AstraZeneca

Statistical Analysis Plan

Study Code	D3461C00005
Edition Number	3.0
Date	20 August 2018

A Multicentre, Randomised, Double-blind, Placebo-controlled, Phase 3 Study Evaluating the Efficacy and Safety of Two Doses of Anifrolumab in Adult Subjects with Active Systemic Lupus Erythematosus

A Multicentre, Randomised, Double-blind, Placebo-controlled, Phase 3 Study Evaluating the Efficacy and Safety of Two Doses of Anifrolumab in Adult Subjects with Active Systemic Lupus Erythematosus

Study Statistician [redacted] 21 Aug 2018
Date

A Multicentre, Randomised, Double-blind, Placebo-controlled, Phase 3 Study Evaluating the Efficacy and Safety of Two Doses of Anifrolumab in Adult Subjects with Active Systemic Lupus Erythematosus

Global Product Statistician [redacted]

21 Aug 2018
Date

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ADA	Anti-drug antibodies
AE	Adverse event
AESI	Adverse event of special interest
ALT	Alanine transaminase
AST	Aspartate transaminase
AUC	Area under the curve
BDR	Blind data review
BICLA	British Isles Lupus Assessment Group-based Composite Lupus Assessment
BILAG	British Isles Lupus Assessment Group
BMI	Body mass index
C-SSRS	Columbia–Suicide Severity Rating Scale
C3	Third component of complement
C4	Fourth component of complement
CH50	Total haemolytic complement
CI	Confidence interval
CLASI	Cutaneous Lupus Erythematosus Disease Area and Severity Index
CMH	Cochran-Mantel-Haenszel
CSR	Clinical Study Report
DBL	Database lock
dsDNA	Double-stranded deoxyribonucleic acid
ECG	Electrocardiogram
EDV	Early Discontinuation Visit
EQ-5D-5L	EuroQoL 5 dimensions
FACIT-F	Functional Assessment of Chronic Illness Therapy-FATIGUE
GGT	Gamma glutamyl transferase
IFN	Interferon
IV	Intravenous
IXRS	Interactive voice/web response system
LOCF	Last observation carried forward
LTE	Long-term extension

Abbreviation or special term	Explanation
MACE	Major adverse cardiovascular events
MAR	Missing at Random
MCS	Mental Component Score
MedDRA	Medical Dictionary for Regulatory Activities
nAb	Neutralising antibodies
NRS	Numeric rating scale
OCS	Oral corticosteroids
PCS	Physical Component Score
PD	Pharmacodynamic
PGA	Physician's Global Assessment
PHQ-8	Personal Health Questionnaire Depression Scale-8
PK	Pharmacokinetic(s)
PtGA	Patient Global Assessment
Q	Question
Q-Q	Quartile-quartile
Q4W	Every 4 weeks
QoL	Quality of life
SAE	Serious adverse event
SD	Standard deviation
SDI	Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index
SE	Standard Error
SF-36-v2 (acute)	Short Form 36 version 2 (acute recall)
SLE	Systemic lupus erythematosus
SLEDAI-2K	Systemic Lupus Erythematosus Disease Activity Index 2000
SOC	Standard of care
SRI(X)	Systemic Lupus Erythematosus Responder Index of ≥X
TELVC	Treatment emergent laboratory/vital signs changes
TFL	Tables, Figures and Listings
ULN	Upper limit of normal
VAS	Visual analogue scale
WHO-DD	World Health Organization Drug Dictionary

Abbreviation or special term	Explanation
WPAI	Work productivity and Activity Impairment

AMENDMENT HISTORY

Date	Brief description of change
03 May 2017	Updates according to the changes in the table, figure and listing (TFL) shells after review Updates to provide additional details to support programming activities Updates to specify handling and examination of missing data (sensitivity and tipping point analyses) Updates to clarify database lock and unblinding after all subjects completed Week 52.
17 August 2018	Updates following Blind Data Review: Clarifications added, inconsistencies removed, additional analyses added, and minor changes to some analyses. Updated to clarify handling of (partially) missing AE dates Analysis model for longitudinal data changed to repeated measures model. Additional rules for pooling of strata for CMH analysis defined Updates to document handling of site excluded from analysis. Updates to further clarify week 52 lock activities.

1. STUDY DETAILS

1.1 Study objectives

1.1.1 Primary objective

Primary Objective:	Outcome Measures:
To evaluate the effect of anifrolumab 300 mg compared to placebo on disease activity as measured by the difference in the proportion of subjects who achieve a systemic lupus erythematosus (SLE) Responder Index of ≥4 (SRI[4]) at Week 52	Composite variable SRI(4), defined by the following criteria: - Reduction from baseline of ≥4 points in the Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K) and - No new organ systems affected as defined by 1 or more British Isles Lupus Assessment Group (BILAG-2004) A or 2 or more BILAG-2004 B items compared to baseline using BILAG-2004 and - No worsening from baseline in subjects' lupus disease activity defined by an increase ≥0.30 points on a 3-point Physician's Global Assessment (PGA) visual analogue scale (VAS) and - No discontinuation of investigational product or use of restricted medications beyond the protocol-allowed threshold[a] before assessment

[a] Any medication classified as restricted medications as described in Appendix A

1.1.2 Secondary objectives

Key Secondary Objectives:	Outcome Measures:
To evaluate the effect of anifrolumab 300 mg compared to placebo on:	
The proportion of subjects with SRI(4) at Week 52 in the interferon (IFN) test-high sub-group	SRI(4) (see outcome measure for primary objective)
The proportion of subjects who achieve an Oral corticosteroids (OCS) dose ≤7.5 mg/day at Week 40, which is maintained through Week 52 in the sub-group of subjects with baseline OCS ≥10 mg/day	Maintained OCS reduction defined by the following criteria: - Achieve an OCS dose of ≤7.5 mg/day prednisone or equivalent by Week 40 and - Maintain an OCS dose ≤7.5 mg/day prednisone or equivalent from Week 40 to Week 52 and - No discontinuation of investigational product or use of restricted medications beyond the protocol-allowed threshold[a] before assessment

The proportion of subjects with a ≥50% reduction in Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI) activity score at Week 12 in the sub-group of subjects with baseline CLASI activity score ≥10	50% reduction in CLASI activity score compared to baseline defined by the following criteria: - Achieve ≥50% reduction of CLASI activity score at Week 12 compared to baseline and - No discontinuation of investigational product or use of restricted medications beyond the protocol-allowed threshold[a] before assessment
The proportion of subjects with SRI(4) at Week 24	SRI(4) (see outcome measure for primary objective)
The annualised flare rate through 52 weeks	Annualised flare rate with flare defined as either 1 or more new BILAG-2004 A or 2 or more new BILAG-2004 B items compared to the previous visit
Other Secondary Objectives:	**Outcome Measures:**
To evaluate the effect of anifrolumab 150 mg compared to placebo on disease activity as measured by the difference in the proportion of subjects who achieve SRI(4) at Week 52	SRI(4) (see endpoint for primary objective)
To assess the difference between anifrolumab 300 mg and placebo on measures of disease activity including levels of SRI response other than 4, British Isles Lupus Assessment Group-based Composite Lupus Assessment (BICLA), the individual components of SRI, and the number of swollen and tender joints at Week 52, as well as SRI and BICLA over time	SRI(4), SRI(5), SRI(6), SRI(7), SRI(8), BICLA response, BILAG-2004, SLEDAI-2K, PGA, Major Clinical Response, Partial Clinical Response, and joint count
To assess the difference between anifrolumab 300 mg and placebo on measures of organ damage, ie, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI) at Week 52	SDI
To assess the difference between anifrolumab 300 mg and placebo on subject-reported health status, health-related quality of life, and other subject-reported outcome measures of fatigue, pain, patient global assessment, and work productivity at Week 52	Short Form 36 version 2 (acute recall) (SF-36-v2 [acute]), Pain numeric rating scale (NRS), Functional Assessment of Chronic Illness Therapy-FATIGUE (FACIT-F), Patient Global Assessment (PtGA), Lupus quality of life (QoL), EuroQoL 5 dimensions (EQ-5D-5L), Work Productivity and Activity Impairment (WPAI)-Lupus, and Medical Resource Use Questionnaire

To evaluate the pharmacokinetics (PK), immunogenicity, and pharmacodynamics (PD) of anifrolumab	Anifrolumab concentration and PK parameters, anti-drug antibodies (ADA), 21-gene type I IFN gene signature, anti-double-stranded deoxyribonucleic acid (dsDNA) antibodies, third component (C3), fourth component (C4), and total haemolytic (CH50) complement levels

[a] Any medication classified as restricted medications as described in Appendix A

1.1.3 Safety objective

Safety Objective:	**Outcome Measures:**
To evaluate the safety and tolerability of anifrolumab	Adverse events (AEs) (including AEs of special interest [AESIs]), vital signs, physical examination, 12-lead electrocardiograms (ECG), modified SELENA Flare Index based flares, clinical laboratory tests (haematology, clinical chemistry, urinalysis), Columbia Suicide Severity Rating Scale (C-SSRS), and Personal Health Questionnaire Depression Scale-8 (PHQ-8)

1.1.4 Exploratory objective

Exploratory Objective:	**Outcome Measures:**
[redacted]	[redacted]

1.2 Study design

This is a Phase 3, multicentre, multinational, randomised, double-blind, placebo-controlled study to evaluate the efficacy and safety of an intravenous (IV) treatment regimen of anifrolumab versus placebo in adult subjects with moderately to severely active, SLE who are receiving standard of care (SOC) treatment. The study will be performed in adult subjects aged 18 to 70 years of age.

Approximately 450 subjects receiving SOC treatment will be randomised in a 1:2:2 ratio to receive a fixed IV dose of 150 mg anifrolumab, 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. Investigational product will be administered as an IV infusion via an infusion pump over a minimum of 30 minutes, Q4W.

Randomisation will be stratified using the following factors:

- SLEDAI-2K score at screening (<10 points versus ≥10 points)
- Week 0 (Day 1) OCS dose (<10 mg/day versus ≥10 mg/day prednisone or equivalent)
- Results of the IFN test (high versus low)

This study includes:

- **A Screening Period**: Up to 30 days
- **Treatment Period**: A 52-week double-blind treatment period with investigational product administration Q4W from Week 0 to Week 48 for a total of 13 doses
- **At Week 52**, subjects will have two options:
  - If eligible, enrol into the long-term extension (LTE) study in which anifrolumab 300 mg or placebo will be administered Q4W

    OR
  - Continue in the current study for another 8 weeks to complete a 12-week safety follow-up after the last dose of investigational product (last dose of investigational product will be given in Week 48)

The total study duration could be up to approximately 64 weeks for subjects who do not enrol into the LTE study (including screening period) and up to approximately 56 weeks (including screening period) for those subjects who do enrol into the LTE study.

Database lock (DBL) and unblinding will occur after the last subject has completed Week 52 (Visit 14/EDV). Blinding of subjects and investigators will be maintained to the greatest extent possible after the DBL at Week 52 until Last Subject Last Visit in the LTE study.

1.2.1 Steroid burst

From Week 0 (Day 1) to Week 12, subjects may receive **only** 1 burst of corticosteroids for an increase in SLE disease activity or to control non-SLE related disease (eg, asthma or chronic obstructive pulmonary disease exacerbation). Subjects receiving more than 1 burst during the first 12 weeks of treatment will be considered non-responders for subsequent assessments of disease activity, regardless of the reason for the burst (SLE or non-SLE activity). More details are given in Sections 3.3.4 in the clinical study protocol.

1.2.2 Protocol-specified steroid tapering

Steroid tapering to a target OCS dose of ≤7.5 mg/day **must** be attempted in all subjects with a baseline OCS dose ≥10.0 mg/day. This will commence at Week 8 and continue stepwise until the target is reached, unless at least 1 of the following criteria is met:

- SLEDAI-2K activity which is worsened compared to baseline in major organ systems (renal, central nervous system, cardiopulmonary, vasculitis, fever, thrombocytopenia, or haemolytic anaemia, or gastrointestinal activity)
- Newly-affected organ system(s) based on the SLEDAI-2K, excluding serological abnormalities (dsDNA antibodies, hypocomplementemia)
- Moderate to severe skin disease as reflected by a CLASI activity score of ≥10
- Moderate to severe arthritis disease as reflected by an active joint count of ≥8 tender and/or swollen joints

Investigators will not be required, but may continue, to taper OCS dose beyond the target of 7.5 mg/day up to Week 40 based on disease activity. If a subject has an increase in disease activity secondary to OCS tapering, they may increase the dose up to a maximum of the baseline OCS therapy dose from Week 8 up to Week 40 without the subject being considered a non-responder for subsequent assessments of disease activity. Subjects who require OCS dose above their baseline level may continue in the study but will be considered non-responders for subsequent assessments of disease activity.

Steroid tapering will not be permitted after Week 40.

1.3 Number of subjects

A total of 450 subjects receiving SOC treatment will be randomised 1:2:2 to treatment with 150 mg anifrolumab, 300 mg anifrolumab or placebo.

The sample size is primarily driven by the need to acquire an adequate safety database size, as well as the ability to assess key secondary endpoints. The primary endpoint is the difference in proportion of subjects achieving SRI(4) at Week 52 comparing anifrolumab 300 mg to placebo. With assumed proportions of SRI(4) of 39% and 63% in the placebo and anifrolumab 300 mg groups, respectively, 180 subjects/arm yields more than 99% power to reject the hypothesis of no difference using a 2-sided alpha of 0.05. The minimal detectable difference in SRI(4) between anifrolumab 300 mg versus placebo is approximately 10% with this sample size.

Estimates of power for 2 key secondary endpoints are listed below. These calculations assume that the primary endpoint is met, and the testing of the key secondary endpoints is therefore allowed. Each endpoint is tested using a weighted Holm procedure, and the alpha given by the assigned weight in the first step of the algorithm:

- Difference in proportion of subjects achieving SRI(4) at Week 52 in the IFN test-high sub group: Given 75% of subjects are IFN test-high; proportions of SRI(4) in the IFN test-high subgroup of 35% and 61% in the placebo and anifrolumab treatment groups, respectively; a 2-sided alpha of 0.04 yields 98% power.

- Difference in proportion of subjects who achieve an OCS dose ≤7.5 mg/day at Week 40, which is maintained through Week 52 in the sub-group of subjects with baseline OCS ≥10 mg/day: Given 60% of subjects have an OCS dose of at least 10 mg at baseline; proportions of subjects tapering the OCS dose of 32% and 59% in the placebo and anifrolumab treatment groups, respectively; a 2-sided alpha of 0.004 yields 87% power.

The assumptions of the effect sizes and sizes of subgroups used for the calculations above are based on the observed results in the interim analyses of study CD IA MEDI 546-1013.

2. ANALYSIS SETS

2.1 Definition of analysis sets

2.1.1 All subjects

This analysis set will comprise all subjects screened for the study and will be used for reporting of disposition and screening failures.

2.1.2 Full analysis set

The full analysis set will be used as the primary population for reporting efficacy and safety data. This comprises all subjects randomised into the study who receive at least 1 dose of investigational product and will be analysed according to randomised treatment (modified Intention-To-Treat). Any major deviations from randomised treatment will be listed and considered when interpreting the safety data.

2.1.3 Pharmacokinetic analysis set

All subjects who received anifrolumab and who had at least 1 quantifiable serum PK observation post first dose, will be included in the PK analysis dataset. All PK summaries will be based on this analysis set.

2.2 Protocol deviations

Protocol deviations that may greatly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being will be summarized.

The final list of protocol deviations will be finalized and documented prior to unblinding the study data, and will be based on deviations as defined in the Protocol Deviation Guidance.

During the blind data review (BDR) meeting, protocol deviations will be classified as important or other depending on the impact of the deviation to the analysis. Only important protocol deviations will be listed and tabulated in the Clinical Study Report (CSR).

3. PRIMARY AND SECONDARY VARIABLES

Baseline is defined as the last measurement prior to randomisation and dose administration on Day 1. If the Day 1 value is missing or is invalid or is collected after administration of investigational product, the latest assessment prior to dose administration on Day 1 will serve as baseline.

At some sites, the device to collect subject reported outcomes was not synchronized with the real time. Therefore, baseline for the following subject reported outcomes will be derived based on the date only, ignoring the time of the assessment.

- Short Form 36 version 2 (acute recall)
- Pain numerical rating scale
- Functional Assessment of Chronic Illness Therapy – FATIGUE
- Patient Global Assessment
- Lupus quality of life scale
- EuroQol 5 dimensions
- Work productivity and Activity Impairment – Lupus
- Medical Resource Questionnaire
- C-SSRS
- Personal Health Questionnaire Depression Scale

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.

Values of BILAG-2004, SLEDAI-2K, CLASI, and PGA assessments will be adjudicated values.

3.1 Primary outcome variable

The primary endpoint used to evaluate the effect of anifrolumab 300 mg compared to placebo on disease activity is the difference in proportion of subjects achieving SRI(4) at Week 52, where a subject achieves SRI(4) if all of the following criteria are met:

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

  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 subjects' lupus disease activity defined by an increase ≥0.30 points on a 3-point PGA VAS

- No permanent premature discontinuation of investigational product (according to eCRF form "Discontinuation of Investigational Product")
- No use of restricted medications beyond the protocol-allowed threshold on or before the date of last Week 52 assessment used to derive SRI(4)

  Restricted medications are defined in Appendix A.

If any of the criteria cannot be evaluated at Week 52 (eg, due to missing values) that criterion will be imputed using last observation carried forward (LOCF) and SRI(4) derived based on the complete data. This applies only if Week 48 data is not missing, otherwise the subject will be defined as not achieving SRI(4) at Week 52.

In addition, the individual conditions of SRI(4) (SLEDAI-2K reduction ≥4 points, no new organ systems affected, no worsening of subjects' lupus disease activity, no permanent discontinuation of investigational product, and no use of rescue medication) will be assessed at Week 24 and Week 52 by treatment.

3.2 Key secondary outcome variables

3.2.1 SRI(4) at Week 52 in IFN test-high subjects

The key secondary endpoint used to evaluate the effect of anifrolumab 300 mg compared to placebo on disease activity in the IFN test-high subgroup is the difference in proportion of subjects achieving SRI(4) at Week 52 in subjects classified as IFN test-high at baseline. SRI(4) is defined in Section 3.1.

3.2.2 Oral corticosteroid management

The key secondary endpoint used to evaluate the effect of anifrolumab 300 mg versus placebo on the ability to reduce the OCS dose in subjects with baseline OCS ≥10 mg/day prednisone or equivalent is the difference in proportion of subjects with maintained OCS reduction. 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

  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 4.1.3) will be used as date of Week 40. If no such assessment falls into the respective time window, then the target date for the timepoint will be used instead.
- 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 (ie, no dose greater than the dose at Week 40) between Week 40 and Week 52.

  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 4.1.3) will be used as date of Week 52. If no such assessment falls into the respective time window, then the target date for the timepoint will be used instead.

- No permanent premature discontinuation of investigational product (according to eCRF form "Discontinuation of Investigational Product")
- No use of restricted medications beyond the protocol-allowed threshold on or before the date of Week 52 (as given above)

  Restricted medications are defined in Appendix A.

If any of the conditions cannot be evaluated at Week 52 (eg, due to missing values) the subject is defined as not reaching a maintained OCS reduction.

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

3.2.3 Skin lesions

The key secondary endpoint used to evaluate the effect of anifrolumab 300 mg versus placebo on inflammatory cutaneous lupus lesions in subjects with baseline CLASI activity score ≥10 is the difference in proportion of subjects with an at least 50% reduction in CLASI activity score at Week 12. An at least 50% reduction in CLASI activity score is defined as meeting all the following criteria:

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

  The CLASI activity score will be derived as the sum of all single activity scores (13 locations of erythema, 13 locations of scale/hypertrophy, mucous membrane lesions, recent hair loss, and non-scarring alopecia). A ≥50% reduction is reached if the percentage change is ≤-50%.
- No permanent premature discontinuation of investigational product before assessment (ie, duration of exposure ≥ study day of date of CLASI assessment at Week 12)
- No use of restricted medications beyond the protocol allowed threshold on or before the date of CLASI assessment at Week 12

  Restricted medications are defined in Appendix A.

If any of the criteria cannot be evaluated at Week 12 (eg, due to missing values) that criterion will be imputed using LOCF and an at least 50% reduction in CLASI activity score derived based on the complete data. This applies only if Week 8 data is not missing, otherwise the subject will be defined as not reaching a ≥50% reduction in CLASI activity score at Week 12.

3.2.4 SRI(4) at Week 24

The key secondary endpoint used to evaluate the early effect of anifrolumab 300 mg compared to placebo on disease activity is the difference in proportion of subjects achieving SRI(4) at Week 24. SRI(4) at Week 52 is defined in Section 3.1. SRI(4) at Week 24 will be derived in the same way with using assessments at Week 24 instead of Week 52 (with allowed LOCF of Week 20 data) and no permanent premature discontinuation of investigational product before assessment (ie, duration of exposure ≥ study day of date of last Week 24 assessment used to derive SRI(4)).

3.2.5 Flares

The key secondary endpoint used to evaluate the effect of anifrolumab 300 mg 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 (ie, 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).

The occurrence of a new flare will be checked for each available visit versus the previous 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 will be 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 Week 52 (date of BILAG-2004 assessment at Week 52) or up to the date of last available BILAG-2004 assessment up to and including week 52 in case of premature study discontinuation and will be derived as date of Week 52/ date of last BILAG-2004 assessment minus date of first administration of investigational product + 1.

For the sensitivity analyses of flares while on treatment, the flare exposure time is the time between day of first dose of investigational product and day of last dose of investigational product plus 28 days, both inclusive. All flares occurring within the flare exposure time will be considered for this analysis (even if occurring after the week 52 visit timepoint).

3.3 Other secondary outcome variables

3.3.1 Assessment of disease activity

3.3.1.1 SRI(4) of anifrolumab 150 mg treatment group

The effect of anifrolumab 150 mg compared to placebo on disease activity will be assessed by the difference in proportions of subjects achieving SRI(4) at Week 52. SRI(4) is defined in Section 3.1.

3.3.1.2 Supportive SRI variables

In addition to the endpoint described in Section 3.1, the difference between anifrolumab 300 mg and placebo in SRI at Week 52 will be assessed using levels other than 4, ie, SRI(5), SRI(6), SRI(7), and SRI(8), where SRI(X) (X=5, 6, 7, or 8) is defined by the proportion of subjects who meet the following criteria:

- Reduction from baseline of ≥X points in the SLEDAI-2K
  Reduction from baseline will be derived as described in Section 3.1.
- 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 subjects' lupus disease activity defined by an increase ≥0.30 points on a 3-point PGA VAS
- No permanent premature discontinuation of investigational product (according to eCRF form "Discontinuation of Investigational Product")
- No use of restricted medications beyond the protocol-allowed threshold on or before the date of last Week 52 assessment used to derive SRI(X)

  Restricted medications are defined in Appendix A.

If any of the criteria cannot be evaluated at Week 52 (eg, due to missing values) that criterion will be imputed using LOCF and SRI(X) derived based on the complete data. This applies only if Week 48 data is not missing, otherwise the subject will be defined as not achieving SRI(X) at Week 52.

The difference between anifrolumab 300 mg and placebo in the proportion of subjects achieving SRI(X), X=4, 5, 6, 7, or 8 will also be assessed longitudinally over time up to Week 52.

3.3.1.3 BICLA

The effect of anifrolumab 300 mg versus placebo on disease activity will also be assessed using the difference in proportion of subjects meeting the criteria for BICLA response at Week 52, where a subject is a BICLA responder if 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 as defined as an increase from baseline of >0 points in SLEDAI-2K

  Increase from baseline corresponds to the change from baseline. The derivation of the SLEDAI-2K is described in Section 3.1.
- No worsening from baseline in subjects' lupus disease activity defined by an increase ≥0.30 points on a 3-point PGA VAS;
- No permanent premature discontinuation of investigational product (according to eCRF form "Discontinuation of Investigational Product")
- No use of restricted medications beyond the protocol-allowed threshold on or before the date of last Week 52 assessment used to derive BICLA response

  Restricted medications are defined in Appendix A.

If any of the criteria cannot be evaluated at Week 52 (eg, due to missing values) that criterion will be imputed using LOCF and BICLA response derived based on the complete data. This applies only if Week 48 data is not missing, otherwise the subject will be defined as a BICLA non-responder at Week 52.

The difference between anifrolumab 300 mg and placebo in the proportion of subjects achieving BICLA will also be assessed longitudinally over time up to Week 52.

3.3.1.4 Supportive outcome variables of the Individual components of SRI and BICLA

Individual conditions

Individual conditions of BICLA (reduction of all BILAG-2004 A and B, no worsening from baseline in SLEDAI-2K, no worsening in subjects' lupus disease activity, no permanent discontinuation of investigational product, and no use of rescue medication) are defined in Section3.3.1.3. The individual conditions will be assessed at Week 24 and Week 52 by treatment.

SLEDAI-2K

SLEDAI-2K (calculated as described in Section 3.1) will be evaluated using the difference in mean change from baseline longitudinally over time to Week 52.

Scores for the SLEDAI organ systems will be derived in the same way as SLEDAI-2K but using the scores for the respective items only. The SLEDAI organ systems are defined as follows:

- Central nervous system: seizure, psychosis, organic brain syndrome, visual disturbance, cranial nerve disorder, and lupus headache
- Vascular: CVA (cerebrovascular accident) and vasculitis
- Musculoskeletal: arthritis and myositis
- Renal: urinary casts, hematuria, proteinuria, and pyuria
- Mucocutaneous: rash, alopecia, and mucosal ulcers
- Cardiovascular system and respiratory: pleurisy and pericarditis
- Immunology: low complement and increased DNA binding
- Haematological and fever: fever, thrombocytopenia, and leukopenia

For each SLEDAI organ system, the proportion of subjects with an improvement (ie, a SLEDAI organ system score less than the corresponding score at baseline) at Week 24 and Week 52, respectively, will be assessed for subjects with an organ system involvement at baseline (ie, a SLEDAI organ system score greater than 0).

Physician's Global Assessment

The difference between anifrolumab 300 mg and placebo in the mean change from baseline in PGA (measured on a VAS ranging from 0 to 3) will be assessed by visit up to Week 52.

BILAG-2004

BILAG-2004 grades will be presented by organ system. Furthermore, a BILAG-2004 global score will be 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.

Clinical Response

In addition, the effect of anifrolumab 300 mg versus placebo on Major Clinical Response and Partial Clinical Response will be evaluated where these variables are defined as:

- Difference in proportion of subjects who achieve Major Clinical Response, ie, a subject with BILAG-2004 C scores or better at Week 24 with no new BILAG-2004 A or BILAG-2004 B scores and maintenance of response with no new BILAG-2004 A or B scores between Week 24 and Week 52.
- Difference in proportion of subjects who achieve Partial Clinical Response, ie, a subject 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.

If BILAG-2004 grades are missing or incomplete at any relevant visit, the missing data will be imputed using LOCF and Major and Partial Clinical Response derived based on the complete data. This applies only if data of the previous visit is not missing, otherwise (ie, in case of at least 2 subsequent visits with missing data) the subject will be defined as non-responder.

3.3.1.5 Active, swollen and tender joints

The endpoints used to evaluate the effect of anifrolumab 300 mg versus placebo on active, swollen, and tender joints are:

- Difference in change from baseline to Week 52 in the number of active, swollen, and tender joints, respectively;
- Difference in proportion of subjects with at least 8 swollen and at least 8 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 proportion of subjects with at least 8 swollen and at least 8 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.

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

An at least 20% reduction and at least 50% reduction, respectively, is reached if the percentage change is ≤-20% and ≤-50%, respectively, no restricted medications beyond the protocol allowed threshold were used on or before the assessment, and there was no permanent premature discontinuation of investigational product (according to eCRF form "Discontinuation of Investigational Product"). To achieve at least a 20% reduction and at least a 50% reduction, respectively, the reduction in the number of joints needs to be reached in swollen and tender joints separately.

If the change from baseline in the number of joints cannot be evaluated at Week 52 (eg, due to missing values) the change from baseline will be imputed using LOCF. This applies only if

Week 48 data is not missing, otherwise the subject will be defined as not achieving a reduction.

Furthermore, the change from baseline in the number of active, swollen and tender joints, respectively, will be explored longitudinally up to Week 52 for all subjects.

3.3.1.6 Change in SDI

The endpoint used to evaluate the effect of anifrolumab 300 mg versus placebo on irreversible damage in SLE subjects is the difference in mean change in SDI global score from baseline to Week 52.

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.

Furthermore, the number and percentage of subjects 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'.

3.3.1.7 Supportive outcome variables for the assessment of OCS use

The total daily OCS dose will be assessed longitudinally over time up to Week 52. For the derivation of OCS dose at a specific visit, the target date for the respective visit (as described in Section 4.1.3) will be used.

The standardised area under the curve (AUC) of OCS dose up to Week 52 will be calculated as follows:

For each single daily dose, the duration of the single dose will be calculated as end date – start date + 1. If the start date is before Day 1, Day 1 will be used instead. If the end date is after the date of Week 52 (date of Visit 14), the date of Visit 14 will be used instead. The AUC for each single dose will be derived by the daily dose (mg/day) multiplied with the duration (days). The AUC up to Week 52 is the sum of the AUCs of the single doses. The AUC of OCS dose will only be calculated if all necessary data are available up to Visit 14 (ie, daily prednisone equivalent OCS dose can be calculated). For subjects who discontinued the study before Visit 14, the AUC will be calculated up to the date of study discontinuation. If a subject does not receive any OCS dose (ie, no corresponding medication documented) the AUC for this subject will be set to 0. The standardised AUC will be derived as AUC divided by the available days (date of Visit 14 / date of early discontinuation – date of Day 1 + 1) multiplied by 364 (52 weeks).

3.3.1.8 Supportive outcome variables for the assessment of skin lesions

For all subjects with a CLASI activity score ≥10 at baseline the reduction of CLASI will be compared between Week 12 and Week 52. An at least 50% reduction in CLASI activity score at Week 12 is defined in Section 3.2.3. The reduction at Week 52 is defined in the same way with using assessments at Week 52 (with allowed LOCF of Week 48 data) instead of Week 12 and no permanent premature discontinuation of investigational product (according to eCRF form "Discontinuation of Investigational Product"). Maintenance of effect in CLASI activity score is defined as an at least 50% reduction in CLASI activity score at Week 12 and Week 52.

The difference between anifrolumab and placebo in the mean change from baseline in CLASI activity as well as CLASI damage score will be evaluated longitudinally over time up to Week 52. The CLASI damage score will be calculated as the sum of the scores for dyspigmentation for 13 locations, scarring/atrophy/panniculitis for 12 locations, and scarring of the scalp. If dyspigmentation lasts more than one year, the sum of the dyspigmentation scores from the 13 locations will be multiplied with 2 in the above formula.

3.3.1.9 Supportive outcome variables for the assessment of flares

In addition to the variable in Section 3.2.5, the annualised rate of flares will also be evaluated with a supportive definition of flares. The flare rate and annualized flare rate for the following definition of flares will be derived similarly as described in Section 3.2.5.

- Flare versus baseline:
  A flare is defined as either 1 or more new BILAG-2004 A or 2 or more new BILAG-2004 B compared to baseline and a change versus the previous visit. This implies that a new BILAG-2004 A score maintained in subsequent visits will not be counted as a new flare. However, a change of the situation in a subsequent visit (eg, one of the 2 new BILAG-2004 B scores of the previous visit improves to C but another worsens to B or an occurrence of a new BILAG-2004 B score in addition to the BILAG-2004 A score defining the previous flare) will lead to a new flare.

In addition, time to flare for the first flare will be assessed for flares as defined in Section 3.2.5 and the flares versus baseline. The time to flare will be derived as date of first flare minus date of first administration of investigational product. If the subject did not have a flare, the time to flare will be censored at the end of the flare exposure time (as defined in Section 3.2.5).

3.3.1.10 Lupus Low Disease Activity State

The endpoint used to evaluate the effect of anifrolumab versus placebo on disease activity state is the difference in the proportion of subjects with response in Lupus Low Disease Activity State (LLDAS) at Week 52. A subject will be considered to be an LLDAS responder at a specific visit if all the criteria below are met:

- SLEDAI-2K score ≤4, with no activity in major organ systems and no hemolytic anemia
    - Any subject 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 subject 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 [38° C], excluding infectious cause)
    - Hemolytic anemia will be assessed through BILAG-2004 evaluation at the visit. Any subject with present evidence of active hemolysis as assessed by BILAG-2004 will be excluded from LLDAS at that visit.
- No new lupus disease activity compared with the previous assessment: any patient with present activity (item score >0) in at least one new SLEDAI-2K item (irrespective of organ system) compared with the previous visit (even if the previous visit was unscheduled) won't attain LLDAS.
- PGA VAS score (scale 0–3) ≤1: any patient with PGA score greater than 1 at a study visit won't attain LLDAS at that visit.
- OCS dose ≤7.5 mg /day prednisone or equivalent: any subject 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.
- Well-tolerated standard maintenance doses of immunosuppressive drugs and approved biological agents as allowed and specified in the Clinical Study Protocol: no use of restricted medications beyond the protocol allowed threshold used on or before the

assessment. It will be assumed that any discontinuation of permitted immunosuppressives in the period starting on the day after the previous scheduled visit (if not missed) and ending on the day of the visit is due to toxicity and the subject 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 permitted immunosuppressives were discontinued at the day of the visit or any day of the 28-days period preceding that visit. The Permitted (under the restrictions in Appendix A) immunosuppressives are: Methotrexate, Azathioprine, Mycophenolate mofetil, Mycophenolic acid, and Mizoribine.

- No permanent discontinuation of IP.

If any of the criteria for SLEDAI-2K, BILAG or PGA cannot be evaluated at particular visit (eg, due to missing values) that criterion will be imputed using LOCF and LLDAS derived based on the complete data. This applies only if the data at the previous visit is not missing, otherwise the subject will be defined as not achieving LLDAS at Week 52.

Further, to define the "day of a visit" in the above, the date of the last SLEDAI-2K, BILAG 2004 or PGA assessment used for analysis in the time window of the respective timepoint (as described in Section 4.1.3) will be used. If no assessment falls within a defined window, the corresponding visit will be considered as missing.

The difference between anifrolumab and placebo in the proportion of LLDAS responders will also be assessed longitudinally over time up to Week 52.

3.3.2 Subject reported outcome variables

3.3.2.1 Short Form 36 version 2 (acute recall)

The difference between anifrolumab 300 mg and placebo in the mean change from baseline in SF-36-v2 (acute) domain scores (Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional, and Mental Health), Physical Component Score (PCS) and Mental Component Score (MCS) will be assessed by visit up to Week 52.

Responder criteria are defined for each of the 8 domain scores, the PCS and MCS as an increase from baseline to Week 52 above the following thresholds (based on US general population data from 2009, Maruish, 2011), no restricted medications beyond the protocol allowed threshold on or before the assessment, and no permanent premature discontinuation of investigational product (according to eCRF form "Discontinuation of Investigational Product"):

Table 1 Threshold values for the SF-36-v2 (acute) responder criteria

Domain/Component	Meaningful Change Threshold
Physical Functioning	4.3
Role Physical	3.4
Bodily Pain	6.2

Domain/Component	Meaningful Change Threshold
General Health	7.2
Vitality	6.2
Social Functioning	6.9
Role Emotional	4.5
Mental Health	6.2
PCS	3.4
MCS	4.6

If the change from baseline in the respective score cannot be evaluated at Week 52 (eg, due to missing values) the change from baseline will be imputed using LOCF. This applies only if Week 48 data is not missing, otherwise the subject will be defined as a non-responder.

3.3.2.2 Pain Numerical Rating Scale

The difference between anifrolumab 300 mg and placebo in the mean change from baseline in pain NRS will be assessed by visit up to Week 52.

3.3.2.3 Functional Assessment of Chronic Illness Therapy-FATIGUE

The difference between anifrolumab 300 mg and placebo in the mean change from baseline in FACIT-F total score will be assessed by visit up to Week 52. The FACIT-F total score will be derived using the 13 single item scores ranging from 0 (not at all) to 4 (very much). To get a total score for which higher values indicate better quality of life, most of the single items will be reversed, ie, the final item score used for the total score will be derived as 4 – documented item score. Item scores An5 (I have energy) and An7 (I am able to do my usual activities) will not be reversed, ie, the final item score is the same as the documented item score. The FACIT-F total score will be derived as the sum of the 13 final item scores.

A response in FACIT-F is defined as an improvement from baseline to Week 52 of >3 points (ie, change from baseline >3), no restricted medications beyond the protocol allowed threshold used on or before the assessment, and no permanent premature discontinuation of investigational product (according to eCRF form “Discontinuation of Investigational Product”). If the change from baseline cannot be evaluated at Week 52 (eg, due to missing values) the change from baseline will be imputed using LOCF. This applies only if Week 48 data is not missing, otherwise the subject will be defined as a non-responder.

3.3.2.4 Patient Global Assessment

The difference between anifrolumab 300 mg and placebo in the mean change from baseline in PtGA (measured on a VAS ranging from 0 to 100 mm) will be assessed by visit up to Week 52.

3.3.2.5 Lupus quality of life scale

The difference between anifrolumab 300 mg 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 2.

Table 2 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).

Domain scores will be derived when at least 50% of the items are answered. The mean raw domain score is then calculated by totalling 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.

3.3.2.6 EuroQol 5 dimensions

The EQ-5D-5L is comprised of the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The proportion of subjects 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, will be explored over time.

EQ-5D-5L value sets are not available for all countries, therefore the UK value set will be used for all subjects 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 will be calculated (Van Hout, 2012).

3.3.2.7 Work productivity and Activity Impairment - Lupus

The WPAI scores levels (percentages) and changes from baseline will be explored by visit. The following will be presented:

- Number of subjects employed
  = number of subjects with yes in response to Question 1 (Q1)
  Percentage of all subjects 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

3.3.2.8 Medical Resource Use Questionnaire

The following variables will be explored:

- Number of subjects with health care visits
  Of subjects with health care visits
  - Number of specialist visits
  - Number of primary care visits
- Number of subjects with emergency department visits
  Of subjects 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 subjects with hospital visits

Of subjects 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 ICU

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

3.4 Assessment of study population

3.4.1 Demographic and baseline characteristic variables

Demographic characteristics (including geographic region, age, sex, ethnicity and race) and baseline characteristics (including height, weight, body mass index [BMI] and disease characteristics) will be assessed.

The clinical SLEDAI-2K score will be derived as the sum of the scores for the SLEDAI-2K items vasculitis, arthritis, myositis, rash, alopecia, mucosal ulcers, pleurisy, and pericarditis.

3.4.2 Medical history

Medical histories will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA). Medical history will be evaluated separately for past and current conditions as given in the eCRF.

3.4.3 Prior and concomitant medications

Prior medications are reported according to the eCRF completion guidelines (ie, dependent of the relevance of the medication, the intake during the last few weeks before the first administration of IP, all medication taken during lifetime, or anything in-between will be reported). All medications will be coded using the latest version of World Health Organization Drug Dictionary (WHO-DD).

Any medications taken by the subject prior to the first dose date of investigational product will be 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 investigational product up to Week 52 (Visit 14/EDV), inclusive, will be considered concomitant medication. Any medication started prior to the first dose of investigational product and ended after the first dose up to Week 52 (Visit 14/EDV) / was ongoing will be 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 investigational

product (ie, start date on or before the date of first dose and end date on or after date of first dose or ongoing). The medications will be presented in the following categories:

- Anti-malarial
  defined as medications with an ATC code level 4 of P01BA (aminoquinolones) and P01AX (Other agents against amoebiasis and other protozoal diseases)
  - Any Anti-malarial
  - Anti-malarial Only
  - Anti-malarial in combination with OCS and/or immunosuppressants
- Azathioprine
  defined as medications with a preferred term of azathioprine or azathioprine sodium
- Methotrexate
  defined as medications with a preferred term of methotrexate or methotrexate sodium
- Mycophenolate
  defined as medications with a preferred term of mycophenolate mofetil, mycophenolate sodium or mycophenolic acid
- Mizoribine
  defined as medications with a preferred term of mizoribine
- OCS
  - Any OCS
  - OCS only
  - OCS in combination with immunosuppressants and/or anti-malarial
  - Time on Corticosteroids up to randomization
- NSAID
  defined as medication with an ATC code level 3 of M01A
- Other SLE medication
  defined as SLE medications not covered within the above categories

In case of missing or partial dates, the imputation rules as given in Appendix C will be applied to classify medication as prior and/or concomitant, as appropriate.

3.4.4 Exposure of investigational product

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

$$\text{Duration of exposure (days)} = (\text{Last dosing date} + 28 \text{ days}) - \text{first dosing date} + 1.$$

The total subject years 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 will be counted per subject. Furthermore, the number of subjects with an infusion will be assessed in 4-weekly categories (ie, 4 weeks, 8 weeks, 12 weeks, …, 48 weeks).

The time to discontinuation of investigational product is the same as the duration of exposure. However, for subjects continuing in the LTE study, the time to discontinuation will be censored at the last dosing date + 28 days.

3.5 Safety variables

The following safety data will be collected: vital signs, physical examination, 12-lead ECG, haematology, clinical chemistry, urinalysis, cushingoid features, C-SSRS, PHQ-8, modified SELENA Flare Index based flares, and AEs (including AESIs).

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

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

3.5.1 Adverse events

Adverse events experienced by the subjects 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 investigational product.
  AEs occurring during screening will only be listed.
- AEs occurring during treatment:
  An AE during treatment is defined as an AE with a date of onset ≥ day of first dose of investigational product and ≤ date of last dose of investigational product + 28 days.
- AEs occurring during follow-up:
  An AE during follow-up is defined as an AE with a date of onset > date of last dose of investigational product + 28 days and ≤ date of last dose of investigational product + 84 days.
- AEs occurring during treatment and follow-up:
  The period “during treatment and follow-up” combines the periods “during

treatment" and "during follow-up". An AE during treatment and follow-up is defined as an AE with a date of onset ≥ day of first dose of investigational product and ≤ date of last dose of investigational product + 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 investigational product + 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. Major acute cardiovascular events (MACEs) will be determined according to the assessments of the Cardiovascular Event Adjudication Committee.

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 (investigator) is defined as an AE with a preferred term of "Infusion related reaction". An infection is defined as an AE within the SOC infections and infestations. Opportunistic infections are defined as given by the investigator and non-opportunistic infections are all infections not marked as opportunistic by the investigator. Hypersensitivity is defined according to the narrow MedDRA SMQ Hypersensitivity.

Herpes zoster is further classified according to the information given on the Herpes zoster log as follows:

Category	**Rash [Y/N]**	**Episode status of HZ Event [localized/ disseminated]**	**Specify Disseminated [cutaneous/ systemic]**	**Any organ involvement [Y/N]**
Cutaneous (localised) herpes zoster	Y	Localized	[no rule]	N
Cutaneous disseminated herpes zoster	Y	Disseminated	Cutaneous	[no rule]
Visceral disseminated herpes zoster	[no rule]	Disseminated	Systemic	[no rule]

Adverse events during treatment will also be presented by time intervals of the first onset of the event. For this analysis, repeated events with the same preferred term will not be

considered (ie, if a subject has more than one event with the same preferred term, only the event with the earliest date of onset will be used). For partial or missing dates, the rules as described above will be used, ie. the AE will be 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 investigational product and < date of first administration of investigational product plus 84 days
- Week 12 to < Week 24:
  AEs with date of onset ≥ date of first administration of investigational product plus 84 days and < date of first administration of investigational product plus 168 days
- Week 24 to < Week 36:
  AEs with date of onset ≥ date of first administration of investigational product plus 168 days and < date of first administration of investigational product plus 252 days
- Week 36 to < Week 48:
  AEs with date of onset ≥ date of first administration of investigational product plus 252 days and < date of first administration of investigational product plus 336 days
- ≥ Week 48:
  AEs with date of onset ≥ date of first administration of investigational product plus 336 days

The event rate per 100 subject years is defined as

Number of subjects with an event / [sum of total exposure time in days/ (365.25*100)].

The exposure in a time period for each subject will be calculated as end of period – start of period + 1 (eg, date of last dose of investigational product + 28 days - day of first dose of investigational product + 1 day for summary of AEs during treatment). If a subject discontinued from the study during a period, switched to the LTE study 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 herpes zoster AESI, an alternative event rate per 100 subject years will be derived as

Number of subjects with an event / [sum of time at risk in days/ (365.25*100)].

The time at risk is defined as time (including start and end date) from start of period (eg, date of first administration of investigational product for events during treatment) to the date of first event, death, withdrawal of consent, or end of period, whatever comes first. If an AE has a (partially) missing onset date, the missing parts will be considered on treatment, unless available information indicates otherwise and will be imputed with the earliest possible date given the available information before calculating the alternative event rate. This alternative

event rate may also be calculated for other AESIs if suggested by data. This will be discussed during the BDR meeting and the decision will be made before unblinding the data.

The time to first onset of herpes zoster during treatment will be derived as date of first onset of herpes zoster – date of first administration of investigational product + 1. AEs with an onset date before the date of first administration of investigational product and AEs with an onset after 28 days after the date of last administration of investigational product will not be considered for the time to first onset of herpes zoster during treatment. If an AE has a (partially) missing onset date, the missing parts will be considered on treatment, and the missing parts will be imputed with the earliest possible date given the available information before calculating time to event. If a subject has no herpes zoster during treatment, the time to first onset will be censored at the date of last administration of investigational product + 28 days.

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 investigational product at the respective visit and < date of administration of investigational product at the following visit (or ≤ date of last dose of investigational product + 28 days for the last visit with investigational product).

3.5.1.1 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 haematological and other laboratory abnormalities, and certain events that led to intervention (other than those already classified as serious), dose reduction, or significant additional treatment.

3.5.2 Laboratory variables

The parameters haematology, serum chemistry, urinalysis (outlined in Table 5 in Section 5.3.10 of the clinical study protocol) and of fasting lipid profile (high density lipoprotein cholesterol, low density lipoprotein cholesterol, and triglycerides) will be explored.

Laboratory data will be reported in SI units. Changes from baseline in haematology, clinical chemistry and lipid profile variables will be calculated.

Absolute values will be compared to the reference range as given in Appendix D 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.

Treatment emergent laboratory/vital signs changes (TELVC) will be defined for post-baseline values according to the reference ranges given in Appendix D and Appendix F.

Urinalysis data will be categorised as negative (0), positive (+), or strongly positive (++, +++, or >+++) at each time-point. Treatment-emergent changes will also be assessed. Treatment-emergent changes of urinalysis data are defined as

- Negative/Trace/+ at baseline to ++, +++, ++++ at any on-treatment value OR
- Increase of from baseline of at least ++ at any on-treatment value.

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) will be calculated for each data point. Multiple = Value / ULN, ie, if the ALT value was 72 IU/L (ULN = 36) then the multiple would be 2. Subjects 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) will be flagged:

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

3.5.3 ECGs

The outcome of the overall evaluation of 12-lead ECG measurements by the central reading will be assessed as normal or abnormal. It is the investigator's judgment whether the findings/results on the central ECG laboratory report are clinically relevant or not. The combination of both judgments leads to the following categories used for analysis:

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

If the overall evaluation by the investigator and the central ECG report don't match, the investigator's judgement will be used. In case of repeated measurements (triplicates) at a visit, the worst category at the respective visit will be used for the analysis.

Changes from baseline of the following variables will be explored:

- Heart rate (beats per minute)
- QRS duration (ms)
- PR interval (ms)
- RR interval (ms)
- QT (ms)
- QTcB (ms)

- QTcF (ms)

Potentially Clinical Significant post-baseline values or changes from baseline are defined in Appendix E for some parameters, more than one criterion is given. The proportion of subjects meeting each criterion will be explored.

In case of repeated measurements (triplicates) at a visit, the mean of all available values at the respective visit will be used for the analyses of the continuous ECG variables and the determination of Potentially Clinical Significant values.

3.5.4 Modified SELENA Flare Index based flares

- [redacted]

3.5.5 Physical examination

Weight (kg) will be explored using the difference in mean change from baseline longitudinally over time.

3.5.6 Vital signs

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)

Changes from baseline will be calculated.

Where applicable, absolute values will be compared to the reference ranges given in Appendix F 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 will be classified as TELVC according to reference ranges given in Appendix F.

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

3.5.7 Cushingoid features

The presence of cushingoid features (moon face, buffalo hump, purple or violaceous striae, central obesity, hirsutism, acne, easy bruising, and fragile skin) will be explored by visit.

3.5.8 C-SSRS

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

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 Behaviour
- Category 7 – Aborted Attempt
- Category 8 – Interrupted Attempt
- Category 9 – Actual Attempt (non-fatal)
- Category 10 – Completed Suicide

The Suicidal Ideation or Behaviour Score will be derived from the C-SSRS categories as the maximum suicidal ideation or behaviour category (1-10 on the C-SSRS) present at the assessment. The score will be derived at each assessment for each subject. Non-suicidal self-injurious behaviour will be assigned if no ideation or behaviour is present.

Composite variables based on the above re-ordered categories are defined for assessments during screening, during treatment, and during follow-up, 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 behaviour: A "yes" answer at any time in the respective study period, to any one of the 5 (re-ordered) suicidal behaviour questions (Categories 6-10) on the C-SSRS.
- No suicidal ideation or behaviour: No "yes" answer at any time in the respective study period to any one of the 10 (re-ordered) suicidal ideation and behaviour 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 during follow-up, 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 investigational product.

- During treatment Period:
  Assessments with a date > date of first dose of investigational product and ≤ date of last dose of investigational product + 28 days.
- Follow-up Period:
  Assessments with a date > date of last dose of investigational product + 28 days and ≤ date of last dose of investigational product + 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.

3.5.9 Personal Health Questionnaire Depression Scale-8

The PHQ-8 assesses symptoms of depression over the last 2 weeks. The difference between anifrolumab and placebo in the mean change from baseline in PHQ-8 total score will be assessed by visit up to Week 52. The PHQ-8 total score will be derived as the sum of the 8 single item scores ranging from 0 (not at all) to 3 (nearly every day).

3.6 Pharmacokinetics, pharmacodynamics, and immunogenicity

3.6.1 Immunogenicity variables

Anti-drug antibodies assessments will be conducted utilising a tiered approach (screen, confirm, titre).

The following ADA results will be evaluated as proportion of subjects together with corresponding titre summaries.

- Subjects 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.
- Subjects who are ADA positive at baseline only
  Subjects without any post-baseline measurements are also counted in this category.
- Subjects who are ADA positive at both baseline and post-baseline.
  Only subject 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).
- Subjects who are ADA positive post-baseline only (treatment induced ADA)
  Subjects without a baseline measurement are also counted in this category.
- Subjects who are persistently positive
  Persistently positive is defined as treatment induced ADA (subject 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.

- Subjects who are transiently positive
  Transiently positive is defined as at least one treatment induced (subject is ADA negative at baseline) ADA positive measurement, but not fulfilling the conditions for persistently positive.
- Subjects 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 subjects with treatment-induced ADA) for each subject in the treatment group. Baseline ADA titers are not included in this assessment.
- Subjects who are ADA positive by visit
- Subjects who are ADA positive at a post-baseline measurement for the first time by visit (Subjects that are ADA negative at baseline).

For the presentation of ADA results at a single time point (eg, baseline or by visit summaries), the corresponding titre summary will be based on the titres of the positive samples for that particular visit. For the presentation of ADA results across visits (eg, any post-baseline), the corresponding titre summaries will be based on the maximum titre of all positive samples for each subject. Titres of positive measurements reported as ≤30 (limit of detection) will be imputed as 30. Titre values reported as "< 30" are negative and will not be imputed.

The presence of neutralising 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 subjects with ADA results available (at any time) who are nAb positive at any time
- nAb incidence: Proportion of subjects with ADA results available at post-baseline timepoints who are nAb positive at post-baseline timepoints only (ie. negative at baseline).
- nAb persistently positive: Proportion of subjects with ADA results available at post-baseline 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 post-baseline assessment (for subjects who are nAb negative at baseline).
- Proportion of subjects who are nAb positive by visit

3.6.2 Pharmacokinetic variables

Due to the limited sampling schedule, the PK assessment will be primarily based on the observed steady-state serum trough (predose) 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 IFN 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.

3.6.3 Pharmacodynamic outcome variables

The outcome variable for anti-dsDNA, C3, C4, and CH50 complement levels will be the mean change from baseline longitudinally up to Week 52 in subjects with positive anti-dsDNA at baseline and abnormal complement level at baseline, respectively. An abnormal complement level is defined as complement level below lower limit of normal.

The suppression of the 21-gene type I IFN gene signature as percent suppression of fold change, relative to a pooled normal control, from baseline levels will be derived after week 52 DBL by the Biometrics and Information Sciences (B&I) group at AstraZeneca.

3.7 Exploratory variables

[redacted]

4. ANALYSIS METHODS

4.1 General principles

4.1.1 Data excluded from analysis

In the rare event that a site needs to be closed during the course of the study for quality reasons (eg, suspicion of fraud, non-compliance), all subjects of this site will be excluded from all analyses and summaries. In this situation, the SAP will be amended prior to database lock and details will be added documenting decisions about site closure and data handling (including details of important data to be presented in separate listings).

[redacted]

Therefore, the 3 randomised subjects (E7857501, E7857502, and E7857503) as well as 1 screen failure (E7857504) at site 7857 will not be considered in analyses and summaries. Separate listings for exposure, AEs, PK, ADA, medical history, concomitant medication, and SLEDAI-2K will be provided for the treated subjects.

4.1.2 Database lock and unblinding

Database lock at Week 52

When the last subject reaches Week 52 visit, all data available in the database will be extracted and undergo through the data cleaning process. Data as detailed in Table 3 will be part of the database lock (hereafter referred to as Week 52 lock). Note that data for follow up visits and from the concomitant medication, AE, SAE and AESI log pages may not be completely clean at the time of the Week 52 lock. When all relevant data have been coded, validated, signed, and locked, a clean file of this data will be declared.

Once Week 52 lock is declared, the study will be un-blinded. Blinding of subjects and investigators will be maintained to the greatest extent possible after Week 52 lock and until Last Subject Last Visit in the LTE study. No changes to locked data, will be accepted after completion of Week 52 lock, and no analyses will be repeated for the results up to, and including, Week 52.

Table 3 Data included in Week 52 lock

Subjects	Week 52 lock
Subjects who completed the study at the time of the week 52 lock	All data included in week 52 lock
Subjects who are still ongoing in follow-up at the time of the week 52 lock	All visit data up to and including Week 52 is included in week 52 lock. Follow-up data, concomitant medications, AE, SAE and AESI log forms are not part of the lock.

Database lock at Last Patient Last Visit

After completion of the Last Subject Last Visit, all remaining data not locked as part of the Week 52 lock will undergo lock per protocol stipulated criteria. For subjects active in follow-up at the time of the Week 52 DBL, data from assessments after Week 52 will be listed separately. Database lock conducted after study Last Subject Last Visit will not include nor result in any changes to data locked at the Week 52 time point.

4.1.3 Visit windows

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 4 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	420	407≤ Study Day

For assignment of data to time points using the visit windows, study day will be defined as (Date of assessment – date of first administration of investigational product) +1. Using this definition, the day of first dose of investigational product will be Day 1 and the scheduled visit date of Week 4 will be study day 29 (=28+1) for example.

If multiple readings are recorded within a single visit window, the following rules will be followed.

- If there are two or more observations within the same visit window, then the non-missing one closest to the scheduled visit will be used in the analysis.
- If two observations are equidistant from the scheduled visit, then the non-missing observation with the earlier collection date will be used in the analysis.
- If two observations are collected on the same day, then the non-missing one with the earlier collection time will be included in the analysis.

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” will be used for all measurements before the first administration of IP. If the time of measurement at Study Day 1

is not available, it is assumed to be prior to first administration of IP. For measurements at Study Day 1 with a time of measurement after the start of IP administration, the visit label "Day 1" will be used (indicating a post-Baseline measurement).

For overall analyses not based on any particular study visit, data will be listed and/or analysed, including any repeat or unscheduled visits, unless otherwise specified.

4.1.4 Presentation of results

All analyses will use SAS® version 9.3 or higher. Summary tables will be presented by treatment group (anifrolumab 150 mg, anifrolumab 300 mg, and placebo). All available data (with the exception as described in Section 4.1.1) for each analysis set will be used in the analyses. However, for visit based analyses, the variables will be summarized based on the scheduled days with adjusted analysis-defined visit windows as given in Section 4.1.2. If not stated otherwise, these summaries will be restricted to data up to Week 52. Data (including derived variables) will be presented in listings sorted by treatment group and subject number. A separate document will be produced containing the template table, listing, and figure shells.

Unless otherwise noted, categorical data will be presented using counts and percentages with the denominator for percentages being the number of subjects in the analysis set by treatment group. Percentages will be rounded to one decimal place; except 100% which will be displayed without any decimal places. Percentages will not be displayed for zero counts.

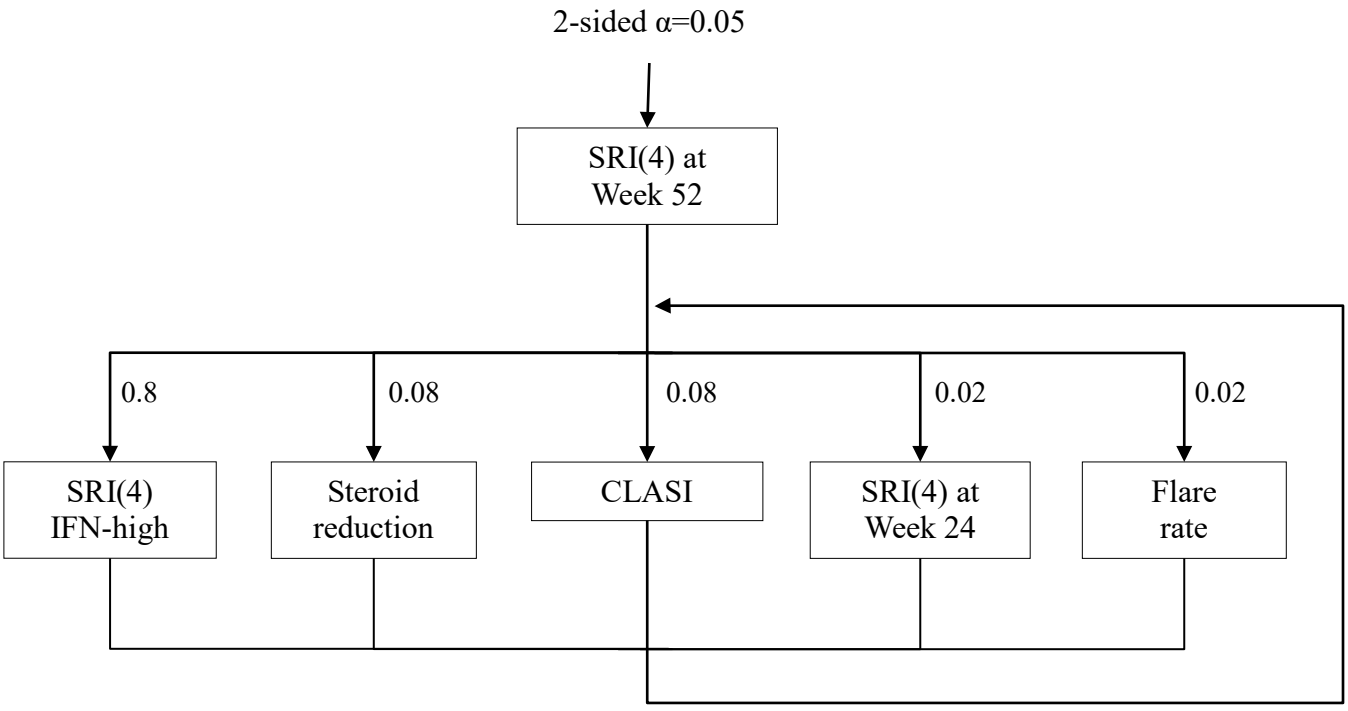
Unless otherwise noted, continuous variables will be summarized using the number of observations (n), mean, standard deviation (SD), median, minimum, and maximum. The minimum and maximum values will be displayed to the same level of precision as the raw data, the mean and median to a further decimal place and the SD to two additional decimal places. CV% (used for PK tables) will be presented with 1 decimal place.

If not stated otherwise, data of all treatment arms will be included in the statistical models. If a model restricted to 2 treatment arms is used (eg, to estimate the response rate with the Cochran-Mantel-Haenszel [CMH] approach), 2 separate models including anifrolumab (300 mg and 150 mg, respectively) and placebo will be calculated. The model including anifrolumab 300 mg will be used for the estimates of the placebo group.

4.1.5 Testing strategy to account for multiplicity considerations

To account for multiplicity to test the primary and five key secondary endpoints, a testing strategy will be followed to control the overall type I error rate in the strong sense. The primary endpoint, ie, the difference in proportion of subjects achieving SRI(4) at Week 52 comparing anifrolumab 300 mg to placebo, will be tested at an alpha level of 0.05. If the observed p-value is ≤0.05, a statistically significant difference in SRI(4) between the treatment groups at Week 52 will be concluded, and the alpha of 0.05 will be preserved for testing of the key secondary endpoints. If the observed p-value is >0.05, no statistically significant difference between treatment groups will be declared, and no formal testing of the key secondary endpoints will be carried out.

Figure 1 Alpha recycling



If the primary endpoint is statistically significant, then the five key secondary endpoints will be tested and the weighted Holm procedure (Zhang, 1997 and Burman, 2009) will be used in order to strongly control the family-wise error rate at the 2-sided 5% level. The procedure applies alpha recycling according to the weights given in Figure 1. The weights were chosen based on a combination of estimated power for the individual key secondary endpoints and their relative clinical importance

In a first step of the weighted Holm procedure, the five key secondary endpoints (SRI(4) in the IFN test-high subgroup, steroid reduction, CLASI reduction, SRI(4) at Week 24, and flare rate) will be tested at alpha levels of 0.04, 0.004, 0.004, 0.001, and 0.001, respectively. If 1 or

more of the hypotheses will be rejected at these levels, the corresponding alpha will be distributed to the endpoints not rejected according to the assigned weights. As an example, the null hypothesis for SRI(4) for the IFN test-high sub-group will be rejected with a p-value ≤0.04 and all other hypotheses will not be rejected due to p-values >0.004 and 0.001, respectively. The alpha level of 0.04 will be distributed to the remaining four key secondary endpoints (steroid reduction, CLASI, SRI(4) at Week 24, and flare rate) resulting in new alpha levels of 0.02, 0.02, 0.005, and 0.005, respectively. If any of the tests of these four endpoints results in a p-value smaller or equal to the corresponding new alpha level, the corresponding alpha level will again be distributed across the remaining endpoints.

This corresponds to the following mathematical notation:

We have 5 tests, and each hypothesis has a weight w_i for i = 1, …, 5 with sum(w_i) = 1

Let $p_{wi} = p_i/w_i$. Order the weighted p-values as $p_{(w1)} \leq p_{(w2)} \leq ... \leq p_{(w5)}$, being $H_{(wi)}$ the hypothesis associated with p-value $p_{(wi)}$ and $w_{(i)}$ the corresponding weights.

The weighted Holm's method can be described as

1. If $p_{(w1)} > \alpha$, fail to reject all 5 hypotheses
2. If $p_{(w1)} \leq \alpha$, reject $H_{(w1)}$, and move to $H_{(w2)}$
3. If $p_{(w2)} > \alpha/(1-w_{(1)})$, fail to reject $H_{(wi)}$ for i ≥2
4. If $p_{(w2)} \leq \alpha/(1-w_{(1)})$, reject $H_{(w2)}$, and move to $H_{(w3)}$
5. Proceed with all remaining hypotheses until the first j such that $p_{(wj)} > \alpha/[1-\text{sum}(w_{(k)})]$ for k = 1,…, j-1

The adjusted p-values $p_{(i)}$ corresponding to hypothesis $H_{(wi)}$ will be calculated as $p_{(i)} = \max_{l<i}(r_{(l)})$, where $r_{(i)} = p_{(wi)}\left[1 - \sum_{k=1}^{i-1} w_{(k)}\right]$. (Wright, 1992) If $p_{(i)} \geq 1$ its value will be presented as >0.999.

4.1.6 Missing Data

Subjects who discontinue investigational product will be asked to come to each visit for the scheduled assessments through Week 52 (Visit 14/EDV). The definition of the primary variable includes 2 criteria that correspond to a non-responder imputation for subjects who prematurely discontinue from investigational product, or who receive restricted medications beyond the protocol-allowed threshold. This is the same for the definitions of key secondary binary variables.

Based on the results from the Phase 2 study (CD IA MEDI 546-1013), the majority of missing data is expected to be due to permanent discontinuation of IP, which for the binary responder endpoints will be imputed as non-responders per the endpoint definitions. It is expected that at each visit between 4% and 7% of subjects will have missing data due to a reason other than

early discontinuation of IP (eg, missing visit [intermediate missing]), based on the results from the Phase 2. Given the expected small amount if intermediate missing data, it is expected to have a limited impact in the overall conclusions.

For binary responder efficacy endpoints, any component with missing value will be imputed using last observation carried forward (LOCF) if only a single (non-consecutive) visit has missing data for that component. In the event of two or more consecutive visits with missing data for the same component, LOCF will be used for the first missing value of each sequence, after which the data will be imputed as non-responders for the specific responder endpoint. The responder endpoint is derived based on the imputed values. If a component (eg, SLEDAI-2K) is based on several data points, LOCF will be done for the single data points.

To examine the impact of missing data, including the impact of non-responder imputation due to permanent discontinuation of IP, on the primary and key secondary endpoints tipping point analyses will be performed. These analyses will vary the assumptions about outcomes among the subsets of subjects on the treatment arms who discontinue IP early. For the primary endpoint an extra sensitivity analysis will be performed to examine the impact of intermediate missing data.

For the derivation of the SLEDAI-2K total score, any laboratory items with missing value will be imputed using LOCF if only a single (non-consecutive) visit has missing data for that item. In the event of two or more consecutive visits with missing data for the same item, LOCF will be used for the first missing value of each sequence, after which the data will remain missing. Laboratory items of SLEDAI-2K are urinary casts, hematuria, proteinuria, pyuria, low complement, increased DNA Binding, thrombocytopenia, and leukopenia.

Missing safety data will generally not be imputed. However, safety assessment values of the form of <x (ie, below the lower limit of quantification (LLOQ)) or >x (ie, above the upper limit of quantification) will be imputed as x in the calculation of summary statistics but displayed as <x or >x in the listings.

Anifrolumab serum concentrations reported as below the LLOQ will be imputed with LLOQ/2 for analysis. Anti-dsDNA values reported as "<x" will be treated as missing values in summary statistics and considered as NEGATIVE results for all analyses.

Details about imputation of partial or missing dates are given in Appendix C.

4.1.7 Examination of model assumptions

Model assumptions for linear mixed effect models will be checked with graphical displays (plot of residuals versus predicted values, a histogram with normal density overlaid, and a quartile-quartile (Q-Q) plot showing the residual quantiles versus quantiles of a normal distribution). If the model assumptions are not met, appropriate data transformations or the use of non-parametric approaches will be discussed during the BDR meeting. The decision about an appropriate approach will be made before unblinding the data.

The assumption of proportionality for Cox proportional hazard models will be assessed by producing complementary log-log plots presenting log (-log (estimated survivor function)) versus log (time). (In the presence of non-proportionality, the hazard ratio will be interpreted as an average hazard ratio over the flare exposure time.) If these raise concerns, a time dependent covariate would be fitted to assess the extent to which this represents random variation. This will be discussed during the BDR meeting and the decision about an appropriate approach will be made before unblinding the data.

4.2 Analysis methods

4.2.1 Analysis of the primary variable

The estimand of primary interest is the difference in the proportions of response between anifrolumab 300 mg and placebo at Week 52 in the full analysis set, where the response is captured with a composite binary endpoint and is defined by sufficient change from baseline in disease activity as measured by SLEDAI-2K, no worsening in BILAG and PGA and ability to adhere to the planned course of the treatment. The intercurrent events: discontinuation of IP and receipt of restricted medications are unfavourable outcome. Therefore, patients treated with restricted medications beyond protocol allowed threshold, and those who discontinued IP for other reasons, will be non-responders. This estimand answers a clinically relevant question comparing the number of patients able to both complete the study treatment and to achieve adequate response without further medication being required. The response is measured by the primary efficacy endpoint, defined as the difference in the proportion of patients who achieved SRI(4) at Week 52, comparing the anifrolumab 300 mg to the placebo groups.

This is measured by the primary efficacy endpoint, defined as the difference in the proportion of subjects achieving SRI(4) at Week 52 comparing the anifrolumab 300 mg to the placebo group. The null hypothesis is that the proportion of subjects achieving SRI(4) on anifrolumab 300 mg is equal to that on placebo. The alternative hypothesis is that the proportion of subjects achieving SRI(4) on anifrolumab 300 mg is not equal to that on placebo, ie,

H0: difference in proportion achieving SRI(4) (anifrolumab 300 mg vs Placebo) = 0
Ha: difference in proportion achieving SRI(4) (anifrolumab 300 mg vs Placebo) ≠ 0.

The proportion of subjects achieving SRI(4) in the anifrolumab 300 mg treatment group will be compared to that in the placebo group using a Cochran-Mantel-Haenszel (CMH) approach (Stokes, 2012) stratified by:

- SLEDAI-2K score at screening (<10 points versus ≥10 points)

  If different measurements are available for re-screened subjects, the value at re-screening will be used.
- Week 0 (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.
- Results of a type 1 IFN test (high versus low)

SLEDAI-2K score at Screening and Week 0 (Day 1) OCS strata will be derived programmatically from the data recorded in the database. Type 1 IFN test stratum will be taken as recorded at randomisation by the Interactive voice/web response system (IXRS). Strata with low counts will be pooled with adjacent strata prior to the analysis. If a sub-stratum within the IFN-low stratum has less than 20 subjects (in the pooled treatment group), then all IFN-low sub-strata will be pooled together. If a sub-stratum within the IFN-high stratum has less than 20 subjects (in the pooled treatment group), then within the IFN-high 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 IFN-high stratum still has less than 20 subjects (in the pooled treatment group), then all IFN-high sub-strata will be pooled together. If all IFN-low sub-strata are pooled together and all IFN-high sub-strata are pooled together and any of these strata has less than 20 subjects (in the pooled treatment group) then all strata are pooled together.

All pooling of strata will be done separately for the anifrolumab 300 mg versus placebo comparison and the anifrolumab 150 mg versus placebo comparison.

The analysis can be described as follows:

A. There are *nij* subjects in each stratum, where *i* is the stratum, and *j* is the treatment group. The number of subjects achieving SRI(4) is *xij*. The proportion of subjects achieving SRI(4) is denoted as *pij* = *xij* / *nij*..

B. For each stratum, the difference in proportion of subjects achieving SRI(4) is calculated as *di* = *piA* – *piP*, where *A* and *P* denote the different treatment groups (anifrolumab 300 mg and placebo, respectively).

C. Weights for each stratum, *wi*, are calculated as *niP* * *niA* / (*niA* + *niP*). The weighted difference is calculated as

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

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

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

$$Var(d_i) = \frac{p_{i.}(1-p_{i.})n_i}{w_i(n_i-1)}$$ and

$$p_{i.} = \frac{x_{i.}}{n_{i.}} = \frac{x_{iA}+x_{iP}}{n_{iA}+n_{iP}}$$

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.

$$Var(d_i) = \frac{p^*_{iA}(1-p^*_{iA})}{n_{iA}} + \frac{p^*_{iP}(1-p^*_{iP})}{n_{iP}}$$ 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 $2\left(1 - Prob\left(|\frac{WD}{SE}|\right)\right)$, where Prob() is the distribution function of the standard normal distribution.

F. The 95% CI for the weighted proportion ($\sum_i w_i p_{ij}/w$) in a treatment group j can be generated using a normal approximation and assuming independence between strata, where p_{ij}* are used as above.
$s_{ij}^2 = Var(p_{ij}) = p_{ij}^{*}(1 - p_{ij}^{*})/n_{ij}$
$s_j^2 = \sum_i w_i^2 s_{ij}^2/w^2, w = \sum_i w_i$

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

The estimated treatment effect (ie, the difference in response rate for anifrolumab 300 mg 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 (including anifrolumab 150 mg) will be presented.

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

A bar plot showing the estimated response rates for subjects with SRI(4) at Week 52 by treatment group (including CIs, number of subjects included in the analysis, and the p-value for the test of differences between anifrolumab 300 mg and placebo) will be provided.

In addition, the individual conditions of SRI(4) (SLEDAI-2K reduction ≥4 points, no new organ systems affected, no worsening of subjects' lupus disease activity, no permanent discontinuation of investigational product, and no use of rescue medication) at Week 24 and Week 52 will be summarised with counts and percentages by treatment group.

All analyses on the primary variable will be conducted with the full analysis set.

Sensitivity analysis - impact of premature discontinuation of IP

Permanent premature discontinuation of IP is part of the SRI(4) definition. In order to examine the impact of different discontinuation rates between treatment groups (Anifrolumab 300 mg and Placebo), a tipping point analyses will be performed. This analysis will vary the assumptions about outcomes among the subsets of subjects on the treatment arms who prematurely discontinue IP and can be described as:

- The proportions of subjects achieving SRI(4) will be analysed 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 randomisation 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, subjects who prematurely discontinue IP having not received restricted medications prior to discontinuation of IP are by definition imputed as non-responders. For this sensitivity analysis, these subjects will be altered from non-responder to responder in an iterative manner.
- At each step of the analysis one of these subjects switches from not achieving SRI(4) to achieving SRI(4), 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 number of subjects assumed to be achieving SRI(4) for placebo and anifrolumab 300 mg, respectively. The region where the conclusion changes, will be considered as the tipping point.
- The grid will be divided in 3 regions, limited in the top by the expected number of subjects with missing values in the anifrolumab 300 mg arm which would have achieved SRI(4) if not prematurely discontinuing IP (based on the proportion of subjects achieving SRI(4) in subjects who completed IP in the anifrolumab 300 mg arm):
    - Likely – Its right limit is the expected number of subjects with missing values in the placebo arm which would have achieved SRI(4) if not prematurely discontinuing IP (based on the proportion of subjects achieving SRI(4) in subjects who completed IP in the placebo arm);
    - Uncertain – its right limit is the expected number of subjects with missing values in the placebo arm which would have been responders if not prematurely discontinuing IP (based on the proportion of subjects achieving SRI(4) in subjects who completed IP in the anifrolumab 300 mg arm);
    - Unlikely – It is the region to the right of the uncertain region.

Sensitivity analysis – impact of intermediate missing data

To examine the impact of intermediate missing data, the following sensitivity analysis with multiple imputations will be performed. Intermediate missing values for SLEDAI-2K, BILAG-2004, and PGA (ie, missing for other reason than early discontinuation of IP) will be imputed separately for each SRI(4) component.

Intermediate missing values of SRI(4) will be imputed based on the imputed values of the BILAG-2004, PGA and SLEDAI-2K components. For each outcome and visit, 100 imputations will be generated by randomised treatment group. The procedure will be initiated with a random seed of 12345. For analysis, each imputed dataset will be analysed separately, and the single estimates will be combined using PROC MIANALYZE. Each component will be imputed as follows:

- SLEDAI-2K total score will be 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 burning iterations to be used, with 100 iterations between each imputation. From the resulting dataset, the imputations of intermediate missing values only will be extracted (ie, imputations of missing values due to early discontinuation of IP will not be considered).
- BILAG-2004 will be imputed as a binary variable reflecting the SRI(4) criterion, ie, "No new organ systems affected as defined by 1 or more BILAG-2004 A or 2 or more BILAG-2004 B items compared to baseline". This binary BILAG variable for subject i at visit t will be modelled as $BILAG_{i(t)} \sim Binomial(\pi_{i(t)})$ 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 will be specified for θ_0, θ_1 and θ_2. 10,000 burn-in iterations will be used, followed by 10,000 main iterations. Imputations will then be taken from every 100th iteration of the main chain (ie, after burn-in). From the resulting dataset, the imputations of intermediate missing values only will be extracted (ie, imputations of missing values due to early discontinuation of IP will not be considered).
- PGA will be imputed in the same way as SLEDAI-2K total score, but using MAXIMUM=3 (instead of MAXIMUM=105).

4.2.2 Key secondary outcome variables

All analyses of the key secondary outcome variables will be conducted with the full analysis set.

4.2.2.1 SRI(4) at Week 52 in IFN test-high subjects

The key secondary endpoint used to evaluate the effect of anifrolumab 300 mg compared to placebo on disease activity in the IFN test-high subgroup is the difference in proportion of subjects achieving SRI(4) at Week 52 in subjects classified as IFN test-high.

All analyses as described in Section 4.2.1 for the primary endpoint, except the sensitivity analysis with multiple imputation, will be performed for the IFN test-high subgroup. For the stratified CMH analyses, the stratification factors will be reduced to SLEDAI-2K score at screening (<10 points versus ≥10 points) and Week 0 (Day 1) OCS dose (<10 mg/day versus ≥10 mg/day prednisone or equivalent).

A supplemental analysis of SRI(4) in subjects classified as IFN test-low will be conducted. The same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab and placebo. The stratification factors will be reduced to SLEDAI-2K score at screening (<10 points versus ≥10 points) and Week 0 (Day 1) OCS dose (<10 mg/day versus ≥10 mg/day prednisone or equivalent).

4.2.2.2 Oral corticosteroid management

The key secondary endpoint used to evaluate the effect of anifrolumab 300 mg versus placebo on the ability to reduce the OCS dose in subjects with baseline OCS ≥10 mg/day prednisone or equivalent is the difference in proportion of subjects with maintained OCS reduction.

The same CMH approach as described in Section 4.2.1 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. For the stratified CMH analyses, the stratification factors will be reduced to SLEDAI-2K score at screening (<10 points versus ≥10 points) and results of a type 1 IFN test (high versus low).

A bar plot showing the estimated response rates for subjects with maintained OCS reduction by treatment group (including CIs, number of subjects included in the analysis, and the p-value for the test of differences between anifrolumab 300 mg and placebo) will be provided.

The tipping point analysis to assess the impact of premature discontinuation of IP as described in Section 4.2.1, will also be performed for maintained OCS reduction.

4.2.2.3 Skin lesions

The key secondary endpoint used to evaluate the effect of anifrolumab 300 mg versus placebo on inflammatory cutaneous lupus lesions in subjects with baseline CLASI activity score ≥10 is the difference in proportion of subjects with an at least 50% reduction in CLASI activity score at Week 12.

The same CMH approach as described in Section 4.2.1 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 (ie, for each post-baseline visit up to Week 52) based on this analysis will be created.

A bar plot showing the estimated response rates for subjects with a ≥50% reduction in CLASI activity score at Week 12 by treatment group (including CIs, number of subjects included in the analysis, and the p-value for the test of differences between anifrolumab 300 mg and placebo) will be provided.

The tipping point analysis to assess the impact of premature discontinuation of IP described in Section 4.2.1, will also be performed for an at least 50% reduction in CLASI activity score.

A further sensitivity analysis will be provided if at least 10 subjects in the anifrolumab 300 mg or placebo treatment arm have a burst and taper of OCS or IM 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.

4.2.2.4 SRI(4) at Week 24

The key secondary endpoint used to evaluate the early effect of anifrolumab 300 mg compared to placebo on disease activity is the difference in proportion of subjects achieving SRI(4) at Week 24.

The analysis of SRI(4) at Week 24 is covered with the longitudinal analysis described Section 4.2.1 for the primary variable. An additional bar plot showing the estimated response rates for subjects with SRI(4) at Week 24 by treatment group (including CIs, number of subjects included in the analysis, and the p-value for the test of differences between anifrolumab 300 mg and placebo) will be provided.

The tipping point analysis to assess the impact of premature discontinuation of IP described in Section 4.2.1, will also be performed for SRI(4) at Week 24.

4.2.2.5 Flares

The key secondary endpoint used to evaluate the effect of anifrolumab 300 mg versus placebo on flares is the difference in annualized flare rates through Week 52 (as defined in Section 3.2.5).

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. The response variable in the model will be the number of flares over the 52-week treatment period (ie, up to Visit 14/EDV). The model will include 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 3.2.5) will be used as an offset variable in the model to adjust for subjects having different exposure times. The estimated treatment effect and the corresponding 95% CI, as well as the 2-sided p-value (for the comparison of anifrolumab 300 mg versus placebo only) will be presented.

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

Sensitivity analysis – multiple imputation

To examine the sensitivity of the results of the main analysis to deviations from the underlying assumptions, an additional analysis will be performed using the controlled multiple imputation method (Keene, 2014). As with the main analysis, the sensitivity analysis includes all data until subjects complete the study/ withdraw from the study regardless of if they discontinue from randomised treatment.

For this method, the number of flares after withdrawal from study will be imputed conditional upon the observed number of flares prior to the withdrawal, a post-withdrawal model assumption, the baseline covariates included in the main analysis model and the time the subject would have remained in the study if not withdrawn (ie, date of first administration of IP + 364 days – date of last available BILAG-2004 assessment).

The method involves first fitting the main analysis (ie, negative binomial regression model as described above) to the observed data. For each imputed dataset, first an independent sample is drawn from the approximate posterior distribution of the model parameters. This consists of sampling new regression coefficients from a multivariate normal distribution, with mean equal to the observed data maximum likelihood estimate, and covariance matrix corresponding to the maximum likelihood covariance estimate. A new (log) shape parameter is drawn from a normal distribution in the same way, which is exponentiated to give a draw of the shape parameter k.

For each generated set of model parameters, the number of flares after withdrawal from study are imputed for each subject by drawing randomly from its conditional distribution given the subject's number of flares before withdrawal. For a given subject let T_1 and T_2 denote the time before and after withdrawal. Let $Y_1$and Y_2 denote the observed number of flares before withdrawal and unobserved number after withdrawal, respectively, which is to be imputed. Let λ_1 and $\lambda_2$denote the assumed rates before and after withdrawal. Let $\psi_1 = T_1\lambda_1$ and $\psi_2 = T_2\lambda_2$. Then, using SAS's RAND negative binomial parametrisation, the conditional distribution of Y_2 given Y_1 is negative binomial with 'number of successes' parameter $k + Y_1$ and 'probability of success' parameter equal to

$$\frac{k+\psi_1}{k+\psi_1+\psi_2}.$$

The imputed number of flares is then combined with the observed flares and data is analysed using the main analysis methodology. This analysis is repeated multiple times and the results combined using Rubin's formulae (Fleming, 2011, Ratitch, 2013).

Sensitivity analysis – tipping point analysis

First a MAR analysis will be performed where for each subject 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 randomised treatment group and baseline covariates.

A tipping point analysis will then be performed, 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 favourable to unfavourable; ie, 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 BDRM and adapted as necessary.

Sensitivity analysis – flares while on treatment

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. The response variable in the model will be the number of flares while on treatment (ie, up to last administration of investigational product + 28 days). The model will include 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 3.2.5) will be used as an offset variable in the model to adjust for subjects having different exposure times. The estimated treatment effect and the corresponding 95% CI will be presented.

4.2.3 Analysis methods for other secondary efficacy variables

All analyses of other secondary outcome variables will be conducted with the full analysis set.

4.2.3.1 Assessment of disease activity

SRI(4) of anifrolumab 150 mg treatment group

The effect of anifrolumab 150 mg compared to placebo on disease activity will be assessed by the difference in proportions of subjects achieving SRI(4) at Week 52.

The same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab 150 mg and placebo and the corresponding response rates (weighted proportions).

This will be included in the primary analysis table.

Supportive SRI variables

SRI(5), SRI(6), SRI(7), and SRI(8) will be evaluated only for subjects with baseline SLEDAI-2K ≥5 points, ≥6 points, ≥7 points, and ≥8 points, respectively.

The same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab 300 mg and placebo as well as the response rates (weighted proportions) for all treatment groups. Longitudinal presentations of the results over time (ie, for each post-baseline visit up to Week 52) based on this analysis will be created.

BICLA

The same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab 300 mg and placebo as well as the response rates (weighted proportions) for all treatment groups for BICLA response. Longitudinal presentations of the results over time (ie, for each post-baseline visit up to Week 52) based on this analysis will be created.

A bar plot showing the estimated response rates for subjects with BICLA response at Week 52 by treatment group (including Cis, and number of subjects included in the analysis) will be provided.

In addition, the individual conditions of BICLA (reduction of all baseline BILAG-2004 A and B scores and no worsening in other organ systems, no worsening from baseline in SLEDAI-2K, no worsening of subjects' lupus disease activity, no permanent discontinuation of investigational product, and no use of rescue medication) will be summarised with counts and percentages by treatment group at Week 24 and Week 52.

Supportive outcome variables of the Individual components of SRI and BICLA

Change from baseline in SLEDAI-2K and PGA will be analysed using repeated measures model with fixed effects for baseline value, treatment group, visit, treatment*visit interaction and stratification factors. 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 differences between anifrolumab 300 mg and placebo, and associated 2-sided CIs.

For each SLEDAI organ system (central nervous system, vascular, musculoskeletal, renal, mucocutaneous, cardiovascular system and respiratory, immunology, and haematological and fever) the number and percentage of subjects with an improvement at Week 24 and Week 52, respectively, will be given for subjects with corresponding organ system involvement at baseline.

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

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

For each of the organ systems, a bar plot will be provided showing the distribution of the scores at baseline, Week 24, and Week 52 by treatment group. The figures will include the percentages and number of subjects.

Observed values and changes from baseline in BILAG-2004 global score will be presented by visit with descriptive statistics.

For Major Clinical Response and Partial Clinical Response, the same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab 300 mg and placebo as well as the response rates (weighted proportions) for all treatment groups.

Active, swollen and tender joints

Change from baseline in the number of active, swollen and tender joints, respectively, will be analysed using the same repeated measures models as described for the analysis of SLEDAI-2K and PGA in 4.2.3.1Thereby, 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.

For subjects with at least 8 swollen and at least 8 tender joints at baseline, the same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab 300 mg and placebo in the proportions of subjects achieving an at least 20% and an at least 50% reduction in swollen and tender joints, respectively. Longitudinal presentations of the results over time (ie, for each post-baseline visit up to Week 52) based on this analysis will be created.

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 3.3.1.6.

Supportive outcome variables for the assessment of OCS use

Additionally, the daily OCS dose will be displayed graphically by a longitudinal plot presenting the means and corresponding SDs by visit. The number of subjects will also be included in the graph. Subjects without an OCS dose will be included with a value of 0. The graphical display will be repeated for subjects with baseline OCS ≥10 mg/day only.

A shift table will be provided for all subjects showing the daily OCS dose at baseline versus the daily OCS dose at Week 52. For this presentation, the daily OCS dose will be categorised as 0, >0 to ≤5mg, >5 to ≤7.5mg, >7.5 to ≤10mg, >10 to ≤15mg, >15 to ≤20mg, >20 to ≤30mg, >30 to ≤40mg, >40 mg, and missing.

Furthermore, the standardised AUC up to Week 52 will be summarized by treatment group for all subjects as well as for subjects with baseline OCS ≥10 mg/day.

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 3.2.2) versus achieving SRI(4) at Week 52 (as defined in Section 3.1) will be provided.

Supportive outcome variables for the assessment of skin lesions

For subjects with baseline CLASI activity score ≥10, a shift table for an at least 50% reduction in CLASI activity score at Week 12 versus an at least 50% reduction at Week 52 will be presented to investigate the maintenance of effect in CLASI activity score.

Change from baseline in CLASI activity score and CLASI damage score, respectively, will be analysed using the same repeated measures models as described for the analysis of SLEDAI-2K and PGA in section 4.2.3.1 Thereby, 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.

Supportive outcome variables for the assessment of flares

The analysis of flares using a negative binomial regression model as described in Section 4.2.2.5 will be repeated for flares versus baseline.

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

The time to first flare will be analysed as a supportive analysis to the assessment of reduction of flares to explore the extent to which treatment with anifrolumab 300 mg delays the time to first flare compared to placebo. The analyses of time to first flare will be provided for the key secondary outcome variable and flares versus baseline. Subjects in the anifrolumab 150 mg treatment group will not be included in the analyses of time to first flare.

Cox proportional hazard models (using a profile likelihood approach with ties=Efron) including the covariates of treatment and the stratification factors will be used to estimate the treatment effect. The estimated hazard ratios and corresponding CIs will be presented for the effect of the treatment group.

Furthermore, the time to first flare (key secondary outcome variable only) will be presented as Kaplan-Meier plot including the number of subjects at risk at each visit.

Lupus Low Disease Activity State

The same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab 300 mg and placebo as well as the response rates (weighted proportions) for all treatment groups for LLDA response.

Longitudinal presentations of the results over time (ie, for each post-baseline visit up to Week 52) based on this analysis will be created.

4.2.3.2 Subject reported outcome variables

Short Form 36 version 2 (acute recall)

Changes from baseline in in SF-36-v2 (acute) domain scores (Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional, and Mental Health), PCS and MCS, respectively, will be analysed using the same repeated measures models as described for the analysis of SLEDAI-2K and PGA in section 4.2.3.1.Thereby, 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.

For the responder analyses, the same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab 300 mg and placebo as well as the response rates for all treatment groups for each of the 8 domain scores, PCS and MCS, respectively. Longitudinal presentations of the results over time (ie, for each post-baseline visit up to Week 52) based on these analyses will be created.

Pain Numerical Rating Scale

Change from baseline in NRS will be analysed using the same repeated measures models as described for the analysis of SLEDAI-2K and PGA in section 4.2.3.1.Thereby, 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.

Functional Assessment of Chronic Illness Therapy-FATIGUE

Change from baseline in FACIT-F total score will be analysed using the same repeated measures models as described for the analysis of SLEDAI-2K and PGA in section 4.2.3.1. Thereby, 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.

The same CMH approach as described in Section 4.2.1 for the primary endpoint will be used to estimate the treatment difference between anifrolumab 300 mg and placebo as well as the response rates for all treatment groups for FACIT-F response. Longitudinal presentations of the results over time (ie, for each post-baseline visit up to Week 52) based on this analysis will be created.

Patient Global Assessment

Change from baseline in PtGA will be analysed using the same repeated measures model as described for the analysis of SLEDAI-2K and PGA in section 4.2.3.1.Thereby, 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.

Lupus quality of life 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 analysed using the same repeated measures models as described for the analysis of SLEDAI-2K and PGA in section 4.2.3.1.Thereby, 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.

EuroQol 5 dimensions

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

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

Work productivity and Activity Impairment - Lupus

At each scheduled visit, the number and percentage of subjects employed will be 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) will be presented longitudinally with descriptive statistics.

Medical Resource Use Questionnaire

The medical resource use, as described in Section 3.3.2.8, will be summarised by treatment. Results of “visit related to an increase in lupus related activity” and “cause of visit/ hospitalization” will be presented as number and percentage of respective visits.

4.2.4 Presentation of study population

4.2.4.1 Subject disposition

The number and percentage of subjects completing the study and completing the study up to and including Week 52 (Visit 14/EDV) and the number and percentage of subjects withdrawing from the study including reason for withdrawal, will be summarized by treatment group for the full analysis set. Additionally, the number and percentage of subjects enrolled to the LTE study will be presented.

The number of subjects completing treatment with investigational product up to and including Week 48 (ie, with last administration of investigational product at Visit 13) and the number of subjects permanently discontinuing the treatment with investigational product including reason for withdrawal, will be summarized by treatment group for the full analysis set.

Furthermore, the number and percentage of subjects remaining on investigational product, discontinued investigational product but remain on study and withdrawn from study will be summarized by visit up to Week 48 for the full analysis set. The number and percentage of subjects withdrawn from study at Week 52 and the number of subjects still in study at weeks 56 and 60 will also be presented.

A summary of number and percentage of subjects in each region, each country and each site by treatment group and overall will be provided for the full analysis set..A summary of number of subjects in each population will be provided for the all subjects analysis set. Important protocol deviations will be summarised for the full analysis set.

4.2.4.2 Demographic and baseline characteristic

Demography and baseline characteristics will be presented by descriptive statistics by treatment group as well as overall for the full analysis set. Additionally, the stratification factors (as calculated from the data for SLEDAI-2K score at screening and Week 0 [Day 1] OCS dose and as recorded at randomization by IXRS for the type 1 IFN test) and the cardiovascular risk will be presented by descriptive statistics for the full analysis set.

An additional table will be prepared summarising miss-stratifications as per the IXRS at randomization for the SLEDAI-2K score at screening and Week 0 (Day 1) OCS dose.

Past and current medical history will be summarized separately by MedDRA primary system organ class and preferred term. Furthermore, the number and percentage of subjects with the cardiovascular risk factors will be presented by risk factor.

Prior and concomitant medications will be summarized separately by WHO-DD Anatomical Main Group (ATC level 1), and preferred term. Disease related treatments at baseline will be summarized by the categories given in Section 3.4.3. Concomitant medication beyond protocol allowed threshold will be summarized by preferred term.

4.2.4.3 Exposure

Exposure will be summarized by treatment group for the full analysis set.

Summary statistics will be provided for the duration of exposure. The number and percentage of subjects treated ≥4 weeks, ≥8 weeks, and up to ≥52 weeks in 4-weekly intervals will be provided. Furthermore, the total subject years of exposure will be presented.

The number and percentage of subjects with infusions will be presented by total number of infusion (ie, 1, 2, …, 13) and by visit (ie, Day 1, Week 4, Week 8, Week 12, …, Week 48).

Furthermore, the time to discontinuation of investigational product will be presented as Kaplan-Meier plot including the number of subjects at risk (ie, still on investigational product).

4.2.5 Analysis methods for safety variables

Safety variables will be summarised by treatment group for the full analysis set.

4.2.5.1 Adverse events

All AE summaries will be presented for anifrolumab 150 mg, anifrolumab 300 mg, total anifrolumab (ie, including both anifrolumab treatment groups) and placebo.

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

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

For summaries during follow-up, only subjects with any study documentation after the date of last dose of investigational product + 28 days will be considered.

An overall summary of subjects with at least one AE in the following categories will be presented:

- Any AE
- Any AE with outcome = death
- Any SAE (including events with outcome = death)
- Any AE leading to discontinuation of investigational product
- Any AE related to investigational product
- Any AE of severe intensity
- Any AESI
  - Any AESI of non-opportunistic serious infections
  - Any AESI of opportunistic infections
  - Any AESI of anaphylaxis
  - Any AESI of malignancy
  - Any AESI of herpes zoster
  - Any AESI of tuberculosis
  - Any AESI of influenza
  - Any AESI of vasculitis
  - Any AESI of major acute cardiovascular events
- Any other significant AE

- The number and percentage of subjects with at least one AE (ie, multiple occurrences of an AE in 1 subject will only be counted once) will be summarised by MedDRA primary system organ class and preferred term for the following AE categories. These summaries will also include the event rate per 100 subject years,

unless otherwise specified. Event rates will generally not be included in summaries presented for AEs during treatment and follow-up.Any AE

- Any AE above reporting threshold of 2%
  This summary will be presented by preferred term only for AEs during treatment only.
- Any AE with outcome = death
  This summary will be presented for AEs during treatment and follow-up only.
- Any SAE (including events with outcome = death)
  This summary will not be presented for AEs during follow-up.
- Any AE leading to discontinuation of investigational product
  This summary will be presented for all AEs irrespective of the study period. The end of exposure for the calculation of event rates is defined as the maximum of date of last dose of investigational product + 28 days and the date of the AE leading to discontinuation. AEs with (partially) missing start date information will be considered on-treatment unless the available information indicates otherwise and will be imputed with the earliest on-treatment date possible given the available start and stop date information for this analysis.
- Any AE by relationship to investigational product (yes, no)
  (multiple occurrences of an AE in 1 subject 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 will not be presented for AEs during treatment and follow-up. This summary will not include the event rate per 100 subject years.
- Any AE by maximum intensity (mild, moderate, severe)
  (ie, multiple occurrences of an AE in 1 subject will only be counted once with the maximum intensity in this AE)
  This summary will not be presented for AEs during treatment and follow-up. This summary will not include the event rate per 100 subject years.
- Any AESI
  This summary will not be presented by system organ class but by AESI category (non-opportunistic serious infections, opportunistic infections, anaphylaxis, malignancy, herpes zoster, tuberculosis, influenza, vasculitis, and MACEs).
  This summary will not be presented for AEs during follow-up.
- Any other significant AE
  This summary will be presented for AEs during treatment and follow-up only.
- Any AE by time interval for the first onset of event
  This summary will be presented for AEs during treatment only.

Cardiovascular outcome events (during treatment and during treatment and follow-up) as determined by the Cardiovascular Event Adjudication Committee will be presented separately, summarising the number of AEs submitted for adjudication and the outcomes of the adjudication including MACE classification. Site reported cardiovascular AEs, cardiovascular SAEs and their corresponding adjudicated outcomes will be listed.

The time to first onset of a non-opportunistic serious infection during treatment and time to first onset of herpes zoster during treatment will be presented as Kaplan-Meier plot including the number of subjects at risk at each visit.

Furthermore, the alternative event rates per 100 subject years for herpes zoster (and possible other AESIs) will be summarized for events during treatment and follow-up, during treatment (overall and by time intervals), and during follow-up. The following subcategories will also be considered in this summary: SAE (including events with outcome of death), AE leading to discontinuation, and AE by maximum intensity (mild, moderate, and severe).

The number and percentage of subjects with at least one anaphylaxis, hypersensitivity, and infusion-related reaction (investigator), respectively, will be summarised overall as well as for the following respective subcategories: SAE (including events with outcome of death), AE leading to discontinuation, and AE by maximum intensity (mild, moderate, and severe). Furthermore, these AE categories will be presented graphically over time as percentage of subjects with a respective AE at a respective visit according to the definition as given in Section 3.5.1.

Infections, opportunistic infections, and non-opportunistic infections will be summarised with the same subcategories as given above.

Key subject information for subjects with an AE with outcome of death, subjects with serious AEs, subjects with an AE leading to discontinuation of investigational product, subjects with AESIs, and subjects with a cardiovascular event, respectively, will be provided (all for AEs occurring during treatment and follow-up only).

4.2.5.2 Laboratory variables

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

Shift plots (scatter plots) presenting baseline values versus minimum post-baseline values (neutrophils, lymphocytes, monocytes and haemoglobin only) and maximum post-baseline values (creatinine, creatinine kinase, blood urea nitrogen, alanine aminotransferase, aspartate aminotransferase, bilirubin), respectively, will be provided. A diagonal line indicating no change and horizontal and vertical reference lines indicating the limits of the reference ranges will also be displayed on the shift plots.

If any laboratory variables show any unusual features (high or low values or a general shift in the data points) at specific visits then shift plots of these data may also be produced. This will be discussed and agreed upon during the BDR meeting.

For each laboratory parameter with available criteria, the number and percentage of subjects with TELVC values will be summarized by visit. Additionally the number and percentage of subjects with at least one TELVC value will be presented. Percentages will be based on

subjects with a measurement at baseline and at least 1 subsequent measurement of the variable (at the respective visit).

The number and percentage of subjects with laboratory values below, within or above the corresponding reference range (see Appendix D) will be presented as shift tables from baseline to maximum and minimum on-treatment value, respectively.

Urinalysis will be summarized as shift tables from baseline to the last on-treatment value for each parameter. Furthermore, the number and percentage of subjects with treatment-emergent changes will be summarized by parameter. Percentages for the summary of treatment-emergent changes will be based on subjects with a measurement at baseline and at least 1 subsequent measurement of the variable (at the respective visit).

In order to identify potential Hy's Law cases, maximum on treatment total bilirubin will be plotted against maximum on treatment ALT, expressed as multiples of ULN. This plot will be repeated to show maximum on treatment total bilirubin against maximum on treatment AST, expressed as multiples of ULN. These plots will be produced on a log scale and reference lines will be included at 2xULN for total bilirubin and at 3xULN for ALT/AST.

4.2.5.3 ECGs

The number and percentage of subjects with normal, abnormal, not clinically significant, and abnormal, clinically significant ECG results will be presented as a shift table from baseline to Week 52.

Observed values and changes from baseline of ECG values will be summarised by parameter and visit using descriptive statistics. The summary statistics presented will be minimum, 1st quartile, median, 3rd quartile, maximum, mean, and SD.

For each ECG parameter with available criteria, the number and percentage of subjects with Potentially Clinical Significant post-baseline values and Potentially Clinical Significant changes from baseline, respectively, will be presented by parameter and criterion.

4.2.5.4 Modified SELENA Flare Index based flares

[redacted]

4.2.5.5 Physical examination

Observed values and changes from baseline of body weight will be summarized. The summary statistics presented will be minimum, 1st quartile, median, 3rd quartile, maximum, mean, and SD.

4.2.5.6 Vital signs

Observed values and changes from baseline of pulse, systolic blood pressure, diastolic blood pressure, respiration rate, and body temperature, respectively, will be summarised by visit.

The summary statistics presented will be minimum, 1st quartile, median, 3rd quartile, maximum, mean, and SD.

For each parameter with available criteria, the number and percentage of subjects with TELVC values will be summarized by visit. Additionally, the number and percentage of subjects with at least one TELVC value will be presented. Percentages will be based on subjects with a measurement at baseline and at least 1 subsequent measurement of the variable (at the respective visit).

For each parameter, the number and percentage of subjects with values below, within or above the corresponding normal range will be presented as shift tables from baseline to each post-baseline visit.

4.2.5.7 Cushingoid features

The number and percentage of subjects will be explored for each feature by visit. For subjects with baseline OCS ≥10 mg/day prednisone or equivalent, the summary will be repeated by maintained OCS reduction at Week 52 (yes versus no) as defined in Section 3.2.2.

4.2.5.8 C-SSRS

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

Subjects with a suicidal ideation or behaviour at any time will be presented in a listing.

4.2.5.9 Personal Health Questionnaire Depression Scale-8

Observed values and changes from baseline in PHQ-8 total score will be presented longitudinally with descriptive statistics.

4.2.6 Analysis methods for pharmacokinetics, pharmacodynamics, and immunogenicity

If not stated otherwise, all analyses of pharmacodynamics and immunogenicity will be performed for the full analysis set by treatment group.

4.2.6.1 Analysis method for immunogenicity variables

An overall summary of ADA response will be provided by treatment group. The proportion of subjects with ADA results as described in Section 3.6.1 will be presented. Summary statistics will be provided for titer of ADA positive subjects by timepoint. Descriptive Statistics will include n (number of reportable titers), minimum, quartiles (1st quartile, median, 3rd quartile)

and maximum. For the summary of subjects who are ADA positive by visit or at a single visit point, the denominator for percentages is the number of subjects with an adequate sample that provides a reportable result of positive or negative for the respective visit. For the summary of overall ADA categories (eg, ADA positive at any time), percentages will be based on subjects with at least one ADA result during the study.

The percentage of subjects 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 separately for each anifrolumab treatment group.

Variables of proportions regarding neutralizing antibodies will be summarised presenting the number and percentage of the categories given in Section 3.6.1.

Key subject information for subjects with an ADA positive result at any time and for subjects who sero-converted post-baseline 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 and PD endpoints are described in Sections 4.2.6.2 and 4.2.6.3. Potential association of ADA with safety and efficacy will be explored by subgroup analyses. Subgroup analyses for efficacy and safety are defined in Section 4.2.7. Further subgroup analyses for safety will be defined in the analysis plans for the integrated summary of safety.

4.2.6.2 Analysis methods for pharmacokinetic variables

All analyses of PK data will be performed for the PK analysis set by treatment group.

Anifrolumab serum concentrations will be summarised using descriptive statistics by treatment group, visit, and time point (pre-dose/post-dose) reporting n (number of non-missing values), m (number of non-zero values), geometric mean, geometric CV%, arithmetic mean, SD, arithmetic CV%, median, minimum and maximum. If applicable, this summary will be repeated including individual concentrations excluded from descriptive statistics.

Serum concentration-time profiles of anifrolumab will be generated as plots of mean values (including SD) by treatment group and time point in a semi-log scale and a linear scale. This will be presented for all values and additionally for values measured before administration (C_{trough}).

Individual concentrations excluded from descriptive statistics will be listed.

The potential influence of demographic covariates such as body weight, race, sex, and age will be explored. Furthermore, potential correlation between anifrolumab concentrations and IFN 21 gene PD signature will be explored. These analyses will be performed by MedImmune's Clinical Pharmacology and DMPK group.

Impact of ADA on PK will be explored. Anifrolumab serum concentration will be summarised by ADA result (positive at any time/negative) and treatment group. Serum

concentration-time profiles of anifrolumab will further be generated by ADA result (positive/negative at any time) as plots of mean values (including SD) by treatment group and time point in a linear scale. This will be presented for all values.

4.2.6.3 Analysis methods for pharmacodynamic variables

For subjects who are positive for anti-dsDNA at baseline and abnormal complement level at baseline, respectively, observed values and changes from baseline of anti-dsDNA, C3, C4, and CH50 complement levels will be summarised by visit.

For IFN 21 gene high subjects at baseline (defined as IFN 21 gene PD signature fold change >2 at baseline), IFN 21 gene PD signature as percent suppression of fold change, relative to a pooled normal control, from baseline levels will be summarised by time point. The median absolute deviation (MAD) will also be included in this summary. Furthermore, the median percent suppression of fold change (including MAD) over time will be presented as a line plot for the same population.

The impact of ADA on IFN 21 gene PD signature will be explored. For IFN 21 gene high subjects at baseline, line plots of IFN 21 gene PD signatures as percent suppression fold change will be generated by ADA category (ADA negative, ADA positive, ADA persistently positive, ADA titre > median, ADA nAb positive). IFN 21 gene PD signatures as percent suppression of fold change will further be summarised descriptively by ADA category (ADA negative/positive at any time), timepoint and treatment group for the same population.

4.2.7 Exploratory variables

[redacted]

4.3 Subgroup analysis

To explore the uniformity of the detected overall treatment effect, subgroup analyses on the primary and key secondary endpoints will be performed for the following factors:

- SLEDAI-2K score at screening (<10 points versus ≥10 points)
- OCS dose at baseline (<10 mg/day versus ≥10 mg/day prednisone or equivalent)
- IFN test (IFN test [high versus low])
- Sex (male versus female)
- Age (≥18 to 65 versus ≥65 years)
- Onset of disease (adult versus paediatric onset)
- BMI (≤30 versus >30 kg/m^2)
- Race (white; black or African American; Asian, native Hawaiian or other Pacific Islander; American Indian or Alaska native; other)
- Ethnicity (Hispanic/Latino versus no Hispanic/Latino)

- ADA result (positive at any time; negative; persistently positive; ADA positive with titre >median of maximum titre)

Table 4 gives an overview of subgroup analyses to be performed. Where necessary (eg, for subgroup analysis based on stratification factors), the model factors will be reduced. If not stated otherwise, the subgroup analysis will be supressed if any of the sub-populations in any treatment group will consist of less than 25 subjects. Subjects in the anifrolumab 150 mg treatment group will not be included in any subgroup analysis.

For the primary endpoint SRI(4) at Week 52, a forest plot will be used to summarise the estimates of the treatment effect for the applicable subgroups.

A listing of adverse events in ADA positive (at any time) subjects will also be presented.

Table 5 Overview of subgroup analyses

Analysis	SLEDAI-2K	OCS at baseline	IFN test	Sex	Age	Disease Onset	BMI	Race	Ethnicity	ADA result
SRI(4) – CMH [a]	X	X	X	X	X	X	X	X	X	X
SRI(4) in IFN test-high – CMH	X	X								X
Maintained OCS reduction – CMH	X		X							
CLASI activity reduction – CMH	X	X	X							
Flare rate (key efficacy) – binominal regression	X	X	X							
AE overview										X
SAE Summary by SOC and PT										X

[a] Subgroup analysis will also be conducted if resulting in a sub-population of <25 subjects.

5. INTERIM ANALYSES

No interim analysis is planned for this study.

6. CHANGES OF ANALYSIS FROM PROTOCOL (NOT APPLICABLE)

7. REFERENCES

Burman, 2009
Burman CF, Sonesson C, Guilbaud O, 2009. A recycling framework for the construction of Bonferroni-based multiple tests. Stats in Med. 2009;28:739–761.

Fleming, 2011
Fleming TR. Adressing 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.

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

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

Stokes, 2012
Stokes M.E., Davis C.S., Koch G.G., 2012. Categorical Data Analysis Using SAS®, Third Edition. SAS Institute, 2012.

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

Ware, 2007
Ware J.E., Jr., Kosinski M., Bjorner J.B., Turner-Bowker, D.M., Gandek, B. and Maruish, M.E., 2007. User's Manual for the SF-36v2TM Health Survey (2nd Ed). Lincoln, RI, QualityMetric Incorporated.

Wright, 1992
Wright, S. P. (1992). Adjusted P-values for simultaneous inference. Biometrics, 48, 1005–1013.

Zhang, 1997
Zhang J, et. al. 1997. Some Statistical Methods for Multiple Endpoints in Clinical Trials. Controlled Clinical Trials 18:204-221

8. APPENDIX

Appendix A: Medication beyond the protocol-allowed threshold

Appendix B: Derivation of prednisone equivalent daily dose

Appendix C: Derivation rules for imputation of partial or missing dates

Appendix D: Reference ranges and TELVC for laboratory values

Appendix E: ECG: Potentially Clinical Significant post baseline values

Appendix F: Reference ranges and TELVC for vital signs

Appendix A

MEDICATION BEYOND THE PROTOCOL-ALLOWED THRESHOLD

Medications beyond the protocol-allowed threshold leading to classify a subject as a non-responder will be identified through a medical review before unblinding the study. Details including programming rules are given in the document “TULIP Supplemental Process Description Restricted Medication”.

Appendix B

DERIVATION OF PREDNISONE EQUIVALENT DAILY DOSE

The daily prednisone equivalent OCS dose will be calculated using the total daily dose in the Prior and Concomitant Medication/Treatment form. If the medication is taken less than daily the documented total daily dose 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 (eg, 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.

All oral corticosteroids taken at the respective day (ie, at the date of the respective visit) will be considered for the derivation of the daily dose. The identification of OCS medication is described in the document "TULIP Supplemental Process description OCS".

The following table gives examples of prednisone equivalent doses (eg, 50 mg cortisone is equivalent to 10 mg prednisone). A complete list is given in the document "TULIP Supplemental Process description OCS".

Conversion to prednisone equivalent dose

Medication (preferred term)	Equivalent dose
Prednisone	10 mg
Cortisone	50 mg
Hydrocortisone	40 mg
Methylprednisolone	8 mg
Prednisolone	10 mg
Triamcinolone	8 mg
Betamethasone	1.2 mg
Dexamethasone	1.5 mg
Budesonide	2.25 mg
Deflazacort	12 mg

Appendix C

DERIVATION RULES FOR IMPUTATION OF PARTIAL OR MISSING DATES

If any medications reported are not able to be determined as prior or concomitant due to missing or partial start dates and/or stop dates, the following imputation rules will be implemented:

- If the year is present but the month and day are missing, then 01JAN will be imputed for the start date and 31DEC for the stop date.
- If the year and month are present but the day is missing, then 01 will be imputed for the start date and the last day of the month for the stop date.
- If the start date is completely missing and the end date is prior to the first dose date of investigational product the medication will be considered prior.
- If the start date is completely missing and the end date is missing or on or after the first dose of investigational product the medication will be considered prior and concomitant.
- If the end date is completely missing and the start date is on or after the first dose of investigational product the medication will be treated as concomitant.
- If the end date is completely missing and the start date is prior to the first dose date of investigational product the medication will be considered prior and concomitant.

Appendix D

REFERENCE RANGES AND TELVC FOR LABORATORY VALUES

Parameter	Unit	Low value	Low decrease	High value	High increase
Haematology					
Haemoglobin	g/L	≤60	NA	≥200	NA
		≤70 and decrease from BL ≥15			
Haematocrit	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
Platelet Count	10E9/L	≤20		≥600	NA
		≤50 and decrease from BL ≥25			
INR		NA	NA	≥4.5	NA
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
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

Parameter	Unit	Low value	Low decrease	High value	High increase
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					
Urine protein/ creatinine ratio	g/mmol	NA	NA	≥0.395	NA
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

Appendix E

ECG: POTENTIALLY CLINICAL SIGNIFICANT POST BASELINE VALUES

ECG parameter	Unit	Low value	Low decrease	High value	High increase
RR interval	ms	<500	NA	>1500	NA
PR interval	ms	NA	NA	≥240	NA
QRS duration	ms	≤60	NA	≥140	NA
QT	ms	≤300	NA	≥500	≥60
QTcF	ms	≤300	NA	≥500	≥30, ≥60
				≥500 and increase from baseline ≥30, ≥500 and increase from baseline ≥60	
QTcB	ms	≤300	NA	≥500	≥30, ≥60
				≥500 and increase from baseline ≥30, ≥500 and increase from baseline ≥60	

Appendix F

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	