary Endpoint – Joint count

As a supportive analysis, the following endpoints will also be explored using the same approach as described in 4.2.6.4:

- Difference in the proportion of participants with at least 6 swollen and at least 6 tender joints at baseline who achieve at least a 20% reduction from baseline in both the number of swollen and tender joints at Week 52;
- Difference in the proportion of participants with at least 8 swollen and at least 8 tender joints at baseline who achieve at least a 20% reduction and at least a 50% reduction from baseline in both the number of swollen and tender joints at Week 52.

4.2.7 Other Secondary Endpoint – Pharmacokinetics

4.2.7.1 Definition

Due to the limited sampling schedule, the PK assessment will be primarily based on the observed steady-state serum trough (pre-dose) concentrations, C-trough. Maximum concentrations after the first dose and the dose at Week 48 will also be evaluated. Analyses will be done by results of the baseline IFN 4-gene test (high versus low).

Individual concentrations will be reviewed for exclusion from descriptive statistics by identifying outliers and reviewing dosing information and sample collection times. Analysis to determine if the identified concentrations should be excluded includes visual inspection of PK-time profiles and comparison of descriptive stats with identified concentrations excluded and included.

4.2.7.2 Derivations

Not applicable.

4.2.7.3 Handling of Dropouts and Missing Data

Anifrolumab serum concentrations reported as below the LLOQ (lower limit of quantitation) are marked as below the limit of quantification (BLQ) and be imputed with LLOQ/2 for analysis.

4.2.7.4 Primary Analysis of Other Secondary Endpoint – Pharmacokinetics

If not stated otherwise, all analyses of PK data is performed for the PK analysis set by IFN 4-gene test status (high/low/overall).

Anifrolumab serum concentrations are summarized using descriptive statistics by IFN 4-gene test status, visit and time point (pre-dose/post-dose) reporting n (number of non-

missing values), nBLQ (number of BLQ values), geometric mean, geometric CV%, arithmetic mean, SD, arithmetic CV%, median, minimum and maximum. If applicable, this summary is repeated including individual concentrations excluded from descriptive statistics.

Serum concentration-time profiles of anifrolumab are generated as plots of mean values (including SD) by IFN 4-gene test status over time points in a semi-log scale and a linear scale. This is presented for all values and additionally for values measured before administration (C_{trough}).

Individual concentrations excluded from descriptive statistics are listed.

Impact of Anti-drug antibodies (ADA) on PK will be explored. Anifrolumab serum concentration is summarized by ADA result (positive at any time (i.e., prevalence)/negative) and treatment group. Serum concentration-time profiles of anifrolumab is further generated by ADA result (positive at any time /negative) in a linear scale; ADA positive group presents each serum concentration-time profile while ADA negative presents mean values (including SD). Note ADA negative is defined as no positive ADA results during the study (neither at baseline nor post-baseline).

4.2.7.5 Supplementary Analyses of the Secondary Endpoint

Not applicable.

4.2.8 Other Secondary Endpoint – Immunogenicity

4.2.8.1 Definition

ADA assessments will be conducted utilizing a tiered approach (screen, confirm, titre). The following ADA results will be evaluated as proportion of participants together with corresponding titre summaries.

- ADA prevalence, defined as participants who are ADA positive at any time (including baseline)
  Note that further use of ADA positive at any time include baseline measurements even if not mentioned explicitly.

- Participants who are ADA positive at baseline only
  participants without any post-baseline measurements are also counted in this category.

- Participants who are ADA positive at both baseline and post-baseline.
  Only participants with baseline and post-baseline measurements will be counted in this category.

- ADA incidence, defined as ADA positive post-baseline (ADA negative at baseline) or a post-baseline increase in pre-existing baseline ADA titres by ≥4-fold during the study period (boost).
- Participants who are ADA positive post-baseline only (treatment induced ADA)
  Participants without a baseline measurement are also counted in this category.
- Participants who are persistently positive
  Persistently positive is defined as treatment induced ADA (participant is ADA negative at baseline) detected at ≥ 2 post-baseline assessments with at least 16 weeks (112 days) between the first and last positive measurement or a treatment induced ADA detected at the last available assessment.
- Participants who are transiently positive
  Transiently positive is defined as at least one treatment induced (participant is ADA negative at baseline) ADA positive measurement, but not fulfilling the conditions for persistently positive.
- Participants with an ADA positive titre >median of maximum titre
  Median of maximum titre is defined as the median of the maximum post-baseline titre (in participants with treatment-induced ADA) for each participant in the treatment group. Baseline ADA titers are not included in this assessment.
- Participants who are ADA positive by visit
  If a participant has any ADA positive result during a visit window, they are considered ADA positive for that visit.

For the presentation of ADA results at a single time point (e.g., baseline or by visit summaries), the corresponding titre summary will be based on the titres of the positive samples for that particular visit. If a participant has multiple ADA positive results during a visit window, the maximum titre of all the positive samples within this visit window will be used for titre summary at that particular visit. For the presentation of ADA results across visits (e.g., any post-baseline), the corresponding titre summaries will be based on the maximum titre of all positive samples for each participant. Titres of positive measurements reported as ≤ x (limit of detection) will be imputed as x. Titre values reported as "< x" are negative and will not be imputed.

The presence of neutralizing antibodies (nAb) will be tested in all ADA-positive samples using a ligand binding assay. The following variables will be evaluated:

- nAb prevalence: Proportion of participants with ADA results available (at any time) who are nAb positive at any time
- nAb incidence: Proportion of participants with ADA results available at post-baseline timepoints who are nAb positive at post-baseline timepoints only (i.e., nAb negative at baseline or ADA negative at baseline).

- nAb persistently positive: Proportion of participants with ADA results available at postbaseline timepoints who are nAb positive at 2 or more post-baseline assessments (with at least 16 weeks between first and last positive) or positive at the last postbaseline assessment (for participants who are nAb negative at baseline or ADA negative at baseline).
- Proportion of participants who are nAb positive by visit

4.2.8.2 Derivations

Not applicable.

4.2.8.3 Handling of Dropouts and Missing Data

Not applicable.

4.2.8.4 Primary Analysis of Other Secondary Endpoint – immunogenicity

If not stated otherwise, all analyses of ADA are performed for the safety analysis set by IFN 4-gene test status (high/low/overall).

The proportion of participants with ADA results as described in Section 4.2.9.1 will be presented. Summary statistics will be provided for titer of ADA positive participants by timepoint. Descriptive statistics will include n (number of reportable titers), minimum, quartiles (1st quartile, median, 3rd quartile) and maximum. For the summary of participants who are ADA positive by visit or at a single visit point, the denominator for percentages is the number of participants with an adequate sample that provides a reportable result of positive or negative for the respective visit. For the summary of overall ADA categories (e.g., ADA positive at any time), percentages will be based on participants with at least one ADA result during the study.

The percentage of participants with positive ADA results by timepoint will be presented graphically in a line plot. ADA titres-time profiles (line plots of median and quartiles by time point) will be generated for the anifrolumab treatment group.

Variables of proportions regarding neutralizing antibodies will be summarized presenting the number and percentage of the categories given in Section 4.2.9.1.

Key subject information for participants with an ADA positive result at any time and for participants with ADA incidence and had ADA titres increase by at least 4-fold from onset to maximum titre will be provided.

The analyses of impact of ADA on PK were described in section 4.2.7.4 and on PD endpoints are described in section 4.2.9.44.2.9.4. Potential association of ADA with safety and efficacy may be explored by subgroup analyses. Endpoint wise subgroup analyses for efficacy and safety may be performed.

4.2.8.5 Supplementary Analyses of the Secondary Endpoint

Not applicable.

4.2.9 Other Secondary Endpoint – Pharmacodynamics

4.2.9.1 Definition

The outcome variable for C3, C4, and CH50 complement levels are the mean change from baseline longitudinally up to Week 52 in participants with abnormal complement level at baseline. An abnormal complement level is defined as complement level below lower limit of normal.

The outcome variable for anti-dsDNA antibodies will be proportion of participants with positive, negative and missing anti-dsDNA test results.

The IFN 21-gene Pharmacodynamic (PD) signature is calculated as the mean expression level of 21 IFN-inducible genes in SLE participants compared to the expression level in control samples from healthy individuals. The IFN 21-gene PD signature following treatment is compared to the IFN 21-gene signature at baseline for each participant at multiple timepoints after treatment. The "percent suppression of fold change" is presented as the "percent of baseline IFN 21-gene PD signature" in all summaries, figures and listings.

Percent of baseline IFN 21-gene PD signature is derived by central lab and transferred back to AstraZeneca after the primary and the final CDL. The derivation follows the description in Appendix 7.1.

4.2.9.2 Derivations

Not applicable.

4.2.9.3 Handling of Dropouts and Missing Data

Not applicable.

4.2.9.4 Primary Analysis of Other Secondary Endpoint – Pharmacodynamics

If not otherwise stated, all analyses of pharmacodynamics are performed for the FAS by treatment group. [redacted]

For participants who are abnormal complement level at baseline, observed values and changes from baseline of C3, C4, and CH50 complement levels will be summarized by visit.

For anti-dsDNA, the number and percentage of participants with result of negative, positive, and missing is presented as shift tables from baseline to each post-baseline visit.

For baseline IFN 4-gene test high participants, IFN 21-gene PD signature as percent of baseline IFN21 PD signature is summarized by time point. The median absolute deviation (MAD) is also included in this summary. Furthermore, the median percent of baseline IFN21 PD signature (including MAD) over time is presented as a line plot for the same population.

The impact of ADA on IFN 21-gene PD signature is explored. For baseline IFN 4-gene test high participants, line plots of IFN 21-gene PD signatures as percent of baseline IFN21 PD signature is generated by ADA category (ADA negative, ADA positive, ADA persistently positive, ADA titre > median, ADA nAb positive). IFN 21-gene PD signatures as percent of baseline IFN21 PD signature is further be summarized descriptively by ADA category (ADA negative/positive at any time), timepoint and treatment group for the same population.

4.2.9.5 Supplementary Analyses of the Secondary Endpoint

Not applicable.

4.2.10 Exploratory Endpoint – individual components of BICLA and SRI response, BILAG-2004, SLEDAI-2K and PGA

4.2.10.1 Definition

The individual assessments used to derive BICLA and SRI(4) are: BILAG-2004, SLEDAI-2K, PGA assessed at Week 24 and 52 by treatment. Subjects who are BILCA and SRI(4) responders are those with no PEOT, no RM and no death per composite strategy.

SLEDAI-2K

SLEDAI-2K (derived as the sum of the scores for all items) is evaluated using the difference in mean change from baseline longitudinally over time to Week 52. Scores for the SLEDAI-2K organ systems are derived in the same way as SLEDAI-2K but using the scores for the respective items only. The SLEDAI-2K organ systems are defined in Section 4.1.6.1.

BILAG-2004

The detail description of BILAG-2004 parameter is presented in Section 4.1.6.1.

Physician's Global Assessment

The PGA is defined as in Section 4.1.6.1. However, the variable is considered on a continuous scale and the mean change from baseline in PGA (measured on a VAS ranging from 0 to 3) is assessed by visit up to Week 52.

Values of BILAG-2004, SLEDAI-2K, PGA, including adjudication of BILAG scoring are assessed.

4.2.10.2 Derivations

Not applicable.

4.2.10.3 Handling of Dropouts and Missing Data

If any of the BILAG organ system grade results cannot be evaluated at any visit up to Week 52 (e.g., due to missing values), the data for that criterion is to be imputed using LOCF to determine the current BILAG organ system grade. This applies only once. For example, if any of the Week 52 organ system grade result is missing and Week 48 data is not missing, the missing Week 52 organ system grade result will be imputed using the corresponding Week 48 data. Otherwise the participant's BILAG-2004 is missing at Week 52. Local laboratory results will not be requested. If lab data is missing for 2 sequential visits, N/A is indicated for the organ system grade.

LOCF records are not used in the repeated measures models.

4.2.10.4 Primary Analysis of Exploratory Endpoints – individual components of BICLA response and SRI response

Individual components of BICLA and SRI(4) and rate of no ICEs at weeks 24 and 52 together will be presented in a table by treatment group.

Change from baseline in SLEDAI-2K and PGA will be analyzed using repeated measures models with fixed effects for baseline value, treatment group, visit, treatment*visit interaction and stratification factors, and participant as random effect. Covariance parameters will be estimated using restricted Maximum Likelihood method and Kenward Rogers denominator degrees of freedom will be used for the tests of fixed effects. An unstructured covariance matrix will be used. In case of convergence issues, the following alternative structures will be used for fitting (in that order): Heterogeneous Toeplitz, heterogeneous AR(1), Heterogeneous CS, Homogeneous CS. Results will be presented for each visit in terms of the adjusted means for each treatment group, estimates of treatment differences, and associated 2-sided CIs.

For each SLEDAI organ system (central nervous system, vascular, musculoskeletal, renal, mucocutaneous, cardiovascular system and respiratory, immunology, and hematological and fever) the number and percentage of participants with an improvement at Week 24 and

Week 52, respectively, will be given for participants with corresponding organ system involvement at baseline.

For BILAG-2004, the number and percentage of participants with a score of A, B, and C, D, or E, respectively, will be given by visit for the following organ systems:

- Constitutional
- Mucocutaneous
- Neuropsychiatric
- Musculoskeletal
- Cardiorespiratory
- Gastrointestinal
- Ophthalmic
- Renal
- Haematology

Observed values and changes from baseline in BILAG-2004 global score will be presented by visit with descriptive statistics. Additionally, a shift table presenting BILAG-2004 improvement over time is presented.

For each BILAG organ system and regardless of involvement at baseline, improvement over time is summarized as the number and percentage of participants with involvement at baseline (A, B, or C), no involvement at baseline (E), or no involvement at baseline but had previous experience (D). No change is defined as the same BILAG-2004 score as baseline. Improved is defined as BILAG-2004 score change from A to B/C/D, or from B to C/D, or from C to D. Worsened is defined as BILAG-2004 score change from B/C to A, or C to A/B. New is defined as BILAG-2004 score change from D or E to A/B/C.

Change from baseline in PGA will be analyzed using the same repeated measures models as described above for the analysis of SLEDAI-2K total score. Thereby, an unstructured covariance matrix is used. In case of convergence issues, the following alternative structures is used for fitting (in that order): Heterogeneous Toeplitz, heterogeneous AR(1), Heterogeneous CS, Homogeneous CS.

4.2.11 Exploratory Endpoint – SRI(X)

4.2.11.1 Definition

In addition to the secondary efficacy endpoint, the assessment of disease activity is evaluated as exploratory endpoint based on the difference in proportions between anifrolumab and placebo in SRI(X) response at Week 52, where SRI(X) (X=5, 6, 7, or 8) is considered. For further description of SRI(X), please refer Section 4.2.2.1.

4.2.11.2 Derivations

Not applicable.

4.2.11.3 Handling of Dropouts and Missing Data

If any of the criteria mentioned on Section 4.2.2.1 cannot be evaluated at any visit up to Week 52 (e.g., due to missing values), the data for that criterion is imputed using LOCF and SRI(X) derived based on the complete data. This applies only once. Missing data after LOCF will be imputed as non-responder. For example, if any of the Week 52 criteria data is missing and Week 48 data is not missing, the missing Week 52 criteria data will be imputed using the corresponding Week 48 data. Otherwise, the participant is defined as not achieving SRI(X) at Week 52.

4.2.11.4 Primary Analysis of Exploratory Endpoint – SRI(X).

Similar analyses as detailed in Section 4.2.2.4 are performed. The difference between anifrolumab and placebo in the proportion of participants achieving SRI(X), X= 5, 6, 7, or 8 will also be assessed longitudinally over time up to Week 52. 95% CIs will be presented.

4.2.12 Exploratory Endpoint – BICLA and SRI(4) dual response

4.2.12.1 Definition

BICLA and SRI(4) dual response is reached if both the criteria for BICLA response listed in Section 4.2.1.1 and SRI(4) response listed in Section 4.2.2.1 are met. If any of the criteria is not met, the participant will be considered as a non-responder for dual response.

4.2.12.2 Derivations

Not applicable.

4.2.12.3 Handling of Dropouts and Missing Data

The rules specified in Sections 4.2.1.3 and 4.2.2.3 will be applied on BICLA and SRI(4) responses respectively.

4.2.12.4 Primary Analysis of Exploratory Endpoint – BICLA and SRI(4) dual response

Similar analyses as detailed in Section 4.2.2.4 are performed. 95% CIs will be presented.

4.2.13 Exploratory Endpoint – Major Clinical Response and Partial Clinical Response

4.2.13.1 Definition

The effect of anifrolumab versus placebo on Major Clinical Response and Partial Clinical Response is evaluated as where these variables are defined as:

- Difference in proportion of participants who achieve Major Clinical Response, i.e., a participant with BILAG-2004 C scores or better at Week 24 with no new BILAG-2004 A or BILAG-2004 B scores compared to baseline and maintenance of response with no new BILAG-2004 A or B scores between Week 24 and Week 52.

- Difference in proportion of participant who achieve Partial Clinical Response, i.e., a participant with a maximum of 1 BILAG-2004 B score or better at Week 24 and maintenance of response without a new BILAG-2004 A or more than 1 new BILAG-2004 B item out to Week 52.

4.2.13.2 Derivations

Not applicable.

4.2.13.3 Handling of Dropouts and Missing Data

If any of the BILAG-2004 grades cannot be evaluated at any visit up to Week 52 (e.g., due to missing values), the data for that grade is imputed using LOCF and the Major and Partial Clinical Response are derived based on the complete data. This applies only once. Missing data after LOCF will be imputed as non-responder. For example, if any of the Week 52 BILAG-2004 grades are missing and Week 48 data is not missing, the missing Week 52 BILAG-2004 grade will be imputed using the corresponding Week 48 data. Otherwise, the participant is defined as Major and Partial Clinical non-responder at Week 52.

4.2.13.4 Primary Analysis of Exploratory Endpoints – Major Clinical Response and Partial Clinical Response

For Major Clinical Response and Partial Clinical Response, the same CMH approach as described in Section 4.2.1.5 错误!未找到引用源。for the primary endpoint is used to estimate the treatment difference between anifrolumab and placebo as well as the response rates. 95% CIs will be presented.

4.2.14 Exploratory Endpoint – LLDAS

4.2.14.1 Definition

A participant is considered as a Lupus Low Disease Activity State (LLDAS) responder at a specific visit if all the criteria below are met:

- SLEDAI-2K ≤ 4, with no activity in major organ systems (renal, CNS, cardiopulmonary, vasculitis) or fever;
    - Any participant with a SLEDAI-2K score greater than 4 won't attain LLDAS at that visit.

- Major organ activity will be assessed at each visit using the SLEDAI-2K items. Any participant with at least one of the following SLEDAI-2K items present (item score >0) at the visit won't attain LLDAS at that visit:
  - Any in central nervous system:
    - Seizure
    - Psychosis
    - Organic brain syndrome
    - Visual disturbance
    - Cranial nerve disorder
    - Lupus headache
  - Any in vascular organ system:
    - Cerebrovascular accident(s)
    - Vasculitis
  - Any in renal organ system:
    - Urinary casts
    - Hematuria
    - Proteinuria
    - Pyuria
  - Any in cardiovascular system and respiratory:
    - Pleurisy
    - Pericarditis
  - Fever (Fever >38 degrees C [38o C], excluding infectious cause)

- No new lupus disease activity compared with the previous assessment (SLEDAI-2K);

- Any participant with present activity (item score >0) in at least one new SLEDAI-2K item (irrespective of organ system) compared with the previous visit won't attain LLDAS.

- PGA ≤ 1 (scale 0–3): any participant with PGA score greater than 1 at a study visit won't attain LLDAS at that visit.

- A current prednisone (or equivalent) dose ≤ 7.5 mg daily;
  - Any participant on OCS treatment with a dose greater than 7.5 mg/day prednisone or equivalent taken on any day of the period starting on the day after the previous scheduled visit (if not missed) and ending on the day of the visit, won't attain LLDAS at that visit. In case the previous visit was missed, LLDAS won't be attained at the visit of evaluation if OCS treatment with a dose greater than 7.5 mg/day prednisone or equivalent was taken at the day of the visit or any day of the 28-days period preceding that visit.

- Standard maintenance doses of immunosuppressive drugs and approved biological agents. This criterion on immunosuppressive medications will be handled through the intercurrent event of RM.

Further, to define the "day of a visit" in the above, the date of the last SLEDAI-2K or PGA assessment used for analysis in the time window of the respective timepoint (as described in Section 3.3.2) is used. If no assessment falls within a defined window, the corresponding visit is considered as missing.

4.2.14.2 Derivations

Not applicable.

4.2.14.3 Handling of Dropouts and Missing Data

If any of the criteria for SLEDAI-2K, BILAG or PGA cannot be evaluated at any visit up to Week 52, (e.g., due to missing values), the data for that criterion is imputed using LOCF and LLDAS derived based on the complete data. This applies only once. Missing data after LOCF will be imputed as non-responder. For example, if any of the Week 52 criteria data is missing and the data at Week 48 is not missing, the missing Week 52 criteria data will be imputed using the corresponding Week 48 data. Otherwise, the participant is defined as not achieving LLDAS at Week 52.

4.2.14.4 Primary Analysis of Exploratory Endpoints – LLDAS

The same CMH approach as described in Section 4.2.1.4 for the primary Exploratory is used to estimate the treatment difference between anifrolumab 300 mg and placebo as well as the response rates for all treatment groups for LLDAS response. Longitudinal

presentations of the results over time (i.e., for each post-baseline visit up to Week 52 during double blind treatment period) based on this analysis are created.

4.2.15 Exploratory Endpoint – Change in SDI

4.2.15.1 Definition

The endpoint used to evaluate the effect of anifrolumab versus placebo on irreversible damage in SLE participants is the difference in mean change in SDI global score from baseline to Week 52.

The SDI is defined for 12 organ systems as described in Section 4.1.6.1.

Furthermore, the number and percentage of participants with changes in damage will be explored at Week 52. A damage according to the SDI score is defined as an SDI global score ≥1. Accordingly, no damage is defined as an SDI global score =0. Post-baseline categories used for the presentation of change in damage will be 'no change', '+ 1 point', '+ 2 points', and '+ 3 or more points'.

4.2.15.2 Derivations

Not applicable.

4.2.15.3 Handling of Dropouts and Missing Data

Not applicable.

4.2.15.4 Primary Analysis of Exploratory Endpoint – Change in SDI

Observed values and changes from baseline in SDI global score will be presented by visit with descriptive statistics. This summary will be repeated by SDI global score at baseline (0 and ≥1).

Additionally, a shift table presenting damage according to the SDI score at baseline versus damage at Week 52 will be presented according to the categories given in Section 4.2.7.4.

4.2.16 Exploratory Endpoint – Lupus QoL

4.2.16.1 Definition

Lupus QoL

The difference between anifrolumab and placebo in the mean change from baseline in lupus QoL domain scores (physical health, pain, planning, intimate relationships, burden to others, emotional health, body image, and fatigue) will be assessed by visit up to Week 52.

The number of items in each domain and the item numbers that refer to that domain are tabulated in Table 3.

Table 3 Lupus QoL domains

Domain	Number of Items	Item Numbers
Physical Health	8	1 to 8
Pain	3	9 to 11
Planning	3	12 to 14
Intimate Relationships	2	15, 16
Burden to Others	3	17 to 19
Emotional Health	6	20 to 25
Body Image	5	26 to 30
Fatigue	4	31 to 34

Each item of the lupus QoL has a five-point Likert response scale (0=all the time, 1=most of the time, 2=a good bit of the time, 3=occasionally, and 4=never).

4.2.16.2 Derivations

Domain scores will be derived 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. The mean raw domain score will be transformed to the domain scores (ranging from 0 as worst QoL to 100 as best QoL) as mean raw domain score divided by 4 and multiplied by 100.

4.2.16.3 Handling of Dropouts and Missing Data

Not applicable.

4.2.16.4 Primary Analysis of Exploratory Endpoint – Lupus QoL

Lupus quality of life (QoL) scale

Change from baseline in each of the lupus QoL domain scores (physical health, pain, planning, intimate relationships, burden to others, emotional health, body image, and fatigue) will be analyzed using repeated measures models with fixed effects for baseline value, treatment group, visit, and treatment*visit interaction stratification factors, and participant as random effect. Covariance parameters will be estimated using restricted Maximum Likelihood method and Kenward Rogers denominator degrees of freedom will be used for the tests of fixed effects. Thereby, an unstructured covariance matrix is used. In case of convergence issues, the following alternative structures is used for fitting (in that order): Heterogeneous Toeplitz, heterogeneous AR(1), Heterogeneous CS, Homogeneous CS. Longitudinal presentations of the results over time (i.e., for each post-baseline visit up to Week 52 during double blind treatment period) based on this analysis is created.

4.2.17 Exploratory Endpoint – EuroQol 5 dimensions (EQ-5D-5L)

4.2.17.1 Definition

The EQ-5D-5L is comprised of the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The proportion of participants in each EQ-5D-5L health state (no problems, slight problems, moderate problems, severe problems and unable to/extreme problems), by dimension, as well as the VAS Score and single summary utility index, including changes from baseline, are explored over time.

EQ-5D-5L value sets are not available for all countries, therefore the UK value set is used for all participants in the study. By using the crosswalk link function and the individual responses to the EQ-5D-5L descriptive system, index values for the EQ-5D-5L is calculated (Van Hout, 2012).

4.2.17.2 Derivations

Not applicable.

4.2.17.3 Handling of Dropouts and Missing Data

Not applicable.

4.2.17.4 Not applicable. Primary Analysis of Exploratory Endpoint – EQ-5D-5L

The 5 dimensions of EQ-5D-5L (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) are summarized by timepoint as number and percentage of participants with the different health states.

Observed values and changes from baseline in VAS Score and single summary utility index, respectively, are presented longitudinally with descriptive statistics.

4.2.18 Exploratory Endpoint – Work productivity and Activity Impairment (WPAI)-Lupus

4.2.18.1 Definition

The Work Productivity and Activity Impairment (WPAI) is a validated, self-administered questionnaire consisting of 6 questions, assessing the impact of disease on productivity. The WPAI yields 4 types of scores: absenteeism (work time missed), presenteeism (VAS [scored from 0 to 10] rating of impairment while working), work productivity loss (overall work impairment /absenteeism plus presenteeism), and activity impairment (VAS [scored from 0 to 10] rating of daily activities, other than work at a job) (Reilly et al, 1993).

4.2.18.2 Derivations

The following variables are derived and presented:

Number of participants employed = number of participants with yes in response to

Question 1 (Q1)

Percentage of all participants employed

Number of work hours missed due to problems associated with lupus as responded in Q2

Absenteeism due to lupus (%) = [Q2 / (Q2+Q4)] x 100

Presenteeism due to lupus (%) = [Q5 / 10] x 100

Work Productivity Loss (%) = (Q2 / (Q2+Q4) + [(1 - Q2 / (Q2+Q4)) x (Q5/10)]) x 100

Activity impairment (%) = [Q6 / 10] x 100

4.2.18.3 Handling of Dropouts and Missing Data

Not applicable.

4.2.18.4 Primary Analysis of Exploratory Endpoint – WPAI-Lupus

At each scheduled visit, the number and percentage of participants employed are presented. Observed values and changes from baseline in hours missed due to lupus and in WPAI scores (absenteeism, presenteeism, work productivity loss and activity impairment) are presented longitudinally with descriptive statistics.

4.2.19 Exploratory Endpoint – Medical Resource Use Questionnaire

4.2.19.1 Definition

The following variables are explored:

- Number of participants with health care visits
  Of participants with health care visits
  - Number of specialist visits
  - Number of primary care visits

- Number of participants with emergency department visits
  Of participants with emergency department visits:
  - Number of emergency department visits
  - Of emergency department visits:
    - Visit related to an increase in lupus related activity
    - Cause of emergency department visit

- Number of participants with hospital visits
  Of participants with hospital visits:

- Number of hospital visits
- Of hospital visits:
    - Visit related to an increase in lupus related activity
    - Cause of hospitalization
- Length of hospital stay
- Total number of days in intensive care unit (ICU)

For all these variables, all available data between first administration of IP and Week 52 will be summed up.

4.2.19.2 Derivations

Not applicable.

4.2.19.3 Handling of Dropouts and Missing Data

Not applicable.

4.2.19.4 Primary Analysis of Exploratory Endpoint – Medical Resource Use Questionnaire

The medical resource uses are summarized by treatment. Results of "visit related to an increase in lupus related activity" and "cause of visit/ hospitalization" are presented as number and percentage of respective visits.

4.3 Pharmacodynamic Endpoint

See Section 4.2.9.

4.4 Pharmacokinetics

See Section 4.2.7.

4.5 Immunogenicity

See Section 4.2.8.

4.6 Safety Analyses

The following safety data are collected: vital signs, physical examination, 12-lead ECGs, clinical laboratory tests (including haematology, clinical chemistry, urinalysis and lipid profile), C-SSRS, Cushingoid features and AEs (including AEs of specific interest [AESIs]).

Safety analyses are performed using the safety analysis set. Unless otherwise specified, safety data are presented using descriptive statistics during the treatment and follow-up period.

Change from baseline to each post-treatment time point where scheduled assessments were made to calculate for relevant measurements.

If not stated otherwise, on-treatment values are defined as values with an assessment date after the first administration of IP and on or before the date of last administration of IP + 28 days.

4.6.1 Exposure

4.6.1.1 Definition and Derivations

The duration of exposure to the IP per subject is defined as the number of days between the start and the end dates of IP plus the dosing frequency interval:

Duration of exposure (days) = (Last dosing date + 28 days) – first dosing date + 1.

The total year of exposure is the sum of duration of exposure (days) of all subjects in the respective treatment group divided by 365.25 (days/year).

The total number of infusions are counted per participant. Furthermore, the number of participants with an infusion are assessed in 4-weekly categories (i.e., 4 Weeks, 8 Weeks, 12 Weeks, …, 52 Weeks). An infusion will be counted if it was given in the corresponding time window. Missed infusions are not counted. Additionally, the number of participants with duration of exposure longer than each of the 4-weekly categories (i.e., ≥4 weeks, ≥8 weeks, ≥12 weeks, …, ≥52 weeks) will be presented. As the duration of exposure takes into account the total exposure time from the last dose + 28 days (as defined above), any missed or delayed infusions will contribute to exposure time, but not to the number of infusions. For example, if a participant interrupted treatment and received no infusion at week 4, he/she will be counted as "No infusion" at week 4, but is counted as having "Exposure >= 8 weeks" if he/she received an infusion at any time after Week 4.

The time to discontinuation of IP is the same as the duration of exposure.

4.6.1.2 Presentation

Exposure is summarized by treatment group for the safety analysis set.

Summary statistics are provided for the duration of exposure. The number and percentage of participants treated ≥4 weeks, ≥8 weeks, and up to ≥52 weeks in 4-weekly intervals are provided. Furthermore, the total patient years of exposure is presented.

The number and percentage of participants with infusions are presented by total number of infusion (i.e., 1, 2, …, 13) and by visit (i.e., Day 1, Week 4, Week 8, Week 12, …, Week 48).

Furthermore, the time to discontinuation of IP is presented as Kaplan-Meier plot including the number of participants at risk (i.e., still on IP).

4.6.2 Adverse events

4.6.2.1 Definition and Derivations

Adverse events experienced by the participants will be collected throughout the entire study and will be coded using the latest version of MedDRA.

Adverse event data will be categorized according to their onset date into the following study periods:

- AEs occurring during screening:
  An AE during screening is defined as an AE with a date of onset ≥ date of first screening visit (S1) and < date of the first dose of IP.
  AEs occurring during screening will only be listed.

- AEs occurring during treatment period:
  An AE during treatment period is defined as an AE with a date of onset ≥ day of first dose of IP and ≤ date of last dose of IP + 28 days.

- AEs occurring during treatment period and follow-up period:
  An AE during treatment period and follow-up period is defined as an AE with a date of onset ≥ day of first dose of IP and ≤ date of last dose of IP + 84 days.

- AEs occurring after follow-up:
  An AE after follow-up is defined as an AE with a date of onset > date of last dose of IP + 84 days.
  AEs occurring after follow-up will only be listed.

If an AE has a missing onset date, then unless the stop date of the AE indicates otherwise, this will be considered as an AE during treatment. Similarly, if an AE has a partial onset date, then unless the partial onset date or the stop date indicates otherwise, this will be considered an AE during treatment.

Adverse events of special interest are marked as such in the eCRF.

Adverse events with missing intensity will be assumed to be severe. Events with missing relationship to study medication per the investigator will be assumed to be related. If no information about seriousness is available, the AE will be considered serious.

An infusion-related reaction (as assessed by the investigator) is defined as an AE with a PT of “Infusion related reaction”.

An infection is defined as an AE within the SOC infections and infestations.

Opportunistic infections and serious non-opportunistic infections are identified by the investigator.

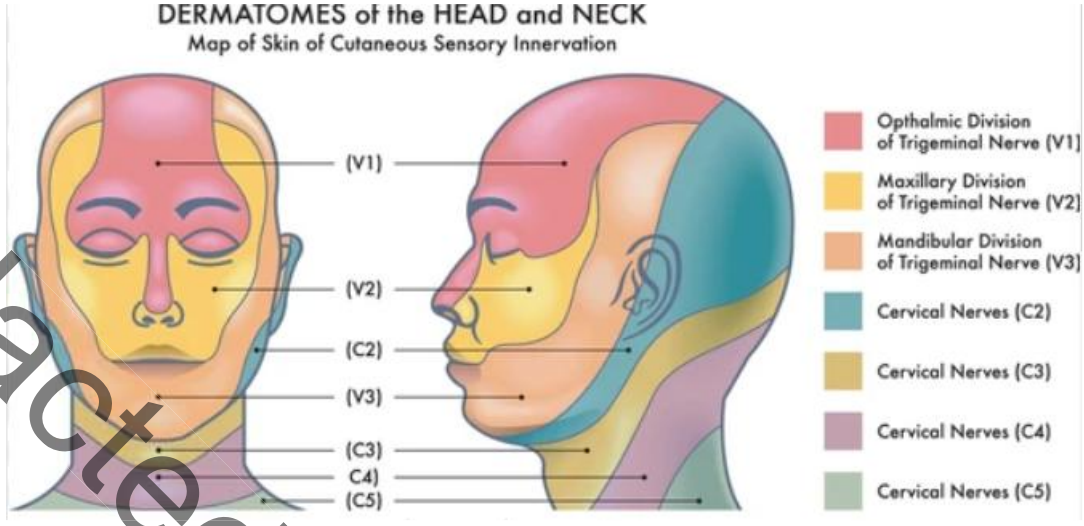
Hypersensitivity is defined as adverse events with MedDRA PT = "Hypersensitivity" and Lower level term (LLT) = "Hypersensitivity reaction".

Herpes zoster is further classified according to the information given on the eCRF form "HERPZOS" as follows:

HZ Category	Cutaneous/Skin Involvement	Any organ involvement other than cutaneous/skin	Dermatome rule
Cutaneous Localized*	Yes	No	[≤ 3 contiguous unilateral dermatomes selected]
Cutaneous Disseminated	Yes	No	[> 3 contiguous unilateral dermatomes selected] or [non-contiguous dermatomes selected] or [dermatomes on left and right selected and/or bilateral selected]
Visceral Disseminated	Yes/No	Yes	[No rule]

* Cutaneous Localized is defined as below:

- CUTANEOUS/SKIN INVOLVEMENT = "YES"
- All affected dermatomes must be either "RIGHT" or "LEFT" not a mix of either or "BOTH"
- No more than 3 dermatomes are affected
- There is no other organ involvement
- There is an unbroken chain of affected dermatomes in the order that is presented in the image below.
    - The order is: V2, [(V1, V3) or (V3, V1) or (V1) or (V3)], C2, C3, C4, C5, C6, C7, C8, T1, T2, T3,… – T12, L1 – l5, S1 – S5



Adverse events during treatment are also presented by time intervals of the first onset of the event. For this analysis, repeated events with the same PT are not considered (i.e., if a participant has more than one event with the same PT, only the event with the earliest date of onset is used). For partial or missing dates, the rules as described above are used, i.e., the AE is considered as an AE during treatment unless the available information indicates otherwise and as occurring in the earliest possible time interval given the available (start and stop) date information. The following time intervals are defined:

- Day 1 to < Week 12:

AEs with date of onset ≥ date of first administration of IP and < date of first administration of IP plus 84 days

- Week 12 to < Week 24:

AEs with date of onset ≥ date of first administration of IP plus 84 days and < date of first administration of IP plus 168 days

- Week 24 to < Week 36:

AEs with date of onset ≥ date of first administration of IP plus 168 days and < date of first administration of IP plus 252 days

- Week 36 to < Week 48:

AEs with date of onset ≥ date of first administration of IP plus 252 days and < date of first administration of IP plus 336 days

- ≥ Week 48:

AEs with date of onset ≥ date of first administration of IP plus 336 days

Exposure-adjusted incident rate (EAIR) per 100 patient years is defined as

Sum of subjects with the event/[sum across subjects' time at risk for the event /(365.25*100)]

Where time at risk is time to first event for subjects who experienced the event during the analysis period and time during the analysis period for subjects who do not experience the event.

The time to first onset of an event during a specified period is derived as date of first onset of the event during the specified period – date of first administration of study intervention + 1. Only AEs with an onset date during specified period will be counted. If a subject has no such event during the specified period, the time to first onset is censored at the end date of the period.

The time during the analysis period for subjects who do not experience the event will be calculated as end of period – start of period + 1 (e.g., date of last dose of IP + 28 days - day of first dose of IP + 1 day for summary of AEs during treatment period; date of last dose of IP + 84 days - day of first dose of IP + 1 day for summary of AEs during treatment and follow-up period). If a participant discontinued from the study during a period or had the last follow-up visit earlier than expected, the date of study discontinuation/end of study will be used as end of the respective period.

For the by timepoint analysis of anaphylaxis, hypersensitivity, and infusion related reactions, an AE at a visit is defined as an AE with a date of onset ≥ day of administration of IP at the respective visit and ≤ date of administration of IP at the following visit (or ≤ date of last dose of IP + 28 days for the last visit with IP).

Other significant adverse events

During the evaluation of the AE data, a medically qualified expert will review the list of AEs that were not reported as serious AE (SAEs) or AEs leading to discontinuation.

Based on the expert's judgment, significant AEs of particular clinical importance may, after consultation with the AstraZeneca Global Patient Safety Physician, be considered other significant AEs and reported as such in the CSR.

Examples of these are marked hematological and other laboratory abnormalities, and certain events that led to intervention (other than those already classified as serious), dose reduction, or significant additional treatment.

4.6.2.2 Presentation

If not stated otherwise, all summaries described below are presented separately for

- AEs during treatment period
- AEs during treatment and follow-up period.

An overall summary of participants with at least one AE in the following categories is presented.

- Any AE
- Any SAE with outcome = death
- Any SAE (including events with outcome = death)
- Any AE leading to discontinuation of IP
- Any AE related to IP
- Any AE of severe intensity
- Any AESI
  - Any AESI of serious non-opportunistic infections
  - Any AESI of opportunistic infections
  - Any AESI of herpes zoster
  - Any AESI of tuberculosis including latent tuberculosis
  - Any AESI of malignancy
  - Any AESI of major adverse cardiovascular events (MACEs)
- Any other significant AE.

If not stated otherwise, the number and percentage of participants with at least one AE (i.e., multiple occurrences of an AE in 1 participant will only be counted once) are summarized by MedDRA primary SOC and PT for the following AE categories. If not stated otherwise, the EAIR per 100 patient years will be presented for AEs during treatment. A 95% confidence interval for the difference between Anifrolumab and Placebo will also be constructed based on Mienttinen and Nurminen method (Liu et. al., 2006). EAIR will generally not be included in summaries presented for AEs during treatment and follow-up period.

- Any AE
- Any AE above reporting threshold of 2%

This summary is presented by PT.

- Any SAE with outcome = death

This summary is presented for AEs during treatment and follow-up period only. This summary will not include the EAIR per 100 patient years.

- Any SAE (including events with outcome = death)

- Any AE leading to discontinuation of IP

This summary is presented for all AEs during treatment period only.

- Any AE by relationship to IP (yes, no)

(Multiple occurrences of an AE in 1 participant will only be counted once as related if at least one AE is related and as not related if all occurrences are not related).

This summary is presented for AEs during treatment and follow-up period only and will not include the EAIR per 100 patient years.

- Any AE by maximum intensity (mild, moderate, severe)

(Multiple occurrences of an AE in 1 participant will only be counted once with the maximum intensity in this AE).

This summary is presented for AEs during treatment and follow-up period only and will not include the EAIR per 100 patient years.

- Any AESI

This summary will not be presented by SOC but by AESI category

(Serious non-opportunistic infections, opportunistic infections, herpes zoster, tuberculosis including latent tuberculosis, malignancy and MACEs [Cardiovascular outcome events are reported by investigators]).

- Any other significant AE

This summary is presented for AEs during treatment and follow-up period only. This summary will not include the EAIR per 100 patient years.

- Any AE by time interval for the first onset of event

This summary is presented for AEs during treatment period only. This summary will not include the EAIR per 100 patient years.

The time to first onset of a serious non-opportunistic infection during treatment period and time to first onset of herpes zoster during treatment period is presented as Kaplan-Meier plots including the number of participants at risk at each visit.

Furthermore, the number and percentages of subjects with herpes zoster (and possible other AESIs) are summarized for events during treatment and follow-up period, during treatment period (overall and by time intervals). The following subcategories are also considered in this summary: SAE (including events with outcome of death), AE leading to discontinuation of IP, and AE by maximum intensity (mild, moderate, and severe).

The number and percentage of participants with at least one anaphylaxis, hypersensitivity, and infusion-related reaction (as reported by the investigator), respectively, is summarized overall as well as for the following respective subcategories: SAE (including events with outcome of death), AE leading to discontinuation of IP, and AE by maximum intensity (mild, moderate, and severe).

Infections, opportunistic infections and serious non-opportunistic infections, are summarized with the same subcategories as given above.

Key subject information for participants with an AE with outcome of death, participants with serious AEs, participants with an AE leading to discontinuation of IP, participants with AESIs, and ADA positive participants with AEs, respectively, is provided during treatment and follow-up period.

4.6.3 Clinical Laboratory

4.6.3.1 Definition and Derivations

The parameters of haematology, serum chemistry, urinalysis and of fasting lipid profile (high density lipoprotein cholesterol, low density lipoprotein cholesterol, and triglycerides) are explored.

Laboratory data are reported in international system of units (SI). Changes from baseline in haematology, clinical chemistry, urinalysis and lipid profile variables are calculated.

Absolute values are compared to the reference range as given in Appendix 1.1 and classified as low (below range), normal (within range or on limits) or high (above range). All values (absolute and change) falling outside the reference ranges are flagged.

Treatment emergent laboratory changes (TELVC) are defined for post-baseline values according to the reference ranges given in Appendix 1.1.

Urinalysis data are categorized as negative (0), positive (+), or strongly positive (++, +++, or >+++) at each time-point.

For the liver function tests: aspartate transaminase (AST), alanine transaminase (ALT), Alkaline phosphatase, gamma glutamyl transferase (GGT) and total bilirubin, the multiple of the upper limit of the normal (ULN) range (see Appendix 1.1) are calculated for each data point. Multiple = Value / ULN, i.e., if the ALT value was 72 IU/L (ULN = 36) then the multiple would be 2. Participants meeting both of the following biochemical criteria for Hy's law (potential Hy's Law) at any point during the study (not necessarily at the same time) are flagged:

- AST ≥3x ULN and/or ALT ≥3x ULN
- Total bilirubin ≥2xULN

4.6.3.2 Presentation

Observed values and changes from baseline of laboratory data for haematology, clinical chemistry, continuous urinalysis, and fasting lipid profile are summarized by visit. The summary statistics presented are minimum, 1st quartile, median, 3rd quartile, maximum, mean, and SD.

For each laboratory parameter with available criteria, the number and percentage of participants with TELVC values are summarized. Additionally, the number and percentage of participants with at least one TELVC value are presented. Percentages are based on participants with a measurement at baseline and at least 1 subsequent measurement of the variable.

The number and percentage of participants with laboratory values below, within or above the corresponding reference ranges are presented as shift tables from baseline to maximum and minimum on-treatment value, respectively.

Urinalysis is summarized as shift tables from baseline categorization to the last post-baseline value categorization for each parameter. Furthermore, the number and percentage of participants with treatment-emergent changes are summarized by parameter. Percentages for the summary of treatment emergent changes are based on participants with a measurement at baseline and at least 1 subsequent measurement of the variable. The three parameters Urine Glucose, Ketone Bodies and qualitative Erythrocytes should be added for presentation of TELVC for urinalysis. The criteria for TELVC for urinalysis data are defined as follows:

- Negative/positive at baseline to ++, +++, ++++ at any post-baseline value OR,
- Increase from baseline of at least ++ at any post-baseline value.

In order to identify potential Hy's Law cases, maximum on-treatment ALT and AST are summarized as number and percent versus maximum on-treatment total bilirubin.

All clinical safety laboratory data is listed, and any out-of-range laboratory measurements are flagged in the listing.

4.6.4 Electrocardiogram (ECGs)

4.6.4.1 Definition and Derivations

12-lead ECGs for all participants at all centers are conducted at the centers using standardized ECG machines throughout the study. The Investigator judges the overall interpretation as normal or abnormal. If abnormal, it is decided as to whether the abnormality is clinically significant or not clinically significant and the reason for the abnormality are recorded on the eCRF, if the Investigator considers it clinically significant. The following categories are used for analysis:

- Normal,
- Abnormal, not clinically significant,
- Abnormal, clinically significant.

In case of repeated measurements (triplicates) at a visit, the worst category at the respective visit will be used for the analysis.

4.6.4.2 Presentation

The number and percentage of participants with normal, abnormal, not clinically significant, and abnormal, clinically significant ECG results are presented as a shift table from baseline to last observation.

4.6.5 Vital signs

4.6.5.1 Definition and Derivations

The following variables will be explored:

- Pulse (beats per minute)
- Systolic blood pressure (mmHg)
- Diastolic blood pressure (mmHg)
- Respiration rate (breaths per minute)
- Body temperature (°C)
- Body weight (kg)

Changes from baseline will be calculated.

Where applicable, absolute values are compared to the reference ranges given in Appendix 7.3 and classified as low (below range), normal (within range or on limits) or high (above range). All values (absolute and change) falling outside the reference ranges will be flagged.

Post-baseline values are classified as TELVC according to reference ranges given in Appendix 7.3.

For infusion visits, only measurements before start of IP are considered for by visit presentations. In case of multiple measurements before the start of IP, the first measurement is used for by visit presentations, but all measurements will be considered for the TELVC classification.

4.6.5.2 Presentation

Observed values and changes from baseline of pulse, systolic blood pressure (SBP), diastolic blood pressure (DBP), respiration rate, body temperature, and body weight, respectively, are summarized by visit. The summary statistics presented are minimum, 1st quartile, median, 3rd quartile, maximum, mean, and SD.

For each parameter with available criteria, the number and percentage of participants with TELVC values are summarized. Percentages are based on participants with a measurement at baseline and at least 1 subsequent measurement of the variable.

For pulse, SBP and DBP, the number and percentage of participants with values below, within or above the corresponding normal range is presented as shift tables from baseline to maximum and minimum on-treatment value, respectively.

4.6.6 C-SSRS

4.6.6.1 Definition and Derivations

The C-SSRS is an assessment tool that evaluates suicidal ideation and behavior.

Two different versions of the questionnaire were used:

- Baseline/Screening version, assessing the last 12 months prior to the assessment
- Since Last Visit Version, assessing the time since last visit.

The following outcomes are C-SSRS categories and have binary responses (yes/no). The categories have been re-ordered from the actual scale in an increasing order of severity from 1 to 10 to facilitate the definitions of the comparative variables.

- Category 1 – Wish to be Dead
- Category 2 – Non-specific Active Suicidal Thoughts
- Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
- Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
- Category 5 – Active Suicidal Ideation with Specific Plan and Intent
- Category 6 – Preparatory Acts or Behavior
- Category 7 – Aborted Attempt
- Category 8 – Interrupted Attempt
- Category 9 – Actual Attempt (non-fatal)

- Category 10 – Completed Suicide

Composite variables based on the above re-ordered categories are defined for assessments during screening, during treatment and follow-up period, respectively as follows:

- Suicidal ideation: A "yes" answer at any time in the respective study period to any one of the 5 (re-ordered) suicidal ideation questions (Categories 1-5) on the C-SSRS.
- Suicidal behavior: A "yes" answer at any time in the respective study period, to any one of the 5 (re-ordered) suicidal behavior questions (Categories 6-10) on the C-SSRS.
- No suicidal ideation or behavior: No "yes" answer at any time in the respective study period to any one of the 10 (re-ordered) suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

The total number of attempts, total number of interrupted attempts, and total number of aborted attempts will be derived for the different study periods by summing up all respective attempts during screening, during treatment and follow-up period, respectively.

For summary of C-SSRS data, the following period definitions will be used:

- Screening Period:

  Assessments with a date ≥ date of first screening visit (S1) and ≤ date of the first dose of IP.

- During treatment and follow-up period:

  Assessments with a date > date of first dose of IP and ≤ date of last dose of IP + 84 days.

If the wrong questionnaire version (Baseline/Screening rather than since last visit or vice versa) was used for an assessment, the assessment will be assigned to a period based on assessment date, regardless of the version that was completed.

4.6.6.2 Presentation

The number and percentage of participants with suicidal ideation (overall and by maximum category), suicidal behavior (overall and by maximum category), and no suicidal ideation or behavior will be given for assessments during screening, during treatment and follow-up period, respectively.

Furthermore, descriptive statistics on the total number of attempts, total number of interrupted attempts, and total number of aborted attempts will be summarized for attempts during screening, during treatment and follow-up period, respectively.

4.6.7 Cushingoid features

4.6.7.1 Definition and Derivations

Cushingoid features include moon face, buffalo hump, purple or violaceous striae, central obesity, hirsutism, acne, easy bruising, and fragile skin.

4.6.7.2 Presentation

The number and percentage of participants are explored for each feature by visit during the treatment period. For participants with baseline OCS ≥10 mg/day prednisone or equivalent, the summary is repeated by maintained OCS reduction at Week 52 (yes versus no) as defined in Section 4.2.3.1).

4.7 China Subpopulation Analyses

Approximately 80% of the participants are estimated to be recruited from mainland China. The efficacy and safety in China subpopulation (defined as all participants from sites in mainland China) were analyzed to facilitate a benefit-risk assessment for regulatory submission in China.

Efficacy are summarised for the China subpopulation, which consists of all randomized participants from sites in mainland China in the Full Analysis Set as defined in Section 3.2.2. Selected key analyses will be conducted and stratified analyses will not include the stratification factor region for the China subpopulation. These analyses are considered descriptive/supportive only, therefore no significance testing will be performed.

Safety and tolerability are summarised for the China subpopulation, which consists of all dosed participants from sites in mainland China in the Safety Analysis Set as defined in Section 3.2.3.

The PK data is summarised for the China subpopulation, which consists of all dosed participants from sites in mainland China in the Pharmacokinetic Analysis Set as defined in Section 3.2.4.

5 INTERIM ANALYSIS

No interim analysis is planned for this study.

6 REFERENCES

Clowse et al, 2017
Clowse ME, Wallace DJ, Furie RA, Petri MA, Pike MC, Leszczyński P, 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 Rheumatol 2017 Feb;69(2): 362–75.

Fleming, 2011
Fleming TR. Addressing Missing Data in Clinical Trials. Ann. Intern. Med. 2011. 154:113-117.

Keene, 2014
Keene ON, Roger JH, Hartley BF, Kenward MG, 2009. Missing data sensitivity analysis for recurrent event data using controlled imputation. Pharmaceutical Statistics. Published online 16 June 2014 in Wiley Online Library.

Liu et al. 2006
Liu GF, Wang J, Liu K, Snavely DB. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Statistics in Medicine 2006; 25: 1275–1286.

Schmidli et al (2014)
Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics. 2014 Dec;70(4):1023-32. doi: 10.1111/biom.12242. Epub 2014 Oct 29. PMID: 25355546.

Ratitch, 2013
Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharmaceutical Statistics 2013; 12: 337-347.

Van Hout, 2012
Van Hout B., Janssen M.F., Feng Y.S., Kohlmann T., Busschbach J., Golicki D., Lloyd A., Scalone L., Kind P., Pickard A.S. 2012. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health. 2012 Jul-Aug; 15(5): 708-15

US Food and Drug Administration, 2018
BLA 125370/s-064 and BLA 761043/s-007 Multi-disciplinary Review and Evalaution Benlysta® (belimumab) for Intravenous Infusion in Children 5 to 17 Years of Age with SLE. 2018 Oct.

7 APPENDIX

7.1 Appendix A - Derivation of IFN21 parameters

There are two IFN parameters to be calculated during the study; percent of baseline (BL) IFN 21-gene PD signature and the gene signature fold change.

The following steps should be implemented in the derivation of these parameters.

1. Calculate the mean Ct values for each gene and the three housekeeper genes (18S, ATCB and GAPDH) on the ThermoFisher TLDA platform (test and normal control samples) within a particular subject.

2. Calculate the mean Ct value of the three housekeeper genes;

$$HK_{mean} = mean(Ct_{18S}, Ct_{ATCB}, Ct_{GAPDH})$$

3. Calculate ΔCt values by subtracting the mean Ct value of the three housekeeper genes from each test gene:

$$\Delta Ct_{gene} = Ct_{test\ gene} - HK_{mean}$$

4. Calculate the ΔCt values for the normal control genes, as in Step 3.

5. Calculate -ΔΔCt values as:

$$-\Delta\Delta Ct_{gene} = \Delta Ct_{normal\ control} - \Delta Ct_{test\ gene}$$

6. Linearize the -ΔΔCt values for each gene as follows:

$$linearized_{gene} = 2^{-\Delta\Delta Ct}$$

7. Calculate the percent of baseline IFN 21-gene PD signature for each gene for each post-BL visit (Visit X) as:

$$\%\ of\ BL\ IFN21 gene\ PD\ signature_{gene} = 100 - \left(\left(\frac{2^{-\Delta\Delta Ct_{BL}} - 2^{-\Delta\Delta Ct_{visit\ x}}}{2^{-\Delta\Delta Ct_{BL}}}\right) \times 100\right)$$

$$= \left(\frac{2^{-\Delta\Delta Ct_{visit\ x}}}{2^{-\Delta\Delta Ct_{BL}}}\right) \times 100$$

The final % of baseline IFN 21-gene PD signature for each timepoint, for each participant, is the median of the individual gene % of baseline IFN 21-gene PD signature values calculated in Step 7. There will be one value per participant per post-Visit 1 visit:

$$Overall\ \%\ \text{of baseline IFN 21 gene PD signature} = median(\%\ \text{of BL IFN 21 gene PD signature}_{gene})$$

The fold change for each timepoint and each participant is the median of the linearized values for each gene, calculated in Step 6:

$$Fold\ Change = median(2^{-\Delta\Delta Ct})$$

7.2 Appendix B - Reference Ranges and TELVC for Laboratory Values

Parameter	Unit	Low value	Low decrease	High value	High increase
			Haematology		
Hemoglobin	g/L	≤60	NA	≥200	NA
		≤70 and decrease from BL ≥15			
Hematocrit	V/V	≤0.18	NA	≥0.64	NA
		≤0.21 and decrease from BL ≥15%			
WBC	10E9/L	≤2, <1	NA	≥20	NA
Neutrophils	10E9/L	<0.5	NA	≥20	NA
		<1.0 and decrease from BL ≥0.5			
Lymphocyte	10E9/L	≤0.5, ≤0.25	NA	≥10.0	NA
Monocytes	10E9/L	NA	NA	≥1.4, ≥5.0	NA
Eosinophils	10E9/L	NA	NA	≥1.5, ≥5.0	NA
Basophils	10E9/L	NA	NA	≥1.0, ≥2.0	NA
INR		NA	NA	≥4.5	NA
Platelet	10E9/L	≤20	NA	≥600	NA
		≤50 and decrease from BL ≥25			
			Biochemistry		
ALT	IU/L	NA	NA	≥3 x ULN, ≥5 x ULN	NA
AST	IU/L	NA	NA	≥3 x ULN, ≥5 x ULN	NA
ALP	IU/L	NA	NA	≥3 x ULN NA	NA
CK	IU/L	NA	NA	≥500, ≥2000	NA
GGT	IU/L	NA	NA	≥5 x ULN	NA
Total bilirubin	µmol/L	NA	NA	≥2 x ULN	NA
Albumin	g/L	≤20	NA	≥100	NA
		≤25 and decrease from BL ≥10		≥70 and increase from BL ≥10	
BUN	mmol/L	NA	NA	≥18	NA

Creatinine	umol/L	NA	NA	≥140, ≥190	NA
Sodium	mmol/L	≤132	NA	≥152	NA
Potassium	mmol/L	≤3	NA	≥5.5	NA
Chloride	mmol/L	≤90	NA	≥120	NA
Fasting Glucose	mmol/L	≤2.5	NA	≥7.0, ≥11.1	NA
Total Cholesterol	mmol/L	NA	NA	≥7.25	NA
Urinalysis					
Ketone Bodies/ Qualitative Erythrocytes/ U-Glucose					Negative/positive at baseline to ++, +++, ++++ at any post-baseline value Increase from baseline of at least ++ at any post-baseline value
Urine protein/ creatinine ratio	g/mmol	NA	NA	≥0.395	NA
Urine protein					Increase >= ++ >= ++ with BL negative, TRACE, or +
Fasting lipid profile					
HDL	mmol/L	≤0.8	NA	NA	NA
LDL	mmol/L	NA	NA	≥5.2	NA
Triglycerides	mmol/L	NA	NA	≥3.6, ≥5.4	NA

7.3 Appendix C - Reference Ranges and TELVC for Vital Signs

Parameter	Unit	Low value	Low decrease	High value	High increase
Pulse	Beats per minute	≤50	NA	≥120	NA
		≤50 and decrease from BL ≥20		≥120 and increase from BL ≥20	
Systolic blood pressure	mmHg	≤90	NA	≥160	NA
		≤90 and decrease from BL ≥20		≥160 and increase from BL ≥20	
Diastolic blood pressure	mmHg	≤50	NA	≥100	NA
		≤50 and decrease from BL ≥10		≥100 and increase from BL ≥10	

7.4 Appendix D - Restricted and prohibited medications – Programmatic Rules

The process of identifying restricted and prohibited medication use outside of the protocol-defined guidance consists of both a programmatic identification of possible infractions of the guidance and a manual review for confirmation. The programmatic identification and associated responsibilities will be completed by the AZ Programming Team. The manual review is conducted by an AstraZeneca led Medication Review Team (MRT). The process and the remit of the MRT is outlined in the Restricted and Prohibited Medication Review Charter.

This appendix describes the programming rules to identify potential restricted medications during the study. The identified medications will be flagged for adjudication if the condition(s) is/are fulfilled and the MRT will decide if the medication is a restricted medication and hence fulfil the criteria of a intercurrent events.

Also, see details on restricted medications in CSP version 3, Section 7.7.1.

A 1 Definitions

If the condition(s) cannot be evaluated due to missing information (e.g., unknown dose or frequency), the medication will be flagged for adjudication.

All medications will be programmatically classified as “restricted (RM)” or “prohibited (PM)”. Those that do not meet the criteria will be left blank.

A 1.1 Time period of interest

An intercurrent event of restricted medication use can occur from the start date of IP, up until last dose of IP + 28 days.

Medications started prior to first dose of IP that are ongoing with no end date reported will be reviewed.

A 1.2 Start and end date of Restricted Medication Use

The start/end date of a restricted medication use is defined as start/end date of the medication as recorded in the eCRF. The start/end date will be available for each medication in the review log.

In the event of overlapping or additional dose regimen intake, the start/end date of the restricted medication use is defined as the start/end date of the overlapping or additional dose which leads to the restricted medication use.

Missing data handling related to the dates of prior or concomitant medications are detailed in Section 3.3.1.3.

A 1.3 Tapering and Duration

In cases of dose increases burst and tapers with expected tapering requirements (e.g. including the criterion "for a period >14 days (in a row)"), multiple doses might lead to an overall duration greater than 14 days. The start date of restricted medication use will be the start date of the first increased dose of medication day.

Example 1:

Baseline OCS dose starts as 5 mg/day. During the treatment period or after date of first administration of investigational product (IP), a dose increase started on Day x of 40mg/day. Medication dose is tapered down to 30mg/day at from Day 8 to Day 12 and 20mg/day from Day 13 to Day 16.

Given the combined duration of medication intake is greater than 14 days and the dose is above the baseline dose, the start date of restricted medication use is defined as Day x at the beginning of the increased or burst dose (40 mg/day).

Day x		Day x+8		Day x+12		Day x+16
40mg						
		30mg				
				20mg		

A 1.4 Visit Window for Analysis

For visit-based analysis, the same visit windows as defined in Section 3.3.2 will be used.

A 1.5 Calculation of total daily dose

A 1.5.1 Conversion factors from oral intake frequency to daily intake

If the medication is taken less than daily the documented total daily dose in the Prior and Concomitant Medication/Treatment form reflects the dose at the day the medication is taken. To calculate the total daily dose to be used for statistical analyses, the documented total daily dose needs to be divided according to the given medication dose frequency (e.g., for medication given every week the documented daily dose needs to be divided by 7). If the total daily dose is missing for a corticosteroid, the available data of this medication will be reviewed by the medical team. If possible, the total daily dose will be derived manually from the available information. OCS administered PRN are not considered in the calculation of the daily dose.

A 1.5.2 Oral Prednisone Equivalent

All systemic steroids will be converted to a prednisone equivalent dose. Systemic steroids are defined as medications with coding ATCDTXT = "CORTICOSTEROIDS" with one of the following routes of administration: ORAL, IV, IM, SC and Inhaled. All systemic steroids

will be converted to a prednisone equivalent dose by multiplying the total daily dose (in mg) of the respective medication with its corresponding conversion factor in the table below. All converted systemic steroid doses given on the same day will be summed to create the total daily dose. The oral corticosteroids (OCS) total daily dose will only include corticosteroids with the route = ORAL.

The following table gives examples of prednisone equivalent doses (e.g., 50 mg cortisone is equivalent to 10 mg prednisone).

Conversion factors to oral prednisone equivalent dose

Medication (preferred term)	Conversion factor	Numerical Conversion factor [a]
Betamethasone	10/1.2	8.3333
Betamethasone dipropionate	10/1.2	8.3333
Betamethasone sodium phosphate	10/1.2	8.3333
Betamethasone valerate	10/1.2	8.3333
Betrospam	10/1.2	8.3333
Budesonide	10/2.25	4.4444
Celestona bifas	10/1.2	8.3333
Cortisone	1/5	0.2
Cortisone acetate	1/5	0.2
Cronolevel	10/1.2	8.3333
Decadron (nos)	10/1.5	6.6667
Deflazacort	1/1.2	0.8333
Dexamethasone	10/1.5	6.6667
Dexamethasone acetate	10/1.5	6.6667
Dexamethasone sodium phosphate	10/1.5	6.6667
Fludrocortisone	0	0
Fluocortolone	3	3
Hydrocortisone	1/4	0.25
Hydrocortisone acetate	1/4	0.25
Hydrocortisone sodium succinate	1/4	0.25
Meprednisone	5/4	1.25
Methylprednisolone	5/4	1.25
Methylprednisolone acetate	5/4	1.25
Methylprednisolone sodium succinate	5/4	1.25
Paramethasone	5/2	2.5
Prednisolone	1	1
Prednisolone acetate	1	1
Prednisolone sodium phosphate	1	1

Prednisolone sodium succinate	1	1
Prednisone	1	1
Prednisone acetate	1	1
Triamcinolone	5/4	1.25
Triamcinolone acetate	5/4	1.25
Triamcinolone acetonide	5/4	1.25

[a] The prednisone equivalent total daily dose (in mg) will be calculated as total daily dose (in mg) of the respective medication multiplied with the conversion factor.

A 1.6 All Other Concomitant Medications During Double-Blind Treatment Period

The total daily dose for other medications is derived in the same way as other corticosteroids except the standardization to an equivalent dose.

A 1.7 Calculation of Methotrexate baseline dose

The methotrexate dose at Day 1 will be determined as the average weekly dose between ICF sign and Day 1. This will only be used in the determination of whether the methotrexate dose during the study is beyond the protocol limit.

A 2 [redacted]

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

		[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

		[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

		[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

A 3 [redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

[redacted]