STATISTICAL ANALYSIS PLAN AMENDMENT 1

Study: SL0023

Product: Dapirolizumab pegol

A Multi-center, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Dose-Ranging Study Followed by an Observational Period to Evaluate the Efficacy and Safety of Dapirolizumab Pegol in Subjects with Moderately to Severely Active Systemic Lupus Erythematosus

SAP/Amendment Number	Date
Final SAP	19 December 2017
Amendment 1	04 September 2018

Confidentiality Stateme

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

Ab	Antibody
ACR	American College of Rheumatology
ADA	Anti-Drug Antibody
AE	adverse event
AESI	adverse event of special interest
AEOI	adverse event of interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANA	antinuclear antibody
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
anti-dsDNA	anti-double-stranded deoxyribonucleic acid antibody
anti-ENA	extractable nuclear antigen antibody
anti-RNP	anti-ribonucleoprotein antibody
anti-SM	anti-Smith antibody
anti-SSA	Sjögren's syndrome antibody A
anti-SSB	Sjögren's syndrome antibody B
aPL	antiphospholipid antibody
aPTT	activated partial thromboplastin time
AST	aspartate aminotransferase
AUC	area under the curve
AUC_{τ}	area under the concentration-time curve over the dosing interval
AUCMB	area under the curve minus baseline
BICLA	BILAG 2004-based Composite Lupus Assessment
BILAG 2004	British Isles Lupus Assessment Group Disease Activity Index 2004
BLQ	below level of quantification
BP	blood pressure
BST	BILAG 2004 System Tally
C3	complement 3
C4	complement 4

CLASI	Cutaneous Lupus Erythematosus Disease Area and Severity Index
C_{max}	maximum plasma concentration
CS	Completer Set
CSR	Clinical Study Report
C-SSRS	Columbia Suicide Severity Rating Scale
CTCAE	Common Terminology Criteria for Adverse Events
C_{trough}	predose plasma concentration
DBP	diastolic blood pressure
DEM	Data Evaluation Meeting
DMC	Data Monitoring Committee
DNA	deoxyribonucleic acid
DORIS	Definitions of Remission in SLE
DZP	dapirolizumab pegol
ECG	electrocardiogram
eCRF	electronic Case Report Form
eGFR	estimated glomerular filtration rate
ES	Enrolled Set
EWV	Early Withdrawal Visit
FAS	Full Analysis Set
FDA	Food and Drug Administration
GCP	Good Clinical Practice
GEE	Generalized Estimating Equation
GGT	gamma-glutamyltransferase
HIV	human immunodeficiency virus
HRQoL	health-related quality of life
hsCRP	high sensitivity C-reactive protein
ICH	International Council for Harmonization
IEC	Independent Ethics Committee
Ig	immunoglobulin
IMP	investigational medicinal product
IPD	Important Protocol Deviation
IRB	Institutional Review Board

iv	intravenous(ly)
LDH	lactate dehydrogenase
LLDAS	Lupus Low Disease Activity State
LLN	lower limit of normal
LLOQ	lower limit of quantification
LOQ	limit of quantification
MCAR	missing completely at random
MCP-Mod	Multiple Comparison Procedure–Modeling
MMRM	Mixed Models for Repeated Measurement
mRNA	messenger ribonucleic acid
n	number of observations
N	number of subjects
PBO	placebo
PD	pharmacodynamics(s)
PDILI	potential drug-induced liver injury
PEG	polyethylene glycol
PGA	Physician's Global Assessment of Disease
PK	pharmacokinetic(s)
PR	pulse rate
PRO	patient-reported outcome
PtGA	Patient's Global Assessment of Disease
RBC	red blood cell
RF	rheumatoid factor
RS	Randomized Set
SAE	serious adverse event
SAP	Statistical Analysis Plan
SBP	systolic blood pressure
sBST	simplified BILAG 2004 System Tally
SD	standard deviation
SFU	Safety Follow-up
SJC	swollen joint count
SLE	systemic lupus erythematosus

SLE-E	Systemic Lupus Erythematosus Emotional Instrument
SLE-F	Systemic Lupus Erythematosus Fatigue Instrument
SLE-M	Systemic Lupus Erythematosus Mobility Instrument
SLE-P	Systemic Lupus Erythematosus Pain Instrument
SLE-S	Systemic Lupus Erythematosus Symptom Inventory Instrument
SLEDAI-2K	Systemic Lupus Erythematosus Disease Activity Index 2000
SLICC	Systemic Lupus International Collaborating Clinics
SMQ	Standardized MedDRA Queries
S2K RI-50	SLEDAI-2K Responder Index-50
SS	Safety Set
TB	tuberculosis
TEAE	treatment-emergent adverse events
TJC	tender joint count
ULN	upper limit of normal
uPCR	urine protein-creatinine ratio
VAS	visual analog scale
WBC	white blood cell
WOCF	worst observation carried forward

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide all information that is necessary to perform the interim and final statistical analysis of SL0023 to support the interim and final clinical study reports (CSRs). It also defines the summary tables, listings and figures to be included in the clinical study reports according to protocol.

The SAP is based on the following study documents: Protocol amendment 1: 14 Dec 2016.

The content of this SAP is compatible with the International Conference on Harmonization (ICH)/ Food and Drug Administration (FDA) E9 Guidance documents.

2 PROTOCOL SUMMARY

This study is a 2-part study consisting of a randomized, double-blind, placebo (PBO)-controlled, parallel-group, dose-ranging period (24-week Double-Blind Treatment Period; Part 1) followed by a 24-week Observational Period (Part 2) in adult subjects with moderately to severely active systemic lupus erythematosus (SLE) who are receiving stable standard-of-care medications (i.e., corticosteroids, immunosuppressants, and/or antimalarials) at study entry.

The primary objective of SL0023 is to assess the dose-response for the efficacy of intravenous (iv) dapirolizumab pegol (DZP; 3 dose groups) at Week 24 of the Double-Blind Treatment Period (Part 1). Secondary objectives are to assess the efficacy of the individual dose regimens of iv DZP at Week 24 and to assess the safety and tolerability of iv DZP. Other objectives are to assess the efficacy of iv DZP at additional time points; to assess the corticosteroid-sparing effect of iv DZP; to assess the pharmacokinetics (PK) of iv DZP and polyethylene glycol (PEG); to assess the pharmacodynamics (PD) of iv DZP; to assess the immunogenicity of iv DZP and PEG; to assess the effects of iv DZP on health-related quality of life (HRQoL), symptoms, fatigue, mobility, and pain; to perform exploratory analyses of the effects of iv DZP with transcriptomic and proteomic biomarkers; to assess durability of the clinical response after withdrawal of iv DZP; and to assess PD after withdrawal of iv DZP (potentially including gene transcription signature).

The primary efficacy variable is the British Isle Lupus Assessment Group Disease Activity Index 2004 (BILAG 2004)-based Composite Lupus Assessment (BICLA) responder rate across 3 doses of DZP and PBO at Week 24. The secondary efficacy variable is the BICLA responder rate in the individual dose groups at Week 24. Several other efficacy variables assessing disease activity, corticosteroid-sparing effects, HRQoL, symptoms, fatigue, mobility, and pain are also included.

Safety variables include adverse events (AEs) and serious AEs (SAEs), subject withdrawals due to AEs, vital sign parameters, electrocardiograms (ECGs), and safety laboratory tests. Several variables evaluating PK, PD, biomarkers, and immunogenicity will also be assessed.

The study consists of a Screening Period of up to 4 weeks, a 24-week Double-Blind Treatment Period (Part 1), and a 24-week Observational Period (Part 2). At the start of Part 1 of the study, eligible subjects will be randomized (1:1:1:1) to 1 of 4 treatment arms (DZP 6mg/kg, 24mg/kg, or 45mg/kg, or PBO) and stratified in accordance with corticosteroid dose (≤10mg/day or >10mg/day prednisone equivalent) determined at Screening. Study drug will be administered by iv infusion every 4 weeks during Part 1.

Subjects who withdraw early from the 24-week Double-Blind Treatment Period (Part 1) will enter an 8-week Safety Follow-up (SFU) Period (which ends 12 weeks after the final dose of study drug). Subjects who complete the 24-week Double-Blind Treatment Period (Part 1) will continue into a 24-week Observational Period (Part 2), during which subjects will not receive study drug but will be able to receive standard-of-care treatment, as indicated. The maximum duration of the entire study (Parts 1 and 2) per subject will be approximately 52 weeks.

The study is planned to enroll at least 267 subjects in order to randomize approximately 160 subjects (40 subjects per treatment group). Eligible subjects are males or females, ≥18 years of age, diagnosed with SLE according to the Systemic Lupus International Collaborating Clinics (SLICC) Classification Criteria and having moderately to severely active disease despite stable standard-of-care treatment with at least 1 of the following, either alone or in combination: corticosteroids (at doses of ≤40mg/day prednisone or equivalent), and/or antimalarials, and/or immunosuppressants.

An initial biomarker analysis may be performed on all evaluable biomarker data from Part 1 of the study when 25% to 35% of the randomized subjects have completed Visit 7 (Week 12) of the Double-Blind Treatment Period. Data from subjects who have prematurely withdrawn from the study prior to Visit 7 will be included. The initial biomarker analysis will focus on transcriptomic biomarker variables, prespecified in a Biomarker Plan, in order to provide guidance for the DZP development program, and will not affect the conduct of the study. A final analysis of biomarker data will be conducted at the end of the study.

After the last subject completes Part 1 of the study, the data for those subjects will be secured and an interim Clinical Study Report will be written for the Double-Blind Treatment Period (Part 1) of the study. This interim CSR will not include any data from the Observational Period (Part 2).

2.1 Study objectives

2.1.1 Primary objective

The primary objective of SL0023 is to assess the dose-response for the efficacy of iv DZP (3 dose groups) at Week 24 of the Double-Blind Treatment Period (Part 1) in adult subjects with moderately to severely active SLE receiving stable standard-of-care treatment.

2.1.2 Secondary objectives

The secondary objectives of SL0023 are:

- To assess the efficacy of the individual dose regimens of iv DZP at Week 24
- To assess the safety and tolerability of iv DZP

2.1.3 Other objectives

The other objectives of the study are:

- To assess the efficacy of the individuals dose regimens of iv DZP at additional time points
- To assess the corticosteroid-sparing effect of iv DZP
- To assess the PK of iv DZP and PEG
- To assess the PD of iv DZP

- To assess the immunogenicity of iv DZP and PEG
- To assess the effects of iv DZP on HRQoL, symptoms, fatigue, mobility, and pain
- To perform exploratory analyses of the effects of iv DZP with transcriptomic and proteomic biomarkers
- To assess durability of the clinical response after withdrawal of study drug
- To assess PD after withdrawal of study drug (potentially including gene transcription signature)

2.2 Study variables

The measures employed in this study include variables to explore clinical disease activity and to assess organ system function, which may be compromised as a result of disease activity in subjects with SLE.

2.2.1 Efficacy variables

2.2.1.1 Primary efficacy variable

The primary efficacy variable is the BICLA responder rate across 3 doses of DZP and PBO at Week 24 (see Section 8.1.1 for a detailed description of this variable).

2.2.1.2 Secondary efficacy variable

The secondary efficacy variable is the BICLA responder rate in the individual dose groups at Week 24.

2.2.1.3 Other efficacy variables

Other efficacy variables being evaluated by visit will include all scheduled visits for that particular variable.

Clinical assessments of disease activity

- BICLA responder rates by visit
- Time to first maintained BICLA response (defined as BICLA response at 2 consecutive study visits)
- Percentage of subjects with maintained BICLA response from Week 12 to Week 24
- Percentage of subjects with maintained BICLA response from Week 24 to Week 48
- Number and percent of subjects with BILAG 2004 improvement (all BILAG 2004 Grade A improved to B, C, or D and all BILAG 2004 Grade B improved to C or D) by visit
- Absolute and change from Baseline in total BILAG 2004 score by visit
- BILAG 2004 Systems Tally (BST)
- Absolute and change from Baseline in S2K RI-50 score by visit
- Absolute and change from Baseline in SLEDAI-2K score by visit
- Change from Baseline in Physician's Global Assessment of Disease (PGA) and Patient's Global Assessment of Disease (PtGA) scores by visit

- SLE Responder Index-4 (SRI-4) responder rates by visit
- SLE Responder Index-5 (SRI-5) responder rates by visit
- SLE Responder Index-6 (SRI-6) responder rates by visit
- Percentage of subjects in each category in the Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index score at Week 24
- Percentage of subjects in each category in the Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index score at Week 48
- Percent change from Baseline in tender joint count (TJC) by visit
- Percent change from Baseline in swollen joint count (SJC) by visit
- Absolute and change from Baseline in Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI) by visit
- Time to flare up to Week 24 (with the occurrence of a flare determined by the Investigator)
- Time to flare after Week 24 (with the occurrence of a flare determined by the Investigator)
- Time to flare from Baseline up to Week 48 (with the occurrence of a flare determined by the Investigator)
- Time to flare up to Week 24 (with the occurrence of a flare defined as: a new sustained BILAG 2004 Grade A/B [present at ≥2 consecutive visits] or start of new induction therapy [either corticosteroid dose increase to ≥0.5 mg/kg/day prednisone equivalent or initiation of cyclophosphamide, rituximab, iv Ig, MMF, MTX, Azathioprin, or plasma exchange])
- Time to severe flare up to Week 24 (with the occurrence of a severe flare defined as a BILAG severe flare)
- Time to severe flare after Week 24 (with the occurrence of a severe flare defined as a BILAG severe flare)
- Time to severe flare from Baseline up to Week 48 (with the occurrence of a severe flare defined as a BILAG severe flare)
- Time to moderate flare up to Week 24 (with the occurrence of a moderate flare defined as a BILAG moderate flare)
- Time to moderate flare after Week 24 (with the occurrence of a moderate flare defined as a BILAG moderate flare)
- Time to moderate flare from Baseline up to Week 48 (with the occurrence of a moderate flare defined as a BILAG moderate flare)
- Time to mild flare up to Week 24 (with the occurrence of a mild flare defined as a BILAG mild flare)
- Time to mild flare after Week 24 (with the occurrence of a mild flare defined as a BILAG mild flare)

- Time to mild flare from Baseline up to Week 48 (with the occurrence of a mild flare defined as a BILAG mild flare)
- Time to severe flare (Variant 1) up to Week 24 (with the occurrence of a severe flare defined as a BILAG severe flare - Variant 1: a new BILAG 2004 Grade A since the previous visit)
- Time to severe flare (Variant 1) after Week 24 (with the occurrence of a severe flare defined as a BILAG severe flare - Variant 1: a new BILAG 2004 Grade A since the previous visit)
- Time to severe flare (Variant 1) from Baseline up to Week 48 (with the occurrence of a severe flare defined as a BILAG severe flare - Variant 1: a new BILAG 2004 Grade A since the previous visit)
- Time to moderate/severe flare (Variant 1) up to Week 24 (with the occurrence of a moderate/severe flare defined as a BILAG moderate/severe flare - Variant 1: a new BILAG 2004 Grade A or 2 new BILAG 2004 Grade Bs since the previous visit)
- Time to moderate/severe flare (Variant 1) after Week 24 (with the occurrence of a moderate/severe flare defined as a BILAG moderate/severe flare - Variant 1: a new BILAG 2004 Grade A or 2 new BILAG 2004 Grade Bs since the previous visit)
- Time to moderate/severe flare (Variant 1) from Baseline up to Week 48 (with the occurrence of a moderate/severe flare defined as a BILAG moderate/severe flare - Variant 1: a new BILAG 2004 Grade A or 2 new BILAG 2004 Grade Bs since the previous visit)
- Variables assessing corticosteroid-sparing effects
- Absolute daily corticosteroid dose (in prednisone equivalent)
- Percentage of subjects with a daily corticosteroid dose (prednisone-equivalent) of 7.5mg prednisone equivalent or less and BICLA response at Week 24
- Percentage of subjects with a daily corticosteroid dose (prednisone-equivalent) of 7.5mg prednisone equivalent or less and a maintained BICLA response from Week 12 to Week 24
- Percentage of subjects with a daily corticosteroid dose (prednisone-equivalent) of 7.5mg prednisone equivalent or less and SRI-4 response at Week 24
- Time-weighted area under the curve in corticosteroid dose (prednisone-equivalent) for the period covering Baseline to Week 24
- Percentage of subjects with a daily corticosteroid dose (prednisone-equivalent) of 7.5mg or less prednisone equivalents by visit

  Percentage of subjects with no concomitant corticosteroid treatment by visit

Patient-reported outcome variables

- Change from Baseline in Lupus Quality of Life (LupusQoL) questionnaire scores by visit
- The following exploratory patient-reported outcome (PRO) variables being developed by UCB will be assessed at sites in English- and Spanish-speaking countries (approximately 45% to 50% of the projected study population):
- SLE-Symptom Inventory Instrument (SLE-S) by visit

- SLE-Fatigue Instrument (SLE-F) by visit
- SLE-Mobility Instrument (SLE-M) by visit
- SLE-Pain Instrument (SLE-P) by visit
- SLE-Emotional States Instrument (SLE-E) by visit

2.2.2 Safety variables

The following safety variables will be evaluated in the course of the study:

- Incidence of adverse events (AEs), serious adverse events (SAEs), AEs leading to discontinuation of study drug and/or study, and adverse event of interest (AEOIs) including and not limited to infusion reactions, thromboembolic events, neurological events and malignancies
- Vital signs
- 12-lead ECG
- Hematology:
  - Hemoglobin
  - Hematocrit
  - Mean corpuscular volume (MCV)
  - Red blood cell (RBC) count
  - Mean corpuscular hemoglobin concentration (MCHC)
  - Mean corpuscular hemoglobin (MCH)
  - White blood cell (WBC) count, WBC differential (basophils, eosinophils, lymphocytes, monocytes, and neutrophils)
  - Platelet count
- Clinical chemistry:
  - Liver function tests (bilirubin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), gamma-glutamyltransferase (GGT), and lactate dehydrogenase (LDH))
  - Creatinine
  - Urea nitrogen
  - Electrolytes (sodium, potassium, calcium, and phosphate)
  - Total cholesterol and triglycerides
  - Total protein and albumin
  - Glucose
  - Lipase

- Creatine phosphokinase

- Additional laboratory assessments
  - Coagulation and hemostasis tests: prothrombin time, activated partial thromboplastin time, prothrombin intl. normalized ratio, and fibrinogen (fibrinogen was optional, ordered by the sites as needed)
  - Urinalysis: pH, bilirubin, protein, blood, glucose, nitrite, ketone, leukocytes, urobilinogen
  - Urine chemistry and microscopy, as applicable:
    a. Protein, albumin, creatinine (for protein:creatinine ratio (mg/mmol) and albumin:creatinine (mg/mmol) ratios)
    b. Microscopy of urine sediment for RBCs, RBC casts, WBCs, and WBC casts
  - Erythrocyte sedimentation rate (ESR)

- C-SSRS assessment of suicidal ideation and behavior

2.2.3 Pharmacokinetic/pharmacodynamic variables

The following PK variables will be evaluated:

- For plasma samples: Plasma concentrations of DZP and PEG
- For urine samples: Urine amount of PEG

The following PD variable will be evaluated:

- Whole blood mRNA signature profiling to characterize potential changes in gene expression relevant to the inflammatory and immune response to DZP in SLE

2.2.3.1 Immunological variables

- Lupus Autoantibody Profile: anti-double-stranded deoxyribonucleic acid (anti-dsDNA) antibodies, antinuclear antibody (ANAs), extractable nuclear antigen antibody (anti-ENA), anti-Smith antibody (anti-SM), Sjögren's syndrome antibody A (anti-SSA), Sjögren's syndrome antibody B (anti-SSB), anti-ribonucleoprotein antibody (anti-RNP), rheumatoid factor (RF), and antiphospholipid (aPL) antibodies (anticardiolipin antibodies, lupus anticoagulant, and anti β2 glycoprotein antibodies)
- Ig (total Ig, IgG, IgM, IgA, and IgE)
- High sensitivity C-reactive protein (hsCRP)
- C3 and C4 levels

2.2.4 Immunogenicity variables

The following immunogenicity variable will be evaluated:

- Anti-DZP and anti-PEG antibodies

2.3 Study design and conduct

SL0023 is a 2-part study consisting of a randomized, double-blind, PBO-controlled, parallel-group, dose-ranging period (24-week Double-Blind Treatment Period; Part 1) followed by a 24-

week Observational Period (Part 2) in adult subjects with moderately to severely active SLE who are receiving stable standard-of-care medications (i.e., corticosteroids, immunosuppressants, and/or antimalarials) at study entry. Subjects (male or female, ≥18 years of age) who enroll in this study must have moderately to severely active SLE, defined as having BILAG 2004 Grade A level disease activity in ≥1 body/organ system or BILAG 2004 Grade B in ≥2 body/organ systems if no BILAG 2004 Grade A level disease is present and SLEDAI-2K score ≥6 at Screening.

A target size of at least 267 subjects has been selected in order to randomize approximately 160 subjects (40 subjects per treatment group). The study will be conducted at approximately 70 sites.

The study consists of a Screening Period of up to 4 weeks, a 24-week Double-Blind Treatment Period (Part 1), and a 24-week Observational Period (Part 2). Subjects will be randomized (1:1:1:1) to 1 of 4 treatment arms (DZP 6mg/kg, 24mg/kg, or 45mg/kg, or PBO) and stratified in accordance with Screening corticosteroid dose (≤10mg/day or >10mg/day prednisone equivalent). Study drug will be administered by iv infusion every 4 weeks during Part 1 of the study.

Subjects who withdraw during the 24-week Double-Blind Treatment Period (Part 1) will enter an 8-week SFU Period (which ends 12 weeks after the final dose of study drug). Subjects who complete the 24-week Double-Blind Treatment Period (Part 1) will continue into a 24-week Observational Period (Part 2), during which subjects will not receive study drug but will be able to receive standard-of-care treatment, as indicated.

The maximum duration of the entire study (Parts 1 and 2) per subject will be approximately 52 weeks, including a Screening Period of up to 4 weeks, a 24-week Double-Blind Treatment Period (Part 1), and a 24-week Observational Period (Part 2).

The end of the study is defined as the date of the last visit of the last subject in the study.

After the last subject completes Part 1 of the study, the data for those subjects will be secured and an interim CSR will be written for the Double-Blind Treatment Period (Part 1) of the study. A complete final CSR will be written for the entire study following completion of the Observational Period (Part 2) of the study.

A schedule of study assessments is provided in the Section 13.1.

2.4 Determination of sample size

The sample size of this study is governed by determining whether there is a dose-response relationship across the 3 doses of DZP and PBO using the Multiple Comparison Procedure–Modelling (MCP-Mod) methodology. The primary endpoint of the study is a binary indicator of whether the subject responds to treatment.

Based upon prior experience in SLE, a PBO responder rate for this study of 25% is postulated. The desired clinical effect is 29% above PBO, resulting in a responder rate of 54%. With power of 80%, a (multiple-comparison corrected) one-sided type-1 error of 5% and all dose groups of the same size, 112 subjects are required to complete Part 1 of the study.

Adjusting for an expected dropout rate of 30% results in the need to randomize 160 subjects (40 subjects per treatment group) and assuming a screening failure rate of approximately 40%, it is anticipated to enroll approximately 267 subjects in this study.

3 DATA ANALYSIS CONSIDERATIONS

3.1 General presentation of summaries and analyses

All statistical analyses will be performed by UCB. Computations and generation of outputs will be performed using SAS® Version 9.4.

All original and derived parameters will be listed and described using summary statistics (number of observations (n), mean, standard deviation (SD), median, minimum and maximum, unless otherwise stated) or frequency counts (number of subjects (N) and percentages).

Unless otherwise stated, listings will be sorted by treatment (PBO, then DZP 6mg/kg, 24mg/kg, or 45mg/kg), subject number within treatment group (not randomization number), parameter (if applicable) and visit (if applicable; including timing relative to dosing if applicable). For listings including non-randomized subjects, the non-randomized subjects will be shown first in the listing, ordered by subject number. All listings will include repeated and unscheduled measurements. Such measurements will appear in chronological order together with the scheduled visits, i.e., a repeated measurement will appear directly after the visit and time relative to dosing, for which the repeat measurement was performed. If imputed values will be included in the listings, these values have to be flagged. In all the listings dates will be presented in the format 'YYYY-MM-DD' and times will be presented in 24h clock format as 'hh:mm' or 'hh:mm:ss' where appropriate.

For continuous variables, descriptive statistics will include number of subjects (n), mean, standard deviation (SD), median, minimum, and maximum. Decimal places for descriptive statistics will always apply the following rules:

- "n" will be an integer.
- Mean, SD, and median will use 1 additional decimal place compared to the original data.
- Minimum and maximum will have the same number of decimal places as the original value.

All descriptive statistics will be presented by treatment where applicable, including columns for all treated subjects and all subjects on demographics, baseline characteristics, and adverse event displays, and using the available data for the study population as observed. Summaries of concentration data will be based on geometric mean and geometric standard deviation instead of arithmetic versions. All tabulations will be sorted by treatment, parameter and visit (including time relative to dosing if applicable, unless otherwise stated). Only scheduled visits and times relative to dosing will be included in the tabulation.

Categorical data will be summarized by visit and treatment group, including columns for all treated subjects and all subjects on demographics, baseline characteristics, and adverse event displays, using the number and percent of subjects in each category. Percentages will be based on the corresponding population size (i.e., the denominator of percentages should match the sample size in the column header), unless otherwise noted via footnote in the applicable summary table. Percentages will be presented to 1 decimal place. For data points with n=0 (i.e., no subjects in the applicable category), no value for percentage of subjects will be displayed.

Missing observations will be included in the denominator unless otherwise specified on the table. For the shift tables, the percentages are based on those with non-missing at both time points.

Statistical tests of efficacy variables will be presented as 2-sided p-values, except for p-values from the 1-sided MCP-Mod testing of the primary efficacy variable. P-values will be rounded to 4 decimal places. P-values less than 0.0001 will be presented as "<0.0001" and p-values greater than 0.9999 will be presented as ">0.9999."

Handling of multiplicity for the primary efficacy variable is discussed in Section 4.7. All other variables will be considered exploratory and will be assessed at a nominal 2-sided α=0.05 significance level.

3.2 Analysis time points

The timeframe for this SAP encompasses the entire study, which extends from Screening to 28 weeks after last dose, i.e. 24 weeks of Double-Blind Treatment Period and 24 weeks of Observational Period (up to Week 48). Details of the study periods are given below. All analyses described in this SAP will be run for the Double-Blind Treatment Period (up to Week 24) for the interim analysis after Part 1 is completed (unless otherwise specified, e.g., analyses of time to flare after Week 24, or summaries of AE groupings that include AEs beyond Week 24). All analyses described in the SAP will be run again for the entire study, Double-Blind Treatment Period (up to Week 24) + Observational Period (up to Week 48), for the final analysis after Part 2 is completed. However, for a few summaries and analyses it is specifically stated that they are run only for the period up to Week 24 or the period after Week 24, and as such they will only include those study periods even for the final analysis.

The following study periods will be included in the analysis:

- Screening Period: up to 28 days (Day -28 through Day -1).
- Double-Blind Treatment Period (Part 1): 24 weeks, starts with the randomization (1:1:1:1) of eligible subjects to 1 of 4 treatment arms (DZP 6mg/kg, 24mg/kg, or 45mg/kg, or PBO), stratified in accordance with screening corticosteroid dose (≤10mg/day or >10mg/day prednisone equivalent). Study drug will be administered by iv infusion every 4 weeks during Part 1, starting on Day 1.
- Observational Period (Part 2): 24 weeks, subjects who complete the 24-week Double-Blind Treatment Period will continue into this 24-week Observational Period. During Part 2, subjects will not receive study drug but will be able to receive standard-of-care treatment, as indicated.
- Early Withdrawal Visit (EWV)/Safety Follow-up (SFU): Subjects who withdraw prior to Week 32 (Visit 12) will complete an EWV (Visit 10 assessments if withdrawn prior to Visit 10; Visit 16 assessments if withdrawn after Visit 10) as well as Visits 11 and 12 for SFU assessments at 8 and 12 weeks after their final dose of study drug (as applicable). Subjects who withdraw after Week 32 (Visit 12) do not require additional SFU visits and will complete the Visit 16/EWV assessments only.

The time points for individual assessments are provided in Table 13–1~~Table 13–1~~.

Analyses conducted by visit will include each planned study visit at which the assessment was scheduled. Data for planned study visits will be based on the nominal time point and will not

include any unplanned visits. Usually, efficacy data will not be obtained at unscheduled visits, with the exceptions: labs in BILAG and SLEDAI-2K. Here labs that are obtained up to 2 weeks after the original visit are used to replace missing lab data.

The minimum post-baseline and maximum post-baseline values included in certain analyses by visit are not specific time points; rather they are the minimum and maximum observed non-missing values per subject during SL0023, inclusive of all scheduled and unscheduled visits.

Analyses which are presented by visit may also include an Early Withdrawal time point for subjects prematurely discontinuing the study. For subjects withdrawing on the same calendar day as a planned study visit, the assessments planned at that study visit will be summarized with the planned study visit and additional assessments not planned for the regular study visit, will be mapped to the next planned timepoint for this assessment. For subjects not withdrawing on the same calendar day as a planned study visit, the data will also be mapped to the next planned study visit. These Early Withdrawal assessments will then be included as observed data with the visits to which they are mapped in all analyses, except in the case of mNRI and strict NRI imputations, in which a subject is counted as a non-responder at the visit at which they discontinue study treatment or withdraw from the study and at all subsequent visits.

Subject data listings will include all time points for collected data, including all assessments (planned or unplanned). Data for derived time points Last Visit and Early Withdrawal will be included in subject data listings.

3.3 Relative study days

Relative day will accompany visit dates, sample dates, onset/start dates, and resolution/stop dates in data listings, and will be calculated as follows:

- If the event start (stop) date occurred prior to first infusion, then relative day is calculated as start (stop) date minus first dose date. The relative day in this situation should be preceded by a '-'.
- If the event start (stop) date occurred on or after the first infusion and through the date of the last infusion, relative day is calculated as start (stop) date minus first dose date + 1.
- If the event start (stop) date occurred after the last/latest infusion, the relative day is calculated as event start (stop) date minus date of last/latest infusion. The relative day in this situation should be preceded by a '+'.

3.4 Definition of baseline values

Unless otherwise specified in the SAP, the last value obtained prior to the first infusion of study drug (Visit 2) will be used as the baseline value. If a baseline measurement is missing, and a screening value available, the screening value will be utilized as Baseline instead.

For vital signs, a measurement prior to, or at the moment of, the start of the first infusion will be available for each visit. The last measurement prior to, or at the moment of, the start of the first infusion of study drug will be used as the baseline value. An additional analysis will be performed using the pre-infusion values as baseline values and determine the changes from pre-infusion to post-infusion for each infusion visit. At each post-baseline visit, the pre-infusion value will be one taken prior to, or at the moment of, the start of the infusion of study drug.

For baseline assessments, if the BILAG 2004 is completed, but the grade for 1 or more individual system(s) is missing, the grade for the system will be set to a grade of BILAG 2004 D if no value can be imputed from the Screening Visit. This is a conservative approach for subjects on active doses, since no improvement can occur following a baseline grade of BILAG 2004 D level disease.

3.5 Protocol deviations

Important protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being.

Important protocol deviations are defined in the Important Protocol Deviations (IPD) document. Sources for identifying protocol deviations will include the clinical database and a cumulative protocol deviation log from the clinical sites. Protocol deviations will be classified at blinded data cleaning and data evaluation meetings. Any exclusion of subjects from analysis sets or individual data from summary outputs will be documented in the meeting minutes.

For subjects enrolled under different versions of the protocol, the programming of protocol deviations should be performed according to the version of the IPD relevant to the specific protocol. The IPD will provide further details with regards to which version of the protocol the specific IPD version is applicable to. The version of the protocol to be followed for each subject will be provided by the Project Manager.

3.6 Analysis sets

3.6.1 Enrolled Set

The Enrolled Set (ES) will consist of all subjects who have given informed consent.

3.6.2 Randomized Set

The Randomized Set (RS) will consist of all subjects randomized into the study.

3.6.3 Safety Set

The Safety Set (SS) will consist of all subjects who are randomized and have received at least 1 dose (any amount) of study drug. Safety variables will be analyzed using the SS.

It is expected that subjects will receive treatment as randomized; hence, safety analyses will be based on the randomized treatment group. However, if after unblinding it is determined that subjects received the incorrect treatment (i.e., not as per the randomization schedule), then for safety analyses subjects will be allocated to the actual treatment they received (as-treated) for the data summaries, assuming that the incorrect treatment was consistently received throughout the study.

3.6.4 Full Analysis Set

The Full Analysis Set (FAS) will consist of all subjects who have received at least 1 full dose of study drug and have at least one post-Baseline efficacy measurement during the Double-Blind Treatment Period, with the exception of 5 patients who were randomized at site 321, who will be excluded from the FAS. Based on medical review of data during the course of the study, inconsistencies and implausibilities were identified and could neither be adequately explained by source data nor feedback by site personal. These findings put the adequately medically qualified

assessment of the disease activity at screening (determination of eligibility), baseline and post-baseline visits in question. After careful consideration of the problems identified, UCB feels that these findings may seriously compromise the ability to draw valid conclusions from the clinical trial. For this reason, all 5 subjects enrolled at Site 321 will be excluded from the FAS. However, these 5 subjects will be included for a sensitivity analysis which will be done in addition.

The FAS is the primary analysis set for efficacy analyses. In case of mistreatment, subjects will be analyzed as randomized for the FAS. The FAS will be used for the Interim Analysis of the study (up to Week 24). The FAS will also be used for the Final Analysis of the study (through Week 48).

3.6.5 Per Protocol Set

The per Protocol Set (PPS) will consist of subjects in the FAS who have received at least one full dose of study drug and have no important protocol deviations during the Screening and Double-Blind Treatment Period that may influence the validity of the data for the primary efficacy variable, the BICLA.

Analyses of the primary and secondary efficacy variable, as well as BICLA responder by visit and SRI-4 response by visit, will be repeated using the PPS for the Double-Blind Treatment Period only.

3.6.6 Pharmacokinetic Per Protocol Set

The Pharmacokinetic Per Protocol Set (PK-PPS) will consist of all subjects in the SS with one full dose of study drug who have provided at least 1 PK sample, and for whom no important protocol deviations affecting the PK variables have been documented. This analysis set will be used for all PK analyses.

3.6.7 Completer Set

The Completer Set (CS) will consist of all subjects in the FAS who did not permanently discontinue study drug during the Double-Blind Treatment Period (Part 1), completed Week 24 of Part 1, and continued into the Observational Period (Part 2) of the study. The CS will only be used for the Final Efficacy Analysis of the study (through Week 48). Note that completion of Week 24 requires that the subject has completed the efficacy scales, BILAG, SLEDAI, and PGA, which are required for determining BICLA response.

3.7 Treatment assignment and treatment groups

For all analyses (where applicable) the results will be presented by treatment group:

- SOC + PBO iv q4w
- SOC + DZP 6mg/kg iv q4w
- SOC + DZP 24mg/kg iv q4w
- SOC + DZP 45mg/kg iv q4w

Selected parameters (where stated in the following sections) may be summarized by treatment and subgroup (Definition of subgroups, s. Section 4.10). The columns ‘SOC + DZP All Subjects’ and ‘All Subjects’ will be included in the tables for disposition, demographic, baseline

characteristics, prior and concomitant medications, corticosteroid dose at screening and baseline, exposure and AEs.

3.8 Center pooling strategy

Disposition tables will be displayed by region. The results of the secondary efficacy variable, the BICLA response rates in the individual dose groups at Week 24, will be displayed by country and region. All other results will be summarized across centers and will not be stratified by center. There will be no specific pooling strategy.

3.9 Coding dictionaries

All AEs, concomitant diseases, and medical history will be coded for analysis according to the Medical Dictionary for Regulatory Activities (MedDRA)® coding dictionary, version 19.1. Prior and concomitant medications will be coded for analysis using the World Health Organization Drug dictionary (WHO-DD), version SEP/2015. The version number of coding dictionary will be displayed within the table.

3.10 Changes to protocol-defined analyses

- A wording clarification (vs. the protocol wording) was made in the definition of a BICLA responder point 3, relating to PGA:

  No worsening in the Physician's Global Assessment of Disease Activity **(PGA)** compared to study entry ("no worsening" is defined as **either no worsening or worsening** <10% **of the full 100mm VAS**, equivalent to **less than** a 10 mm increase **in the PGA compared to study entry score**); and

- For CLASI, the protocol stated that associated symptoms, including itch, pain, and fatigue would be recorded separately on a 1 to 10 VAS by the subjects. However, the VAS assessments were not collected.

- The following endpoints have been included as other efficacy variables – clinical assessment of disease activity:

  – Percentage of subjects in each category in the Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index score at Week 48.

  – Time to severe flare up to Week 24 (with the occurrence of a severe flare defined as a BILAG severe flare)

  – Time to moderate flare up to Week 24 (with the occurrence of a moderate flare defined as a BILAG moderate flare.

  – Time to mild flare up to Week 24 (with the occurrence of a mild flare defined as a BILAG mild flare.

  – Time to severe flare (Variant 1) up to Week 24 (with the occurrence of a severe flare defined as a BILAG severe flare - Variant 1: a new BILAG 2004 Grade A since the previous visit)

- Time to moderate/severe flare (Variant 1) up to Week 24 (with the occurrence of a moderate/severe flare defined as a BILAG moderate/severe flare - Variant 1: a new BILAG 2004 Grade A or 2 new BILAG 2004 Grade Bs since the previous visit)
- Time to severe flare after Week 24 (with the occurrence of a severe flare defined as a BILAG severe flare)
- Time to moderate flare after Week 24 (with the occurrence of a moderate flare defined as a BILAG moderate flare.
- Time to mild flare after Week 24 (with the occurrence of a mild flare defined as a BILAG mild flare.
- Time to severe flare (Variant 1) after Week 24 (with the occurrence of a severe flare defined as a BILAG severe flare - Variant 1: a new BILAG 2004 Grade A since the previous visit)
- Time to moderate/severe flare (Variant 1) after Week 24 (with the occurrence of a moderate/severe flare defined as a BILAG moderate/severe flare (Variant 1): a new BILAG 2004 Grade A or 2 new BILAG 2004 Grade Bs since the previous visit)
- Time to severe flare from Baseline up to Week 48 (with the occurrence of a severe flare defined as a BILAG severe flare)
- Time to moderate flare up to Week 48 (with the occurrence of a moderate flare defined as a BILAG moderate flare.
- Time to mild flare up to Week 48 (with the occurrence of a mild flare defined as a BILAG mild flare.
- Time to severe flare (Variant 1) from Baseline up to Week 48 (with the occurrence of a severe flare defined as a BILAG severe flare - Variant 1: a new BILAG 2004 Grade A since the previous visit)
- Time to moderate/severe flare (Variant 1) from Baseline up to Week 48 (with the occurrence of a moderate/severe flare defined as a BILAG moderate/severe flare - Variant 1: a new BILAG 2004 Grade A or 2 new BILAG 2004 Grade Bs since the previous visit)
- Time to flare up to Week 24 (with the occurrence of a flare defined as: a new sustained BILAG 2004 Grade A/B [present at ≥2 consecutive visits] or start of new induction therapy [either corticosteroid dose increase to ≥0.5mg/kg/day prednisone equivalent or initiation of cyclophosphamide, rituximab, iv Ig, MMF, MTX, Azathioprin, or plasma exchange])

- The endpoint of time to flare after Week 24:

  Time to flare after Week 24 (with the occurrence of a flare defined as: a new sustained BILAG 2004 Grade A/B [present at ≥2 consecutive visits] after withdrawal of study drug or start of new induction therapy [either corticosteroid dose increase to ≥0.5mg/kg/day prednisone equivalent or initiation of cyclophosphamide, rituximab, iv Ig, or plasma exchange] after withdrawal of study drug)

was removed from the analysis because after Week 24 BILAG was only assessed once every other visit (every 8 weeks) which would not allow for a valid assessment of sustained BILAG Grade A/B at consecutive visits.

- The specification of patient-reported outcome variables being developed by UCB (SLE-S, SLE-F, SLE-M, SLE-P, SLE-E) was edited versus the protocol to remove reference to the "Change from Baseline…scores". These endpoints do not have continuous numerical scores and as such, no change from Baseline score can be calculated. Only descriptive frequency summaries per item will be provided, as stated in Section 8.3.3. The full analysis of these exploratory PRO variables is defined in a stand-alone mixed method analytical plan outside the SAP.
- A new nomenclature was included for SRI-50. To avoid confusion with the SRI response SRI-50 was renamed to S2K RI-50.
- Definition of SRI-4, SRI-5, SRI-6 was changed to:
  - No more than 1 shift from BILAG 2004 Grade C, D, or E to B post-Baseline

  The protocol wording for this criteria was "No more than 1 shift from BILAG 2004 Grade C, D, or E to B, or from B to A, post-Baseline". However, the allowance of a shift from B to A was an error and was in conflict with the second criterion of the definition which states "No shift from BILAG 2004 Grade B, C, D, or E to A post-Baseline".
- Immunological variables were only included in Section 2.2.3.1 'Immunological variables', not additionally in Section 11.3.2 'Additional laboratory evaluations', according to protocol.
- Additional endpoints were included in the SAP, see Section 8.4.
- In protocol section 15.6, it states that as an additional sensitivity analysis for SLEDAI-2K total score, if items are missing "rather than imputing missing component values from previous visit values, when a total score cannot be computed due to 1 or more missing individual items, the total score will be imputed by extrapolating from the score based on the nonmissing items at the visit. In this scenario, provided that at least 75% of the individual items have a nonmissing response, the total of nonmissing items will be computed. This total will then be extrapolated to the equivalent value based on the total maximum value of the SLEDAI-2K instrument."

  It was decided that there was not strong rationale for this approach, because the individual items on the SLEDAI-2K scale refer to all different organ systems. Responses for one organ system do not indicate what the response in another organ system may have been. Therefore, extrapolating non-missing responses from some organ systems to complete the SLEDAI-2K total score for missing responses in other organ systems is not a good approach, and this sensitivity analysis was not done. SLEDAI-2K total score will be analyzed with imputation for missing individual items using values from the previous visit, using MMRM and ANCOVA. Additionally, SLEDAI-2K total score will be analyzed using a worst observation carried forward approach for missing individual items and missing total scores.

4 STATISTICAL/ANALYTICAL ISSUES

4.1 Adjustments for covariates

Randomization will be stratified by Screening corticosteroid dose (prednisone-equivalent). Accordingly, corticosteroid strata will be included in statistical models. For the primary efficacy variable, the corticosteroid strata will be accounted for as detailed in Section 8.1.2.2. For the secondary efficacy variable a GEE analysis controlling for corticosteroid strata will be performed. The odds ratio of the responder rates at Week 24 will be estimated and presented together with the associated 95% confidence interval.

For the analysis of other efficacy variables, statistical models will include treatment group as a factor and control for corticosteroid strata. GEE and MMRM models will include visit and treatment*visit in the models. For models of continuous endpoints, the Baseline value will also be included in the model as a covariate.

4.2 Handling and definition of the corticosteroid prednisone equivalent dose

For the purposes of further analysis and tabulations, all corticosteroid doses (including both prior and concomitant corticosteroids taken for any indication) will be converted to prednisone equivalent doses. The conversion table for systemic corticosteroids is presented in Table 4–1~~Table 4–1~~. Systemic corticosteroids are those with a route of administration of oral, intravenous, or intramuscular. All others (including topical, ocular, nasal, subcutaneous, intra-articular, etc.) are non-systemic. All non-systemic steroids will have a prednisone equivalent dose = 0mg/day. Budenoside is considered non-systemic even if taken orally, and will always be assigned a prednisone equivalent dose = 0mg/day.

The prednisone equivalent dose will be included in the listing. For example, if the total daily dose of triamcinolone is 8mg, the equivalent total daily dose of prednisone will be 10mg.
If a new corticosteroid is reported that is not in Table 4–1~~Table 4–1~~, the appropriate prednisone-equivalent conversion will be determined and the prednisone equivalent dose will be included in the listing.

Formatte

Formatte

Table 4–1: Prednisone equivalent doses of systemic corticosteroids

Steroids	Dose equivalent to 5mg/day Prednisone
Cortisone	25
Hydrocortisone	20
Deflazacort	6.5
Prednisolone	5
Prednisone	5
Methylprednisolone	4
Triamcinolone	4
Dexamethasone	0.75

Table 4–1: Prednisone equivalent doses of systemic corticosteroids

Betamethasone	0.6

Note: All non-systemic corticosteroids (including topical, ocular, nasal, subcutaneous, intra-articular, etc.) will be assigned a prednisone equivalent dose = 0mg/day.

Cortisol (hydrocortisone) is the standard of comparison for glucocorticoid potency.

4.3 Handling of dropouts or missing data

Imputation of missing data will be done as described for missing dates and/or times for concomitant medication and AE data (Section 6.4 and Section 11.2, respectively), and for the BILAG 2004 body/organ system score, for the SLEDAI-2K total score, for the Physician's Global Assessment of Disease and the efficacy variables, as described below in this section. All calculations and statistical analyses will be performed using the available data (including data for subjects who discontinued the study treatment).

A modified non-responder imputation (mNRI) for missing data will be used for the primary efficacy variable (responder rate at Week 24 across 3 doses of DZP and PBO according to a combined response index (BICLA)). If Week 24 assessment was done but the value for a single component (BILAG 2004, SLEDAI-2K, or Physician's Global Assessment of Disease) is not available, then the missing component value will be imputed from the respective previous visit value prior to computing the BICLA treatment response variable (limited to 1 visit back). In the case where the predose laboratory result is missing (at Screening and at Baseline and there is no other unscheduled predose laboratory result), the BILAG 2004 body/organ system Baseline score for the body system requiring the missing laboratory result will be set to a score of BILAG 2004 D level disease, see Section 3.4.

If the subject has 2 or more missing components for the visit being analyzed, then the BICLA treatment response value is considered to be unknown (ie, left as missing and therefore considered a non-responder in the primary analysis approach, and subject to the missing data approach specified for other sensitivity analyses). In the case where a subject has a missing component, but already is known to have failed to achieve responder status on the basis of the non-missing components (e.g., SLEDAI-2K is missing but the non-missing BILAG 2004 may not have demonstrated sufficient improvement), the missing value does not need to be imputed to know the subject's treatment response classification which is non-responder. In addition, subjects who prematurely withdraw from the study or who discontinue treatment prematurely (while still remaining in the study) are categorized as a non-responder on that visit and all subsequent visits for the primary analysis, although other approaches will be taken in sensitivity analyses.

For the BILAG 2004 and SLEDAI-2K, if the assessment was done at a visit, but individual item scores/grades within the scale are missing, the respective individual item scores/grades will be imputed from the respective previous individual item scores/grades (limited to 1 visit back), in order to obtain a total score (or total assessment of improvement or worsening for BILAG) for the scale at the current visit, which can then be used as a component of the BICLA response variable. If the values of the individual item scores/grades are also missing at the previous visit,

then they cannot be imputed for the current visit, and the scale score is then missing for the current visit.

The analysis of the secondary and other efficacy responder rates will use a similar approach for handling of dropouts or missing data. Note that for analysis of a responder rate at Week 4, if imputation from the respective previous visit is required due to missing data, a value from Visit 2 may be used.

For the other continuous efficacy variables that are scales made up of multiple individual items (BILAG 2004 total score, SLEDAI-2K total score, CLASI Activity Score, CLASI Damage Score, and SLICC/ACR Damage Index), observed case data will include limited imputation as follows. Assessments completed with one or more individual scores missing will impute individual scores/grades from the previous visit (limited to 1 visit back). This will be done so that one missing item score will not render the entire scale total score incalculable. Note that for analysis at Week 4, if imputation from the respective previous visit is required due to missing data, a value from Visit 2 (Visit 1 for SLICC/ACR Damage Index) may be used. Beyond that, no direct imputation will be done for a missing assessment at a visit, and the efficacy response will be considered to be unknown (ie, left as missing and subject to the analysis approach applied to continuous efficacy variables, mixed effects models for repeat measurement (MMRM) and supportive Analysis of Covariance (ANCOVA) using observed case data). For other continuous efficacy variables, analyses will be based on observed case data with no imputation of any individual items or total scores, unless otherwise noted for the individual endpoint in Section 8.3.1.

As stated above, subjects who prematurely withdraw from the study or who discontinue treatment prematurely (while still remaining in the study) are categorized as a non-responder on that visit and all subsequent visits for the primary analysis. However, there is no date collected for discontinuation of study medication when a patient remains in the study. Additionally, the date collected for study withdrawal is often long after the decision to withdraw from the study due to completion of follow-up visits and documentation. Therefore, the following algorithm will be used to determine the date at which a subject is considered a non-responder in efficacy analyses due to study withdrawal or study treatment discontinuation:

If a subject has withdrawn from the entire study (without having previously withdrawn study drug), the subject should be considered a non-responder starting on the ***earliest*** of these dates:

- the withdraw date given in the Subject Disposition CRF (i.e., stored in SDTM DS dataset)
- if withdrawal during the double-blind period and subject did attend the next scheduled infusion visit after the last infusion received, then the date of the next scheduled infusion visit after the last infusion, at which the subject should have had an infusion but did not
- if withdrawal during the double-blind period and the next scheduled infusion visit was missed/skipped after the last infusion, then the last infusion date+28 days
- If withdrawal during the observational period, then date of last visit attended

If a subject has withdrawn *only* from study drug, then the subject should be considered a non-responder starting on the *earlier* of these dates:

- the date of the next scheduled visit at which the subject should have had an infusion but did not, if the subject did attend the next scheduled infusion visit
- the last infusion date+28 days (in case the next scheduled infusion visit was missed/skipped after the last infusion)

The above proposed algorithm is for determining at what point the subject is considered a non-responder for efficacy analyses.

It also needs to be determined when a subject Discontinued Study Drug or Withdrew from the Study relative to study visits for the summary of Discontinuation Reasons by Visit Table. Specifically, at each visit, it needs to be determined whether a subject discontinued study drug or withdrew from the study prior to that visit.

If a subject has withdrawn from the entire study, on the Discontinuation Reasons by Visit Table the subject is counted as withdrawing from the study prior to visit beginning with:

- the next scheduled visit after the last visit attended (if the withdraw date in DS >= the last visit attended)
- the next visit attended after the withdraw date in DS (if the withdraw date in DS < the date of the last visit attended)

If a subject has withdrawn *only* from study drug, on the Discontinuation Reasons by Visit Table the subject is counted as discontinuing study drug prior to visit beginning with:

- the next scheduled visit at which the subject should have had an infusion but did not

4.4 Handling of Retests

All original and retested assessments will be listed. In case of retest of a variable, the following rules are applied:

- The last non-missing value of baseline assessments is used for statistical analyses.
- The nominal visit value of post-dose assessments (discharge included) is used for statistical analyses. Retests of post-dose assessments will not be used in analyses and summaries, but will be included in listings.

These rules imply that:

- When more than 1 assessment is available for a specific study moment, these assessments are organized as first the original assessment, assigned to the nominal visit, and then the retested assessment(s), assigned to unscheduled visits. The date and time of measurement are used to sort the data.
- For BP, when selecting the last non-missing value for Baseline, a complete set of both SBP and DBP is selected for statistical analyses.

For laboratory assessments:

- Not assessable values are excluded from the statistical analyses.

- When selecting the last non-missing value for Baseline, a complete set of WBC counts from the same laboratory sample assessment is selected (neutrophil count, eosinophil count, basophil count, monocyte count, atypical lymphocyte count and lymphocyte count). The rounded sum of these counts should be equal to 100, if the individual counts are expressed in %.

4.5 Interim analyses and data monitoring

A Data Monitoring Committee (DMC) will be formed to monitor the safety of the study. The composition and operation of the DMC will be defined in the DMC Charter. The presentation and analysis of data for the DMC meetings is described within a separate SAP.

In Part 1 (double-blind part) of the study 3 Data Evaluation Meetings (DEMs) are planned for assessment of aggregate data throughout the study:

DEM-1 will take place after 15 to 25% of the enrolled subjects have complete data of primary variable. Objectives: Evaluation of aggregate data in order to identify systemic data errors and potential trends in important protocol deviations, and review of aggregate data to detect outliers and to evaluate relevant data distributions.

DEM-2 will take place after 50 to 60% of the enrolled subjects have complete data of primary variable. Objectives: Evaluation of aggregate data in order to identify systemic data errors and potential trends in important protocol deviations, discussion of dry-run tables, listings and figures, and finalization of the content of outputs defined in the SAP.

DEM-3 will take place after 100% of the enrolled subjects have complete data of primary variable, all data are entered and approximately 95% to 100% of the data are clean. Objectives: confirmation of statistical assumptions, discussion of any outstanding analysis topics and finalization of rules for exclusion from analysis set.

DEM-3a will take place after 100% of the enrolled subjects have complete data through the end of the Observational Period of Early Withdrawal, all data are entered and approximately 95% to 100% of the data are clean. Objectives: Final evaluation of data to ensure readiness for final database lock.

After the last subject completed Part 1 of the study an interim CSR will be written for the Double-Blind Treatment Period (Part 1) of the study. A complete final CSR will be written for the entire study following completion of the Observational Period (Part 2) and will include all information of the entire study (Double-blind and Observational Period). Final tables should include all tables run for the interim analysis, updated to include all study visits where appropriate. In general, the interim analysis will cover the visits up to Week 24 and the final analysis will cover all visits up to Week 48. All efficacy tables of the final analysis will be run on the FAS and CS. Tables that only cover Week 24 to Week 48 will be run on the CS.

4.6 Multicenter studies

The results of the secondary efficacy variable, the BICLA response rates in the individual dose groups at Week 24, will be displayed by country and region. The other data from all centers will be pooled for the purposes of the final analyses. There will be no formal statistical evaluation of the effect of center on the results obtained.

4.7 Multiple comparisons/multiplicity

Multiple Comparison Procedure-Modeling (MCP-Mod) methodology (Bretz et al, 2005) will be used to test the dose-response relationship across 3 doses of DZP and PBO for the BICLA responder rate. Monotonic dose-response trends (as well as non-monotonic dose-response trends in one sensitivity analysis) will be tested using the optimal contrasts as determined by the MCP-Mod methodology such that the overall Type I error rate is controlled at 0.05 (one-sided testing). Analyses of secondary and other efficacy variables will not be adjusted for multiple comparisons, and as such, p-values are considered descriptive and caution should be exercised in interpreting them.

4.8 Use of an efficacy subset of subjects

Not applicable.

4.9 Active-control studies intended to show equivalence

Not applicable.

4.10 Examination of subgroups

The primary efficacy variable will be summarized by treatment group and subgroups of interest, as defined below:

- Region (Western/Central Europe, Eastern Europe, Latin America, North America)
- Demographic factors at Screening:
  - Age: < median age, ≥ median age (Median age based on the Safety Set)
  - Weight: ≤ 80kg, > 80kg
  - BMI: < median BMI, ≥ median BMI (Median BMI based on the Safety Set)
  - Gender: female, male
  - Race: White vs. non-White
  - Ethnicity: Not Hispanic/Latino vs. Hispanic/Latino
- Baseline disease activity:
  - High SLEDAI-2K at Baseline: <10, ≥10
  - SLEDAI-2K at Baseline: < median, ≥ median (Median SLEDAI-2K based on the Safety Set)
  - BILAG 2004 at Baseline (severely active patients): Subjects with at least 1 BILAG 2004 Grade A, subjects with no BILAG 2004 Grade A at Baseline (Visit 2) [a]
  - BILAG 2004 at Baseline (multi-organ moderately to severely active): Subjects with >2 BILAG Grade As or Bs, subjects with ≤2 BILAG 2004 Grade As or Bs at Baseline (Visit 2)
  - Lower BILAG 2004 activity at Baseline: Subjects with maximum 1 BILAG 2004 Grade B (BILAG C, D or E allowed) at Baseline, subjects with at least 2 BILAG 2004 Grade Bs or ≥1 BILAG 2004 Grade A at Baseline

  - No significant disease activity at Baseline: Subjects with only BILAG 2004 Grade C, D or E in all organ systems, subjects with at least 1 BILAG 2004 Grade B in one organ system
  - Musculoskeletal only: BILAG 2004 Grade A in musculoskeletal at Baseline and no BILAG 2004 Grade A or B in another organ system, all other
  - BILAG 2004 score at Baseline: ≤12, >12
  - CLASI Activity Score at Baseline: <4, ≥4
- SLE associated lab parameters
  - Anti ds DNA status at Baseline: anti-dsDNA in Farr assay ≤ 10iU (negative/ indeterminate), > 10iU (positive)
  - ANA titer at Screening: ≤1:160, >1:160 (For ANA titer, 1:80 < 1:160, i.e., it is the denominators that are assessed to determine whether the ANA titer is < or > a cutoff.)
  - Any low complement at Baseline: Subjects with C3 **or** C4 < LLN, subjects with both C3 and C4 ≥LLN
  - Both low complement at Baseline: Subjects with C3 **and** C4 < LLN, subjects with either C3 or C4 ≥LLN
  - Anti-dsDNA and low complement at Baseline: Anti-dsDNA Farr > 10iU and either C3 or C4 < LLN, all other
- Concomitant medications at Baseline (Visit 2)
  - Corticosteroid dose (prednisone-equivalent) randomization strata: ≤10mg/day, >10mg/day
  - Corticosteroid dose (prednisone-equivalent) high dose/low dose: ≤7.5mg/day, >7.5mg/day
  - Corticosteroid vs. no dose: 0 mg/day, >0mg/day
  - Anti-malarial at Baseline: yes, no
  - Immunosuppressants at Baseline: yes, no
- Other
  - Unblinded cases: Subjects unblinded during the study for any reason (e.g., qualifying safety event)

All subgroup summaries will be provided for BICLA by visit for the FAS for the interim analysis, and for the FAS and the CS for the final analysis. Descriptive statistics by subgroup will be presented for the BICLA responder by visit. There will be no statistical analyses within subgroups. A patient data listing including all subgroups will be provided.

Furthermore, subgroup summaries will be provided for treatment-emergent AEs (TEAEs). Incidences of TEAEs will be presented by the following subgroup:

- aPL status positive at any time during the study (Screening or Week 24 for Interim Analysis assessment of the Double-Blind Period; and Screening, Week 24, or Week 48 for Final Analysis assessments including AEs during the Observational Period) vs. negative at all time points assessed during the study. aPL positive is defined as subject has ever experienced at least 1 of the following immunologic criteria during the study: positive test result for anti-β2 glycoprotein (IgG>ULN or IgM>ULN), medium or high titer anticardiolipin antibody level (IgG>ULN or IgM>ULN), positive lupus anticoagulant test (time (noted as DRVVT in the laboratory dataset)>ULN **and** ratio (noted as DRVVTRT in the laboratory dataset)≥ULN).

There will also be subgroup summaries provided (for the final analysis) for BICLA by visit for the FAS and CS and for TEAEs by the following subgroups:

- anti-DZP: any positive [pre or post] during study vs. negative at all time points analyzed,
- anti-PEG: any positive [pre or post] during study vs. negative at all time points analyzed.

5 STUDY POPULATION CHARACTERISTICS

5.1 Subject disposition

Subject disposition will be listed for all subjects screened and will include the following information: subject status (screen failure, study completed or discontinued, study drug completed or discontinued, study drug discontinued, but study continued), date of informed consent, date of randomization, date of first and last dose of study drug (including relative day for last dose, as described in Section 3.3), date of premature study and/or study drug termination (if applicable) and primary reason for study and/or study drug termination (if applicable). The date of premature study withdrawal is taken from the CRF, however the date of study drug termination will be determined as per the algorithm in Section 4.3. The listing will also include the date and reason for breaking the randomization code (if applicable) as well as the date of final contact for the subject and the previous subject number, if a subject was re-screened. For screen failures the date and reason for screen failure will be listed.

Subject disposition will be summarized for all subjects screened including the reasons for screen failure, the number of subjects re-screened and the number of subjects randomized. Subject disposition will also be summarized for the Randomized Set by treatment and overall, including the reasons for discontinuation of study and/or study drug and the number of subjects with premature termination of study drug but not early withdrawn from the study. Details of discontinuation due to AEs will be tabulated. Reasons for discontinuation of study and/or of treatment and visit of discontinuation will be listed.. Discontinuations will be assigned to visits as per the algorithm in Section 4.3.

A Kaplan Meier figure will display the Kaplan Meier estimates for subjects who were neither withdrawn from the study nor permanently withdrawn fromstudy drug vs. time. The time until discontinuation of study drug or premature withdrawal from the study will be calculated using the date at which a subject is considered a non-responder in efficacy analyses due to study withdrawal or study treatment discontinuation determined as per the algorithm in Section 4.3. The reason for this approach is that the figure presents the time to study drug discontinuation or premature withdrawal, so it should use the time to whichever event occurs first. Similarly, the

date at which a subject is considered a non-responder in efficacy analyses due to study withdrawal or study treatment discontinuation is the earliest date of either study treatment discontinuation or study withdrawal as per the Disposition CRF. If a subject has not prematurely withdrawn from the study nor discontinued study drug by Week 24 (for the interim analysis), she/he should be censored on the Week 24 visit date (or the planned Week 24 visit date if the visit was missed). If a subject has not prematurely withdrawn from the study nor discontinued study drug by Week 48 (for the final analysis), she/he should be censored on the day of the Week 48 visit date (or final visit date if Week 48 visit was missed).

A subject who is rescreened will be counted as 2 screened subjects in disposition tables that provide a count of screened subjects. That way each reason for screen failure will be tabulated as well. The disposition tables will indicate how many subjects were rescreened and thus counted twice in the total number screened.

Study eligibility criteria will be listed and a separate listing of subjects who did not meet the eligibility criteria will be presented. Only failed criteria will be included in the former listing.

Finally, a listing of study visit dates will be presented by subject including the relative study day (calculated as described in Section 3.3) for each visit.

To further support public disclosure, a summary of the number and percentage of subjects in each study defined Analysis Set, by treatment and overall, will be produced for the Final Analysis only based on the Enrolled Set.

5.2 Protocol deviations

Important protocol deviations will be identified and classified by the deviation types defined in the Important Protocol Deviations (IPD) document, which will include at least the deviation types, listed below. A listing of all important protocol deviations identified at the Data Evaluation Meeting will be presented for all subjects in the Randomized Set, and will include deviation type and description. The number and percentage of subjects in the Randomized Set with important protocol deviations will be summarized overall and by treatment.

The following deviation types are defined in the IPD:

- Inclusion criteria
- Exclusion criteria
- Withdrawal criteria
- Study drug compliance
- PK blood sampling deviations
- PK urine sampling deviations
- Prohibited concomitant medication use

Deviation types will be classified according to the following categories as defined in the IPD:

- Important deviations that may significantly affect a subject's rights or safety
- Important deviations for the PPS
- Important deviations for the PK-PPS

Subjects are excluded from the PPS if they have an important protocol deviation that may impact the results of the primary efficacy analysis. Subjects are excluded from the PK-PPS if they have an important protocol deviation that may impact the results of the PK analysis. The denominator for the percentages will be the number of subjects in the Randomized Set for each treatment or overall.

The assignment of subjects to each of the analysis sets will be listed for all subjects screened. In addition, a listing of all screened subjects who did not meet study eligibility criteria will be presented.

6 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

6.1 Demographics

Tables with descriptive statistics and listings will be given for the demographic variables age (at time of informed consent), gender, weight and height at Screening, race, ethnicity, and body mass index (BMI). Demographic characteristics will be summarized for the FAS, the SS and the CS by treatment and overall. Age will be summarized as a continuous variable and as a categorical variable based on the following categories: <65 years, ≥65 - <75 years, ≥75 years. Age will also be summarized by additional categories required by EudraCT and clinicaltrials.gov for public disclosure.

6.2 Other Baseline characteristics

Systemic lupus erythematosus history (BILAG 2004 SLE History) will be listed for subjects in the FAS including the date of first diagnosis and whether the subject has ever experienced an SLE-related sign or symptom in each body system (constitutional, mucocutaneous, neuropsychiatric, musculoskeletal, cardiorespiratory, gastrointestinal, ophthalmic, renal, hematology and vasculitis). The number of subjects who ever experienced an SLE sign or symptom (BILAG 2004 SLE History) will be tabulated for each Body System for the FAS and CS as well.

Procedure history will be listed chronologically for all subjects in the SS including the reported term for the procedure, the date of the procedure and the information whether the procedure is related to SLE.

Information about the SLE disease status at Screening and at Baseline will be summarized in tables for the SS, FAS, and CS. Baseline is already defined in Section 3.4 as the last value obtained prior to the first infusion of study drug (Visit 2). For the purpose of the display of Baseline Characteristics at Screening only, Screening will be defined as the last value obtained prior to Visit 2, i.e., it will be the last value obtained at Visit 1 or at an unscheduled visit prior to Visit 2.

The information about SLE disease status at Screening and at Baseline will include:

- the frequencies of British Isles Lupus Assessment Group Disease Activity Index 2004 (BILAG 2004) grades in each system,
- the descriptive statistics for the BILAG 2004 total score,
- the number of subjects with BILAG 2004 Total Score <20 or with BILAG 2004 Total Score ≥20,

- the descriptive statistics for the total score of the SLEDAI-2K (with and without inclusion of labs as a component of the SLEDAI-2K score) (where labs are excluded, the items excluded from the SLEDAI-2K score are: 'low complement ', 'anti-DNA', 'urinary casts', 'hematuria', 'proteinuria', 'pyuria', 'thrombocytopenia', and 'leukopenia'),
- the number of subjects with SLEDAI-2K <10 or with SLEDAI-2K ≥10,
- the number of subjects with SLEDAI-2K (calculated with dsDNA positivity defined as anti-dsDNA ≥ 5 iU) < 10 or with SLEDAI-2K (calculated with dsDNA positivity defined as anti-dsDNA ≥ 5 iU) ≥ 10,
- ,
- the descriptive statistics for the Physician's Global Assessment of Disease (PGA) score,
- the descriptive statistics for the Patient's Global Assessment of Disease (PtGA) score,
- the results for the Systemic Lupus International Collaborating Clinics (SLICC) Damage Index: the number of subjects with a SLICC score '0' (no organ damaged), a SLICC score '1' (one organ damaged) or a SLICC score '>1' (multiple organ damaged).

Furthermore, the following disease specific baseline parameters will be included in the table:

- the descriptive statistics for the time since diagnosis,
- the number of flares in the 6 months prior to Screening,
- At Screening and at Baseline: the number of subjects
  - fulfilling screening eligibility criteria: either ≥1 BILAG Grade A or ≥2 BILAG Grade Bs
  - who do not fulfill screening eligibility anymore but still are evaluable: 1 BILAG Grade B only, no BILAG Grade A
  - who do not fulfill screening eligibility and are not evaluable: no BILAG Grade A, no BILAG Grade B
  - with at least 1 organ system severely active: ≥1 BILAG Grade A
  - with only 1 (severely) involved organ system: 1 BILAG Grade A and no other organ system with BILAG Grade A, Grade B,
  - with multi-organ involvement: ≥2 organ systems with BILAG Grade A or Grade B
  - with multiple severely active organs: ≥2 organ systems with BILAG Grade A
  - with multiple organ involvement but no severely active: ≥2 organ systems with BILAG Grade B, no organ system with BILAG Grade A,
- the number of subjects with aPLs in central lab (at least ever experienced 1 of the following immunologic criteria: positive test result for anti-β2 glycoprotein I, medium or high titer anticardiolipin antibody level, positive lupus anticoagulant test), for each criterion and combined,

- the number of subjects with a thromboembolic event in the medical history (based on Thromboembolic Standardized MedDRA Queries (SMQs) plus the additional item "antiphospholipid syndrome" and tabulated overall),
- the number of subjects with anti-dsDNA (based on data from central lab [0 to <5iU (negative), ≥5 to ≤ 10iU (intermediate-positive), > 10iU (positive), ≥ 5 iU (intermediate-positive and positive)] and also based on data from the SLICC classification form which reflects whether a patient has ever experienced anti-dsDNA positivity in the past),
- the number of subjects with ANA titer classified as ANA titer < 1:80, ANA titer ≥ 1:80 and ≤ 1:160, or ANA titer > 1:160 (For ANA titer, 1:80 < 1:160, i.e., it is the denominators that are assessed to determine whether the ANA titer is < or > a cutoff.),
- the number of subjects with low complement (either low complement 3 (C3), low complement 4 (C4), either C3 or C4, or both), and
- the number of subjects with renal dysfunction (defined as estimated GFR<60mL/min/1.73m^3),
- the number of subjects with uPCR >113 mg/mmol
- the number of subjects on corticosteroids
- the number of subjects on antimalarials
- the number of subjects on immunosuppressants
- the number of subjects on corticosteroids and antimalarials
- the number of subjects on immunosuppressants and antimalarials
- the number of subjects on corticosteroids and immunosuppressants
- the number of subjects on corticosteroids, antimalarials and immunosuppressants
- the number of subjects in each region, and
- the number of subjects in each country.

All baseline characteristics will be listed.

6.3 Medical history and concomitant diseases

Medical history and ongoing medical conditions will be listed for all subjects in the SS including the reported term, start date (month and year only), end date (month and year only) (or ongoing if applicable) and the information whether the reported condition is related to SLE. A glossary of all medical conditions will also be presented including the reported term, the preferred term and the System Organ Class.

The classification of medical history and concomitant diseases will be done according to the MedDRA coding system (version 19.1) using system organ class (SOC), high level term and preferred term. These data will be summarized using frequency tables for subject count by classified data.

6.4 Prior and concomitant medications

Any medication with end date and time before first administration of study drug will be considered and labeled as 'prior medication'.

Medication started and not stopped before first study drug administration date will be considered and labeled as 'concomitant medication'. Medication for which the start date and time is between the start of study drug administration and the end of the Safety Follow-up (SFU) Period (12 weeks after the final dose of study drug) will be considered and labeled as 'concomitant medication'.

'Observational Period medication' will be defined as any medication taken during the period starting at the Week 24 visit and running through the end of the study. In case of early withdrawal, medications taken during the period starting at 4 weeks after the final dose of study drug and running through the end of the study will be considered as 'Observational Period medication'. As such, some medications will be considered as 'concomitant medication' and as 'Observational Period medication'.

Any medications with incomplete start and end dates/times will be handled according to the following rules for classification as prior and concomitant and for the calculation of relative study days. Such imputations will only be performed for these classifications and calculations; in the listings all data will be shown as recorded on the eCRF.

Imputation of Partial Start Dates:

- If only the month and year are specified and the month and year of first dose is not the same as the month and year of the start date, then use the 1st of the month.
- If only the month and year are specified and the month and year of first dose is the same as the month and year of the start, then use the date of first dose.
- If only the year is specified, and the year of first dose is not the same as the year of the start date, then use January 1 of the year of the start date.
- If only the year is specified, and the year of first dose is the same as the year of the start date, then use the date of first dose.
- If the date is completely unknown, then use the date of first dose.

Imputation of Partial End Dates:

- If only the month and year are specified then use the last day of the month.
- If only the year is specified then use December 31 of that year.
- If the date is completely unknown, do not impute the stop date.

Imputations of missing data will be performed before calculation of relative study days.

Prior, concomitant and Observational Period medications will be listed for subjects in the SS, separately in a listing for SLE medications and an overall listing for all medications. Variables will be included which classify the medication as prior, concomitant and/or Observational Period. Furthermore, the following information will be included in the listings: reported term, dose per intake and dose unit, frequency, formulation, indication (including free text recorded by Investigator for non-SLE indications), start and stop dates (including relative study day), and in

case of missing stop date, the information whether the intake of medication is ongoing. Additionally, for immunosuppressants, anti-malarials, and corticosteroids that are prior medications (stopped before the study), the reason for discontinuation or change of dose will be included in the listings. A listing of subjects with concomitant medical procedures will also be provided.

Furthermore, the number of subjects with prior, concomitant or Observational Period medication will each be analyzed separately using frequency tables based on classified data by SLE-related/not SLE-related (information about (no) SLE-relation will be provided by tick box in CRF). The classification of the medication will be done according to the WHO-DD ATC classification, version Sep. 2015. The number of subjects with SLE-related medication at Baseline will also be tabulated. Specifically, the number of subjects taking corticosteroids, the number of subjects taking antimalarials, and the number of subjects taking immunosuppressants at Baseline will be tabulated in the Baseline Characteristics table. Concomitant medications will be manually reviewed by a physician during the study to assign medications to these groups. This review will be done in a blinded manner for the Interim Analysis.

Additionally, the descriptive statistics of the corticosteroid dose (prednisone-equivalent) at Baseline will be tabulated for the SS, FAS and CS analysis sets. The number of subjects with '≤10mg/day corticosteroids' vs. '>10mg/day corticosteroids', with 0mg/day vs. ≤7.5mg/day corticosteroids vs. >7.5mg/day corticosteroids, and with <10mg/day corticosteroids vs. ≥10mg/day corticosteroids will be displayed.

7 MEASUREMENTS OF TREATMENT COMPLIANCE

The IMP will be administered by designated study site staff. The unblinded monitor will review the pharmacy records at each site, including the Drug Dispensing Record Form. The unblinded monitor will compare the dispensing record and vials to the individual subject's identifiers and visit schedule to assure that the subject received the correct treatment and dose, and that the dosing schedule is correct. The unblinded monitor's report will include details of any missed doses, errors in dose, treatment errors or schedule errors, and reasons for these. Subjects are expected to receive all doses of IMP as detailed in the schedule of assessments. Any subject who deviates from the dosing schedule or misses any scheduled treatment should be reported to the Medical Monitor in a blinded fashion for determination of possible schedule adjustments and continued eligibility.

Date and time of start and end of each IMP administration will be listed. Treatment administration including total volume infused, estimated fraction of dose remaining, planned and actual dose of DZP administered (mg), duration of infusion (min) and compliance (%) (derived as described in Section 11.1.1) will be listed (including PBO when applicable). A summary of total volume given, actual dose received, and infusion duration (min) and compliance by visit and the total duration of exposure will also be provided. Descriptive statistics for the total number of infusions administered, for %doses with correct/incorrect (<80%/ >120%) administration, for %skipped doses and %infusions with a shorter infusion time as planned (<90 min) will be tabulated.

The planned volume to be infused is 150 mL for all subjects and doses. The planned infusion time is approximately 120 min.

8 EFFICACY ANALYSES

The efficacy analysis for the Interim Analysis will be based on all subjects of the FAS. The efficacy analysis for the Final Analysis will be based on the FAS and the CS.

Regarding the BILAG-2004, the SLEDAI-2K total score and the Physician's Global Assessment of Disease, as well as other endpoints, the handling of dropouts and missing data is described in Section 4.3.

All statistical models should be controlling for corticosteroid strata unless the model cannot accommodate a covariate or will not converge with a covariate.

The endpoints BICLA, SRI-4, SRI-5, SRI-6, S2K RI-50 and SLEDAI-2K include the SLEDAI-2K Total Score as part of their derivation. The SLEDAI-2K Total Score will be re-calculated based on central laboratory results: Although the SLEDAI-2K Total Score is captured on the CRF, these values should not be used as the SLEDAI-2K values being analyzed. The SLEDAI-2K includes several items that are related to lab parameters. While the majority of these items include an additional judgment by the assessor, the items connected to 'low complement' and 'increased DNA binding' are directly related to lab parameters. For these parameters the central lab results for the lab sample taken at the same visit will be used to determine if an item is considered to be 'present', based on the following definitions:

- 'low complement' is defined as either complement C3 or complement C4 or both is below LLN

- 'increased DNA binding' is defined as a Farr assay >10 iU

If there is no lab sample taken at the same visit, a lab within 14 days after the visit may be used. If there is no lab within 14 days after the visit, the lab results from one visit back may be used. (Note that this also applies for C3 and C4, such that if one is non-missing and the other is missing, the missing one will be imputed as described here using a lab from up to 14 days after the visit, or if that is not available then lab results from one visit back.) Only in case there is no lab value available at the respective visit, within 14 days after the visit, nor at the previous visit, then the data from the CRF will be used if available. The number of points to be added to the SLEDAI-2K score if an item is present is the same whether the item is determined to be present based on CRF data or central lab data.

8.1 Statistical analysis of the primary efficacy variable

8.1.1 Derivation of primary efficacy variable

The primary efficacy variable, the BICLA responder rate across 3 doses of DZP and PBO at Week 24, is a categorical response variable (yes/no) incorporating the following criteria for achievement of responder status (i.e., all criteria must be met to achieve responder status). Note that "study entry" within these criteria refers to Baseline, which is Visit 2 (or an earlier visit if the Visit 2 assessment was not done).

1. BILAG 2004 improvement, defined as BILAG 2004 Grade As at study entry improved to B/C/D, and BILAG 2004 Grade Bs at study entry improved to C/D, and no BILAG 2004 worsening in other BILAG 2004 organ systems (that had BILAG 2004 Grade C, D, or E at Baseline) such that there are no new BILAG 2004 Grades A nor greater than 1 new BILAG 2004 Grade(s) B; and
2. No worsening in the SLEDAI-2K total score compared to study entry (defined as no increase in SLEDAI-2K total score); and
3. No worsening in the Physician's Global Assessment of Disease Activity (PGA) compared to study entry ("no worsening" is defined as either no worsening or worsening <10% of the full 100mm VAS, equivalent to less than a 10mm increase in the PGA compared to study entry score); and
4. No changes in specified concomitant medications as escape medications according to the following criteria (all criteria must be met):
   - No increase of (or addition of a new) immunosuppressant nor antimalarial doses over Baseline (Visit 2) levels, nor initiation of any medication listed in Table 7-1 in the protocol.
   - No increase in oral corticosteroid dose over Baseline (Visit 2) levels for an SLE related indication.
   - No iv, im, nor intra-articular injections of corticosteroids which are SLE related.

For subjects known to be non-responders due to changes in concomitant medications, subjects will be primarily categorized as non-responders throughout the remainder of the study, regardless of visit attendance or assessments performed. A subject will become a non-responder

due to a change in concomitant medication that is considered to be escape medication, beginning the day after the reported start date of the relevant concomitant medication.

In addition, subjects who permanently discontinue study drug prematurely are primarily categorized as non-responders. Any subject who is withdrawn from the study prior to completing the Visit 9 infusion has effectively permanently discontinued study drug prematurely, and is therefore primarily categorized as a non-responder.

8.1.2 Analysis of the primary efficacy variable

The primary efficacy variable is the responder rate at Week 24 across 3 doses of DZP and PBO according to a combined response index (BICLA); the analysis will use a modified non-responder imputation (mNRI) for missing data, s. Section 4.3. The analysis will be run on the FAS and the PPS.

The Week 24 BICLA dose-response relationship across 3 doses of DZP and PBO will be tested using the Multiple Comparison Procedure–Modeling (MCP-Mod) methodology (Bretz et al, 2005). Monotonic dose-response trends will be tested using the optimal contrasts as determined by the MCP-Mod methodology such that the overall Type 1 error rate is controlled at 0.05 (one-sided testing).

8.1.2.1 MCP-Mod - Candidate models

The primary endpoint of the study is a binary indicator of whether the subject responds to treatment. Four different dose-response models ('candidate models') will be used: a linear model, a logistic model, and 2 Emax models, and the MCP-Mod methodology will control for multiplicity.

The analysis will be performed using the "DoseFinding" library in the R statistical package (Bretz et al, 2005). This library implements design and analysis aspects of the MCP-Mod method.

Three families of models will be considered in the candidate set for representing the expected response μ_d (on the logit scale) at dose d:

- Linear — $\mu_d = e_0 + \delta * d$,

  parameters to estimate: e_0 and δ
- Logistic — $\mu_d = e_0 + e_{max} / \{1 + exp([ED_{50} - d] / \delta)\}$
  parameters to estimate: e_0, e_{max}, ED_{50} and δ
- Emax — $\mu_d = e_0 + e_{max} * d / (ED_{50} + d)$,

  parameters to estimate: e_0, e_{max} and ED_{50}

where, $e0$ is the placebo effect (except for in the logistic model), δ is the slope parameter, Emax is the asymptotic maximum effect and $ED50$ is the dose giving half of the asymptotic maximum effect.

The pre-specified standardized parameters for each candidate model are from the sample size calculations:

Linear model: μ_d=d

Logistic model: μ_d=1/{1+exp[(17.2-d)/7]}, i.e., ED_{50}=17.2 and δ=7

Emax model 1: μ_d=d/(13.5+d), i.e., ED_{50}=13.5

Emax model 2: μ_d=d/(25+d), i.e., ED_{50}=25

8.1.2.2 MCP-Mod - Analysis

The analysis will be performed using a two-step approach that extends MCP-Mod methodology to perform dose-reponse modeling and testing in the context of general parametric models (Pinheiro et al, 2014) as follows:

- Using the R function *glm, a* logistic regression model (with ANOVA-type parametrization), including parameters for dose category and corticosteroid stratum, will be fitted to obtain LSmean LogOdds estimates μ_d for each dose and their associated covariance matrix S. The LSmeans for each dose will be averaged over the effect of corticosteroid stratum.
- Then all subsequent inference using the R function *MCPMOD* depends only on the pre-defined standardized model shapes, the fitted lsmeans μ_d and covariance matrix S as inputs. The random seed used for MCPMOD analysis will be 123493.

If at least one model is statistically significant from the multiple contrast test, there is an indication of a dose-response effect.

MCPMOD will also identify the best candidate model using the model selection criteria, lowest AIC (Akaike Information Criteria). The *MCPMOD* selected model will then be used to produce an estimate of the MCP-Mod target dose, defined as the minimum dose that achieves the protocol design treatment effect, 29% on the probability scale given a placebo response of 25%, expressed on the LogOdds scale, i.e., 1.258955. The *MCPMOD* selected model will also be used to produce estimates of the probability of response, difference in probability of response vs. placebo, and odds ratio vs. placebo at selected doses.

The number of BICLA responders and non-responders at Week 24 will be tabulated. The assessment of each subject as a responder or as a non-responder and the assessment for each criterion the BICLA is based on (see Section 8.1.1) will be tabulated by treatment and subject.

As the primary efficacy endpoint is a composite variable, reasons for failure to achieve treatment response and the exclusive reason for failure to achieve treatment response will also be summarized with descriptive statistics. One table includes the subjects who may have more than one reason for non-response/adherence to medication, a second table will include the number of subjects who have one reason exclusively for non-response/adherence to medication. Analysis of the individual components of the composite endpoint is described in Section 8.3.

8.1.2.3 Sensitivity Analyses of the Primary Efficacy Variable - missing data

Sensitivity analyses to examine the impact of missing data on the primary analysis will be conducted. All sensitivity statistical hypothesis tests will be considered descriptive only and will be conducted at the nominal level α=0.05 (one-sided testing).

The primary analysis will be conducted using the FAS. This analysis will be repeated using the PPS as a sensitivity analysis.

To examine the impact of missing data on the primary analysis, missing treatment response status at Week 24 will be imputed using the following 3 alternate methods:

1. mNRI data set where non-missing study response after discontinuation of study drug is used for subjects who discontinued study treatment but remained in the study
2. Observed cases data set (including only observed data at Week 24 with no imputation of any missing data in BICLA components [including no imputation of any missing individual items in the multiple item scales])
3. Strict NRI data set (all/any missing data in BICLA components [including any missing individual items in the scales] at Week 24 leads to non-responder status on the BICLA for Week 24)

The MCP-Mod procedure described in Sections 8.1.2.1 and 8.2.1.2 will be re-run for the FAS on each of these datasets to examine the selected dose-response model and the *MCPMOD* estimated target dose (defined as the minimum dose that achieves the protocol design treatment effect, 29% on the probability scale given a placebo response of 25%, expressed on the LogOdds scale, i.e., 1.258955) identified by the model.

8.1.2.4 Sensitivity Analyses of the Primary Efficacy Variable - candidate models

A sensitivity analysis will be run to explore the impact of the MCP-Mod candidate models on identifying a statistically significant dose-response model. The primary analysis MCP-Mod using mNRI for missing data will be repeated, exactly as described in Sections 8.1.2.1 and 8.1.2.2, except that there will be 3 additional candidate models included in the analysis, in addition to those included in the primary efficacy analysis. The intent is to explore whether there is an alternate dose-response model, perhaps non-monotonic, that is a better fit for the primary efficacy variable than those considered in the primary efficacy analysis.

Two additional families of models will be considered in the candidate set for representing the expected response μ_d (on the logit scale) at dose d:

- Exponential — $\mu_d = e_0 + e1*[exp(d/\delta)-1]$,

  parameters to estimate: e_0, $e1$ and δ
- Beta — $\mu_d = e_0 + e_{max} * B(\delta_1,\delta_2)(d/D)^{\delta 1}(1-d/D)^{\delta 2}$,

  parameters to estimate: e_0, e_{max}, δ_1 *and* δ_2

where, $e0$ is the placebo effect and Emax is the asymptotic maximum effect. For the Exponential model, $e1$ is the slope parameter, and δ controls the convexity of the model. For the Beta model,

δ_1 and δ_2 are shape parameters, D is a scale parameter that will be fixed at 1.2*maximum tested dose (i.e., 1.2*45=54), and $B(\delta_1,\delta_2)=(\delta_1+\delta_2)^{\delta1+\delta2}/(\delta_1^{\delta1}\delta_2^{\delta2})$.

The pre-specified standardized parameters for each candidate model are from the sample size calculations:

Exponential model: $\mu_d=\exp(d/0.7)-1$, i.e., $\delta=0.7$

Beta model 1: $\mu_d=B(0.3,0.8)(d/D)^{0.3}(1-d/D)^{0.8}$, i.e., $\delta_1=0.3$, $\delta_2=0.8$

Beta model 2: $\mu_d=B(0.6,0.8)(d/D)^{0.6}(1-d/D)^{0.8}$, i.e., $\delta_1=0.6$, $\delta_2=0.8$

This sensitivity analysis will be run using the FAS. The sensitivity statistical hypothesis tests will be considered descriptive only and will be conducted at the nominal level $\alpha=0.05$ (one-sided testing).

8.2 Statistical analysis of the secondary efficacy variable

The secondary efficacy variable is the BICLA responder rate (as defined in Section 8.1.1) in the individual dose groups at Week 24.

The main analysis for the secondary variable will be based on the same mNRI as for the primary variable analyzed by a logistic regression model, modeling treatment response at Week 24 as a function of treatment group, controlling for corticosteroid strata. Least squares (LS) mean proportion responding in each treatment group, as well as LS mean differences in the proportion responding, odds ratios, 95% confidence intervals (CIs) for each, and p-values based on the Wald Chi-square statistic for the pairwise comparisons of each DZP treatment group with PBO will be obtained from the logistic regression model and presented. Additionally, the crude difference in proportion responding and standard Wald asymptotic 95% CI based on the normal approximation to the binomial distribution will be presented for each DZP treatment group vs. PBO. This analysis will be conducted using the FAS.

A test for treatment*corticosteroid interaction will be performed in a separate model, and the interaction will be explored if significant.

8.2.1 Sensitivity Analyses of the Secondary Efficacy Variable.

Sensitivity analyses to examine the impact of missing data on the secondary analysis will be conducted as follows.

First, the analysis of the secondary efficacy variable will be repeated for the PPS.

Then, the same logistic regression model with stratification by corticosteroid strata will be applied to the following data sets including 3 alternate methods of imputation of missing data:

1. mNRI data set where non-missing end of study response is used for subjects who discontinued study treatment but remained in the study.
2. Observed cases data set (including only observed data at Week 24 with no imputation of any missing data in BICLA components [including no imputation of any missing individual items in the multiple item scales])
3. Strict NRI data set (all/any missing data in BICLA components [including any missing individual items in the scales] at Week 24 leads to non-responder status on the BICLA for Week 24)

Furthermore, an estimate of the treatment differences at Week 24 will be obtained from a Longitudinal Generalized Estimating Equation (GEE) model for binary outcome, controlling for corticosteroid strata, and with a treatment by visit interaction.

The Generalized Estimating Equation (GEE) is only valid under MCAR (missing completely at random) assumption, which is generally not reasonable for clinical trials. Therefore, the data used for this analysis, which is the observed data at Week 24 with no imputation of any missing data in BICLA components (including no imputation of any missing individual items in the multiple item scales), will be modified such that informative dropouts are set to non-responder for this analysis, which is permitted within the GEE framework. In general, all other missing values will be left missing in this model. However, as per the definition of BICLA (Section 8.1.1), subjects who prematurely discontinue the study but have had disallowed changes in concomitant medication dosing prior to discontinuing, and hence would have been a non-responder at all subsequent visits if those visits had been done, will be set to non-response for all subsequent time points even if the visit was not done. In addition, subjects who prematurely discontinued the study with primary reason for discontinuation being informative will be set to non-response for all subsequent time points as well, such that, following these imputations, the only missing values in this analysis are values which are inadvertent missing as opposed to clearly informative missing values. Informative reasons for discontinuation include lack of efficacy and adverse events other than pregnancy and/or primagravida.

The number and percent of responders in each treatment group will be presented. LS mean proportion responding in each treatment group and LS mean differences and 95% CIs in proportion responding between each DZP treatment group and PBO will be computed using the GEE model described above. Odds ratios of each DZP treatment group vs PBO along with the 95% confidence interval and p-value (versus PBO) will be presented. An unstructured covariance matrix will be used for this analysis. If the model does not converge, the Toeplitz or compound symmetry covariance matrix (Toeplitz should be tried first and then compound symmetry) will be used. If the GEE model still does not converge with any of the 3 covariance matrices, the reason may be that there are too few responders at Visit 4. In that case, the GEE model will be run including data from Visits 6-10 only, using the unstructured covariance matrix (or Toeplitz or compound symmetry in that order if needed for convergence). In the GEE analysis, missing values are not explicitly imputed, but GEE analysis uses information from the observed data to implicitly impute the unobserved data. All available scheduled post-Baseline assessments are utilized for deriving Week 24 treatment differences (except if Visit 4 has to be excluded to achieve model convergence). Additionally, the crude difference in proportion responding and standard Wald asymptotic 95% CI based on the normal approximation to the binomial distribution will be presented for each DZP treatment group vs. PBO.

8.3 Statistical analysis of the other efficacy variables

The analyses of the other efficacy variables will be performed for all visits through Week 24 in the interim analysis and for all visits through the end of study (Week 48) for the final analysis (unless otherwise noted).

The main analysis of other efficacy responder rates will be based on the same mNRI imputation as for the primary efficacy variable analyzed by a logistic regression model, modeling treatment response as a function of treatment group, controlling for corticosteroid strata, as detailed in Section 8.2 for the secondary efficacy variable. Note that BILAG Improvement and BILAG

substantial/partial/no response are both considered responder endpoints and are analyzed using mNRI imputation as specified here. (However, BILAG substantial/partial/no response will be analyzed using a Cochran-Mantel-Haenszel test due to the 3 ordered levels of response.)

Additionally, an estimate of the treatment differences will be obtained from a Longitudinal Generalized Estimating Equation (GEE) model for binary outcome, controlling for corticosteroid strata, and with a treatment by visit interaction. The details of this GEE model will be the same as that used for the sensitivity analysis of the secondary efficacy variable described in Section 8.2.1. In general, missing values will be left missing in this model, however informative dropouts will be set to non-responder for this analysis, as detailed in Section 8.2.1. All available scheduled post-Baseline assessments will be utilized for deriving treatment differences (except if Visit 4 has to be excluded to achieve model convergence).

Analyses of continuous efficacy variables will be performed using mixed effects models for repeated measurement (MMRM) with change from Baseline as dependent variable and fixed effects for treatment, visit, and treatment*visit as well as corticosteroid strata and baseline values as covariates. These models will use all available post-Baseline data at all visits up to and including Week 24, and the data from multiple visits will be incorporated as repeated measures within each subject. As detailed in Section 3.2, data from EWVs (prior to Week 24) will be recoded to the closest scheduled treatment period visit (after the last non-missing visit value) and will be included among the observed data for analysis. An unstructured covariance matrix (or Toeplitz or compound symmetry in that order if needed for convergence) will be utilized. LS means for changes from Baseline for each treatment group will be presented by visit. LS mean differences in the changes from Baseline for each DZP dose group versus PBO and corresponding 95% CIs and p-values will also be presented for each scheduled time point. The MMRM analysis of continuous efficacy variables will be performed using observed case data. This observed case data will include the limited imputation of missing individual scores from these multiple-item scales (BILAG 2004 total score, SLEDAI-2K total score, CLASI Activity Score, CLASI Damage Score, and SLICC/ACR Damage Index) from the previous visit (limited to 1 visit back), but no imputation when an entire scale is missing. In the MMRM analysis, missing values for total scale scores (i.e., when the entire scale is missing) will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data.

Supportive models of the continuous efficacy variables will involve analysis of covariance (ANCOVA) models with treatment group as a factor and corticosteroid strata and baseline scores as covariates. LS means for changes from Baseline at each visit for treatment groups will be presented. LS mean differences in the changes from Baseline for each DZP dose group versus PBO and corresponding 95% CIs and p-values will also be presented by visit.

These supportive ANCOVA analyses of continuous efficacy variables will be performed using observed case data. This observed case data will include the limited imputation of individual scores from these multiple-item scales (BILAG 2004 total score, SLEDAI-2K total score, CLASI Activity Score, CLASI Damage Score, and SLICC/ACR Damage Index) from the previous visit (limited to 1 visit back), but no imputation when an entire scale is missing. In the ANCOVA analysis, missing values for total scale scores are excluded from the analysis.

For all efficacy variables, DZP treatment arms, when statistically compared with PBO, will use a two-sided α=0.05 level of significance, without any adjustment for multiple comparisons. As

such, these results are considered exploratory in nature and caution should be exercised in interpreting these results.

8.3.1 Clinical assessment of disease activity

In addition to the tables and listing, given in for the primary and secondary efficacy variables, the **BICLA responder rates by visit** will be tabulated and listed. The analysis of the responder rates by visit will use the same approach as for the secondary efficacy variable. Data will be imputed using mNRI and the analysis of each time point will consist of a logistic regression model, modeling treatment response at the applicable study visit as a function of treatment group with corticosteroid strata included in the model as a covariate, s. Section 8.2. P-values based on Wald Chi-square will be generated for the pairwise comparisons of each DZP treatment group with PBO. As an additional analysis approach, a GEE model will be fit, as described in Section 8.2.1. Line graphs will be presented to display the responder rate over time by treatment group. A part of the assessment of the BICLA responder rate will be based on the SLEDAI-2K Total Score as described in Section 8, including the lab criteria for anti-dsDNA positive '≥10iU'. Additionally, as a sensitivity analysis, the table program will be re-run including SLEDAI-2K total scores based on anti-dsDNA positive '≥5iU' as 'positive'.

Analyses of the number of subjects with **maintained BICLA response** (defined as BICLA response at 2 consecutive study visits) will be provided. These analyses will use the primary imputation approach for BICLA, the mNRI data.

The time to maintained BICLA response will be analyzed with statistical survival methods. The time to maintained BICLA response will be defined as the time until the start of the Maintained BICLA response (i.e., time until the first visit with BICLA response that is then maintained for at least one more consecutive visit thereafter). Subjects who do not achieve a maintained BICLA response by Week 24 (interim analysis), Week 48 (final analysis) or early termination visit will be censored at the date of the last assessment during study period. Survival curves will be generated using the Kaplan-Meier product limit estimate. The treatment differences vs PBO will be analyzed using the stratified log-rank statistic, adjusting for corticosteroid strata. Median time to event values for each treatment group will be summarized and Kaplan-Meier survival curves will be presented by treatment group on a single graph.

Additionally, the number and percentage of subjects with maintained BICLA response from Week 12 to Week 24 for the interim analysis, and from Week 24 to Week 48 for the final analysis will be summarized.

The BICLA responder rate is primarily based on the BILAG 2004 disease activity measurement instrument. The **BILAG 2004** was selected as the primary tool for measurement of SLE disease activity on the basis of its comprehensiveness, ability to capture incremental changes in disease activity, and the clinical relevance of its grading system (Tsokos et al, 2007).

The Investigator will assess 97 BILAG 2004 components and record the assessments in the eCRF. Only features attributable to SLE will be recorded and based on the subject's condition in the last 4 weeks compared with the previous BILAG 2004 assessment. The components of the BILAG 2004 Index will assess nine body/organ-based systems:

- Constitutional
- Mucocutaneous

- Neuropsychiatric
- Musculoskeletal
- Cardiorespiratory
- Gastrointestinal
- Ophthalmic
- Renal
- Hematological

Disease activity is graded separately for 9 body systems to 5 different grades (A to E) as follows:

A ("Active")	Severely active disease
B ("Beware")	Moderately active disease
C ("Contentment")	Mild stable disease
D ("Discount")	Inactive now but previously active
E ("Excluded")	Never affected

A shift from BILAG 2004 Grade A or B to a lower (less severe/active) grade indicates a clinically relevant change in disease activity. It is important to note that for the BILAG 2004 improvement, all Grades A and Grades B at Baseline need to improve. If a Grade A from Baseline improves to a B but a single Grade B from Baseline remains as a B, then the subject has not improved in their BILAG 2004.

Shift tables of BILAG 2004 Grades will be presented by body system showing the number and percentage of subjects with Grade A, B, C, D or E at each post-Baseline visit by the Grade A, B, C, D or E at Baseline. The denominator for the percentages will be the number of subjects in the Full Analysis Set. The shift table will include a missing category, presenting the number of subjects with no results at Baseline or subsequent visit. Furthermore, a shift will be presented for the BILAG 2004 Grades at Baseline and at the subsequent visits where the grades are combined in the categories A/B and C/D/E rather than presenting each grade individually. Here, the number and percentage of subjects with Grade A or B at Baseline and with Grade C, D or E at each subsequent visit and conversely, the number and percentage of subjects with Grade C, D or E at Baseline and with Grade A or B at each subsequent visit will be presented. In addition the number and percentage of subjects with no change in the grade classifications by category A/B vs. C/D/E will be presented in the tabulation (subjects with Grade A or B at Baseline and at the subsequent visit; subjects with Grade C, D or E at Baseline and at the subsequent visit).

The responses to the individual BILAG body systems and components and the BILAG 2004 Grades for each body system will be listed by treatment, subject and visit, after grouping subjects by treatment.

The number and percentage of subjects with **BILAG 2004 improvement** will be summarized by treatment group and visit. BILAG 2004 improvement, analyzed on its own as an 'other' endpoint, will still be defined the same as it is when it is part of the BICLA. Specifically, BILAG 2004 improvement is defined as BILAG 2004 Grade As at study entry improved to B/C/D, and BILAG 2004 Grade Bs at study entry improved to C/D, and no BILAG 2004

worsening in other BILAG 2004 organ systems (that had BILAG 2004 Grade C, D, or E at Baseline) such that there are no new BILAG 2004 Grades A nor greater than 1 new BILAG 2004 Grade(s) B. Note that "study entry" within these criteria refers to Baseline, which is Visit 2 (or an earlier visit if the Visit 2 assessment was not done).

BILAG 2004 improvement is considered a responder rate endpoint and analyzed similarly to the secondary efficacy variable, using mNRI imputation. Missing individual items (BILAG Grades) can be imputed from the previous visit (limited to 1 visit back) and subjects with missing values due to premature study discontinuation will be imputed as non-responders (i.e., not improved) for this analysis. P-values based on Wald Chi-square will be generated for pairwise comparisons of each DZP treatment group with PBO at each visit, using a logistic regression model as described in Section 8.2. As an additional approach, a GEE model will be fit as described in Section 8.2.1.

The **BILAG 2004 total score** is defined as the sum of the numeric scores of each of the 9 individual body systems, where letter grades A/B/C/D/E are converted to numeric score equivalents. Numeric equivalents for the version of the BILAG used in this study, the BILAG-2004 Index, are A=12, B=8, C=1, D/E=0 (Yee et al, 2010). Descriptive statistics for absolute values of the BILAG 2004 total score and changes from Baseline will be tabulated by treatment and visit on the observed case values as well as the WOCF imputed values. As detailed in Section 4.3, the observed case data for BILAG 2004 total score will include the limited imputation of individual missing scores (Grades) from this multiple-item scale from the previous visit (limited to 1 visit back), but no imputation when an entire scale is missing. The WOCF imputed values will include imputation of individual missing scores (Grades) from this multiple item scale by using the worst observation of all previous visits for the individual item, and additionally will include imputation of the total score for an entire missing scale using the worst observation of all previous visits for the scale. P-values for pairwise comparisons of change from baseline in each DZP treatment group to PBO will be generated for each visit based on a mixed effects models for repeated measurement (MMRM) run on the observed case values. In this analysis, missing values for total score of the scale will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data, s. Section 8.3. Analyses at each time point measured will be repeated using an ANCOVA model (detailed in Section 8.3) with the observed case values (i.e., missing values for total score of the scale not imputed), as well as an ANCOVA model with WOCF imputed values. Line graphs will be presented to display mean observed BILAG total score and mean observed change from Baseline BILAG total score, with 95% confidence intervals over time by treatment group. Corresponding listings will be provided.

The **BILAG 2004 systems tally** is a new method of representing the BILAG-2004 index scores longitudinally, which combines the flexibility and simplicity of numerical scoring with the clinical intuitiveness and meaningfulness of the BILAG 2004 index categorical score (Yee et al, 2012). It was devised using a data-driven methodology and is based on counts of the number of systems with active/worsening disease and improving disease between two assessments. It has 2 forms: the comprehensive BILAG 2004 systems tally (BST) and the simplified BILAG 2004 systems tally (sBST).

BST comprises 6 components:

1. Number of systems with major deterioration (change of Grade B/C/D/E to A or Grade D/E to B)
2. Number of systems with minor deterioration (change of Grade C to B)
3. Number of systems with persistent significant activity (no change from Grade A or B)
4. Number of systems with major improvement (change of Grade A to C/D or Grade B to D)
5. Number of systems with minor improvement (change of Grade A to B or B to C)
6. Number of systems with persistent minimal or no activity (change of Grade C/D/E to C/D/E)

sBST is a simplification of BST with 3 components:

1. Number of systems with active/worsening disease (systems with major deterioration, minor deterioration and persistent significant activity)
2. Number of systems with improving disease (systems with major improvement and minor improvement)
3. Number of systems with persistent minimal or no activity.

The shifts described above (major deteriorations, etc.) within BST and sBST are based on comparison of BILAG grades at consecutive visits (no comparisons to Baseline, except for the first post-Baseline visit which should be compared to Baseline). Additionally, the shifts are based on 'observed' values of the BILAG only, rather than using any values imputed for missing data. This should be applied based on protocol defined visits at which BILAG is assessed (every 4 weeks).

BST and sBST results should only be considered if there is a non-missing grade for all 9 systems for both time points being compared. If a BILAG 2004 is not done at a visit, then comparisons involving that time point would not be possible. For example, if Week 12 is missing, then Week 8 shifts to Week 12 cannot be assessed, nor can shifts from Week 12 to Week16. Hence a single missing time point will result in two 'paired time points' having missing values.

Where 'two paired time points' or a 'visit pair' is missing due to a missing value for one or more of the 9 body systems at one or both of the visits in the 'visit pair', then the time interval for that 'visit pair' should not be counted in the denominator for AMsBST. That results in the total time for the AMsBST denominator being lower than total time based on subject's last visit.

BST and sBST results will be presented cumulatively for the double-blind treatment period for the FAS, and cumulatively for observational period for the FAS. Cumulative presentation in this context means presentation of the total number of 'visit pairs' across the entire relevant study period with X systems meeting a specific component (e.g., major deterioration) of the BST or sBST, over the denominator of all non-missing 'visit pairs'. sBST results will be analyzed using a GEE model for ordinal outcomes outcome with factors for treatment, corticosteroid strata, study visit, and treatment by study visit interaction. For this GEE analysis, the model will have to be provided with by-visit data (i.e., the number of 'visit pairs' at each visit with X systems meeting a specific component), although the table will display this information cumulatively across all relevant visits.

Adjusted mean sBST has been advocated as a useful summary of sBST over time. In the same way, the component of number of systems with active/worsening disease in sBST might be used

to create a comparable measure of 'adjusted mean sBST - active/worsening' (AMsBST - active/worsening). The formula for AMsBST would be:

$$\frac{\sum_{i=2}^{n} Xiti}{\sum_{i=2}^{n} ti}$$

where

Xi=sBST component of interest (active/worsening systems in this example) at visit i and

t_i = time interval between visit i and visit i-1.

This measure would provide an overall measure of burden of disease activity over period of study. Hence, it could be used to demonstrate the magnitude of treatment effect between treatment arms. Since AMsBST is calculated over the entire duration of study, there would be just one value per subject (and not one per time point) for this variable. AMsBST values will be analyzed using a nonparametric van Elteren test comparing DZP arm to placebo, adjusted for corticosteroid strata.

The **SLEDAI-2K** measures disease activity in the 30 days prior to and at the time point of the assessment, but for both Visits 1 and 2, the SLEDAI-2K will assess disease activity in the past 10 days only, to represent the disease status at study entry (Gladman et al, 2002). It is a global index and includes 24 clinical and laboratory variables marked as present or absent, that are weighted by the type of manifestation, but not by severity. The SLEDAI-2K includes scoring for antibodies (anti-dsDNA positive or negative) and low complement, as well as some renal and hematologic parameters. The total score falls between 0 and 105, with higher scores representing increased disease activity. Details of the SLEDAI-2K calculation are given in Section 8.

The results of the SLEDAI-2K will be listed by treatment, subject and visit, and will include the individual component results (Yes or No) and the overall score for each visit. Listings will be sorted by visit and then by the individual descriptor, such that results for the same visit appear together. Changes from Baseline in total score will be included in the listing.

As described in Section 8, the SLEDAI-2K total score will be calculated using the central lab data to determine the presence or absence of a few items that are directly based on labs. For these few items, both SLEDAI-2K information will be included in the listing - the (not analyzed) values from the eCRF and the (analyzed) values based on the central lab information.

Descriptive statistics for absolute values and changes from Baseline in SLEDAI-2K total score will be presented by treatment and visit on observed case values as well as the WOCF imputed values. As detailed in Section 4.3, the observed case data for SLEDAI-2K total score will include the limited imputation of individual missing scores from this multiple-item scale from the previous visit (limited to 1 visit back), but no imputation when an entire scale is missing. The WOCF imputed values will include imputation of individual missing scores from this multiple item scale by using the worst observation of all previous visits for the individual item, and additionally will include imputation of the score for an entire missing scale using the worst observation of all previous visits for that scale. The calculation of SLEDAI-2K Total Score as described in Section 8, includes the lab criteria for anti-dsDNA positive '>10iU'. Additionally, as a sensitivity analysis, the table program will be re-ran including SLEDAI-2K total scores based on anti-dsDNA positive '≥5iU' as 'positive'. Line graphs will be presented to display

mean observed SLEDAI-2K total score and mean observed change from Baseline SLEDAI-2K total score, with 95% confidence intervals over time by treatment group.

P-values for pairwise comparisons of change from Baseline in each DZP treatment group compared to PBO will be generated for each time point based on a MMRM analysis run on the observed case values. In this analysis, missing values for total score of this scale will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data, s. Section 8.3. Analyses at each time point measured will be repeated using ANCOVA with the observed case values (i.e., missing values for total score of this scale not imputed), as well as using ANCOVA with the WOCF imputed values. Line graphs will be presented to display mean observed SLEDAI-2K total score and mean observed change from Baseline in SLEDAI-2K total score, with 95% confidence intervals over time by treatment group. Corresponding listings will be provided.

The **S2K RI-50** comprises the same 24 descriptors covering 9 organ systems as the SLEDAI-2K and in contrast to SLEDAI-2K captures improvement of at least 50% in disease activity compared with the previous assessment; the S2K RI-50 reflects disease activity over the 30 days prior to the assessment (Touma et al, 2011). The S2K RI-50 score is evaluated as the sum of the scores of the 24 individual descriptors. If the S2K RI-50 was not completed for an individual item at a visit, then if the SLEDAI-2K score for that same item at the same visit = 0, then the S2K RI-50 score for that item at that visit =0. No other imputation will be made for S2K RI-50 missing data (including no imputation using previous visit data).

The determination of S2K RI-50 will be based on the SLEDAI-2K Total Score as described in Section 8 and as such, also needs to use central lab results to determine the item score for 'low complement' and 'increased DNA binding'. This will be done as follows:

- If the Farr assay ≤10 iU then the DNA binding is normal and there is no 'increased DNA binding', so the score for this item is 0. However, if the Farr assay >10 iU, then there is ‘increased DNA binding’ and it must be determined whether it is <50% or ≥50% improved vs. the previous visit as follows:
    - If the Farr assay ≤ (0.5* Farr assay value at previous visit) then there is ≥50% improvement vs. the previous visit, so the score for this item is half the total points for this item on the SLEDAI-2K scale, i.e., 1.
    - If the Farr assay > (0.5* Farr assay value at previous visit) then there is <50% improvement vs. the previous visit, so the score for this item is the total points for this item on the SLEDAI-2K scale, i.e., 2.
    -
- If both the complement C3 ≥ LLN and the complement C4 ≥ LLN, then there is no 'low complement', so the score for this item is 0. However, if either C3 or C4 is < LLN, then there is ‘low complement’ and it must be determined whether it is <50% or ≥50% improved vs. the previous visit as follows:
    - If there is ≥50% increase in the level of C3 or C4 vs. the previous visit (i.e., C3 ≥1.5* C3 value at previous visit or C4 ≥1.5* C4 value at previous visit), without any drop in the value of the other compared to the previous visit

OR

- If one of either C3 or C4 has changed from < LLN to ≥ LLN since the previous visit, without any drop in the value of the other compared to the previous visit

THEN that is considered ≥50% improvement, so the S2K RI-50 score for 'low complement' is half the total points for this item on the SLEDAI-2K scale, i.e., 1.

ELSE

- Else there is <50% improvement, so the S2K RI-50 score for 'low complement' is the total points for this item on the SLEDAI-2K scale, i.e., 2.

For these parameters, 'low complement' and 'increased DNA binding', the central lab results for the lab sample taken at the same visit will be used. If there is no lab sample take on the same day as the respective visit, a lab within 14 days after the visit may be used. If there is no lab within 14 days after the visit, the lab results from one visit back may be used. Only in case there is no lab value available at the respective visit, within 14 days after the visit, nor at the previous visit, then the data from the CRF will be used if available.

The determination of the S2K RI-50 score as described above includes the lab criteria for anti-dsDNA positive '>10iU' as 'positive'. Additionally, as a sensitivity analysis, the table program will be re-run using the lab criteria for anti-dsDNA positive '≥5iU' as 'positive'.

In the protocol, the S2K RI-50 was referred to as the SRI-50. However, since that time a new nomenclature has been adopted for S2K RI-50. To avoid confusion with the SRI-4, SRI-5, and SRI-6 response S2KRI-50 was renamed to S2K RI-50.

The results of the **S2K RI-50** will be listed by treatment, subject and visit and will include the individual component results and the overall score for each post-Baseline visit. Listings will be sorted by visit and then by the individual descriptor, such that results for the same visit appear together. Descriptive statistics will be presented for the S2K RI-50 total score and changes from previous visit in S2K RI-50 total score (including SLEDAI-2K as Baseline values, i.e., as the previous visit for Week 4) by treatment and visit.

P-values for pairwise comparisons of change from Baseline in each DZP treatment group compared to PBO will be generated for each time point based on a mixed effects models for repeated measurement (MMRM) run on the observed case values. The observed case data for S2K RI-50 will include the limited imputation of an individual missing item score as 0 if the SLEDAI-2K score for the same item at the same visit is 0, but will include no other imputation of missing item scores and no imputation when an entire scale is missing. In the MMRM analysis, missing values for the S2K RI-50 total score will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data, s. Section 8.3. Analyses at each time point measured will be repeated using ANCOVA with the observed case values (i.e., missing values the S2K RI-50 total score not imputed). Line graphs will be presented to display mean observed S2K RI-50 total score and mean observed change from Baseline in S2K RI-50 total score, with 95% confidence intervals over time by treatment group.

Physician's Global Assessment (PGA): The Investigator will assess the overall status of the subject's Systemic Lupus Erythematosus signs and symptoms and the functional capacity of the subject using a visual analog scale (VAS): Zero indicates 'very good, asymptomatic and no limitation of normal activities'; 100 indicates 'severe disease' and is at the very end of the scale. This refers to the most severe possible disease, and does not reflect the most severe ever seen in a particular patient, but the most severe disease ever seen in all SLE patients.

The results of the VAS scores will be listed by treatment, subject and visit. The Physician's global assessment of disease activity score (using 100mm VAS) and changes vs Baseline in physician's global assessment of disease will be summarized with descriptive statistics by treatment group at each time point measured. P-values for pairwise comparisons of each DZP treatment group compared to PBO will be generated for each time point based on a MMRM analysis run on the observed case values. The observed case data for PGA contains all observed data and no imputations. In this MMRM analysis, missing values for PGA score will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data, s. Section 8.3. Analyses at each time point measured will be repeated using ANCOVA with the observed case values (i.e., missing values not imputed). Line graphs will be presented to display mean observed physician's global assessment of disease and mean observed change from Baseline in physician's global assessment of disease, with 95% confidence intervals, over time by treatment group. Line graphs will be presented to display mean observed PGA score and mean observed change from Baseline in PGA score, with 95% confidence intervals over time by treatment group).

The **SRI-4, SRI-5**, and **SRI-6** define responders as (i.e., all criteria must be met):

- Reduction in SLEDAI-2K score of ≥4, ≥5, and ≥6 points, respectively
- No shift from BILAG 2004 Grade B, C, D, or E to A post-Baseline
- No more than 1 shift from BILAG 2004 Grade C, D, or E to B post-Baseline
- Increase in Physician's Global Assessment of Disease from Baseline of <10% of the total 100mm VAS (i.e., increase from Baseline of <10mm)

Assessments of SRI-4, SRI-5, and SRI-6 will include an assessment of non-responders based on the same concomitant medication rules as for the BICLA response: In case of escape treatment (s. Section 8.1.1, Derivation of primary efficacy variable), a subject is also considered as a non-responder for SRI-4, SRI-5 and SRI-6 beginning the day after the reported start date of the concomitant medication considered as escape treatment and at all subsequent visits. In case of missing data the same mNRI rules as for BICLA will be applied to SRI-4, SRI-5 and SRI-6.

The number of SRI-4, SRI-5 and SRI-6 responders will be tabulated by visit. A corresponding listing will be provided.

The analysis of the responder rates by visit will use the same approach as for the secondary efficacy variable (s. Section 8.2). At each time point, mNRI imputed data will be analyzed using a logistic regression model, modeling treatment response at the applicable study visit as a function of treatment group with corticosteroid strata included in the model as a covariate. To assess for sensitivity to missing data, as for the secondary efficacy variable (s. Section 8.2.1), the logistic regression model will also be applied to the following data sets including 3 alternate methods of imputation of missing data:

1. mNRI data set where non-missing end of study response is used for subjects who discontinued study treatment but remained in the study.
2. Observed cases data set (observed data at Week 24 with no imputation of any missing data in SRI-4/5/6 components [including no imputation of any missing individual items in the multiple item scales])
3. Strict NRI data set (all/any missing data in SRI-4/5/6 components [including any missing individual items in the scales] at Week 24 leads to non-responder status on the SRI-4/5/6 for Week 24).

As an additional approach, for SRI-4, SRI-5 and SRI-6 a GEE model will be fit, as described in Section 8.2.1 . Line graphs will be presented for SRI-4 to display the responder rate over time by treatment group.

The determination of SRI-4, SRI-5 and SRI-6 will be based on the SLEDAI-2K Total Score as described in Section 8 and includes the lab criteria for anti-dsDNA positive '>10iU'. Additionally, as a sensitivity analysis, the table program will be re-run including SLEDAI-2K total scores based on anti-dsDNA positive '≥5iU' as 'positive'.

A sensitivity analysis based on the PPS will be performed for SRI-4.

Patient's Global Assessment (PtGA): Subjects will rate their global assessment of their SLE disease activity for the day of the visit using a 100 mm visual analog scale (VAS) where 0 is "very good, no symptoms" and 100 is "very poor, very severe symptoms.", i.e., higher scores indicate more severe disease.

The analysis described in the section 'Physician's Global Assessment (PGA)' will be repeated for the Patient's Global Assessment (PtGA) of disease.

The **SLICC/ACR Damage Index** (version 1996) for SLE measures irreversible, accumulated organ damage from either the disease process or disease treatment, which has been present for at least 6 months, in 12 organ systems.

Definition of damage:

- Non-reversible change
- Occurring since onset of lupus
- Not related to active inflammation
- Present for at least 6 months
- Ascertained by clinical assessment/x-rays

The investigator (or qualified designee) will perform the SLICC/ACR Damage Index assessment for the 12 different organ systems, and check if the criterion is present ("Yes" ticked) or not ("No"), or "Unknown," as appropriate. The different organ systems and corresponding ranges of possible score are given in Table 9-1 in the protocol.

Due to the long-term nature of the changes that are measured by the SLICC/ACR Damage Index, all summaries, analyses, and listings of SLICC/ACR Damage Index will be done at the time of the Final Analysis only, and not for the Interim Analysis.

The results of the SLICC/ACR Damage Index will be listed by treatment, subject and visit and will include the individual component results (Yes/No) and the corresponding score. The SLICC/ACR Damage Index Total Score will be categorized into categories '0', '1', '2', and '>2' for summary and analysis. Frequency counts (n and percentage) for the number of subjects with a SLICC/ACR Damage Index Total Score in each category, '0', '1', '2', and '>2' will be presented. Cochran-Mantel-Haenszel analysis of the number of subjects in each score category by treatment will be performed, adjusting for corticosteroid strata. A shift table which presents the shift in the SLICC/ACR Damage Index from Baseline at each visit, including the categories '0', '1', '2', and '>2' will be provided. Observed case data will be used for these analyses, which as detailed in Section 4.3, will include limited imputation of missing individual scores from the previous visit (limited to 1 visit back) if the entire scale is not missing.

Furthermore, tables with descriptive statistics (n, mean, SD, median, min, max) for the SLICC/ACR Damage Index scores and their changes from Baseline will also be provided. P-values for pairwise comparisons of each DZP treatment group compared to PBO will be generated for each time point based on a mixed effects models for repeated measurement (MMRM) run on the observed case values. As detailed in Section 4.3, the observed case data for SLICC/ACR Damage Index scores will include the limited imputation of individual missing scores from this multiple-item scale from the previous visit (limited to 1 visit back), but no imputation when an entire scale is missing. In this MMRM analysis, missing values for SLICC/ACR Damage Index score will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data, s. Section 8.3. Analyses at each time point measured will be repeated using ANCOVA with the observed case values (i.e., missing values not imputed).

The **joint assessment** will be carried out on 28 joints. Tenderness and swelling will be graded on a 2-point scale (Grade 0, Grade 1). For details, see Table 9–2 in the protocol. The total tender joint counts (**TJC**) and swollen joint counts (**SJC**) are the sum of all individual respective tenderness and swelling grades. If the TJC/SJC is completed, but there are missing individual observations in the TJC or SJC, then the non-missing observations are assessed and weighted by dividing by the number of non-missing values and multiplying by 28. If a joint is not evaluable at Baseline, then that joint is set to missing throughout the study. If more than 50% of the tenderness or swelling grades are missing, then no imputation is done and the total TJC or SJC are set to missing.

The total scores are used to assess the percentage change from Baseline. Tables with the descriptive statistics for the absolute values, the percent changes from Baseline, and the changes from Baseline, and the corresponding listing will be provided.

P-values for pairwise comparisons of change from baseline in each DZP treatment group compared to PBO will be generated for each time point based on a mixed effects models for repeated measurement (MMRM) run on the observed case values. The observed case data for TJC and SJC includes the limited imputation involving weighting for missing individual items within an otherwise completed scale as described above, but will include no other imputation of missing item scores and no imputation when an entire TJC or SJC scale is missing. In this MMRM analysis, missing values for TJC or SJC total score will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved

data, s. Section 8.3. Analyses at each time point measured will be repeated using an ANCOVA with observed case values (i.e., missing values for TJC or SJC total score not imputed).

The **CLASI** consists of 2 scores; the first summarizes the activity of the disease and the second is a measure of the damage done by the disease, see Table 9–3 in the protocol (Albrecht et al, 2005). Activity is scored on the basis of erythema, scale/hypertrophy of skin and mucous membranes, acute hair loss, and nonscarring alopecia. Damage is scored in terms of dyspigmentation and scarring, including scarring alopecia. Subjects are asked whether dyspigmentation due to cutaneous lupus erythematosus lesions usually remain visible for more than 12 months, which is taken to be permanent. If so, the dyspigmentation score is doubled. The CLASI Activity and Damage Total scores are then calculated by simple addition based on the extent of the individual skin symptoms. The extent of involvement for each of the skin symptoms is documented according to specific anatomic areas that are scored according to the worst affected lesion within that area for each symptom.

For both the CLASI Activity Total score and the CLASI Damage Total score, tables with the descriptive statistics for the absolute values and the changes versus Baseline and the corresponding listing (including the individual responses and scores) will be provided.

Due to the long-term nature of the changes that are measured by the CLASI Damage Total Score, all summaries, analyses, and listings of the CLASI Damage Total Score will be done at the time of the Final Analysis only, and not for the Interim Analysis. However, the CLASI Activity Total Score will be summarized, analyzed and listed for both the Interim Analysis and the Final Analysis.

For the CLASI Activity Total Score, descriptive statistics for the absolute values and the changes from Baseline will also be presented for subjects with CLASI Activity Total Score > 0 at Baseline, for subjects with CLASI Activity Total Score ≥ 4 at Baseline, and for subjects with CLASI Activity Total Score ≥ 10 at Baseline.

P-values for pairwise comparisons of each DZP treatment group compared to PBO will be generated for each time point based on a mixed effects models for repeated measurement (MMRM) run on the observed case values. As detailed in Section 4.3, the observed case data for CLASI will include the limited imputation of individual missing scores from this multiple-item scale from the previous visit (limited to 1 visit back), but no imputation when an entire scale is missing. In this MMRM analysis, missing values for CLASI scores will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data, s. Section 8.3. Analyses at each time point measured will be repeated using an ANCOVA with observed case values (i.e., missing values not imputed).

In addition, for subjects with CLASI Activity Total Score ≥ 4 at Baseline and for subjects with CLASI Activity Total Score ≥ 10 at Baseline, CLASI Activity Total Score improvement (defined as ≥50% improvement in the CLASI Activity Score compared to Baseline) will be summarized by treatment group and visit. Subjects with missing values will be imputed as non-responders (i.e., not improved) for this analysis. P-values based on Wald Chi-square will be generated for pairwise comparisons of each DZP treatment group with PBO at each visit, using logistic regression model as described in Section 8.3.

The reduction in **lupus flares** and increase in time to lupus flare is considered an appropriate measure to evaluate the efficacy of potential new lupus treatments. Flare will be assessed by multiple definitions:

- flare determined by the Investigator (Clinically significant increase of SLE disease activity of the patient in 1 or more organ systems since the last visit and consideration of a change or increase in the treatment for SLE of your patient)
- flare determined by the Sponsor (A new sustained BILAG 2004 Grade A/B [present at ≥2 consecutive visits] or start of new induction therapy [either corticosteroid dose increase to ≥0.5mg/kg/day prednisone equivalent or initiation of cyclophosphamide, rituximab, iv Ig, MMF, MTX, Azathioprin, or plasma exchange]). For flare assessed by the Sponsor, "a new sustained BILAG 2004 Grade A/B [present at ≥ 2 consecutive visits]" means that a BILAG Grade A or B must be observed that was not present at the previous visit, and then must continue to be observed at the subsequent visit. This criteria could be fulfilled by a new BILAG Grade A, which would be a new BILAG Grade A at one visit, after having been a BILAG Grade B, C, D or E at the previous visit, and then a Grade A at the next consecutive visit as well. This criteria could also be fulfilled by a new BILAG Grade B, which would be a new BILAG Grade A or B at one visit, after having been a BILAG Grade C, D or E at the previous visit, and then a Grade B at the next consecutive visit as well. Or the new BILAG Grade B could be a new BILAG Grade B at one visit, after having been a BILAG Grade C, D or E at the previous visit, and then a Grade A or B at the next consecutive visit as well.
- BILAG severe flare - defined as a new BILAG 2004 Grade A since previous visit in any system due to individual items that are new or worse qualifying for the Grade A (Isenberg et al, 2011). Determination of items that are new or worse qualifying for the Grade A will be according to the supplementary information for the numerical scoring of the BILAG-2004 index (Yee et al, 2010). Programming details will be documented in a programming specification outside the SAP.
- BILAG moderate flare - defined as 2 or more new BILAG 2004 Grade Bs since previous visit in any systems due to individual items that are new or worse qualifying for the Grade B (Isenberg et al, 2011). Determination of items that are new or worse qualifying for the Grade B will be according to the supplementary information for the numerical scoring of the BILAG-2004 index (Yee et al, 2010). Programming details will be documented in a programming specification outside the SAP.
- BILAG mild flare - defined as a single new BILAG 2004 Grade B since previous visit in any system due to individual items that are new or worse and qualify for the Grade B, or 3 or more new BILAG 2004 Grade Cs since previous visit in any systems due to individual items that are new or worse and qualify for the Grade C (Isenberg et al, 2011). Determination of items that are new or worse qualifying for the Grade B or the Grade C will be according to the supplementary information for the numerical scoring of the BILAG-2004 index (Yee et al, 2010). Programming details will be documented in a programming specification outside the SAP.

- BILAG severe flare - Variant 1 (a new BILAG 2004 Grade A since the previous visit). For BILAG severe flare - Variant 1 if a new A is observed at a visit, it *does not* have to be due to new or worsening symptoms.
-
- BILAG moderate/severe flare - Variant 1 (new BILAG 2004 Grade A or 2 new BILAG 2004 Grade Bs since the previous visit). For BILAG moderate/severe flare - Variant 1 if a new A or 2 new Bs are observed at a visit, they *do not* have to be due to new or worsening symptoms.

For each definition of flare (assessed by Investigator, assessed by Sponsor, BILAG severe flare, BILAG moderate flare, BILAG mild flare, BILAG severe flare - Variant 1, and BILAG moderate/severe flare - Variant 1):

- The number and proportion of subjects with >= 1 flare up to and including the visit will be presented by visit, since Baseline to Week 24, and since Week 24 (for the Observational Period).
- Descriptive statistics for the number of flares per affected subjects (i.e., subjects with at least one flare) will be presented overall per study period (i.e., not by visit).
- Descriptive statistics, based on statistical analysis survival methods, will be provided for time to flare up to Week 24, time to flare after Week 24 (for analyses of the Observational Period), and time to flare from Baseline to Week 48 (for analyses of the DB Treatment Period and the Observational Period together; and with one exception, descriptive statistics for the time to flare as assessed by Sponsor will only be provided for time up to Week 24, as explained in Section 3.10). Start time of a flare for all definitions that require a BILAG Grade sustained for 2 consecutive visits will be the date of the first of two consecutive BILAG Grade of increased intensity.

For the Interim Analysis (at the end of the DB Treatment Period (Part 1), **only** the analyses of BILAG severe flare – Variant 1 and BILAG moderate/severe flare – Variant 1 for the DB Treatment Period will be performed. All other analyses of flare described here will be delivered at the time of the Final Analysis and included in the Final CSR, regardless of whether the endpoints are focused on the DB Treatment Period or the Observational Period.

Patients who do not experience a flare by Week 24 (for analyses of the DB Treatment Period), Week 48 (for analyses including the Observational Period) or Early Termination will be censored at the date of last assessment during the study period (i.e., the last visit date or the study discontinuation date). Survival curves will be generated using the Kaplan-Meier product limit estimate. The differences versus PBO will be analyzed using the stratified log rank statistic, adjusting for corticosteroid strata as randomized. Median time to event values for each treatment group will be summarized and Kaplan Meier survival curves will be presented by treatment group on a single graph.

Additional general rules for counting flares are as follows.

- For all flares based on BILAG Grades, if the BILAG assessment was done at a visit, but individual grades within the scale are missing, the respective individual grades will be imputed from the respective previous individual item grades (limited to 1 visit back).

- For all flares based on BILAG Grades, all visits at which flares start will be identified based on BILAG Grades within each organ system individually. Then, no more than one flare will be counted per visit for visits that have >=1 organ system with a flare starting at that visit.
- For definitions of flares that consider 2 Grade B's at the same visit to be a flare, if there is no Grade A at a particular visit, the criteria for having a B will have to be met within each of >=2 organ systems individually at that visit in order to count a flare at that visit.
- Additionally for flare as determined by the Sponsor, flares will be counted at a visit based on the following criteria. However, if the subject also has a flare at the visit based on BILAG Grades, only 1 flare can be counted per visit.
    - Subject has newly begun any systemic corticosteroids for the first time in the study or has increased dose of corticosteroids since the last visit, i.e. from the last visit day+1 up to and including the visit day being evaluated.
    - Subject has had a plasma exchange since the last visit, i.e. from the last visit day+1 up to and including the visit day being evaluated, as recorded in the medical procedures.
    - Subject has newly initiated treatment with cyclophosphamide, rituximab, iv Ig, MMF, MTX, Azathioprin since the last visit, i.e. from the last visit day+1 up to and including the visit day being evaluated. For these treatments, it will not be considered initiation of therapy if the subject had already been taking the same therapy previously at the same or higher dose, and just had a short gap of ≤ 4 weeks in that treatment.

Further details about flare assessment are described in Section 8.4.4 (additional endpoints).

8.3.2 Variables assessed corticosteroid-sparing effects

Subjects will be issued daily diaries at each visit to record concomitant corticosteroid doses taken on a daily basis at home in between visits. All reported corticosteroid doses will be converted to prednisone equivalent doses (s. Section 4.2), and all summaries and analyses of corticosteroid doses will be based on daily prednisone equivalent doses. In order to obtain the daily prednisone equivalent dose, in addition to converting corticosteroid doses as per Section 4.2, the reported frequency of dosing will be accounted for. For example, to obtain a daily dose, a dose reportedly taken 'BID' will have to be doubled, and a dose reportedly taken 'QOD' will have to be halved. If more than one corticosteroid was reportedly taken on a given day, then the daily dose must be the sum of the doses of each of those corticosteroids.

For all summaries and analyses of corticosteroids, only corticosteroids with an indication for SLE will be included. For summaries other than time-weighted AUCMB and total AUC, only oral corticosteroids taken for SLE will be included. For time-weighted AUCMB and total AUC, all corticosteroids taken for SLE will be included regardless of route of administration, although non-systemic corticosteroids are assigned a prednisone equivalent dose of 0mg, so effectively all systemic corticosteroids taken for SLE will be included for AUC and AUCMB.

Descriptive statistics for the absolute daily corticosteroid dose (prednisone-equivalent) and changes from Baseline by visit will be presented by treatment, based on observed case data.

Missing values will not be explicitly imputed and p-values for pairwise comparisons of change from baseline in each DZP treatment group to PBO will be generated for each visit based on a mixed effects models for repeated measurement (MMRM) analysis. In this MMRM analysis, missing values will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data, s. Section 8.3. Analyses at each visit will be repeated using an ANCOVA model with the observed case values (i.e., missing values not imputed). A line graph will be presented to display mean and mean change in daily corticosteroid dose (prednisone-equivalent) from Baseline across visits. All corticosteroid data will be listed by treatment, subject and visit.

Frequency tables will be provided displaying the number of subjects with daily corticosteroid dose (prednisone equivalent) in various categories, and in some cases also with responses on BICLA or SRI-4. For all of these categorical frequency tables, subjects with a missing daily corticosteroid dose at any time point due to premature study discontinuation will have this value imputed from the dose of oral corticosteroids with SLE indication at the time of discontinuation (i.e., LOCF). If a subject discontinues study drug, but stays in the study and reports a new corticosteroid dose (may be in a higher dose or a lower dose) then these values will be included in the analysis of corticosteroid data.

Frequency tables will be provided for subjects with no daily corticosteroid dose (prednisone-equivalent) vs. subjects with a daily corticosteroid dose (prednisone-equivalent) greater than 0mg, and for subjects with a daily corticosteroid dose of 7.5mg/day prednisone equivalent or less, vs. subjects with a daily corticosteroid dose of more than 7.5mg/day prednisone equivalent. Logistic regression will be used to compare the proportion of patients with no concomitant corticosteroid treatment for each treatment group vs. placebo. Logistic regression will also be used to compare the proportion of subjects with a daily corticosteroid dose (prednisone-equivalent) of 7.5mg/day or less, for each treatment group vs. placebo.

Additionally, frequency tables will be provided displaying the proportion of patients from each treatment group in each of following categories: subjects with a daily corticosteroid dose (prednisone-equivalent) of 7.5mg/day or less at Week 24 and BICLA response at Week 24, subjects with a daily corticosteroid dose (prednisone-equivalent) of 7.5mg/day or less at Week 24 and SRI-4 response at Week 24, and subjects with a daily corticosteroid dose (prednisone-equivalent) of 7.5mg/day or less at Week 12 that stays ≤7.5mg/day for rest of time through Week 24 and maintained BICLA response from Week 12 to Week 24. Furthermore, these will each be repeated with a daily corticosteroid dose (prednisone-equivalent) cut-off 10mg/day instead of 7.5mg/day. For each category described here, a logistic regression analysis will be used to compare the proportion of patients in the category for each treatment group vs. placebo.

Cumulative distribution function plots will be presented for the daily corticosteroid dose (prednisone-equivalent) at Baseline, at Week 24 (Interim Analysis) and at Week 48 (Final Analysis) as well as for the changes vs. Baseline at Week 24 and Week 48.

Average change from Baseline in corticosteroid dose (prednisone-equivalent) in mg/day (including all routes of administration) during the entire treatment period will be summarized via time-weighted area under the curve minus baseline (AUCMB), estimated using the trapezoidal rule, for the change from Baseline in corticosteroid dose (prednisone-equivalent) through Week 24/Week 48. Time-weighted AUCMB will be calculated as follows using all available data.

$$time-weighted\ AUCMB\ \approx\ \frac{\sum x\frac{\Delta x}{2}[f(x)+f(x+\Delta x)]}{xfinal-xbaseline}$$

where

x represents the time point of the assessment,

f(x) represents the change from Baseline in corticosteroid dose (prednisone-equivalent), and

Δx represents the elapsed time between time points.

Δx will be set to one day for this analysis, so that the time-weighted AUCMB will be calculated by summing over time intervals of one day in length. This is to increase the precision of the AUCMB calculation and to make the approach consistent across subjects who will have had dose changes at varying intervals of time. Missing values following premature discontinuation of study will not be imputed for this time-weighted AUCMB analysis, so it will be based on the study days up until the subject's last visit.

The descriptive statistics for the time-weighted AUCMB in corticosteroid dose (prednisone-equivalent) will be tabulated.

Additionally, total corticosteroid dose (prednisone-equivalent) in mg (including all routes of administration) during the entire treatment period will be summarized via area under the curve (AUC), estimated using the trapezoidal rule, for the corticosteroid dose (prednisone-equivalent) through Week 24/Week 48. Total AUC will be calculated as follows,

$$total\ AUC\ \approx\ \sum x\frac{\Delta x}{2}[f(x)+f(x+\Delta x)]$$

where

x represents the time point of the assessment,

f(x) represents the change from Baseline in corticosteroid dose (prednisone-equivalent), and

Δx represents the elapsed time between time points.

As with the time-weighted AUCMB, Δx will be set to one day for this analysis, so that the total AUC will be calculated by summing over time intervals of one day in length. This is to increase the precision of the AUC calculation and to make the approach consistent across subjects who will have had dose changes at varying intervals of time. However, for the total AUC analysis, missing values following premature discontinuation of study will be imputed, so it will be based on the total scheduled study days up until Week 24 (i.e. 168 days)/ Week 48 (i.e. 336 days). (The imputation is needed for total AUC since it is not a time-weighted (mg/day) value, but rather a total (mg) value for the study. Thus, in order for it to be compared across subjects, the length of time included needs to be constant for all.) The imputation of missing values will be the same as done for frequency tables of corticosteroid categories. Subjects with a missing daily corticosteroid dose at any time point due to premature study discontinuation will have this value imputed from the dose of oral corticosteroids with SLE indication at the time of discontinuation (i.e., LOCF). However, if a subject discontinues study drug, but stays in the study and reports a new corticosteroid dose (may be in a higher dose or a lower dose) then these values will be included in the analysis of corticosteroid data.

The descriptive statistics for the total AUC in corticosteroid dose (prednisone-equivalent) will be tabulated.

For time-weighted AUCMB and total AUC only, corticosteroid dose (prednisone-equivalent) will be computed using all routes of administration, not only the daily oral corticosteroids. When calculating the area under the curve, the actual date at which each new dose starts must be used (not the first visit date that occurs after the new dose has started).

In the case, when a site has recorded the stop date of one dose equal to the start date of the next dose, it incorrectly appears as if the subject was taking a double dose for 1 day. For all summaries and analyses of daily corticosteroid dose (prednisone-equivalent), including but not limited to time-weighted AUCMB and total AUC, in that special case, the old dose only will be assumed to have been taken on the stop date, and the new dose will be assumed to have begun on start date + 1.

8.3.3 Patient-reported outcome variables

The patient-reported outcome (PRO) questionnaires include LupusQoL, SLE-S, SLE-F, SLE-M, SLE-P, SLE-E, and Subject Experience Interview. The LupusQoL will be collected for all subjects at all sites participating in the study. In addition, the following exploratory PRO variables being developed by UCB will be assessed at sites in English- and Spanish-speaking countries (approximately 45% to 50% of the projected study population): SLE-S, SLE-F, SLE-M, SLE-P, and SLE-E. These newly developed instruments are being piloted in the current study and the data generated from this study will aid in refining the instruments for use in future studies. The Subject Experience Interview will also be assessed at sites in English- and Spanish-speaking countries.

The **LupusQoL** (version 2007) is a disease-specific HRQoL instrument developed with SLE patient qualitative input (McElhone et al, 2007). It consists of 8 domains: physical health (8 items), pain (3 items), planning (3 items), intimate relationships (2 items), burden to others (3 items), emotional health (6 items), body image (5 items), and fatigue (4 items). Scores are calculated independently for each domain (provided at least 50% of the items in the domain are answered) and transformed to a 0 to 100 point scale with higher scores denoting better HRQoL. If more than 50% of the items in the domain are missing, then the domain score is missing. A "Not Applicable" response is treated as unanswered/missing.

LupusQoL scores in each of the LupusQoL domains will be summarized by treatment group at each time point measured, including change from Baseline. Missing values will not explicitly be imputed and p-values for pairwise comparisons of change from Baseline in each treatment group compared to PBO will be generated for each time point measured based on a mixed effects models for repeated measurement (MMRM) run on the observed case values. The observed case data for LupusQoL contains all observed data and no imputations (including no imputations for any individual items). In this MMRM analysis, missing values for LupusQoL scores will not be explicitly imputed, but MMRM analysis uses information from the observed data to implicitly impute the unobserved data, s. Section 8.3. Analyses at each time point measured will be repeated using an ANCOVA model with observed case values (i.e., missing values not imputed). Line graphs will be presented to display mean and mean change from baseline, with 95% confidence intervals, of observed LupusQoL scores over time by treatment group.

The **SLE-S** consists of 26 symptom items. The subject is asked to score each symptom item based on the severity experienced during the past 7 days using the following scale: 1=none; 2=mild; 3=moderate; 4=severe.

The **SLE-F** consists of 55 fatigue items across 3 scales: 16 physical domain items, 18 mental domain items, and 21 fatigability domain items. The subject is asked to score each fatigue item based on how frequently they experienced the item during the past 7 days using the following scale: 1=none of the time; 2=a little of the time; 3=some of the time; 4=most of the time; 5=all of the time.

The **SLE-M** consists of 12 mobility items. The subject is asked to score each mobility item based on how frequently they experienced the item during the past 7 days using the following scale: 1=none of the time; 2=a little of the time; 3=some of the time; 4=most of the time; 5=all of the time.

The **SLE-P** consists of 3 Numeric Rating Scales and 29 items across a Pain and Physical Sensations Scale and a Pain Checklist. The subject is asked to indicate on Numeric Rating Scales the level of pain experienced in the past 7 days at its LEAST, at its WORST, ON AVERAGE, and also RIGHT NOW. The subject is then asked to score each of 12 pain and physical sensations items based on the severity experienced during the past 7 days using the following scale: 1=none; 2=very mild; 3=mild; 4=moderate; 5=severe; 6=very severe. Finally, the subject is asked to indicate the locations of the pain experienced in the past 7 days using a checklist consisting of 17 body locations.

The **SLE-E** consists of seven feelings and mood items. The subject is asked to score each item based on how frequently they experienced the item during the past 7 days using the following scale: 1=none of the time; 2=a little of the time; 3=some of the time; 4=most of the time; 5=all of the time.

The newly exploratory PROs (SLE-S to SLE-E) do not require a full analysis as they are exploratory instruments. All results will be listed. Frequency tables will be provided at item-level, providing the number and percentage of subjects responding at each level of the scale for that item, by visit. Denominators for percentages will be based on the number of subjects who completed the item (question) at the specified visit. Another standalone SAP/psychometric plan (accountable party: Modus Outcomes) will be created, which would detail out all the exploratory and correlation analyses across these new PROs.

Qualitative **subject interviews** will be conducted by a 3rd party vendor to collect the subject's experience with lupus in terms of symptoms and impact on daily activities, and the perceived changes during the course of the study. The extent to which the exploratory PRO items and scales used fully captured their experience will also be assessed. The qualitative output of these interviews will be analyzed and reported on by the team from Modus Outcomes.

8.4 Additional Endpoints and Analyses

The following additional endpoints, not included in the protocol, will be analyzed:

8.4.1 DORIS (Definitions of Remission in SLE) complete remission on treatment

SLEDAI-based remission is defined as:

- Clinical SLEDAI-2K score = 0 (SLEDAI-2K without serology: anti-dsDNA and C3/C4 items) and
- PGA ≤16mm and
- Prednisone equivalent systemic dose for SLE indication ≤5mg per day
- Anti-dsDNA is negative
- Complement C3 or C4 ≥ LLN

Note that the prednisone equivalent systemic dose criteria (of ≤5mg/day) has to be met from the (last (planned) visit day + 1) through to the current visit day. i.e., last (planned) visit day+1, and every day after last visit day + 1 up to and including current visit day.

Additionally, if any of the components of DORIS shown above are missing, then the DORIS result will be missing, unless the non-missing components determine that the subject has failed to achieve DORIS in which case the subject will count as NOT achieving DORIS. However, if the SLEDAI-2K assessment was done at a visit, but individual item scores within the scale are missing, the respective individual item scores will be imputed from the respective previous individual item scores (limited to 1 visit back).

The SLEDAI based definition of remission will be used to evaluate the endpoint

- Number of consecutive visits in remission status

Determine the maximum number of consecutive visits the subject is in remission. If a subject has e.g., 2 consecutive visits with remission, and after a break 3 more consecutive visits with remission, the maximum number of consecutive visits the subjects is in remission will be 3. The analysis of this endpoint will be performed based on DB Treatment Period for interim as well as final analysis. Additionally number of consecutive visits in remission will be analyzed based on entire study for final analysis.

8.4.2 LLDAS (Lupus Low Disease Activity State)

Australian version/definition

- No significant disease activity as per SLEDAI-2K and BILAG 2004:
  - SLEDAI-2K score ≤4
  - The following SLEDAI-2K items: urinary casts, hematuria, proteinuria, pyuria, seizure, visual disturbance, cranial nerve disorder, lupus headache, CVA, vasculitis, pleurisy, pericarditis, fever should not be documented as ‘present’.
  - No evidence for hemolysis (BILAG 2004, Item 96: No)

    Gastrointestinal BILAG 2004 Organ System: no BILAG 2004 disease activity greater than BILAG 2004 Grade D.
  - Renal/Neuropsychiatric/Cardiorespiratory BILAG 2004 Organ Systems: no BILAG 2004 disease activity greater than BILAG 2004 Grade D.
  - Pyrexia and Cutaneous Vasculitis (BILAG 2004 Item 1 and 13): ‘not present’
- No new and/or worsening disease activity

– No worsening in any BILAG system and no new activity as per SLEDAI-2K since last visit.

Definition of worsening in any BILAG 2004 Organ System: At any time, the BILAG 2004 Grade at one visit for a particular BILAG 2004 Organ System is greater in severity than at the previous visit, ie, A is worse than B, C, D, E. B is worse than C, D, E. C is worse than D, E.

Definition of new activity as per SLEDAI-2K: At any time, the SLEDAI-2K score at one visit is higher than at the previous visit.

- PGA ≤33mm
- Prednisone equivalent systemic dose for SLE indication ≤7.5mg per day
- Immunosuppressive therapy: No increase in dose in the past 12 weeks, no dose higher than allowed as per protocol. It will not be considered an increase in dose if the subject had already been taking the same therapy previously at the same or higher dose and just had a short gap of ≤ 1 week in that treatment.

Note that the prednisone equivalent systemic dose criteria (of ≤7.5mg/day) has to be met from the (last (planned) visit day + 1) through to the current visit day, i.e., last (planned) visit day+1, and every day after last visit day + 1 up to and including current visit day.

Additionally, if any of the components of LLDAS shown above are missing, then the LLDAS result will be missing, unless the non-missing components determine that the subject has failed to achieve LLDAS in which case the subject will count as NOT achieving LLDAS. However, if the BILAG-2004 and/or SLEDAI-2K assessment was done at a visit, but individual item scores/grades within the scale are missing, the respective individual item scores/grades will be imputed from the respective previous individual item scores/grades (limited to 1 visit back).

The above definition of low disease activity state would be used to evaluate these endpoints:

- Number of subjects in LLDAS status by visit
- Number of visits in LLDAS status/subject
- Number of subjects with no visits, <25% of visits in LLDAS, ≥25% - <50% of visits in LLDAS, ≥50% - <75% of visits in LLDAS, and ≥75% of visits in LLDAS.

Two Imputation Approaches for the % of Visits in LLDAS:

– Observed Cases – will use only those visits the subject attended in the denominator, so missed visits do not impact the % of visits in LLDAS.

– Not Achieved Imputation – will use all visits in the denominator even if subject dropped out and didn't attend visit, so that is the same as counting the subject as not achieving LLDAS for the missed visits.

The LLDAS analyses described above will be done for interim and final analysis based on DB Treatment Period and for the final analysis based on the entire study in addition.

8.4.3 BILAG 2004 substantial/partial/no response

- Substantial response: all of the following

- All BILAG 2004 Grade As and Bs at Baseline have improved to Grade C or D
- No escape treatment as per BICLA (s. Section 8.1.1, topic 4 of BICLA definition)
- No organ system worsened from BILAG 2004 Grade C or D/E at Baseline to Grade A or B

- Partial response: all of the following
  - At least one organ system improved from BILAG 2004 Grade A or B at Baseline to Grade C or D
  - At least one organ system did not improve from BILAG 2004 Grade A or B at Baseline to Grade C or D
  - No organ system worsened from BILAG 2004 Grade C or D/E at Baseline to Grade A or B
  - No escape treatment as per BICLA (s. Section 8.1.1, topic 4 of BICLA definition)
- Non-response
  - No improvement from BILAG 2004 Grade A or B at Baseline to Grade C or D

  or
  - An organ system has worsened from BILAG 2004 Grade C or D/E at Baseline to Grade A or B

  or
  - Escape treatment as per BICLA (s. Section 8.1.1, topic 4 of BICLA definition)

Based on the above definition of BILAG substantial response, partial response, and non-response, the number and percentage of subjects with response in each categorical response level will be summarized by treatment group and visit. This will be a by-visit display, showing the number and percentage of subjects in each of the 3 categorical response levels at that visit. BILAG substantial/ partial/ no response is considered a responder rate endpoint and therefore analysis is based on mNRI BILAG data. At each visit, a Cochran-Mantel-Haenszel test of the association between treatment and the ordered response levels, using modified ridit scores, and controlling for corticosteroid strata at Baseline will be performed for each treatment group vs. placebo.

8.4.4 Flare assessment

Variants of the investigators flare assessment as per protocol

CRF form “Investigator Assessment of SLE Flare” has 2 questions:

1. Do you consider that the SLE disease activity of the subject in one or more organ systems (involving new or worse clinical signs and symptoms and/or laboratory measurements) increased clinically significantly since the last visit?
2. If ‘yes’, do you consider a change or did you increase the treatment for SLE in your subject?

Variants:

1. Significant worsening only: Only 1st question has to be ticked yes
2. Full definition + 'measurable' worsening: Both questions have to be ticked yes and there is any one or more of these:
   - any increase in SLEDAI-2K from previous visit
   - any BILAG 2004 individual item considered as 'new' or 'worsening'
   - worsening of PGA by ≥10mm compared to previous visit score.

In determining flares as defined above, similar to the determination of other flares in this study, if the BILAG 2004 and/or SLEDAI-2K assessment was done at a visit, but individual item scores/grades within the scale are missing, the respective individual item scores/grades will be imputed from the respective previous individual item scores/grades (limited to 1 visit back). Also, no more than one flare can be counted per visit.

The above definitions of flares would be used to evaluate these endpoints:

- Time to first flare
  - from Baseline up to Week 24 (for analysis of DB Treatment Period)
  - after Week 24 (for analysis of Observational Period)
  - from Baseline up to Week 48 (for analysis of DB Treatment and Observational Period together)
- Proportion of subjects with at least 1 flare since Baseline up to and including the visit will be presented by visit
- Proportion of subjects with at least 1 flare since Week 24 up to and including the visit will be presented by visit (for the Observational Period)
- Descriptive statistics for the number of flares per affected subject (i.e., subjects with at least one flare) overall (per study period)

All analyses of these variants of the investigators flare assessment will **only** be delivered at the time of the final analysis and included in the final CSR, regardless of whether the endpoints are focused on the DB Treatment Period or the Observational Period.

Patients who do not experience a flare by Week 24 (for analyses of the DB Treatment Period), Week 48 (for analyses of the Observational Period) or Early Termination will be censored at the date of last assessment during the study period (i.e., the study discontinuation date). Survival curves will be generated using the Kaplan-Meier product limit estimate.

8.4.5 BICLA variants

Below are 3 alternate ways of defining a BICLA responder. Each alternate definition of BICLA responder should be analysed using the approach described in SAP section 8.2.1.

Modified escape treatment criterion (topic 4 of BICLA definition):

a) No increase of systemic (or addition of a new) immunosuppressant nor systemic antimalarial doses over Baseline (Visit 2) levels for a duration of > 1 week , nor initiation of any medication listed in Table 7-1, s. Section 7.8.3 of Protocol Amendment.

b) No increase in oral corticosteroid dose over Baseline (Visit 2) levels for an SLE related indication

c) No iv, im, injections of corticosteroids which are SLE related

d) No intra-articular injections of corticosteroids which are SLE related in patients with a BILAG A or B in the musculoskeletal BILAG organ system within 12 weeks before the BICLA visit being assessed. (e.g., for a Week 12 BICLA, there must not have been an injection since baseline (visit 2). For a Week 24 BICLA, there must not have been an injection since Week 12.)

e) No increase in systemic corticosteroid dose for any indication within 8 weeks before the BICLA visit

All other items are unchanged.

Modification by withdrawing items:

1. Withdraw SLEDAI-2K and PGA criterion (criterion 2 and 3) from BICLA definition. (If a subject did not attend a visit, they will be counted as a non-responder in mNRI imputations for that visit.)
2. Withdraw PGA criterion (criterion 3, keeping in criterion 2) from BICLA definition.

All other items are unchanged.

9 PHARMACOKINETICS AND PHARMACODYNAMICS

9.1 Pharmacokinetics

The following PK variables will be evaluated:

- For plasma samples: Plasma concentrations of DZP and PEG
- For urine samples: Urine amount of PEG

Individual subject concentrations of DZP and of PEG will be listed and summary statistics per visit and planned time be provided. Plasma concentration vs. time points will be displayed in figures: one figure with the geometric mean plasma concentrations time profile +/- 95% CI (one plot with all dose levels), and a spaghetti plot per dose level, for DZP and for PEG, each. All plots will be provided in linear and log scale. The lower limit of quantification (LLOQ) for DZP concentration in plasma is 0.2µg/mL and for PEG concentration in plasma 20µg/mL. For the summary statistics, if a sample is BLQ, it will be considered as half of the LLOQ. Summary statistics will only be reported if at least 2/3 of the samples are above the LLOQ for a particular time point.

Additionally, figures with the geometric mean plasma concentrations time profiles for DZP and PEG in plasma on the same plot, one plot per dose group, will be provided in linear and log scale.

Additionally, a figure with the geometric mean ± CIs for C_{trough} vs time will be provided by dose level for DZP and for PEG.

Urine collection for the determination of PEG concentrations will be performed at the visits indicated in the schedule of events. For each collection interval the amount of PEG excreted in the urine will be calculated using the following formula:

Amount excreted per interval (µg) = Concentration (µg/mL) x Urine volume (mL)

The urine concentration of PEG and the amount of PEG excreted per collection interval will be tabulated and listed. Urine amount vs. time points will be displayed in figures: one figure with the geometric mean urine amount time profile +/- 95% CI (one plot with all dose levels), and a spaghetti plot per dose level. All plots will be provided in linear and log scale. The LLOQ for PEG concentration in urine is 1 µg/mL.

Dapirolizumab pegol PK data from SL0023 may be combined with other studies in the population PK approach and the analysis will be described in a separate Data Analysis Plan and results reported separately.

If data merit, PK/PD analyses may be conducted for the clinical efficacy endpoints and PD variables of interest. All PK/PD analyses will be described in more detail in a separate Data Analysis Plan, and results will be reported in an appendix to the CSRs.

9.2 Pharmacodynamics

9.2.1 Transcriptomics

The summary and analysis of the 96-gene transcript profiling will be included in a separate report. That report will be an appendix of the final SL0023 CSR. There will be a separate analysis plan for that summary and analysis.

9.2.2 Auto-antibody Measurements

All summaries of auto-antibody measurements will be run on the Safety Set, unless otherwise specified. All auto-antibody measurements will be listed as part of the Immunology Listing.

For anti-dsDNA, the number and percent of subjects with:

- anti-dsDNA (data from central lab)
  - clear cut (>10 iU)
  - intermediate or clear cut (≥5 iU)

will be summarized at each visit. For the categories of subjects with anti-dsDNA>10 iU at Baseline, and with anti-dsDNA≥5 iU at Baseline, subjects achieving a reduction will also be tabulated, where a reduction in anti-dsDNA is defined as a ≥50% decrease from Baseline, or reverting to negative. In the case of a value at the upper LOQ, a reduction to less than 50% of the numerical LOQ will be considered to be a reduction in anti-dsDNA.

Also, a shift table for anti-dsDNA will be created providing the number and percentage of subjects in each anti-dsDNA category at Baseline that have shifted to each anti-dsDNA category post-Baseline at Week 24 and Week 48. The anti-dsDNA categories are shown below.

- anti-dsDNA classified as

- anti-dsDNA < 5 iU
- anti-dsDNA >=5 iU to <=10 iU
- anti-dsDNA >10 iU

Further for ANA titers, a shift table will be created providing the number and percentage of subjects in each ANA titer category at Baseline that have shifted to each ANA titer post-Baseline by visit at which assessments occurred. The ANA titer categories are shown below.

- ANA titer classified as
  - ANA titer < 1:80
  - ANA titer ≥ 1:80 and ≤ 1:160
  - ANA titer > 1:160.

  (For ANA titer, 1:80 < 1:160, i.e., it is the denominators that are assessed to determine whether the ANA titer is < or > a cutoff.)

Values and change from Baseline values will be summarized by visit, as part of the Immunology table. The values and change from Baseline values by visit will be presented for the relevant parameter, for subjects with:

- Anti-dsDNA > 10 iU at Baseline
- Anti-dsDNA ≥ 5 iU at Baseline

9.2.3 Immunological Variables

All summaries of immunological variables will be run on the Safety Set, unless otherwise specified.

All immunological parameters (s. Section 2.2.3.1) will be listed by subject and time point including changes from Baseline for numeric variables and flags for measurements outside the clinical reference ranges (where applicable). Values that are below the lower limit of the reference range will be flagged as 'L' (low) and values that are above the upper limit of the reference range will be flagged as 'H' (high) and listed as well.

Note that for the Interim Analysis, Sjögren's syndrome antibody A (anti-SSA), Sjögren's syndrome antibody B (anti-SSB), and high sensitivity C-reactive protein (hsCRP) will not be presented. These will only be presented in the Final Analysis.

Summary tables and figures will be presented for both absolute values and changes from Baseline.

The number of subjects and percentages will be tabulated by treatment group and visit for subjects with:

- values ≤ upper limit and analogously, with values above upper limit,
- value < lower limit,
- value above the upper limit at Baseline who switch to ≤ upper limit,
- value < lower limit at Baseline who switch to ≥lower limit.

A shift table will be presented for C3/C4 levels showing the number and percentage of subjects with low C3/C4 at Baseline changed to normalized C3/C4 post-Baseline.

Moreover, number and percentages will be provided by visit for subjects with

- low complement
  - low complement 3 (C3),
  - low complement 4 (C4),
  - both C3 and C4 are low
  - either of C3 or C4 are low
- evidence for aPLs:
  - Anti-Beta2 glyoprotein positive (IgG>ULN or IgM>ULN)
  - Anti-cardiolipin positive (IgG>ULN or IgM>ULN)
  - Lupus Anticoagulant increased (time (noted as DRVVT in the laboratory dataset)>ULN **and** ratio (noted as DRVVTRT in the laboratory dataset)>ULN)
  - Any of the 3 above.

The values and change from Baseline values by visit will be presented for the relevant parameter, for subjects with:

- C3 < LLN at Baseline
- C4 < LLN at Baseline
- Anti-Beta2 glyoprotein positive (IgG> ULN or IgM> ULN) at Baseline
- Anti-cardiolipin positive (IgG> ULN or IgM> ULN) at Baseline

1. Lupus Anticoagulant increased (time (noted as DRVVT in the laboratory dataset)>ULN **and** ratio (noted as DRVVTRT in the laboratory dataset)> ULN) at Baseline

10 IMMUNOGENICITY

10.1 Anti-Dapirolizumab antibodies

Anti-dapirolizumab antibody (anti-DZP) will be measured as specified in the schedule of assessments during the double-blind treatment period and the observational period at Day 1, Week 4, Week 8, Week12, Week 16, Week 20, Week 24, Week 28, Week 32 and Week 48.

The Anti-drug antibody (ADA) analysis and determination of anti-DZP antibodies will occur via a dual assay strategy, consisting of first ADA assay (which is sensitive towards target interference) followed by further characterization of confirmed ADA positive samples in a second ADA assay that has been optimized with regards to target tolerance characteristics, which will generate the final ADA result. For both ADA assays a tiered analysis approach will be followed.

Samples will first be evaluated in the screening assay using a false positivity rate of 5%, followed by analysis of screened positive samples in the confirmatory assay (which is a drug

depletion assay) to confirm the true positivity of the samples. Samples that are confirmed as positive will be evaluated in a titration assay to quantify the anti-DZP antibody level and will be reported as titer (= dilution factor). The titer represents the last dilution factor of the sample's titration series still scoring positive in the screening ADA assay.

The screening cut-point, the confirmatory cut-point and titration cut-point will be determined by the bioanalytical laboratory during assay validation.

Results from the first ADA assay will be reported as

- negative (either screening negative or negative in the confirmatory assay as immunodepletion negative)
- or as a final titer result.

Samples that are confirmed as positive (and will be reported with a titer result) from the first ADA assay will be further evaluated in the 2nd target tolerant ADA assay for which the screening result (screening positive or negative), the confirmatory result (confirmed positive or negative) and the titer result will be reported.

Anti-DZP status should be determined for each visit where samples were taken for ADA analysis. At each visit

- Values reported as negative (either screening negative or negative in the confirmatory assay based on the first ADA assay or screening negative or confirmed negative based on the 2nd ADA) are defined as anti-DZP-
- Values reported as titers based on the 2nd ADA assay (positive titer values) are defined as anti-DZP+
  The MRD will be outlined in the Data Transfer Agreement from the laboratory and footnoted where relevant in tables, listings and figures.

In addition, subjects will receive an overall anti-DZP subject classification based on their pre-existing antibody (pre-Ab) status at baseline and their treatment emergent ADA status, and will be classified as follows based on the ADA assay results (with anti-DZP+ samples being defined as outlined above). This analysis is done based on the results from the second ADA assay.
This classification will be done based on the ADA results obtained 1) during the double-blind treatment period for the interim analysis and 2) over the entire study period (i.e. the double-blind treatment period and the observational period for the final analysis).

1. **Pre-Ab negative – treatment emergent ADA negative** = subjects who are negative at baseline visit and are ADA negative at all sampling time points post-treatment (during in the double-blind treatment period or over the entire study period)
2. **Pre-Ab negative – treatment induced ADA positive** = subjects who are negative at baseline visit and having a positive ADA value at any time in post-treatment.
3. **Pre-Ab positive – treatment reduced ADA =** subjects who are ADA positive at baseline visit and are ADA negative at all times post-treatment
4. **Pre-Ab positive – treatment unaffected ADA=** subjects who are ADA positive at baseline visit and are positive at any sampling time point post-treatment with all titer

values of the same or less magnitude as baseline (titers post-treatment are ≤ a predefined fold difference from baseline)

5. **Pre-Ab positive – treatment boosted ADA=** subjects who are positive at baseline and are positive at any sampling time point post-treatment with at least one titer value that is increased compared to baseline (i.e. having post-treatment a titer value ≥ a predefined fold increase from the highest reported titer at baseline)

The fold increase from baseline needed to consider a titer value reported post-treatment to be above the assay variation will be noted in the relevant tables, listings and figures.

Based on the overall anti-DZP subject classification above, the following will be determined and presented:

- Total prevalence of pre-Ab: n/N % number of subjects in category 3, 4 and 5.
- Total incidence of treatment-emergent ADA positive: n/N % number of subjects in category 2 and category 5

10.2 Anti-PEG antibodies

Anti-PEG antibody (anti-PEG) will be measured for each visit as specified in the schedule of assessments during the double-blind treatment period and the observational period at Day 1, Week 4, Week 8, Week12, Week 16, Week 20, Week 24, Week 28, Week32 and Week 48.

Determination of anti-PEG antibodies will occur via a tiered analysis approach, where samples are first being evaluated in the screening assay using a false positivity rate of 5%, followed by analysis of screened positive samples in the confirmatory assay (which is a drug depletion assay) to confirm the true positivity of the samples. Samples that are confirmed as positive will be evaluated in a titration assay to quantify the anti-PEG antibody level and will be reported as titer (= dilution factor). The titer represents the last dilution factor of the sample's titration series still scoring positive in the screening ADA assay.

The screening cut-point, the confirmatory cut-point and titration cut-point will be determined by the bioanalytical laboratory during assay validation.

The screening result (screening positive or negative), the confirmatory result (confirmed positive or negative) and the titer result will be reported.

Anti-PEG status should be determined for each visit where samples were taken for ADA sample analysis. At each visit

- Values reported as negative (either screening negative or negative in the confirmatory assay) are defined as anti-PEG-
- Values reported as titers (positive titer values or reported as ≥ MRD) are defined as anti-PEG+

In addition, subjects will receive an overall anti-PEG subject classification based on their pre-existing antibody (pre-Ab) status and treatment emergent ADA status, and will be classified as

follows based on the ADA assay results (with anti-PEG+ samples being defined as outlined above).
This classification will be done based on the ADA results obtained 1) during the double-blind treatment period for the interim analysis and 2) over the entire study period (i.e. the double-blind treatment period and the observational period) for the final analysis.

1. **Pre-Ab negative – treatment emergent ADA negative** = subjects who are negative at baseline visit and are ADA negative at all sampling time points post-treatment (during in the double blind treatment period or over the entire study period)
2. **Pre-Ab negative – treatment induced ADA positive** = subjects who are negative at baseline visit and having a positive ADA value at any time in post-treatment.
3. **Pre-Ab positive – treatment reduced ADA =** subjects who are ADA positive at baseline visit and are ADA negative at all times post-treatment
4. **Pre-Ab positive – treatment unaffected ADA=** subjects who are ADA positive at baseline visit and are positive at any sampling time point post-treatment with all titer values of the same or less magnitude as baseline (titers post-treatment are ≤ a predefined fold difference from baseline).
5. **Pre-Ab positive – treatment boosted ADA=** subjects who are positive at baseline and are positive at any sampling time point post-treatment with at least one titer value that is increased compared to baseline (i.e. having post-treatment a titer value ≥ a predefined fold increase from the highest reported titer at baseline)

The fold increase from baseline needed to consider a titer value reported post-treatment to be above the assay variation will be noted in the relevant tables, listings and figures.

- Overall subject classification:
    - Total prevalence of pre-Ab: n/N % number of subjects in category 3, 4 and 5.
    - Total incidence of treatment-emergent ADA positive: n/N % number of subjects in category 2 and category 5

10.3 Analyses Done for Both Anti-DZP and Anti-PEG Antibodies

Immunogenicity will only be summarized and analyzed for the final analysis, and results will be reported in the Final CSR only.

Immunogenicity will be assessed through summary tables and figures, and listing of individual results by subject. Note that placebo subjects will not be included for analyses of anti-DZP. However, placebo subjects will be included for analyses of anti-PEG, as subjects may have anti-PEG antibodies pre-existing.

All analyses will be run on the safety population, unless specified otherwise.
All ADA (anti-DZP and anti-PEG) results will be listed.

- The incidence of the ADA+ and ADA- (for anti-DZP+ and anti-DZP- and for anti-PEG+ and anti-PEG-) will be tabulated by treatment group and visit, and incidence reported as proportion of subjects having positive ADA samples at any time point (number and percentage). Missing samples will not be included in the denominator.
- Same table as above tabulated by treatment group and visit and overall anti-DZP subject classification and overall anti-PEG subject classification (defined in sections 10.1 and 10.2).
- Boxplots of the ADA titer for subjects who are ADA+ (anti-DZP+ and anti-PEG+) will be presented by treatment group and visit, with the incidence (i.e., proportion) of subjects having ADA+ samples at each visit (number and percentage) presented in the legend along the X-axis. Visits will be presented along the X-axis. Each treatment group will be displayed on a separate page. Missing samples will not be included in the denominator.
- The number and percentage of subjects with the first occurrence of ADA positivity (both anti-DZP+ and anti-PEG+) at each visit will be provided.
- The number and percentage of subjects in the overall anti-DZP subject classification and overall anti-PEG subject classification (specified in sections 10.1 and 10.2) will be reported and tabulated per dose group and over for all Dapirolizumab treated subjects.
- The time to achieving treatment-emergent ADA positivity, separated by treatment group and overall anti-DZP subject classification / overall anti-PEG subject classification (defined in sections 10.1 and 10.2), will be analyzed based on Kaplan-Meier approach. Subjects will be considered to have an event at time where treatment emergent ADA positive is first achieved. Subjects classified as treatment-emergent ADA negative will be censored at time of last available ADA result. Note subjects presenting a treatment emergent response are subjects classified as i) preAb negative – treatment induced ADA positive or as ii) pre-Ab positive – treatment emergent ADA boosted positive.
- Spaghetti plots of ADA titer (Y-axis) against time (X-axis), separated by treatment group for all ADA positive subjects and for the subjects in each overall anti-DZP subject classification and overall anti-PEG subject classification (defined in sections 10.1 and 10.2).

For the evaluation of impact of immunogenicity on PD and efficacy, the PD and efficacy variables will be determined following interim analysis. The analyses described below will only be done if data merit, and results will be reported in the Final CSR only.

- Spaghetti plots for DZP plasma concentration, PK, and efficacy marker per overall anti-DZP subject classification and overall anti-PEG subject classification (defined in sections 10.1 and 10.2) and per dose group with identification of the different subject categories and on individual subject level. A minimum of 3 subjects in the overall anti-DZP subject classification and overall anti-PEG subject classification is required for summary figures.

- Descriptive statistics for DZP plasma concentration, PK, and efficacy marker (geometric mean +/-SD of PK, PD, efficacy read-outs time profile) per overall anti-DZP subject classification and overall anti-PEG subject classification (defined in sections 10.1 and 10.2) and per dose group. A minimum of 3 subjects in the overall anti-DZP subject classification and overall anti-PEG subject classification is required for summary tables.
- Individual subject plots that present ADA titer, DZP plasma concentration and PD/efficacy marker on the same plot over time. These will be line plots with a separate page for each subject, time on the X axis and a line plotted for each of ADA titer, DZP plasma concentration and PD/efficacy marker on overlaid Y-axes.
- Bar chart of efficacy responder, separated by dose group and overall anti-DZP subject classification and overall anti-PEG subject classification (defined in sections 10.1 and 10.2), the time points of primary efficacy evaluation as a function of ADA titer will be presented graphically. To create this bar chart, ADA titer will be classified into groups as in the box plot described below. Those groups must be determined after seeing the ADA data.
- Box plots of all DZP plasma concentrations for subjects who are ADA+ (anti-DZP+) versus overall anti-DZP subject classification (group 1; group 2) presented on a linear scale.

Immunogenicity is also evaluated as a covariate in the PK/PD modelling analysis. This analysis will be described in a separate PK/PD data analysis plan and results will be reported in an appendix to the CSR.

If data merits, immunogenicity may also be correlated with possible safety findings for reporting in the Final CSR only.

- A summary table of all immune related TEAEs by ADA Status. For this summary, subjects will be categorized by the overall anti-DZP subject classification and overall anti-PEG subject classification (defined in sections 10.1 and 10.2), and will be presented for immune related TEAEs occurring prior to becoming treatment emergent ADA positive, TEAEs occurring after becoming TE ADA positive, and TEAEs for subjects who remained TE ADA negative.

In addition, the immunogenicity results will be correlated with thromboembolic events provided as an individual listing of subjects experiencing thromboembolic events with identification of the subject category for anti-DZP antibodies and anti-PEG antibodies and the onset of ADA (both anti-DZP and anti-PEG antibodies). This will only be done if data merit and results will be reported in the Final CSR only.

11 SAFETY ANALYSES

All safety analyses will be based on the Safety Set.

11.1 Extent of exposure

11.1.1 Presentation of dosing data

Administration of investigational medicinal product (IMP) will be listed by subject. The listing will include the following information for each dose received:

- Date of infusion
- Start and stop time of infusion
- Total volume of IMP given (mL)
- Estimated fraction of dose remaining (%)
- Compliance (%)
- Planned dose (mg)
- Actual dose (mg)
- Percent of planned dose administered (%)
- Whether or not the infusion was stopped (including the reason)
- Whether the infusion was permanently discontinued (including the reason)

For all subjects the planned infusion volume will be 150 mL. The planned infusion time will be approximately 120 min.

For the total volume infused (mL), the duration of infusion (min), the actual dose (mg), the total duration of exposure (days) and the compliance (%) summary statistics will be provided.

The duration of infusion will be calculated as follows:

Duration of infusion (min) = Stop infusion – start infusion

The total duration of exposure will be calculated as follows:

Exposure (days) = (Date of last dose − Date of first dose) + 84

11.1.2 Calculation of planned and actual dose received

The total planned dose given at each visit will then be calculated to 1 decimal place as follows:

Planned dose (mg) = Body weight (kg) x Randomized Dose (mg/kg).

The actual dose (mg) received will be calculated to 1 decimal place, and is based on the infusion volume given (mL) and the actual concentration of DZP in the infusion bag (mg/mL):

Actual dose (mg) = Infusion volume given (mL) x Concentration DZP in infusion (mg/mL).

The concentration of DZP in the infusion IV bag (mg/mL) is based on the number of mL of reconstituted solution from vials of DZP that were included in the IV bag and the total volume of the IV bag. There were 100mg DZP per every 1 mL of reconstituted solution from the vials.

Concentration DZP in infusion (mg/mL) =
{ [Sum across all vials used (amount of reconstituted solution taken from the vial (mL)) x
100 mg/mL (DZP in the reconstituted solution)] / Total volume of IV bag prepared (mL) }

Compliance is based on the volume given (mL) compared to the 150mL that were supposed to be administered for each infusion, thus representing the site's compliance with the dosing regimen, and will be calculated to 1 decimal place for each visit:

Compliance (%) = (Infusion volume given (mL)/150 mL) x 100%.

Percent of planned dose administered (%) is based on the actual dose compared to the planned dose, thus representing both the volume of infusion given as well as the actual contents of the IV bag, and will be calculated to 1 decimal place for each visit:

Percent of planned dose administered (%) = [Actual dose (mg) / Planned dose (mg)] x100%

Study drug exposure and treatment compliance will also be assessed across the duration of the study and tabulated by treatment group for the following parameters:

- Percent of doses (i.e., infusions) with the correct dose administered. Correct dose is defined as any infusion where 80% ≤ (Percent of planned dose administered) ≤ 120%.
- Percent of doses (i.e., infusions) with incorrect dose administered. Incorrect dose is defined as any infusion where (Percent of planned dose administered) <80% or (Percent of planned dose administered) >120%, excluding those cases where the subject may not have received intended dose due to that infusion being permanently discontinued.
- Percent of doses (i.e., infusions) where the length of infusion was too short, excluding cases where the infusion was permanently discontinued. Too short is defined as less than 90 minutes.
- Percent of skipped doses (i.e., infusions) will be summarized. Skipped doses will be identified using missing data on CRF at regularly-scheduled dosing visits.

11.2 Adverse events

An AE is any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product that does not necessarily have a causal relationship with this treatment. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product.

All AEs occurring during the study (i.e., after signature of the informed consent document) will be recorded in the electronic case report form (eCRF). For each AE the following information will be recorded in the eCRF: AE term (verbatim term), date and time of onset, whether or not the AE was an infusion reaction, if yes, whether this infusion reaction fulfills the Sampson criteria (for definition of Sampson criteria, see Sampson et al, 2006), pattern of event, whether or not the AE was classified as a SAE, as an AE of interest or as an AE of special interest, intensity, relationship to IMP, action taken with IMP, other action taken, outcome, date and time of outcome, intensity, and whether the AE led to study drug discontinuation or to study discontinuation.

Adverse events (including serious AEs) are characterized as either pretreatment or treatment-emergent according to the following criteria:

- Pretreatment AEs are the events with onset date and time prior to the first administration of study drug (DZP or PBO).
- Treatment-emergent AEs are those with onset date and time at or after the first administration of study drug, and up to 12 weeks (84 days) after last dose. The events that emerge after the final drug administration up to 12 weeks after last dose, will therefore also be considered as treatment-emergent (e.g., in the case of premature discontinuation or during the FU period (12 weeks after last dose)).

Time to onset of AE (noted as "Time Since First Dose" on AE listings) is calculated as the time between first dose and onset of AE. Time to onset is expressed in days, hours, minutes. Time to onset of AE is not calculated for pretreatment AEs.

The time to onset for each AE (relative to the first dose) will be calculated as follows for all TEAEs:

Time to onset since first dose= (Date:time of AE onset − Date:time of first dose of IMP)

For AEs with time of onset not provided,

Time to onset since first dose will be estimated by:

(Date of AE onset - date of first dose of IMP)

and will be shown in days only (without hours). If the date of AE onset and date of first dose of IMP are the same, Time to onset since first dose will be shown as <1 day.

In addition, the time to onset since the most recent dose prior to the TEAE will be calculated:

Time to onset since most recent dose=

(Date:time of AE onset − Date:time of most recent dose of IMP prior to the TEAE start)

For AEs with time of onset not provided,

Time to onset since most recent dose will be estimated by:

(Date of AE onset - date of most recent dose of IMP prior to the TEAE start)

and will be shown in days only (without hours). If the date of AE onset and date of most recent dose of IMP prior to the TEAE start are the same, Time to onset since most recent dose will be shown as <1 day.

The duration of each AE will be calculated as follows:

AE Duration (days) = (Date:time of outcome − date:time of onset)

For AEs with time of onset and/or time of outcome not provided,

AE duration will be estimated by: (Date of AE outcome - date of AE onset)

and will be shown in days only (without hours). If the date of AE onset and date of AE outcome are the same, AE duration will be shown as <1 day.Any AEs with incomplete onset and outcome (end) dates/times will be handled according to the following rules for classification as treatment-emergent and for calculation of time to onset. Such imputations will only be performed for these classifications and calculations; in the listings all data will be shown as recorded on the eCRF.

- If only the month and year are specified and the month and year of first dose is not the same as the month and year of onset, then use the 1st of the month.
- If only the month and year are specified and the month and year of first dose is the same as the month and year of onset, then use the date of first dose.
- If only the year is specified, and the year of first dose is not the same as the year of onset, then use January 1 of the year of onset.
- If only the year is specified, and the year of first dose is the same as the year of onset, then use the date of first dose.
- If the AE onset date is completely unknown, then use the date of first dose.
- Imputations for missing end dates/times will not be performed for classification as treatment-emergent and for calculation of time to onset as this is not required.

In the listings, AE duration will be presented in hours and minutes for durations of less than 24h. For AEs with duration longer than 24h the duration will be presented in days and hours (where hour will be rounded to the nearest integer). For the calculation of duration for AEs with missing start dates, the dates will be imputed as described above.

In case of uncoded AEs, these AEs should be designated as "UNCODED" at all MedDRA levels, and such AEs will be included in summary tables and subject listings based on this classification.

The listings show the data on AEs for all subjects in the Randomized Set who had at least one AE (pretreatment and/or treatment-emergent). Subjects who had no AEs are not presented on the listings.

All AEs will be listed by subject including the information described above. In addition, all TEAEs will be flagged in the listing.

Pre-treatment AEs will only be listed.

For the AEs, 4 kinds of tables will be provided:

For the Interim Analysis,

- all TEAEs of the DB Treatment Period (all TEAEs through Week 24),

and for the Final Analysis:

- All TEAEs (all TEAEs from first dose to +84 days after last infusion)
- All AEs of Observational period (all AEs from Week 24+1 day through end of study)
- All AEs (all AEs from 1st dose to end of study).

For a subject who withdraws from the study, all TEAEs of the DB Treatment Period will include all AEs through the day of the last DB Treatment Period visit, and all AEs of Observational Period will include all AEs from 1 day after the last DB Treatment Period visit through the end of study.

The incidence of treatment-emergent adverse events (TEAEs) will be summarized by MedDRA system organ class, high level term, and preferred term. In addition the incidence of TEAEs

starting within 1 Day after infusion will be presented. Tables with incidences of classified TEAEs by maximum intensity, by relationship, and by aPL status positive/negative at any time during the study will be provided.

Furthermore, a table including the exposure adjusted incidences will be presented. In detail, the table will include the number of events, number and proportion of subjects with the event (ie, incidence proportion), the exposure-adjusted incidence rate with the associated 95% confidence interval, and the exposure adjusted event rate.

For incidence rate, the numerator will be the total number of subjects experiencing the AE, each subject counted once only even if the subject has the event multiple times. The denominator will be in 100 patient-years. That is, the total summation of individual patient-years at risk up to the first occurrence of the AE for subjects with that AE, and the total patient-years at risk for those subjects not experiencing that AE, divided by 100. Incidence rates will be presented with a 95% exact confidence interval based upon the Chi-Square distribution (Ulm, 1990).

For exposure adjusted event rate, the numerator will be the number of AEs including repeat occurrences in individual subjects. The denominator will be in 100 patient-years. That is, the total summation of individual patient-years at risk (total sum of duration of exposure for all subjects) divided by 100. No confidence interval will be computed.

Tables summarizing the incidence of TEAEs will also be provided for:

- the number of subjects with thromboembolic TEAEs (based on the "Embolic and thrombotic events" SMQ plus the additional item "antiphospholipid syndrome") (this incidence table will be presented by maximum intensity),
- the number of subjects with opportunistic infection TEAEs (based on a review of MedDRA preferred terms by the sponsor),
- the number of subjects with infusion reactions (as indicated by the investigator on the CRF)
- the number of subjects with hypersensitivity TEAEs (based on the narrow SMQ for hypersensitivity) as well as the number of subjects with hypersensitivity TEAEs starting within 1 Day after an infusion.

Incidence tables with the TEAEs leading to study discontinuation and/or permanent withdrawal of study drug, and incidence tables with the serious AEs (SAEs) will also be provided.

Furthermore, a figure with the most frequent AEs sorted by relative risk will be provided.

To further support public disclosure, the following summaries of TEAEs will be produced for the Final Analysis only for the grouping of All TEAEs (all TEAEs from first dose to +84 days after last infusion) only:

- a summary of the incidence of non-serious TEAEs occurring in >5% of subjects within any individual treatment group (incidence is compared to the threshold prior to any rounding),
- a summary of the incidence of drug-related, treatment emergent SAEs

- a summary of the incidence of drug related, TEAEs leading to withdrawal from treatment.

11.2.1 Adverse events of special interest and adverse events of interest

An AE of special interest is any AE that a regulatory authority has mandated to be reported on an expedited basis, regardless of the seriousness, expectedness, or relatedness of the AE to the administration of a UCB product/compound. There are currently no AEs of special interest identified for DZP; however, in accordance with the Sponsor's internal guidelines for all drugs in development, Potential Hy's Law is to be reported as an AE of special interest (AESI). For definition of Potential Hy's law, s. Section 11.3.3.

Any PDILI event that meets the criterion for potential Hy's Law must be reported as an AESI, but not all PDILI events meet Hy's Law. For details of PDILI events, s. Section 11.3.3.

Adverse events of interest (AEOI), regardless of seriousness, as defined for this study are

- Moderate to severe infections, including opportunistic infections and tuberculosis
- Infusion reactions (as indicated by the investigator on the CRF, including hypersensitivity and anaphylaxis). Hypersensitivity AEs will be based on the narrow SMQ for hypersensitivity.
- Thromboembolic events (including but not limited to cardiovascular events, stroke, myocardial infarction, pulmonary embolism, and deep vein thrombosis). Thromboembolic events will be based on the Embolic and Thrombotic Events SMQ plus the additional item "antiphospholipid syndrome".
- Prespecified neurological events: Severe and/or serious headache, positional headache, cranial nerve dysfunction, or signs and symptoms of meningitis (photophobia, neck stiffness)
- Malignancies

All AEOIs will be listed by subject. The incidence of the AEOIs will be summarized by MedDRA system organ class, high level term and preferred term. Subjects who meet Hy's Law Criteria will be listed.

There will be additional summaries of the incidence of TEAEs for these subsets of AEOIs: thromboembolic AEs, hypersensitivity AEs, and opportunistic infections.

11.3 Clinical laboratory evaluations

11.3.1 Routine safety laboratory evaluations

The routine clinical laboratory evaluations are specified in Table 11–1~~Table 11–1~~ .

Formatte

Table 11–1: Routine Clinical Laboratory Evaluations

Category	Panel	Parameters/ Comments
Clinical chemistry	Electrolytes	Sodium and potassium
Clinical chemistry	Minerals	Calcium and phosphate

Table 11–1: Routine Clinical Laboratory Evaluations

Category	Panel	Parameters/ Comments
Clinical chemistry	Metabolic	Glucose, lipase, and creatine phosphokinase
Clinical chemistry	Kidney Function	Urea nitrogen, and creatinine
Clinical chemistry	Proteins	Total protein and albumin
Clinical chemistry	Liver Function	ALP, ALT, AST, GGT, bilirubin and LDH
Clinical chemistry	Lipids	Total cholesterol and triglycerides
Hematology	RBC	RBC count, hemoglobin and hematocrit
Hematology	RBC Indices	Mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration and erythrocyte sedimentation rate
Hematology	Platelets	Platelet count
Hematology	WBC	WBC count
Hematology	WBC Differential	Neutrophils, basophils, eosinophils, lymphocytes, atypical lymphocytes, and monocytes
Hematology		T cells (CD3+), B cells (CD19+)
Urinalysis		pH, protein, glucose, ketone, urobilinogen, bilirubin, blood, nitrite, leukocytes
Urine sediment		RBCs, WBCs, RBC casts and WBC casts

Measurements outside the clinical reference ranges will be assessed by the Investigator. In addition, laboratory values will be graded (where possible) according to CTCAE Version 4.03.

All laboratory data will be listed by subject and time point including changes from Baseline (as defined in Section 3.3), percentage changes from Baseline for numeric variables and flags for measurements outside the normal ranges.

Values that are below the lower limit of the reference range will be flagged as ‘L’ (low) and values that are above the upper limit of the reference range will be flagged as ‘H’ (high). The flag and the CTCAE grade will be included in the listings.

Any laboratory parameters that are given as ‘<xx’ or ‘>xx’ in the database will be imputed with the absolute value of the number without the sign (eg, <2.2 will be imputed as 2.2) for the calculation of the changes from Baseline.

Summary tables will be provided for both absolute values and changes from Baseline. Shift tables for CTCAE grading category will be presented for clinical chemistry and hematology parameters. For each of clinical chemistry and hematology, there will be:

- one table for lab parameters that have a CTCAE grading based on values that are <LLN. This table will show the number and percentage of subjects in their maximum CTCAE grading category compared to Baseline. The maximum grading category will represent the maximum decrease from Baseline.
- one table for lab parameters that have a CTCAE grading based on values that are >ULN. This table will show the number and percentage of subjects in their maximum CTCAE grading category compared to Baseline. The maximum grading category will represent the maximum increase from Baseline.

The denominator for the percentages in both types of CTCAE grading shift tables will be the number of subjects in the Safety Set.

Urinalysis and urine sediment results will be listed and tabulated. Note that for RBCs and WBCs in urine sediment, a few samples were reported as "TNTC" which means "too numerous to count". For summary, these samples will be counted at the upper limit of quantification provided by the lab, 719.9/hpf.

Furthermore, figures will be provided both for the mean value by visit and for the mean changes from baseline by visit for continuous laboratory parameters (as observed).

11.3.2 Additional laboratory evaluations

The additional laboratory evaluations are specified in Table 11–2~~Table 11–2~~ .

Formatte

Table 11–2: Additional Laboratory Evaluations

Parameter	Assessment Visits
Coagulation and hemostasis tests (prothrombin time, activated partial thromboplastin time, prothrombin intl. normalized ratio, and fibrinogen (fibrinogen was optional, ordered by sites as needed))	Visit 1 (Screening), 7 (Week 12), and 10/EWV (Week 24), 12 (Week 32) predose on dosing days
Urinary albumin:creatinine ratio (mg/mmol), Urinary protein:creatinine ratio (mg/mmol)	Visit 1 (Screening), Visit 2 (Day 1), 3 (Week 2), 4 (Week 4), 6 (Week 8), 7 (Week 12), 8 (Week 16), 9 (Week 20), and 10/EWV (Week 24), 11 (Week 28), 12 (Week 32), 14 (Week 40) , 16 (Week 48) predose on dosing days

The urinary albumin:creatinine ratio (mg/mmol) and urinary protein:creatinine ratio (mg/mmol) will be calculated by the laboratory from the corresponding urinary albumin, protein and creatinine measurements obtained at Visit 1 (Screening) to 16/EWV (Week 28), except Visit 5 (Week 6). A summary table showing descriptive statistics for the change from Baseline values by visit for urinary protein:creatinine ratio and urinary albumin:creatinine ratio will be provided for the subset of subjects with a urinary protein:creatinine ratio >56 mg/mmol at Baseline.

HIV test at Screening, hepatitis test at Screening and Tuberculosis test are done as part of the inclusion/exclusion criteria. Therefore they are only listed, but no analysis on this will be done.

The other parameters defined above will be listed by subject and time point including changes from Baseline for numeric variables and flags for measurements outside the clinical reference

ranges (where applicable). Values that are below the lower limit of the reference range will be flagged as 'L' (low) and values that are above the upper limit of the reference range will be flagged as 'H' (high) and listed as well. Summary tables will be presented for both absolute values and changes from Baseline.

11.3.3 Potential drug-induced liver injury

Evaluation of potential drug-induced liver injury (PDILI) consists of the diagnostic testing, continued monitoring and consult with a local hepatologist. The following testing is requested for PDILI:

- If ALT and/or AST ≥3xULN and no symptoms of hepatitis or hypersensitivity then the subject must be discussed with the Study Physician within 24h.
- If ALT and/or AST ≥3xULN and symptoms of hepatitis or hypersensitivity present then the local hepatologist should be consulted by the physician for assessment and management of potential hepatic disease and the subject must be discussed with the Study Physician.
- If ALT and/or AST ≥3xULN and total bilirubin of ≥2xULN then the local hepatologist should be consulted by the physician for assessment and management of potential hepatic disease and the subject must be discussed with the Study Physician.

Hepatitis symptoms include fatigue, nausea, vomiting, right upper quadrant pain or tenderness. Hypersensitivity symptoms include eosinophilia (>5%), rash and fever without clear alternative cause. The subjects who meet the PDILI criteria will be listed.

A listing of subjects who possible meet the Hy's Law criteria, defined as ≥3xULN ALT and/or AST with coexisting ≥2xULN bilirubin in the absence of ≥2xULN ALP, with no alternative explanation for the biochemical abnormality, will be provided.

11.4 Vital signs, physical findings, and other observations related to safety

11.4.1 Vital signs

The following vital signs measurements will be assessed:

- Pulse rate (PR) (bpm)
- Systolic blood pressure (SBP) (mmHg)
- Diastolic blood pressure (DBP) (mmHg)
- Body temperature (ºC)
- Body weight (kg)
- Body height (cm)

Vital signs measurements (absolute values and changes from Baseline) will be summarized and listed by visit and timing relative to dosing including changes from Baseline, as well as changes from pre-infusion to post-infusion time points at infusion visits. The listing will also include details to abnormal values. Vital signs with values below lower limit of normal and a decrease greater than the limit given in the following table or vital signs with values above upper limit of normal and an increase greater than the given value in the table will be classified as abnormal.

A summary of the number and percentage of subjects with a vital sign abnormality at any post-Baseline time point will also be provided by vital sign measurement.

Table 11–3: Reference ranges for vital signs

		Absolute Values		Changes from Baseline	
Parameter	**Unit**	**Lower Limit**	**Upper Limit**	**Decrease**	**Increase**
Pulse rate	bpm	45	90	-15	15
Systolic blood pressure	mmHg	90	150	-15	15
Diastolic blood pressure	mmHg	40	90	-15	15
Body temperature	ºC	35.9	37.6	-	-

11.4.2 Electrocardiogram

Abnormal ECGs will be listed by subject, visit, and timing relative to dosing. Only the Investigator/Cardiologist evaluation will be presented in the listing together with the details of any abnormalities. Only ECGs, evaluated as 'abnormal, not clinically significant' or 'abnormal, clinically significant' will be included.

A summary of the number and percentage of subjects with an abnormal ECG finding observed at any post-Baseline time point will also be provided by each reported ECG abnormal finding.

According to CRF quantitative ECG parameters will not be evaluated.

11.4.3 Physical examination

Abnormal results of the physical examination together with details of abnormalities: abnormality SLE related or not, abnormality clinically significant or not, will be listed by subject and visit.

11.4.4 Other safety variables

11.4.4.1 Assessment of suicidal ideation and behavior

Suicidal ideation and behavior will be assessed using the C-SSRS (Columbia Suicide Severity Rating Scale). This scale will be used to assess suicidal ideation and behavior that may occur at Screening, Visit 10 (Week 24) and Visit 12 (Week 32). At Screening, subjects will be asked to assess suicidal ideation and behavior for their lifetime and also for the past 6 months prior to Screening. At Visit 10 and Visit 12, subjects will be asked to assess suicidal ideation and behavior since the last study visit.

The C-SSRS includes 10 categories with binary responses (yes/no):

Category 1: [redacted]

Category 2: [redacted]

Category 3: [redacted]

Category 4: [redacted]

Category 5: [redacted]

Category 6: [redacted]

Category 7: [redacted]

Category 8: [redacted]

Category 9: [redacted]

Category 10: Completed Suicide

Endpoints based on the above categories are defined as

Suicidal **ideation**: A "yes" answer at any time during treatment to any one of the 5 suicidal ideation questions (Categories 1-5) on the C-SSRS.

Suicidal **behavior**: A "yes" answer at any time during treatment to any one of the 5 suicidal behavior questions (Categories 6-10) on the C-SSRS.

Suicidal **ideation or behavior**: A "yes" answer at any time during treatment to any one of the 10 suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

A listing will be provided including all data for only those subjects who have a 'Yes' response in any category (for any timeframe) at any visit. The listing will indicate each visit at which the scale was administered and the timeframe upon which each C-SSRS category was evaluated.

12 REFERENCES

Albrecht J, Taylor L, Berlin JA, Dulay S, Ang G, Fakharzadeh S, et al. The CLASI (Cutaneous LE Disease Area and Severity Index): an outcome instrument for cutaneous lups erythematosus. J Invest Dermatol. 2005 November;125(5):889-894.

Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics. 2005;61:738-748.

Gladman DD, Urowitz MB, Rahman P, Ibañez D, Tam LS. Accrual of organ damage over time in patients with systemic lupus erythematosus. J Rheumatol. 2003;30:1955-1959.

Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol. 2002;29:288-291.

Isenberg DA, Allen E, Farewell V, D'Cruz D, Alarcon GS, Aranow C, et al. An assessment of disease flare in patients with systemic lupus erythematosus: a comparison of BILAG 2004 and the flare version of SELENA. Ann Rheum Dis. 2011;70:54-59.

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

Pinheiro J, Bornkamp B, Glimm E, Bretz F. Model-based dose finding under model uncertainty using general parametric models. Statist. Med. 2014;33:1646–1661.

Sampson HA, Muñoz-Furlong A, Campbell RL, Adkinson NF Jr, Bock SA, Branum A, et al. Second symposium on the definition and management of anaphylaxis: summary report—Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. J Allergy Clin Immunol. 2006;117(2):391-397.

Touma Z, Gladman DD, Ibañez D, Taghavi-Zadeh S, Urowitz MB. Systemic Lupus Erthematosus Disease Activity Index 200 Responder Index-50 enhances the ability of SLE Responder Index to identify responders in clinical trials. J Rheumatol. 2011;38:2395-2399.

Tsokos GC, Gordon C, Smolen JS, editors. Systemic Lupus Erythematosus: A companion to Rheumatology. Pennsylvania: Mosby Elsevier; 2007.

Yee C-S, Cresswell L, Farewell V, Rahman A, Teh L-S, Griffiths B, et al. Numerical scoring for the BILAG-2004 index. Rheumatology 2010;49:1665-1669.

Yee C-S, Gordon C, Isenberg DA, Griffiths B, Teh L-S, Bruce I, et al. The BILAG-2004 systems tally – a novel way of representing the BILAG-2004 index scores longitudinally. Rheumatology. 2012;51:2099-2105.

13 APPENDICES

13.1 Schedule of study assessments

A schedule of study assessments is presented in Table 13–1.

Gray-shaded visits in Table 13–1~~Table 13–1~~ were not covered by the Interim Analysis.

Formatte

Table 13–1: Schedule of study assessments

Period	Screening	Double-Blind Treatment Period (Part 1)									Observational Period (Part 2)					
Visit	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10/ EWV**	**V11/ SFU [b]**	**V12/ SFU [b]**	**V13**	**V14**	**V15**	**V16/ EWV**
Week [a]	**W-4 to -1**	**1[a]**	**W2**	**W4**	**W6**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28[b]**	**W32[b]**	**W36**	**W40**	**W44**	**W48**
Day [a] (±3 days for Visits 3 to 16)	**D-28 to -1**	**D1[a]**	**D14**	**D28**	**D42**	**D56**	**D84**	**D11**	**D140**	**D168**	**D196[b]**	**D224[b]**	**D252**	**D280**	**D308**	**D336**
Written informed consent	X															
Demographic information	X															
Lifestyle	X															
History: general medical/procedures	X															
TB Signs & Symptoms questionnaire	X	X[c]					X[c]			X						
Chest x-ray [d]	X															
Verification of eligibility	X	X[c,e]														
C-SSRS	X									X		X[f]				X
Physical exam/interim medical history	X	X[c]		X[c]		X[c]	X[c]	X[c]	X[c]	X		X		X		X
12-lead ECG	X			X[c]						X						X
Randomization		X[c]														
Recording of concomitant medications	X	X[c]	X	X[c]	X	X[c]	X[c]	X[c]	X[c]	X	X	X	X	X	X	X
Recording of medical procedures		X[c]	X	X[c]	X	X[c]	X[c]	X[c]	X[c]	X	X	X	X	X	X	X
Adverse events	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Withdrawal criteria check		X[c]	X	X[c]	X	X[c]	X[c]	X[c]	X[c]	X	X[g]	X[g]	X	X	X	
Overnight fast (prior to visit) [h]		X[h]			X[h]		X[h]			X[h]						X[h]
Contact IVRS/IWRS	X	X	X	X	X	X	X	X	X	X						
Study drug administration		X		X		X	X	X	X							
Dispense daily diary	X	X	X	X	X	X	X	X	X	X	X	X[g]	X	X	X	
Collect/review daily diary/corticosteroid taper		X[c]	X	X[c]	X	X[c]	X[c]	X[c]	X[c]	X	X	X	X	X	X	X
Dispense urine collection container [i]	X							X	X	X	X	X[g]	X		X	

Period	Screening	Double-Blind Treatment Period (Part 1)									Observational Period (Part 2)					
Visit	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10/ EWV	V11/ SFU [b]	V12/ SFU [b]	V13	V14	V15	V16/ EWV
Week [a]	W-4 to -1	1[a]	W2	W4	W6	W8	W12	W16	W20	W24	W28[b]	W32[b]	W36	W40	W44	W48
Day [a] (±3 days for Visits 3 to 16)	D-28 to -1	D1[a]	D14	D28	D42	D56	D84	D112	D140	D168	D196[b]	D224[b]	D252	D280	D308	D336
Vital signs [j]:																
Unspecified time during visit	X		X		X					X	X	X	X	X	X	X
Predose; every 15min post start of infusion; every 30min until 2h post end of infusion		X		X		X	X	X	X							
Collection of blood samples for:																
TB test [k]	X															
HIV and hepatitis screening	X															
Serum pregnancy testing (β-hCG) (subjects of childbearing	X															
Hematology [l]	X	X[c]	X	X[c]		X[c]	X[c]	X[c]	X[c]	X	X	X		X		X
Clinical chemistry [l]	X	X[c]	X	X[c]		X[c]	X[c]	X[c]	X[c]	X	X	X		X		X
Additional laboratory assessments:																
Anti-dsDNA antibodies	X	X[c]		X[c]		X[c]	X[c]	X[c]	X[c]	X		X		X		X
ANAs, anti-ENAs (anti-SM, anti-SSA, anti-SSB, anti-RNP), and RF	X									X						X
aPL antibodies (anticardiolipin antibodies, lupus anticoagulant, and anti-β2 glycoprotein-1)	X									X						X
Anti-DZP, anti-PEG antibodies		X[c]		X[c]		X[c]	X[c]	X[c]	X[c]	X	X	X				X
Coagulation and hemostasis tests [m]	X						X[c]			X		X				X
Total Ig, IgG, IgM, IgA, and IgE		X[c]	X	X[c]		X[c]	X[c]	X[c]	X[c]	X		X				X

Period	Screening	Double-Blind Treatment Period (Part 1)									Observational Period (Part 2)					
Visit	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10/ EWV**	**V11/ SFU [b]**	**V12/ SFU [b]**	**V13**	**V14**	**V15**	**V16/ EWV**
Week [a]	**W-4 to -1**	**1[a]**	**W2**	**W4**	**W6**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28[b]**	**W32[b]**	**W36**	**W40**	**W44**	**W48**
Day [a] (±3 days for Visits 3 to 16)	**D-28 to -1**	**D1[a]**	**D14**	**D28**	**D42**	**D56**	**D84**	**D112**	**D140**	**D168**	**D196[b]**	**D224[b]**	**D252**	**D280**	**D308**	**D336**
Serum complement (C3, C4)	X	X[c]	X	X[c]		X[c]	X[c]	X[c]	X[c]	X		X		X		X
HsCRP	X	X[c]	X	X		X	X[c]	X	X	X	X	X		X		X
Whole blood mRNA		X[c]	X	X[c]			X[c]	X[c]		X		X				X
Proteomic signature profile	X		X	X[c]		X[c]	X[c]	X[c]		X		X				X
Cardiovascular proteins, lipids, lipid particles [h]		X[c,h]			X[h]		X[c,h]			X[h]						X[h]
DNA (PGx analysis)		X[c]								X						X
Plasma PK sampling [n]:																
Unspecified time during visit			X		X					X	X	X	X	X		X
Predose; end of infusion; 30min, 1h, 2h post end of infusion		X		X					X							
Predose; end of infusion						X	X	X								
Collection of urine for:																
Urine pregnancy testing (β-hCG) (subjects of childbearing potential)		X[c]	X	X[c]	X	X[c]	X[c]	X[c]	X[c]	X	X	X	X	X	X	X
Urinalysis, chemistry, microscopy [l]	X	X[c]	X	X[c]		X[c]	X[c]	X[c]	X[c]	X	X	X		X		X
Urine PK sampling (collected at home) [i]		X							X	X	X	X	X	X		X
Clinical assessments of disease activity:																
BILAG 2004	X	X[c]		X[c]		X[c]	X[c]	X[c]	X[c]	X		X		X		X
SLEDAI-2K	X	X[c]		X[c]		X[c]	X[c]	X[c]	X[c]	X		X		X		X
S2K RI-50				X[c]		X[c]	X[c]	X[c]	X[c]	X		X		X		X
PGA		X[c]		X[c]		X[c]	X[c]	X[c]	X[c]	X		X		X		X

Period	Screening	Double-Blind Treatment Period (Part 1)									Observational Period (Part 2)					
Visit	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10/ EWV**	**V11/ SFU [b]**	**V12/ SFU [b]**	**V13**	**V14**	**V15**	**V16/ EWV**
Week [a]	**W-4 to -1**	**1[a]**	**W2**	**W4**	**W6**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28[b]**	**W32[b]**	**W36**	**W40**	**W44**	**W48**
Day [a] (±3 days for Visits 3 to 16)	**D-28 to -1**	**D1[a]**	**D14**	**D28**	**D42**	**D56**	**D84**	**D112**	**D140**	**D168**	**D196[b]**	**D224[b]**	**D252**	**D280**	**D308**	**D336**
PtGA		X [c]					X [c]			X						X
SLICC/ACR Damage Index	X									X						X
TJC, SJC	X	X [c]		X [c]		X [c]	X [c]	X [c]	X [c]	X		X		X		X
CLASI	X	X [c]		X [c]		X [c]	X [c]	X [c]	X [c]	X		X		X		X
Investigator Assessment of Flare			X	X [c]	X	X [c]	X [c]	X [c]	X [c]	X	X	X	X	X	X	X
Patient-reported outcomes:																
LupusQoL		X [c]		X [c]		X [c]	X [c]	X [c]	X [c]	X		X		X		X
SLE-S [o]		X [c,o]		X [c,o]		X [c,o]	X [c,o]	X [c,o]	X [c,o]	X [o]		X [o]		X [o]		X [o]
SLE-F [o]		X [c,o]		X [c,o]		X [c,o]	X [c,o]	X [c]	X [c,o]	X [o]		X [o]		X [o]		X [o]
SLE-M [o]		X [c,o]		X [c,o]		X [c,o]	X [c,o]	X [c]	X [c,o]	X [o]		X [o]		X [o]		X [o]
SLE-P [o]		X [c,o]		X [c,o]		X [c,o]	X [c,o]	X [c]	X [c,o]	X [o]		X [o]		X [o]		X [o]
SLE-E [o]		X [c,o]		X [c,o]		X [c,o]	X [c,o]	X [c]	X [c,o]	X [o]		X [o]		X [o]		X [o]
Subject Experience Interview [o]				X [c,o]						X [o]						

ANA=antinuclear antibody; anti-dsDNA=anti-double-stranded deoxyribonucleic acid; anti-ENA=extractable nuclear antigen antibody; anti-RNP=anti-ribonucleoprotein antibody; anti-SSA=Sjögren's syndrome antibody A; anti-SSB=Sjögren's syndrome antibody B; anti-SM=Smith antibody; aPL=antiphospholipid; β-hCG=β-human chorionic gonadotropin; BILAG 2004=British Isles Lupus Assessment Group Disease Activity Index 2004; C3=complement 3; C4=complement 4; CLASI=Cutaneous Lupus Erythematosus Disease Area and Severity Index; C-SSRS=Columbia Suicide Severity Rating Scale; D=Day; DZP=dapirolizumab pegol; ECG=electrocardiogram; EWV=EarlyWithdrawal Visit; HIV=human immunodeficiency virus; hsCRP=high sensitivity C-reactive protein; Ig=immunoglobulin; IGRA=interferon-γ release assay; IVRS/IWRS=interactive voice/web response system; LupusQoL=Lupus Quality of Life questionnaire; mRNA=messenger ribonucleic acid; PEG=polyethylene glycol; PGA=Physician's Global Assessment of Disease; PGx=pharmacogenomics; PK=pharmacokinetic; PtGA=Patient's Global Assessment of Disease; QFT-GIT=Quanti-FERON®- TB GOLD in-Tube test; RF=rheumatoid factor; SFU=Safety Follow-up; SJC=swollen joint count; SLEDAI-2K=Systemic Lupus Erythematosus Disease Activity Index 2000; SLE-E=Systemic Lupus Erythematosus-Emotional States Instrument; SLE-F=Systemic Lupus Erythematosus-Fatigue Instrument; SLE-M=Systemic Lupus Erythematosus- Mobility Instrument; SLE-P=Systemic Lupus Erythematosus-Pain Instrument; SLE-S=Systemic Lupus Erythematosus-Symptom Inventory Instrument; SLICC/ACR=Systemic Lupus International Collaborating

Clinics/American College of Rheumatology Damage Index; S2K RI-50=Systemic Lupus Erythematosus Disease Activity Index-2K Responder Index-50; TB=tuberculosis; TJC=tender joint count; V=Visit; W=Week

[a] Day 1 is the first day of Week 1; for all other weeks, the day shown is the final day of the week indicated.
[b] Subjects who withdraw prior to Week 32 (Visit 12) will complete an EWV (Visit 10 assessments if withdrawn prior to Visit 10; Visit 16 assessments if withdrawn after
Visit 10) as well as Visits 11 and 12 for SFU assessments at 8 and 12 weeks after their final dose of study drug (as applicable). Subjects who withdraw after Week 32 (Visit 12) do not require additional SFU visits and will complete the Visit 16/EWV assessments only.
[c] To be performed before study drug dosing is started.
[d] A chest x-ray should be performed if one is not available within 3 months prior to Visit 1 (Screening).
[e] Subject must have a SLEDAI-2K score without any laboratory values of ≥4 at Visit 2 in order to be eligible to be randomized.
[f] Performed only in subjects who have withdrawn from the study prior to Week 32 (Visit 12) and are undergoing SFU assessments.
[g] Performed only in subjects who are continuing into Part 2.
[h] Subjects should undergo an overnight fast (at least 8 hours) prior to coming to the site for Visits 2, 5, 7, 10, and 16 for the collection of fasting blood samples for cardiovascular proteins, lipids, and lipid particles. Subjects may then eat breakfast after serum samples have been obtained for these assessments and before study drug administration (if applicable).
[i] For urine PK assessments, a urine collection container will be dispensed to the subject at the designated visits for urine collection at home prior to their next visit. Prior to
Visits 9 and 10 only, the subject will collect overnight urine samples, which includes the first void after the subject goes to sleep the night before the visit through the first morning void the morning of the visit. The subject should note the times for the first and last urine collection. For all other noted visits (Visits 2, 11, 12, 13, 14, and 16), the subject will only collect their first morning void.
[j] Height will only be measured at Screening. With the exception of weight, which should be measured predose, all other vital signs (pulse rate, blood pressure, and body temperature) should be measured at the times indicated. Weight should only be measured at Screening and on dosing visits. Vitals signs should be measured within ±5min of the scheduled time point for the every 15min assessments and within ±10min of the scheduled time point for the every 30min assessments. There are no time window
requirements for vital signs assessments when measured predose or at unspecified times during the visit.
[k] Tuberculosis testing should be done using IGRA. The QFT-GIT is the preferred IGRA test (see protocol, Section 13.6). It is recommended that the QFT-GIT be the first test performed at Screening to reduce the number of screening procedures conducted for any QFT-GIT-positive subjects that may need to be withdrawn from the study.
[l] Hematology, clinical chemistry, and urinalysis parameters to be assessed are defined in protocol, Section 13.9.
[m] As defined in the Laboratory Manual.
[n] Blood samples for PK measurements should be drawn within ±10min of the scheduled time point for each post-infusion time point or within 10min after the infusions ends for end of infusion time points (i.e., not prior to the end of the infusion; further details are provided in the Laboratory Manual). There are no time window requirements for blood samples for PK measurements when measured predose or at unspecified times during the visit.
[o] The exploratory SLE-S, SLE-F, SLE-M, SLE-P, and SLE-E Instruments and Subject Interview will be assessed at sites in English- and Spanish-speaking countries only.

13.2 SLEDAI-2K and SRI-50

Table 13–2: SLEDAI-2K and SRI-50

Table 2. SLEDAI-2K Responder Index 50 (SRI-50)© — Definitions. Descriptors are present at the time of the visit or in the preceding 30 days and attributed to lupus.

DESCRIPTOR	SLEDAI-2K DEFINTION	DEFINITION OF SRI-50 IMPROVEMENT
Seizure	Recent onset. Exclude metabolic, infectious or drug causes.	≥50% reduction in frequency of baseline seizure days/month.
Psychosis*	Altered ability to function in normal activity due to severe disturbance in the perception of reality. Include hallucinations, incoherence, marked loose associations, impoverished thought content, marked illogical thinking, bizarre, disorganized, or catatonic behavior. Exclude uremia and drug causes.	≥50% improvement of the psychotic manifestations judged by physician.
Organic brain syndrome*	Altered mental function with impaired orientation, memory, or other intellectual function, with rapid onset and fluctuating clinical features. Include clouding of consciousness with reduced capacity to focus, and inability to sustain attention to environment, plus at least 2 of the following: perceptual disturbance, incoherent speech, insomnia or daytime drowsiness, or increased or decreased psychomotor activity. Exclude metabolic, infectious or drug causes.	≥50% improvement of the organic brain manifestations judged by physician.
Visual disturbance	Retinal changes of SLE. Include cytoid bodies, retinal hemorrhages, serous exudate or hemorrhages in the choroid, or optic neuritis. Exclude hypertension, infection, or drug causes.	≥50% improvement of the retinal exam assessed by physician.
Cranial Nerve disorder§	New onset of sensory or motor neuropathy involving cranial nerves.	≥50% recovery of motor or sensory function in affected nerve within 1 month from the event on the basis of decrease in lupus disease activity or ≥50% decrease of the severity of pain within 1 month from the event on the basis of decrease in lupus disease activity as determined by patient on numerical scale of 1-10 if applicable with no worsening in either.
Lupus headache#	Severe, persistent headache; may be migrainous, but must be nonresponsive to narcotic analgesia.	≥50% decrease of the severity of pain as determined by patient on numerical scale of 1-10.
CVA§	New onset of cerebrovascular accident(s). Exclude arteriosclerosis.	≥50% recovery of motor or sensory function related to CVA within 1 month from the event on the basis of decrease in lupus disease activity as determined by physician without worsening in either.
Vasculitis†	Ulceration, gangrene, tender finger nodules, periungual infarction, splinter hemorrhages, or biopsy or angiogram proof of vasculitis.	≥50% improvement of the vasculitis lesions present with no new lesion or worsening in either. A ≥50% improvement for ulceration or gangrene is defined as ≥50% decrease in the body surface area; for periungual infarction, splinter hemorrhages or tender finger nodules a ≥50% improvement is defined as ≥50% decrease in the total number of involved digits with periungual infarction, splinter hemorrhages and tender finger nodules. Multiple lesions in a single digit, count only one.

Numerical scale: 1is mild and 10 is most severe

To determine body surface area use Rule of Nines for skin scoring: Head 9%, chest 9%, abdomen 9%, back 18%, legs 36%, arms/hands 18% and mucous membrane 1%; physician's palm for 1%.

* Overlap of symptoms will count for only one descriptor either Psychosis or Organic Brain Syndrome.

Lupus headache improvement will count regardless of whether patient is using narcotic analgesia or not though it has to be part of the baseline lupus headache.

§ CVA and Cranial Nerve improvement will count if it occurs within 1 month from the event on the basis of decrease in lupus disease activity as this more likely on the basis of decrease disease activity.

† Vasculitis, Rash and Alopecia; if the total BSA ≤1%, a ≥50% improvement is defined by ≥50% decrease in the activity of the most active lesion by decreasing by 2 grades or ≥50% decrease in the number of lesions or decrease in the size of the biggest lesion with no worsening in either.

Table 2 Continued. SLEDAI-2K Responder Index 50 (SRI-50)© — Definitions.

DESCRIPTOR	SLEDAI-2K DEFINITION	DEFINITION OF SRI-50 IMPROVEMENT
Arthritis	≥2 joints with pain and signs of inflammation (i.e. tenderness, swelling or effusion).	≥50% reduction in the number of joints with pain and signs of inflammation (i.e. tenderness, swelling or effusion).
Myositis	Proximal muscle aching/weakness, associated with elevated creatinine phosphokinase/aldolase or electromyogram changes or a biopsy showing myositis.	≥50% increase in muscles power judged by physician or increase by or 1 grade upon a scale of zero to five or ≥50% decrease in the level of creatinine phosphokinase/aldolase level comparing to previous visit with no worsening in either.
Urinary casts	Heme-granular or red blood cell casts.	Decrease by ≥50% in the total number of casts (heme-granular and red blood cell casts).
Hematuria	>5 red blood cells/high power field. Exclude stone, infection, or other cause.	Decrease by ≥50% in the number of red blood cell /high power field at this visit.
Proteinuria	New onset, recurrent, or persistent proteinuria of more than > 0.5 gram/24 hours.	Decrease by ≥50% in the range of proteinuria.
Pyuria	>5 white blood cells/ high power field. Exclude infection.	Decrease by ≥50% in the number of white blood cells/ high power field.
Rash†	New onset, recurrent, or persistent inflammatory lupus rash. *Activity of skin lesions should be based on the evaluation of the most active lesion.*	Decrease by ≥50% of involved body surface area and/or activity of most active lesion with no worsening in either. Activity of the lesion should be determined by the color of the lesions: 0 –absent 1 – pink, faint erythema 2 – red 3 – dark red/purple/violaceous/crusted/hemorrhagic A ≥50% decrease in the activity of the lesion is defined by decreasing by 2 grades. Dyspigmentation, scarring and atrophy are not active lesions.
Alopecia†	New onset, recurrent, or persistent abnormal, patchy or diffuse loss of hair. *Size of patchy alopecic lesion should be determined based on involved total scalp surface. Total scalp surface is 4.5%.* *Diffuse alopecia is determined* by patient on numerical scale of 1-10. *Activity of alopecia should be based on the evaluation of the most active lesion.*	Decrease by ≥50% of total scalp involved area for patchy alopecic lesion or ≥50% reduction in the diffuse alopecia as determined by patient on numerical scale of 1-10, and/or activity of the most active alopecic lesions with no worsening in either. Activity of the alopecic lesion should be determined by the color of the most active lesion: 0 –absent 1 – pink, faint erythema 2 – red 3 – dark red/purple/violaceous/crusted/hemorrhagic A ≥50% decrease in the activity of the lesion is defined by decreasing by 2 grades.
Mucosal Ulcers	New onset, recurrent, or persistent oral or nasal ulcerations.	Decrease by ≥50% in the number of ulcers at this visit.
Pleurisy	Pleuritic chest pain with pleural rub or effusion, or pleural thickening.	≥50% reduction in the pain severity as determined by patient on numerical scale of 1-10 and/or ≥50% reduction in the amount of fluid (on imaging) with no worsening in either.
Pericarditis	Pericardial pain with at least one of the following: Rub, effusion, electrocardiogram or echocardiogram confirmation.	≥50% reduction in the pain severity as determined by patient on numerical scale of 1-10 and/or ≥50% reduction in the amount of fluid (on imaging) with no worsening in either.
Low Complement	Decrease in CH50, C3, or C4 below the lower limit of normal for testing laboratory.	≥50% increase in the level of any complement or normalization of one of them without a drop in either.
Increased anti-DNA antibodies levels	Increase in the level of anti-DNA antibodies above normal range for testing laboratory.	≥50% reduction in the level of anti-DNA antibodies.
Fever	>38° C. Exclude infectious causes.	≥50% reduction in the degree of fever above normal.
Thrombocytopenia	<100,000 platelets/ x 10^9/L. Exclude drug causes.	≥50% increase in the level of platelets but <100,000 platelets/mm^3.
Leukopenia	<3,000 white blood cells/x 10^9/L. Exclude drug causes.	≥50% increase in the level of white blood cells but <3,000/mm^3.

14 AMENDMENT(S) TO THE STATISTICAL ANALYSIS PLAN (SAP) (IF APPLICABLE)

14.1 AMENDMENT 1

Rationale for the amendment

During programming of analysis datasets and analyses, it was recognized that additional detail for aspects of analysis including endpoint definitions, data imputations, statistical model implementation, and display of results were needed. SAP Amendment 1 was written to clarify and document those details.

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

This document has been reviewed and approved per the Review and Approval of Clinical Documents Standard Operating Procedures. Signatures indicate that the final version of the Statistical Analysis Plan (SAP) or amended SAP is released for execution.