STATISTICAL ANALYSIS PLAN

Study: AS0011

Product: Bimekizumab

A PHASE 3, MULTICENTER, RANDOMIZED, DOUBLE-BLIND,
PLACEBO-CONTROLLED STUDY EVALUATING THE
EFFICACY AND SAFETY OF BIMEKIZUMAB IN THE
TREATMENT OF STUDY PARTICIPANTS WITH ACTIVE ANKYLOSING SPONDYLITIS

SAP/Amendment Number	Date
Final SAP	05 Jun 2019
SAP Amendment 1	05 Aug 2021
SAP Amendment 2	15 Nov 2021

Confidentiality Statement

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADAb	Anti - drug antibody
ADaM	Analysis Dataset Model
AE	Adverse event
ANCOVA	Analysis of covariance
AS	Ankylosing spondylitis
ASAS	Assessment of SpondyloArthritis International Society
ASAS20,40,5/6	Assessment of SpondyloArthritis International Society 20%, 40%, 5 out of 6 response criteria
ASAS-PR	ASAS partial remission
ASDAS-CRP	Ankylosing Spondylitis Disease Activity Score C-reactive protein
ASDAS-MI	Ankylosing Spondylitis Disease Activity Score major improvement
ASQoL	Ankylosing Spondylitis Quality of Life
ASspiMRI-a	Ankylosing Spondylitis spine MRI-activity
ATC	Anatomical Therapeutic Chemical Classification
axSpA	Axial spondyloarthritis
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BASDAI50	Bath Ankylosing Spondylitis Disease Activity Index 50% improvement
BASFI	Bath Ankylosing Spondylitis Functional Index
BASMI	Bath Ankylosing Spondylitis Disease Metrology Index
BLQ	Below the limit of quantification
BKZ	bimekizumab
BMI	Body mass index
BP	Blood pressure

BSA	Body surface area
CI	Confidence interval
COVID-19	Coronavirus disease 2019
CFS	COVID-19 Free Set
CRF	Case report form
CRP	C-reactive protein
CV	Coefficient of Variation
DBTP	Double-blind Treatment Period
DILI	Drug-induced liver injury
DMARD	Disease-modifying antirheumatic drug
DMC	Data Monitoring Committee
EAER	Exposure adjusted event rate
EAIR	Exposure adjusted incidence rate
ECG	Electrocardiogram
eC-SSRS	electronic Columbia-Suicide Severity Rating Scale
ES	Enrolled Set
EQ-5D-3L	EuroQol-5D-3-Level
ET	Early termination
FACIT	Functional Assessment of Chronic Illness Therapy
FAS	Full analysis set
HLA-B27	Human leukocyte antigen B27
HLT	High level term
HRQoL	Health-related quality of life
hs-CRP	High-Sensitivity C-Reactive Protein

IBD	Inflammatory Bowel Disease
IMP	Investigational Medicinal Product
IPD	Important protocol deviation
IWRS	Interactive Web Response System
LLOQ	Lower limit of quantification
LOCF	Last observation carried forward
LSM	Least Square Means
MACE	Major adverse cardiac events
MAR	Missing at random
MASES	Maastricht Ankylosing Spondylitis Enthesitis
MCMC	Markov-Chain Monte Carlo
MCS	Mental Component Summary
MDA	Minimal Disease Activity
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MOS	Medical outcome Study
MRI	Magnetic resonance imaging
MS	Maintenance Set
NAb	Neutralizing anti-drug antibody
NI	Negative immunodepletion
NRI	Non-responder imputation
NRS	Numeric Rating Scale
NS	Negative screen
NSP	Nocturnal spinal pain

OC	Observed cases
OLE	Open label extension
OR	Odds ratio
PCS	Physical Component Summary
PGADA	Patient's Global Assessment of Disease Activity
PhGADA	Physician's Global Assessment of Disease Activity
PHQ-9	Patient Health Questionnaire-9
PI	Positive immunodepletion
PK-PPS	Pharmacokinetics Per-Protocol Set
PPS	Per-Protocol Set
PS	Positive screen
PSO	Psoriasis
PT	Preferred term
Q4W	Every 4 weeks
RD	Risk difference
RS	Randomized set
SAE	Serious adverse event
SAP	Statistical analysis plan
sc	Subcutaneously
SD	Standard deviation
SIB	Suicidal Ideation and Behavior
SJC	Swollen joint count
SF-36	Short-Form 36-item Health Survey
SMQ	Standardized MedDRA Query

SOC	System Organ Class
SPARCC	Spondyloarthritis Research Consortium of Canada
SS	Safety Set
TEAE	Treatment emergent adverse event
TEMA	Treatment-emergent markedly abnormal
TFLs	Tables, Listings and Figures
TJC	Tender joint count
TNFα	Tumor necrosis factor alpha
ULN	Upper limit of normal
VAS	Visual Analog Scale
WHO-DD	World Health Organization Drug Dictionary
WPAI-SHP	Work Productivity and Activity Impairment Questionnaire - Specific Health Problem

1 INTRODUCTION

This statistical analysis plan (SAP) provides the necessary information to perform the final statistical analysis as well as the interim analyses for study AS0011. It also defines the summary tables, listings and figures (TFLs) to be generated for the interim and final clinical study report.

The SAP is based on the Protocol Amendment 4 (16 FEB 2021). All references to study protocol hereafter refer to this version of the protocol, and, unless otherwise specified, the study will be analyzed as described in this version of the protocol. If a future protocol amendment necessitates a substantial change to the statistical analysis of the study data, or if analysis definitions must be modified or updated, this SAP will be amended accordingly.

The content of this SAP is compatible with the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)/Food and Drug Administration (FDA) E9 Guidance documents (ICH E9, 1998).

2 PROTOCOL SUMMARY

2.1 Study objectives

2.1.1 Primary objective

The primary objective is to demonstrate the efficacy of bimekizumab administered subcutaneously (sc) every 4 weeks (Q4W) compared with placebo in the treatment of study participants with active ankylosing spondylitis (AS).

2.1.2 Secondary objectives

The secondary objectives are as follows:

- To assess the efficacy of bimekizumab compared with placebo
- To assess the safety and tolerability of bimekizumab
- To assess the impact of bimekizumab on patient-reported quality of life
- To assess the impact of bimekizumab on spinal mobility
- To assess the impact of bimekizumab on enthesitis and peripheral arthritis

2.1.3 Other objectives

Other objectives are as follows:

- To assess the immunogenicity of bimekizumab
- To assess the pharmacokinetics of bimekizumab
- To assess the maintenance of efficacy of bimekizumab
- To assess the relationship between exploratory biomarkers, drug treatment, and AS disease biology
- To assess the impact of bimekizumab on work productivity
- To assess the effect of bimekizumab on inflammatory changes using magnetic resonance imaging (MRI)

2.2 Study variables

2.2.1 Efficacy variables

2.2.1.1 Primary efficacy variable

The primary efficacy variable is the ASAS40 (Assessment of SpondyloArthritis International Society 40%) response at Week 16.

2.2.1.2 Secondary efficacy variables

The secondary efficacy variables are as follows:

- ASAS40 response at Week 16 in the Tumor Necrosis Factor alpha (TNFα) inhibitor-naïve study participants
- ASAS20 response at Week 16
- Change from Baseline in Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) total score at Week 16
- ASAS partial remission (ASAS-PR) at Week 16
- Ankylosing Spondylitis Disease Activity Score major improvement (ASDAS-MI) at Week 16
- ASAS5/6 response at Week 16
- Change from Baseline in Bath Ankylosing Spondylitis Functional Index (BASFI) at Week 16
- Change from Baseline in nocturnal spinal pain (NSP) score (based on Numeric Rating Scale) at Week 16
- Change from Baseline in Ankylosing Spondylitis Quality of Life (ASQoL) total score at Week 16
- Change from Baseline in the Short-Form 36-item Health Survey (SF-36) Physical Component Summary (PCS) score at Week 16
- Change from Baseline in Bath Ankylosing Spondylitis Disease Metrology Index (BASMI) at Week 16
- Change from Baseline in the Maastricht Ankylosing Spondylitis Enthesitis (MASES) index at Week 16 in the subgroup of study participants with enthesitis at Baseline
- Enthesitis-free state based on the MASES index at Week 16 in the subgroup of study participants with enthesitis at Baseline

2.2.1.3 Other efficacy variables

The following other efficacy variables will be assessed. These variables will be assessed at the time points defined in Table 5-2 of the protocol. All time points not specified in Section 2.2.1.1 or Section 2.2.1.2 of the SAP are considered exploratory.

- ASAS40 response
- Time to ASAS40 response
- ASAS20 response

- Time to ASAS20 response
- ASAS5/6 response
- ASAS-PR
- Change from Baseline in Ankylosing Spondylitis Disease Activity Score C-reactive protein (ASDAS-CRP)
- ASDAS status (eg, inactive disease [ID], low disease [LD], high disease [HD], and very high disease [vHD])
- ASDAS-MI
- Change from Baseline in BASDAI total score
- BASDAI50 response
- Change from Baseline in the average score of Questions 5 and 6 of the BASDAI concerning morning stiffness
- Change from Baseline in BASFI
- Change from baseline in the MASES index in the subgroup of study participants with enthesitis at Baseline
- Enthesitis-free state based on the MASES index in the subgroup of study participants with enthesitis at Baseline
- Change from Baseline in BASMI
- Change from Baseline in Physician's Global Assessment of Disease Activity (PhGADA)
- Change from Baseline in Patient's Global Assessment of Disease Activity (PGADA)
- Change from Baseline in total spinal pain score (based on Numeric Rating Scale)
- Change from Baseline in nocturnal spinal pain score (based on Numeric Rating Scale)
- Change from Baseline in high sensitivity C-reactive protein (hs-CRP)
- Responses to the European Quality of Life-5 Dimensions 3-Level (EQ-5D-3L) version
- Change from Baseline in EQ-5D-3L Visual Analog Scale (VAS) scores
- Change from Baseline in the EQ-5D-3L utility score
- Change from Baseline in the SF-36 PCS score
- Change from Baseline in the SF-36 Mental Component Summary (MCS) score
- Change from Baseline in Sleep Disturbance and Sleep Problems Index II score (Medical Outcomes Study [MOS 12-item] scale)
- Change from Baseline in Functional Assessment of Chronic Illness Therapy (FACIT) - Fatigue subscale score
- Change from Baseline in ASQoL total score

- Change from Baseline in Work Productivity and Activity Impairment Questionnaire–Specific Health Problem (WPAI-SHP)
- Change from Baseline in (44/44) tender joint count (TJC) and swollen joint count (SJC)
- Change from Baseline in Ankylosing Spondylitis spine MRI-activity (ASspiMRI-a) in the Berlin modification score (Note that this endpoint will be analyzed for study participants in the MRI substudy only)
- Change from Baseline in sacroiliac joint Spondyloarthritis Research Consortium of Canada (SPARCC) score (Note that this endpoint will be analyzed for study participants in the MRI substudy only)

2.2.2 Pharmacokinetic variable

The pharmacokinetic (PK) variable is the plasma concentration of bimekizumab. Study participants are asked to provide blood samples for these measurements at Baseline, Week 2, 4, 8, 12, 16, 20, 24, 36, 52, ET (Early Termination) and SFU visits.

2.2.3 Pharmacogenomic variables

Where local regulations permit, additional blood samples will be collected at specific time points from study participants consenting to participate in a pharmacogenomic sub-study.

The pharmacogenomic variables are the blood or blood derivative (eg, serum) concentrations of cytokines and chemokines of relevance to IL-17A/F signaling pathway and axSpA biology and will be assessed at Baseline, at Week 16 and at Week 52.

Stool samples will also be collected at Baseline and Week 52. These samples may be used to assess biomarkers of gut inflammation including but not limited to calprotectin and microbiome testing.

These variables will be described in a separate statistical analysis plan. The nature and format of these sub-study analyses will be determined at a later date.

2.2.4 Immunological variables

Immunological variables allow evaluation of immunogenicity. The immunological variables are:

- The anti-bimekizumab antibody status,
- The treatment-emergent antibody positivity derived from anti-drug antibody (ADAb) assays.
- The neutralizing anti-drug antibody (NAb) status.

The anti-drug antibody status will be assessed at Baseline, Week 4, 8, 12, 16, 20, 24, 36, 52, ET and SFU, and the treatment-emergent antibody positivity at the same post-Baseline timepoints.

Note that the NAb analysis results are not planned to be part of the CSR for the Week 24 interim analysis.

2.2.5 Safety variables

2.2.5.1 Secondary safety variables

Secondary safety variables to be assessed are as follows:

- Incidence of treatment-emergent adverse events (TEAEs)

- Incidence of treatment-emergent serious adverse events (SAEs)
- TEAEs leading to withdrawal from Investigational Medicinal Product (IMP)

2.2.5.2 Other safety variables

Other safety variables to be assessed are:

- Change from Baseline in vital signs (blood pressure [BP], temperature, and pulse rate)
- Standard 12-lead electrocardiogram (ECG) results
- Change from Baseline in clinical laboratory values (hematology, biochemistry and urinalysis)
- Change from Baseline in Patient Health Questionnaire-9 (PHQ-9)

Physical examination findings considered clinically significant changes since the physical examination at the Screening Visit will be recorded as AEs

2.3 Study design and conduct

AS0011 is a Phase 3, multicenter, randomized, double-blind, placebo-controlled study. To be eligible to participate in this study, study participants must meet the following key inclusion criteria:

- Adult study participants with a diagnosis of active AS, determined by documented radiologic evidence (x-ray) fulfilling the Modified New York criteria for AS (1984), including at least 3 months of symptoms,
- Age at symptom onset < 45 years old,
- Study participant must have moderate-to-severe active disease as defined by BOTH of the following at screening AND at Baseline:
  - BASDAI ≥4 **AND** Spinal pain ≥4 on a 0 to 10 Numeric Rating Scale (NRS) (from BASDAI Item 2)

Study participants treatment assignment will be stratified by prior TNFα inhibitor exposure (yes/no) and region (4 categories).

Detailed study entrance criteria are described in the protocol.

The study consists of the following periods:

- a Screening Period (≥14 days to ≤35 days),
- a 52-Weeks Treatment Period, consisting of:
  - a 16-Week Double-Blind Treatment Period, and
  - a 36-Week Maintenance Period
- and a Safety Follow-Up (SFU) Period, 20 weeks after the final dose of IMP (for study participants who complete the study and do not enter the open-label extension (OLE) study or for study participants who discontinue early, including those withdrawn from IMP).

The maximum study duration per study participant will be up to 73 weeks. The end of the study is defined as the date of the last visit of the last study participant in the study.

At the end of the Baseline visit, eligible study participants will be randomized in a 2:1 ratio (stratified by region [Asia, Eastern Europe, North America, Western Europe] and prior TNFα inhibitor exposure [yes, no]) to receive one of the two following blinded treatment groups:

- Bimekizumab 160mg sc Q4W,
- Placebo

Study treatment will be administered by unblinded study personnel at the study site.

Approximately 300 study participants are planned to be randomized (200 study participants in bimekizumab treatment group, and 100 study participants in placebo group). Study participants will remain on their allowable background medication through the study conduct.

The Double-Blind Treatment Period ends after Week 16 assessments. At Week 16, study participants will transition from the double-blind, placebo-controlled treatment into the 36-week Maintenance Period. Study participants receiving placebo during the Double-Blind Treatment Period will be re-allocated to bimekizumab treatment at Week 16 after all study assessments have been completed.

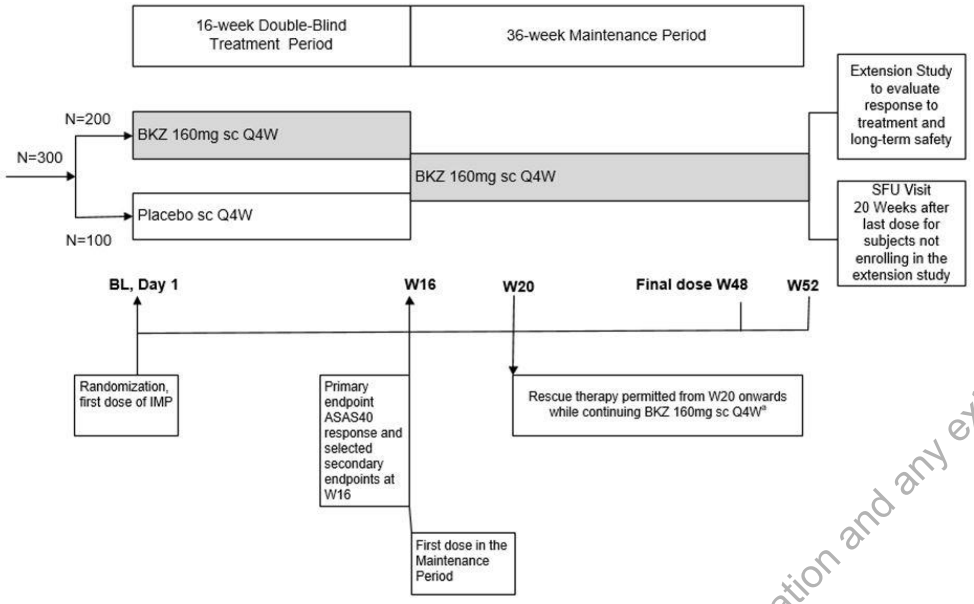
From Week 20 onwards, all study participants will be eligible for nonbiologic rescue therapy.

Prior to the completion of the Maintenance Period, Investigators should discuss treatment options with the study participant. Study participants who are not withdrawn from IMP and who meet the eligibility criteria will be given the opportunity to enter an extension study to continue to receive bimekizumab.

Study participants discontinuing IMP during the Double-Blind Treatment Period or during the Maintenance Period will be encouraged to return for all scheduled visits through Week 52 and the SFU Visit, as applicable.

A schematic representation of the study design is presented in Figure 2–1.

Figure 2–1: Study design



ASAS40=Assessment of SpondyloArthritis International Society 40% response criteria; BKZ=bimekizumab; BL=Baseline; IMP=investigational medicinal product; Q4W=every 4 weeks; sc=subcutaneous; SFU=Safety Follow-Up; W=week

[a] Subjects are eligible for nonbiologic rescue therapy starting at Week 20 with treatment at the discretion of the Investigator while continuing to receive BKZ. Treatment with non-BKZ biologics or prohibited treatment will lead to BKZ discontinuation.

Refer to the study clinical protocol for more detailed information on the study assessments and schedule of procedures.

2.4 Determination of sample size

Approximately 300 study participants will be randomly assigned in a 2:1 ratio to the following treatment groups: bimekizumab 160 mg sc (200 study participants) or placebo sc Q4W (100 study participants). The primary efficacy analysis is based on the bimekizumab dose versus placebo for ASAS40 response at Week 16. All sample size and power calculations were done at a significance level of 0.05 in a two-sided test.

All sample size and power calculations were performed using the software nQuery Advisor® 7.0.

2.4.1 Power calculation for primary endpoint

The sample size assumptions for bimekizumab versus placebo are based on the ASAS40 response data from the Phase 2b bimekizumab study in study participants with active AS (AS0008).

An ASAS40 response at Week 16 of 40% can be assumed for the bimekizumab treatment group. This assumed treatment response is based on the bimekizumab 160mg dose results of AS0008 at Week 12 (46.7%). It was also assumed that the treatment response at Week 12 and Week 16 would be the same (ie, conservative assumption). In addition, the observed ASAS40 response rates were adjusted to account for a higher number of study participants with prior TNFα inhibitor exposure and for a higher number of study participants with early withdrawal. Placebo ASAS40 response at Week 16 was assumed to be 15% and is more conservative than that of AS0008 at Week 12 (ie, ASAS40 response=13.3%).

The sample size for showing statistical superiority of bimekizumab versus placebo was calculated using a 2-sided 2-sample Chi square test with continuity correction (Fleiss et al, 1980). With 200 study participants in the bimekizumab treatment group and 100 study participants in the placebo group, the test for detecting statistical superiority of bimekizumab versus placebo based on ASAS40 response at Week 16 is powered with >99%.

2.4.2 Power calculations for secondary endpoints in the hierarchical testing

The assumptions for power calculations of the secondary endpoints are based on the Week 12 results of AS0008, unless indicated otherwise. It was assumed that the treatment response at Week 12 and Week 16 would be the same (ie, conservative assumption). In addition, the observed data were adjusted in the same way as the primary endpoint to account for a higher number of study participants with prior TNFα inhibitor exposure and for a higher number of study participants with early withdrawal.

All power calculations for binary endpoints were performed using a 2-sided 2-sample Chi-square test with continuity correction (Fleiss et al, 1980). All power calculations for continuous endpoints were performed using a 2-sided 2-group Satterthwaite t-test (Moser et al, 1989) The power calculations provided below represent independent evaluations for each endpoint and do not account for multiplicity.

For ASAS40 response of the TNFα inhibitor-naïve population at Week 16, an ASAS40 response of 48% for the bimekizumab group was assumed. Placebo ASAS40 response at Week 16 is 18%. It is also assumed that 70% of the study participants in each treatment group are TNFα inhibitor-naïve. With those assumptions, the endpoint is powered with >99% at the planned sample size (ie, 140 and 70 study participants in the bimekizumab and placebo group, respectively).

For ASAS20 response at Week 16, an ASAS20 response of 70% for the bimekizumab group was assumed. Placebo ASAS20 response at Week 16 is 30%. With those assumptions, the endpoint is powered with 99% at the planned sample size.

For change from Baseline in BASDAI at Week 16, an adjusted between-treatment difference of -1.38 with an SD=1.8 for bimekizumab and SD=1.4 for placebo was used. With those assumptions, the endpoint is powered with >99% at the planned sample size.

For ASAS-PR at Week 16, an ASAS-PR of 17% for the bimekizumab group was assumed.

Placebo ASAS-PR at Week 16 was assumed to be 3.3%. With those assumptions, the endpoint is powered with 94% at the planned sample size.

For ASDAS-MI at Week 16, the ASDAS-MI data of AS0008 at Week 12 and Week 16 were used and the placebo response was extrapolated to Week 16. The assumed ASDAS-MI are 36.0% and 13.8% at Week 16 for the bimekizumab and placebo group, respectively. With those assumptions, the endpoint is powered with 98% at the planned sample size.

For ASAS5/6 response at Week 16, an ASAS5/6 response of 44% for the bimekizumab group was assumed. Placebo ASAS-PR at Week 16 was assumed to be 5%. With those assumptions, the endpoint is powered with >99% at the planned sample size.

For change from Baseline in BASFI at Week 16, an adjusted between-treatment difference of -2.35 with an SD=1.9 for bimekizumab and placebo was used. With those assumptions, the endpoint is powered with >99% at the planned sample size.

For change from Baseline in nocturnal spinal pain at Week 16, the AS data from the certolizumab pegol study AS001 were used. An adjusted between-treatment difference of -1.88 with an SD=2.9 for bimekizumab and SD=2.4 for placebo was used. With those assumptions, the endpoint is powered with >99% at the planned sample size.

For change from Baseline in ASQoL score at Week 16, an adjusted between-treatment difference of -2.39 with an SD=4.5 for bimekizumab and placebo was used. With those assumptions, the endpoint is powered with >99% at the planned sample size.

For change from Baseline in PCS score of the SF-36 at Week 16, an adjusted between-treatment difference of 4.56 with an SD=8.1 for bimekizumab and SD=5.8 for placebo was used. With those assumptions, the endpoint is powered with >99% at the planned sample size.

For change from Baseline in BASMI at Week 16, an adjusted between treatment difference of -0.55 with an SD=0.8 for bimekizumab and SD=0.7 for placebo was used. With those assumptions, the endpoint is powered with >99% at the planned sample size.

3 DATA ANALYSIS CONSIDERATIONS

3.1 General presentation of summaries and analyses

Statistical analysis and generation of tables, figures, study participant data listings, and statistical output will be performed using SAS® Version 9.3 or higher. The following should be noted in this regard:

- UCB uses SAS in a 64-bit Windows environment, and it is well-documented that in this environment the maximum accuracy of any numeric value is 15 significant digits.
- However, SAS by default does not limit the accuracy of numeric values to 15 significant digits which, in certain instances, may result in inaccurate representation of the data and cause errors when used in subsequent calculations, particularly when comparing a value to a chosen threshold. This, in turn, could potentially result in a change in classification of a study participant from a responder to a non-responder (and vice versa) if these values occur near a threshold used in the evaluation of response (or a critical laboratory value for example).

Therefore, in order to avoid issues caused by inaccurate floating point representation of numeric values, the Analysis Dataset Model (ADaM) variables AVAL (absolute value), CHG (change) or PCHG (percentage change) are rounded to 12 decimal places prior to the comparison to the threshold. This is applied exclusively during the derivation of new response parameters (subsequently retained in the dataset) or critical value variables and does not imply inherent rounding on AVAL, CHG or PCHG variables which are retained unrounded in the final ADaM dataset.

All tables and listings will use Courier New font size 9.

Descriptive statistics will be displayed to provide an overview of the study results.

For categorical parameters, the number and percentage of study participants in each category will be presented. Unless otherwise noted, the denominator for percentages will be based on the number of study participants included in the respective analysis set. Study participants with missing data can generally be accounted for using one of the following approaches:

- Percentages will be summarized based on all study participants in the analysis set and a "Missing" category (corresponding to study participants with missing data for the variable

being summarized) will be included as the last row in the list of categories being summarized.

- Percentages will be based only on those study participants with observed data for the variable being summarized. As the denominator may be different from the number of study participants in the analysis set being considered, the denominator should be displayed in the table. The general format for displaying this will be "n/Nsub (%)".

The approach to be considered will be further specified in the relevant section of the variable of interest.

Unless otherwise noted, all percentages will be displayed to one decimal place. No percentage will be displayed for zero counts, and no decimal will be presented when the percentage is 100%.

Unless stated otherwise, continuous endpoints will be expressed as a change from Baseline which will be calculated as the value at a specific timepoint minus the value at Baseline.

For continuous parameters, descriptive statistics will include number of study participants with available measurements (n), arithmetic mean, standard deviation (SD), median, minimum, and maximum (unless otherwise stated).

For hs-CRP variable, the summary statistics should contain arithmetic mean, geometric mean, geometric coefficient of variation (CV), median, first and third quartile (Q1 and Q3), minimum and maximum. The geometric CV (%) will be calculated using the following formula:

$$CV = \sqrt{e^{SD_{ln}^2} - 1}$$

where SD_{ln} represents the standard deviation of the ln-transformed hs-CRP values.

For bimekizumab PK concentrations, geometric mean, geometric CV, 95% confidence intervals for geometric mean (assuming log-normally distributed data) will calculated if at least ⅔ of the values of interest are above the LLOQ and N>=3. If this is not the case, only median, minimum and maximum will be presented.

Decimal places for descriptive statistics will always apply the following rules:

- "n" will be an integer.
- Mean, SD, SE, and median will use 1 additional decimal place compared to the original raw value from the Case Report From (CRF).
- Coefficient of variance (CV [%]) will be presented with 1 decimal place.
- Minimum and maximum will have the same number of decimal places as the original raw value.

If the number of decimal places reported in the raw data is varied then either the maximum raw number of reported decimal places or 3 will be used, whichever is the lowest, as a guide for the descriptive statistics.

Unless stated otherwise, statistical tests of efficacy variables will be performed 2-sided and 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." Statistical comparison will be performed at the 0.05 level of significance.

The SAS® outputs supportive of any inferential statistics that are part of the hierarchical testing procedure (ie, all inferential statistics associated with the endpoints in figure 4.1) will be provided as a separate PDF document in addition to TFLs. These outputs will be included in the 'Documentation of Statistical Methods' section of the clinical study report.

Unless specified differently in the SAP, the order of treatment groups to be presented in tables from left to right will be:

- Placebo
- Bimekizumab 160mg Q4W

Selected tables may include columns for:

- all study participants, regardless of study treatment
- all study participants on bimekizumab, regardless of whether they shifted from placebo after the Double-Blind Treatment Period or not
- study participants randomized on placebo who shifted to bimekizumab after the Double-Blind Treatment Period

The abbreviation for bimekizumab is BKZ and will be used in tables and listings headers. In the TFLs study participants on bimekizumab 160mg will be displayed as "BKZ 160mg Q4W".

All study participants dosed with bimekizumab will be summarized in some TFLs in a treatment column labeled as "BKZ 160mg Q4W Total".

For all TFLs not dedicated to the Double-Blind treatment period only, the Placebo-randomized study participants who shifted to bimekizumab after the Double-Blind Treatment Period will be summarized in a treatment column labeled as "Placebo / BKZ 160mg Q4W".

Per protocol,

- visit windows of ±3 days are permissible for all visits through Week 16 (included). The visit window is relative to the Baseline Visit.
- visit windows of ±4 days are permissible for all visits after Week 16,
- for the SFU Visit, the visit should occur no more than 3 days prior to the scheduled visit date and within 7 days after the scheduled visit date.

Only data collected at the scheduled visit will be included in the analysis. The only deviations may concern PK data, unscheduled vendor data collected with a few days of delay, and MRI data (see below).

All by-visit summaries will contain nominal (ie, scheduled) visits only. Unscheduled visits will not be mapped to scheduled visits except for some assessments that may occur within 3-day time window of a scheduled visit. In that case, the assessment will be mapped to the corresponding scheduled visit and will be used for the analysis. Where scheduled visit data is available no mapping will be done, but if scheduled visit data is not available while unscheduled visit data within the 3-day time window is available, the unscheduled visit data collected on a day nearest to the scheduled visit day will be mapped and used in the analysis. This will only occur for PK data (see Section 9.1) some vendor data (see Section 10.4.3) and MRI data (see Section 8.3.1.12).

Assessments collected at unscheduled visits will only be listed, except in the case of determining Baseline, treatment-emergent markedly abnormal (TEMA) criteria for laboratory and vital sign parameters. For analyses on TEMA laboratory and vital sign, all post-Baseline (scheduled and unscheduled) values will be summarized in table outputs.

For by-visit tables summarizing efficacy data, the SFU visit will not be included. By-visit tables summarizing safety data should however include the SFU visit.

A complete set of data listings containing all documented data and all derived data (eg, change from Baseline) will be generated.

Unless otherwise stated, listings will be sorted by treatment, study participant number within each treatment group (not randomization number), variable (if applicable) and visit (if applicable; including timing relative to dosing if applicable). For listings including nonrandomized study participants, the nonrandomized study participants will be shown first in the listing, ordered by study participant number. All listings will include repeat and unscheduled measurements; such measurements will appear in chronological order together with the scheduled visits, ie, a repeated measurement will appear directly after the visit and time relative to dosing for which the repeat measurement was performed. 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'.

3.2 General study level definitions

3.2.1 Analysis time points

3.2.1.1 Relative day

- The relative day will be included in some listings and will be calculated as follows. The way that relative day is calculated depends on when the given event occurs relative to the date of first IMP administration. If the event start (stop) date occurred on or after the first dose of study drug administration, but prior to the study drug stop date, relative day is calculated as start (stop) date of the event minus first dose date of study drug dose + 1.

  Relative day 1 is the date of first study drug administration.

- If the event start (stop) date occurred after the last dose of study drug administration, the relative day to the most recent study drug dose is calculated as start (stop) date of the event minus most recent study drug dose date. The relative day in this situation should be preceded by a '+'

- If the event start (stop) date occurred before the first dose of study drug administration, the relative day is calculated as start (stop) date of the event minus first dose date of study drug dose. The relative day in this situation should be preceded by a '-'.

Relative day will only be computed for fully completed dates and will be left blank for partial dates.

For AEs, relative days for start and stop dates will be calculated as the number of days since the first IMP administration.

For events occurring after Week 16 in study participants who have switched from placebo to bimekizumab, an additional relative day will be calculated using the same rules as above but

based on the date of first injection of bimekizumab. This additional relative day will be provided in listings for concomitant medications and safety data.

For non-treatment emergent AEs, relative day of onset will be negative if the event started and stopped before the first dose. A complete date must be established in order to correctly identify the AEs. Section 4.2.2 describes imputation rules in case of missing data for AEs.

3.2.1.2 End date of the Treatment Period

The end of the Treatment Period will be either the date of Week 52 visit (Visit 17) for study participants completing the Treatment Period, or the date of the early termination (ET) visit for study participants who discontinued early during the Treatment Period.

Note that study participants who previously discontinued IMP and are continuing for all scheduled visits through Week 52 will also be considered as having completed the Treatment Period.

If a study participant does not have a Week 52/ET visit, then either the date of the last scheduled or unscheduled visit during the Treatment Period or the date of last known dose of study drug during the Treatment Period, whichever is later, will define the end date of the Treatment Period.

3.2.2 Study periods

3.2.2.1 Study periods for safety and efficacy data

Based on the specific requirements of safety analyses and efficacy analyses, two concepts of the definition of periods need to be considered for the study, which will be defined in detail in the sections below.

3.2.2.1.1 Study periods for safety data

The following study periods are defined for the classification of safety data by study period:

- Pre-treatment Period (Screening Period): Starts at the time of the informed consent date and ends prior to the day and time of the first dose administration of study drug.
- 16-Week Double-Blind Treatment Period (DBTP): Starts on the day of the first IMP administration and ends prior to the day of first IMP administration in the maintenance period (first IMP in the MP typically occurs at Week 16, but for those that miss the Week 16 IMP, first IMP in the MP will be the first IMP administration after Week 16). For study participants that withdraw from the study prior to start of IMP in the maintenance period, the DBTP ends on the day of the ET Visit.

  See Section 3.7 (Treatment assignment and treatment groups) for allocation of AEs that occur on the day of IMP administration.

  For vital sign parameters, the DBTP will start at the day and time of first IMP administration and ends prior to the day and the time of first IMP administration in the MP as there are 3 assessments collected at Baseline and Week 16 (one pre-dose, one 30 minutes post-dose and one 1-hour post-dose).
- 36-Week Maintenance Treatment Period (MP): Starts on the day of IMP administration at the Week 16 visit and ends 28 days (one dosing interval) after the day of last IMP administration before the Week 52 visit. For participants that do not receive IMP at the Week 16 visit, the MP starts on the day of first IMP administration after Week 16. For study participants that

withdraw from the study after Week 16 but before Week 52, the MP ends on the day of the ET Visit. In the event that no ET visit is performed, the MP ends 28 days after the last IMP administration.

For vital sign parameters, the day and the time will be used to define the MP for the same reason that given above.

- Safety Follow-Up Period: The SFU period will be defined for all study participants who complete the final IMP administration of the study at Week 48 and do not enter the OLE study, or for study participants who discontinue early including those withdrawn from IMP.
  - For study participants who complete the final IMP administration of the study at Week 48 and do not enter the OLE study, the SFU period starts 29 days after the day of last IMP administration before the Week 52 visit and ends on the day of the SFU Visit or on the last contact date, whichever happens later.
  - For study participants who discontinue early, the SFU period starts the day after the ET Visit and ends on the day of the SFU Visit or on the last contact date, whichever happens later. In the event that no ET visit is performed, the SFU period starts 29 days after the day of last IMP administration.

  Per study protocol, if the Week 52 Visit or ET visit is done ≥ 20 weeks after the final dose of IMP, the SFU Visit is not required. In that case, the last contact date (Week 52, ET or earlier if applicable) will be used as end of the SFU period.

  Note that an IMP administration is said to be "final" when it refers to the very last IMP taken by the study participant in the study. Otherwise, the IMP administration is said to be "last" for a specific study period.

  Safety analyses will be summarized for the DBTP, MP, and Overall Periods. Summaries for the MP will also include the SFU period for study participants that do not enter the OLE or discontinue early in the MP. The SFU period for study participants that discontinue in the DBTP will only be summarized in the Overall Period tables.

3.2.2.1.2 Study periods for efficacy data

The following study periods are defined for the classification of efficacy data by study period:

- Screening Period: Starts at the time of the informed consent date and ends prior to the day and time of the first IMP administration.
- 16-Week Double-Blind Treatment Period: Starts on the day of the first IMP administration and ends the day of the Week 16 Visit or on the day of the ET Visit for study participants withdrawn from the study before Week 16 visit. If Week 16 Visit is missing or skipped, the DBTP will end on the day of the last visit before Week 16.

  Efficacy assessments from the Week 16 Visit belong to the DBTP as they are conducted prior to the Week 16 IMP administration.
- 36-Week Maintenance Treatment Period: Starts on the day after the Week 16 Visit and ends the day of Week 52 Visit. If the Week 16 Visit is missing, the MP will start the day after the date of the last visit before Week 16. For study participants withdrawn from the study after Week 16 but before Week 52, the MP will end at the ET Visit.

A study participant will be considered to have completed a study period if they complete the last planned study visit for that period.

It is possible for a study participant to discontinue the study after completing the Week 16 efficacy assessments, (and therefore after completing the DBTP), but prior to entering the MP the next day. Therefore, all study participants who discontinue at or after Week 16 and never start the MP will be counted as having discontinued in the overall period only.

3.2.2.2 Coronavirus Disease 2019 (COVID-19) pandemic periods

The following COVID-19 pandemic periods will be defined:

- Prior COVID-19 pandemic period: Period prior to COVID-19 pandemic start date, which is defined as 11-Mar-2020 by the World Health organization (WHO), not including the pandemic start date.
- COVID-19 pandemic period: Period from 11-Mar-2020 through the COVID-19 pandemic end date (which is currently not defined), including the dates of pandemic start and end.
- Post COVID-19 pandemic period: Period after the WHO declaration of the end of the pandemic, beginning on the day after the pandemic end date. The end of the pandemic had not yet been declared at the time of SAP finalization. As a result, this period will only be included in analyses if the WHO declares an end to the pandemic prior to the end of the study.

These periods will be used to classify the timing of the following relative to COVID-19 pandemic:

- study participant enrolment in the study,
- study participant Week 16 Visit,
- AE onset date.

3.2.3 Mapping of assessments performed at Early Termination Visit

If the early termination visit occurs at a scheduled visit as outlined in the schedule of assessments, then no mapping is necessary, and any early withdrawal assessments should correspond to that scheduled visit. Premature study termination visit assessments that occur on a date after a scheduled visit will be assigned to the next scheduled site visit per the protocol following the last visit where assessments were available. The only exception to the above rule is for anti-bimekizumab antibody assessments, in which all early withdrawal visit assessments will be assigned to the next scheduled visit at which anti-bimekizumab antibodies are assessed.

3.3 Definition of Baseline values

A Baseline value for a study participant is defined as the latest measurement for that study participant up to and including the day of administration of first study medication (ie, generally Visit 2). When no treatment information is available, the randomization date will be used as reference in place of the first day of study drug administration. If a Baseline assessment is taken on the same day as first administration of study medication, it is eligible to be used as the Baseline value, even in the case that the time of the assessment is recorded as taking place after the time of first study medication administration. This is considered acceptable as this measurement is still the best representation of the Baseline value of the given assessment since it

is highly unlikely that the study medication could have an impact on any measurement in such a short period of time. However, such cases should be rare as study center personnel are instructed to do all assessments at the Baseline visit prior to administering study medication. One exception to this rule is plasma concentration. If Baseline plasma concentration is measured at a time after the first administration of study medication, then it should not be eligible to be considered as a Baseline plasma concentration. Such cases should be discussed with the quantitative clinical pharmacologist.

For HLA-B27 only, the Baseline value may be based on a value recorded after the first study medication administration because the study medication is not expected to have any impact on the HLA-B27 measurement.

If a Baseline measurement is missing or not collected, and a Screening value is available, the Screening value will be utilized as Baseline instead. If no measurement is available prior to receiving study medication, then the Baseline value is treated as missing. Finally, re-test value can be considered to define a Baseline, as mentioned in Section 3.1.

Baseline values for component scores should be computed using components from the same visit where the relevant measurements were recorded prior to dosing. However, if the Screening visit has all of the components, but the Baseline visit is missing one or more components, the Baseline value for the component score will be calculated using the Screening visit values.

When the time of first dose is derived, it should be based on the first injection of study treatment, regardless of whether or not it is an active treatment.

For study participants randomized to placebo that switched to bimekizumab at the Week 16 visit the laboratory data (only for the shift table analyses, the analyses on TEMA and Hy's Law cases) and the vital sign data (only for the TEMA analyses) will have an additional Baseline for the Maintenance Period that will correspond to the latest measurement on or prior to the first bimekizumab dose in that period. This Baseline will be called "Baseline Week 16". To avoid any confusion in the outputs where both Baseline values will be used, the Baseline relative to the administration of first study medication will be called "Baseline Day 1".

For the MRI substudy, Baseline will also be defined as the latest measurement for that study participant up to and including the day of administration of first study medication. However, the efficacy analyses for the substudy will be performed using two sets of study participants based on the date of the Baseline MRI relative to first study medication administration: 1) all eligible substudy participants with an MRI any time prior to first study medication administration, and 2) eligible substudy participants with an MRI performed any time from 3 weeks prior to the first administration of study medication up to the day of administration of first study medication. Ideally, the Baseline MRI would be performed as close to the first administration of study medication as possible to allow accurate measurement of disease status at Baseline by SPARCC and ASpiMRI-a (Berlin modification) score. However, several factors including rescreening and cross-screening of participants, the COVID-19 pandemic, and other unforeseen issues resulted in several participant MRI scans that were >3 weeks prior to the first administration of study medication. Because there is no generally accepted window for the definition of Baseline for MRI endpoints, two sets of participants were identified to include: 1) all eligible substudy participants, and 2) only participants with MRI's performed close to the first study drug administration. The threshold of 3 weeks prior to the first study drug administration was selected after consultation with imaging experts.

For spine x-ray, the Baseline value may be based on an x-ray performed up to 6 months prior to first administration of study drug.

3.4 Treatment Discontinuation Date and Intercurrent Event Date

The concept of intercurrent event is one of the estimand attributes as defined in Section 8.1.2 and Section 8.1.3. For non-responder imputation (NRI) analyses of binary efficacy endpoints, the intercurrent event date will be defined as the date of treatment discontinuation due to any reason. For multiple imputation (MI) analyses of binary efficacy endpoints, the intercurrent event date will be defined as the date of treatment discontinuation due to AE (including death) or lack of efficacy, while for the multiple imputation analyses of continuous and categorical efficacy endpoints (including the reference-based and hypothetical MI analyses), the intercurrent date will be defined as the date of treatment discontinuation due to any reason. Since treatment discontinuation date is not collected, the treatment discontinuation date is defined as:

> Treatment end date + dosing interval (28 days) + visit window (3 days for DBTP, 4 days for the MP)

With "Treatment end date" derived from the "Study Medication Discontinuation" CRF.

3.5 Protocol deviations

Important protocol deviations (IPD) are deviations from the protocol which potentially could have a meaningful impact on study conduct or on the primary efficacy, key safety, or PK outcomes for an individual study participant. The criteria for identifying IPD will be defined within the appropriate protocol-specific document. IPD will be reviewed as part of the ongoing data cleaning process and data evaluation. IPD will be identified and documented prior to unblinding to confirm exclusion from Per-Protocol Set (PPS) and Pharmacokinetics Per-Protocol Set (PK-PPS).

Since the PPS will be used for a supportive analysis of the primary endpoint which is assessed at Week 16, the exclusion from PPS is limited to the double-blind period, ie, only study participants with IPD or a decrease in dosing or dosing frequency of axSpA background medication due to intolerance/ AE/ side-effects, as permitted per study protocol, with potential impact on primary efficacy endpoint prior to and at Visit 8 (Week 16) can be excluded from the PPS. Study participants with IPD or a protocol-permitted decrease in dosing or dosing frequency of axSpA background medication due to intolerance/ AE/ side-effects, as permitted per study protocol, with potential impact on primary efficacy endpoint after Visit 8 (Week 16) will not be excluded from the PPS.

Also, an IPD will not always necessitate the removal of a study participant from the PPS. Depending on the type of protocol deviation, the study participant may remain in the PPS. The protocol deviation specification document developed by the clinical study team should clearly state which deviations will result in a removal from the PPS. See details in Section 3.6.6.

The impact of the COVID-19 pandemic on study procedures/conduct (eg, missed visits, remote visits, interruption of study treatment) will be documented using the information collected on a dedicated eCRF page.

3.6 Analysis sets

The primary efficacy variable will be analyzed for all study participants in the Randomized Set (RS), and supportive analyses of the primary efficacy variable will be performed on the Full Analysis Set (FAS) and the PPS. All other efficacy variables will be based on the RS

Demographics tables will be performed on the RS as well as on the Safety Set (SS), if SS is different from RS. Safety variables will be summarized on the SS. PK variables will be analyzed for all study participants in the safety set and/or PK-PPS.

The analysis sets are defined in the following sections.

3.6.1 Enrolled Set

The Enrolled Set (ES) consists of all study participants who have given informed consent.

3.6.2 Randomized Set

The Randomized Set (RS) consists of all randomized study participants.

3.6.3 Safety Set

The Safety Set (SS) consists of all randomized study participants who received at least one dose of the IMP.

3.6.4 Maintenance Set

The Maintenance Set (MS) will consist of all study participants who have received at least 1 dose of bimekizumab treatment in the Maintenance Treatment period.

3.6.5 Full Analysis Set

The Full Analysis Set (FAS) consists of all randomized study participants who received at least one dose of the IMP and have a valid measurement of all components of the primary efficacy variable at Baseline.

3.6.6 Per Protocol Set

The Per-Protocol Set (PPS) consists of all study participants in the RS who had no important protocol deviations affecting the primary efficacy variable. Important protocol deviations will be predefined and study participants with important protocol deviations will be evaluated during ongoing data cleaning meetings prior to unblinding of the data. See Section 3.5 for further details. Exclusions from the FAS will be considered important protocol deviations that also result in exclusion from the PPS. Additional exclusions from the PPS due to protocol-permitted decrease in dosing or dosing frequency of axSpA background medication due to intolerance/AE/side-effects may also be possible in case a potential impact on the primary endpoint cannot be excluded.

In addition, if after unblinding it is determined that there are study participants who were dosed with bimekizumab in place of placebo, then these study participants will be removed from PPS. Study participants who received a single dose with placebo in place of bimekizumab will remain in the PPS, but participants that received more than a single dose with placebo (or received one dose with placebo and also missed one or more additional doses, therefore fulfilling the IPD criteria of more than 1 missed dose up to week 12 during double-blind phase) when randomized to bimekizumab will be excluded from the PPS.

3.6.7 Pharmacokinetic Per-Protocol Set

The Pharmacokinetic Per Protocol Set (PK-PPS) consists of all randomized study participants who received at least one dose of the IMP and provided at least one quantifiable plasma concentration post-dose (after first IMP administration) without important protocol deviations that would affect the concentration of all samples obtained.

3.6.8 Immunogenicity Safety Set

ADAb and NAb analyses will use the safety set defined above. In the event that sample exportation in China is not approved, thereby preventing the NAb and ADAb analyses of samples of Chinese participants in the specified laboratories located outside of China, a separate immunogenicity safety set will be defined to include all randomized study participants, excluding China participants, who received at least one dose of the IMP.

3.6.9 COVID-19 Free Set

The COVID-19 Free Set (CFS) will consist of all study participants in the RS who had no COVID-19 impact on the primary efficacy endpoint up to Week 16. This will be defined as study participants neither having:

- a COVID-19 related IPD,
- nor a data entry on the COVID-19 impact eCRF,
- nor a COVID-19 related AE (See Section 10.2.4 for definition).

3.7 Treatment assignment and treatment groups

It is expected that study participants receive treatment as randomized for the Double-Blind Treatment Period, or as assigned per protocol study design (via reallocation) for the Maintenance Period.

Hence safety analyses will be based:

- on the SS as randomized, for the Double-Blind Treatment Period (displaying data under Placebo and BKZ 160mg Q4W columns),
- on the MS as assigned treatment, for the Maintenance Period (displaying data under BKZ 160mg Q4W Total column),
- on the SS as randomized/assigned treatment for the overall period (displaying data under BKZ 160mg Q4W Total column),

However, if after unblinding it is determined that study participants randomized to placebo received bimekizumab at any time during the Double-Blind-Treatment Period, then for safety analyses these study participants will be reallocated to the bimekizumab treatment group. Study participants randomized to bimekizumab will only be reallocated to the placebo treatment group if they never received bimekizumab. For AE analyses, AEs for each study participant will be summarized based on the treatment at the onset of each particular AE. AEs that occur on the day of first IMP administration will be allocated to the treatment received on that day. For study participants randomized on placebo who switch treatment per study design at Week 16 (ie, reallocation to bimekizumab treatment), AEs will be allocated to the treatment the study participant was on the day the event occurred. If the event occurs on the day of the switch, it will be attributed to the initial treatment (ie, Placebo), unless the AE fulfills the criteria for an

anaphylactic reaction in which case it will be attributed to the new treatment (ie, bimekizumab). Specifically, this includes events meeting any of the following criteria:

- Events that fulfil the anaphylactic reaction criteria for acute events (refer to Appendix 1)
- Events, that fulfil the hypersensitivity reaction criteria
- Events with an HLT of 'Administration site reactions NEC'
- Events with an HLT of 'Injection site reaction'

Efficacy analyses will be performed according to randomization and not actual treatment received. That is, if after unblinding it is determined that study participants randomized to placebo received bimekizumab at any time during the Double-Blind-Treatment Period efficacy analyses for these study participants will not be reallocated to the bimekizumab treatment group.

Hence efficacy analyses will be based:

- on the RS as randomized, for the Double-Blind Treatment Period (displaying data under Placebo and BKZ 160mg Q4W columns),
- on the RS as randomized/assigned treatment for the overall period (displaying data under Placebo/BKZ and BKZ 160mg Q4W columns).

The description of all efficacy analyses is detailed in Section 8.

3.8 Center pooling strategy

Geographic regions have been categorized as North America, Western Europe, Eastern Europe, and Asia. Table 3–1 displays the geographic regions with corresponding countries.

Table 3–1: Region definitions

Region	Countries
North America	United States
Western Europe	Belgium, France, Germany, Netherlands, Spain, United Kingdom
Eastern Europe	Bulgaria, Czech Republic, Hungary, Poland
Asia	China, Japan, Turkey

If the percentage of randomized study participants is less than 10% in a region, then regions will be combined to create a new geographic region stratum for efficacy modeling, so that there are no modeling convergence issues across efficacy variables. This new pooled geographic region stratum will then be used for any modeling (including MI, logistic regression and mixed model) including subgroup analyses.

3.9 Coding dictionaries

Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA®) version 19.0. Medications will be coded using the World Health Organization Drug Dictionary (WHO-DD) version SEP2020 or later. Medical procedures will not be coded.

To maintain consistency across studies in the bimekizumab development program, the dictionary versions will be held constant to support consistency of coding in the regulatory submissions.

3.10 Changes to protocol-defined analyses

The following changes from the protocol will be considered:

- The protocol mentions that subgroup analysis using descriptive statistics will be performed on the primary efficacy variable. However OR for the comparison of bimekizumab versus placebo and associated 95 % CI will be calculated. See Section 8.1.3 for further details.
- Race will be analyzed as additional subgroup variable.
- The Maintenance Set has been added as additional analysis set.
- The primary/main analysis of continuous secondary efficacy endpoints which are part of the sequential testing procedure, as well as the components of the primary ASAS40 endpoint, will use a reference-based imputation method.

3.11 Changes related to COVID-19

The impact of the COVID-19 pandemic on study procedures/conduct and on the primary efficacy endpoint and safety endpoints (TEAEs, serious TEAEs, and study withdrawal due to TEAEs) will be investigated and additional analysis outputs will be provided as appropriate. These additional analyses were not planned as part of the protocol as the pandemic was not ongoing at the time of protocol finalization.

These additional analyses will include analyses by period of the COVID-19 pandemic (prior/during/post) as defined in Section 3.2.2.2.

These additional analyses are described in the following sections:

- Study participant disposition (Section 5.1)
- Details of impacted visits and effects on collection and reporting of efficacy data (Section 5.3)
- Protocol deviations (Section 5.3)
- Exposure (Section 10.1.4 and Section 10.1.5)
- Adverse events (Section 10.2)

In addition, the primary analysis for the primary efficacy endpoint will be repeated by timing of the Week 16 Visit relative to the start and end of the COVID-19 pandemic (see Section 8.1.4.8).

4 STATISTICAL/ANALYTICAL ISSUES

4.1 Adjustments for covariates

The primary efficacy analysis that will investigate the treatment effect will be adjusted for:

- Prior TNFα inhibitor exposure (yes, no), and
- Region (Asia, Eastern Europe, North America, Western Europe)

However, prior TNFα inhibitor exposure may be dropped from the model in order to achieve model convergence (ie, in case the likelihood maximization algorithm failed to converge). If convergence is still not obtained, then region may instead or also be dropped from the model. This will be decided at the time of the table creation and clarified in respective table footnotes. Finally, in the case of the continuous secondary endpoints of the testing hierarchy, if the

reference-based multiple imputation model still does not converge, the analysis will be replaced using LOCF as imputation method (see Section 4.2.1.11)

Prior TNFα inhibitor exposure and region are the stratification factors of the randomization. In the event that a study participant is stratified in the incorrect stratum (ie, the stratum recorded in the Interactive voice or web Response System (IXRS) differs from the actual stratum the study participant belongs to), the actual stratum will be used for the analysis.

The secondary analyses will be adjusted on the same categorical factors as the primary analysis. For continuous variables, Baseline value (of the variable of interest) will also be included as covariate where appropriate.

4.2 Handling of dropouts or missing data

4.2.1 Handling of Missing data for efficacy analyses

Different approaches will be used to handle missing data and intercurrent events for all endpoints as summarized in Table 4–1 below. Details of the estimand structures and statistical analyses are described in Section 8 and details of the imputation methods are described in the following sections.

Table 4–1: Missing data handling methods for efficacy endpoints

Variable Priority	Variable Type	Composite Estimand (NRI)	Modified Composite Estimand (MI)	Tipping Point	Treatment Policy	Reference-based (MI)	Hypothetical Estimand (MI)	OC	LOCF
Primary	Binary	P	S[a]	S[a]	S[a]			S	
Secondary included in statistical testing procedure	Binary	P	S[a]					S	
	Continuous					P	S	S	P[g]
Secondary not included in statistical testing procedure	Binary	X	X					X	
	Continuous						X	X	
Other	Binary	X	X					X	
	Continuous					X[e]	X[d]	X[b]	
	Categorical						X[f]	X[c]	

P=Primary method, S=Supportive method, X=Method to be used (no priority designated).
Note: Composite estimand (NRI) refers to the approach in which data preceded by the intercurrent event of study treatment discontinuation for any reason are imputed as non-response, and other missing data are also imputed as non-response.
Note: Modified Composite Estimand (MI) refers to the approach in which data preceded by the intercurrent event of study treatment discontinuation due to AE (including death) or lack of efficacy are imputed as non-response, and other missing data are imputed via a multiple imputation model.
Note: Tipping Point refers to the approach in which missing data and data preceded by the intercurrent event of study treatment discontinuation for any reason are multiply imputed and then adjusted.
Note: Treatment Policy refers to the approach in which intercurrent events are not considered and all missing data are multiply imputed, and all other data are as observed.

Note: Hypothetical Estimand (MI) refers to the approach where outcomes for study participants without an intercurrent event of study treatment discontinuation are as observed, and outcomes for study participants with missing data or with an intercurrent event for any reason are imputed via a multiple imputation model.
[a] Imputation method is applied on continuous data, and binary variable is derived from the continuous variable based on imputed data set.
[b] Required only for by-visit summaries of variables whose value at Week 16 is part of the hierarchical testing procedure.
[c] For variables with multiple categories, data will be summarized as observed with an additional missing row to capture missing data at a given visit.
[d] MI will not be performed for the change from baseline in WPAI-SHP, change from baseline in EQ-5D-3L Utility Scores, change from Baseline in ASspiMRI-a (Berlin modification) score, or change from baseline in SPARCC score where only observed case analysis will be performed.
[e] Performed only for components of the primary endpoint.
[f] MI analysis will only be required for ASDAS status endpoint.
[g] Last Observation Carried Forward (LOCF) will only be used for the secondary continuous endpoints in the hierarchy if the Reference-Based Multiple Imputation analyses do not converge.

In this section, two terms will be defined for binary endpoints:

- Binary endpoints derived based on one continuous measurement (eg, BASDAI50 at Week 16) will be called 'non-composite binary endpoint'.
- Binary endpoints derived based on several continuous measurements (eg, ASAS40 at Week 16), will be called 'composite binary endpoint'.

4.2.1.1 Handling of Missing data for the Primary Endpoint

Because the ASAS40 composite binary primary endpoint is based on 4 different component scores, it is necessary to consider various data scenarios that could impact the calculation of the response. The rules described here are applicable in the context of the calculation of ASAS40 response and may differ from the rules applied for calculating and summarizing the components individually (some values may need to be imputed for component analysis but are not required here to evaluate the ASAS response). It is expected that cases using the following rules will occur infrequently, especially cases of 0 results at Baseline, as study participants are required to have moderate to severe active AS at study entry. The following rules will be applied prior to invoking any imputation analysis at the variable level (including observed case analysis), unless when specified differently:

- If a study participant has any component value that is equal to 0 at Baseline, then the following will be done:
  - If the post-Baseline value is equal to 0, then the absolute and percent improvement for that component will be treated as 0 for purposes of ASAS response calculations.
  - If the post-Baseline value is greater than 0, then the component is treated as having a 20% worsening and a 1-unit absolute worsening from Baseline for purposes of ASAS response calculation.
- If a study participant has any component value that is missing at Baseline, then the absolute and percent improvement for that component will be treated as 0 across all visits for purposes of ASAS response calculation.
- If at least 3 of the 4 components have observed data collected, the ASAS response will be calculated based on observed data. If 3 observed components show improvement, it is assumed that the 4th domain will not worsen.

For the primary analysis based on NRI (see Section 8.1.2 for details), if there are fewer than 3 components with observed data at a given post-Baseline visit, then the study participant will be classified as "non-responder", consistent with the pre-defined approach for handling intercurrent events defined in Section 8.1 For addressing missing components in supportive analyses, the following rules will be implemented for the primary endpoint ASAS40 only:

- Modified Composite Estimand - Multiple Imputation

  Multiple imputation will be applied on missing data, and data that are collected after the date of intercurrent event (IE) as defined in Section 3.4. If there are fewer than 3 components with observed data (considering all observed data after the IE date as missing), each of the components will be imputed separately before deriving the response; the available observed values and the imputed values will be used to derive the response. Study participants with IEs, will then be considered non-responders at all time points after the IE date. Analysis details are presented in Section 4.2.1.4.2 and Section 8.1.4.4.

- The tipping point approach:

  The tipping point approach implemented if the primary endpoint analysis result is statistically significant at alpha=0.05. The tipping point approach will first include a Worst Case imputation:

  – All data after the date of IE for treatment discontinuation due to any reason will be set to missing.

    - If at least 3 of the 4 components have observed data collected and where possible, the ASAS response/non-response will be calculated based on observed data.
    - If fewer than 3 components have observed data, then study participants treated with bimekizumab will be classed as non-responders, while study participants treated with placebo will be classed as responders.

    If the significant result remains, then no further analysis is necessary.

  – If the result changes as a result of the worst-case approach, then imputation with delta adjustments will be made to each missing component before deriving the ASAS endpoint to establish the ‘tipping point’ at which the result changes.

  A similar approach to that described above for the Modified Composite Estimand - Multiple Imputation will be implemented for the Delta Adjustment approach. However, study participants with treatment discontinuation will not be automatically considered non-responders at all timepoints after the IE date. Instead, missing data and data after the IE date for treatment discontinuation for any reason will be imputed and adjusted, and the imputed and adjusted data will be used, together with observed data, to derive the response. These approaches are described in detail in Section 4.2.1.8 and Section 8.1.4.7.

- The treatment policy strategy:

  A similar approach to that described above for the Modified Composite Estimand - Multiple Imputation will be implemented with the difference being that study participants with treatment discontinuation will not be set to missing prior to imputation and will not be automatically considered non-responders at all time points after the IE date. Specifically:

  - If at least 3 of the 4 components have observed data collected and where possible, the ASAS response/non-response will be calculated based on observed data.
  - If there are fewer than 3 components with observed (on and off treatment) data at a given post-Baseline visit, then multiple imputation methods will be applied, where any observed data after treatment discontinuation will be as observed (ie, NOT set to missing) and will not automatically be switched to non-response. Missing data for each of the components will be imputed separately before deriving the response based on imputed components. Details of this approach are presented in Section 4.2.1.9 and Section 8.1.4.5.

- The OC analysis:
  - If at least 3 of the 4 components have observed data collected and where possible, the ASAS response/non-response will be calculated based on observed data for study participants that are still on the original randomized treatment; data collected after the treatment discontinuation for any reason will not be included.
  - If there is partial missing data for ASAS and from the observed data is it clear that the study participant could not possibly be a responder, then the study participant will be coded as a non-responder and will be counted in the denominator of the OC analysis. For example, this could happen if:
    - there are 2 observed components and at least one shows worsening, or
    - there are 2 observed components and both show no response, or
    - there is only one observed component and it shows worsening.
  - In the other cases, the response will be missing and the study participant will not be included in the OC analysis. Details of this analysis are described in Section 4.2.1.10 and Section 8.1.4.6.

4.2.1.2 Handling of Missing Data for Secondary and Other Efficacy Endpoints

Missing data and intercurrent events for secondary and other efficacy endpoints will be handled as summarized in Table 4–1. Details of the imputation methods are described below and details of the statistical analyses and estimands are described in Section 8.2.2, Section 8.2.3, Section 8.3.2 and Section 8.3.3.

For composite binary secondary or other endpoints, the MI will be implemented on each component variable, before deriving the binary endpoint based on the imputed components (See Section 4.2.1.4.2). For each component, multiple imputation will use all visits where the individual component is planned to be collected. However, the binary endpoint will be derived (or presented in by-visit tables) only for those visits where **all** components are planned to be conducted per study protocol. In addition:

- ASAS20, ASAS-PR, and ASAS5/6 will follow the ASAS40 rules for when Baseline is equal to zero or baseline is missing as described in Section 4.2.1.1.
- For ASAS20, if at least 3 of the 4 components have observed data collected, the ASAS20 response will be calculated based on observed data. If 3 observed components show improvement, it is assumed that the 4th domain will not worsen.

- For ASAS5/6:
  - If at least 5 of the 6 components have observed data collected, the ASAS5/6 response will be calculated based on observed data. If 5 observed components show improvement, it is assumed that the 6th domain will not worsen.
  - If there are fewer than 5 components with observed data at a given post-Baseline visit, then imputation methods will be applied, consistent with the pre-defined approach for handling intercurrent events. See Section 8.2.1.1 for further details.

For multiple imputation of non-composite binary secondary or other endpoints, the standard multiple imputation (MI) methods (See Section 4.2.1.4.1) will be implemented on the raw score (eg, BASDAI for the BASDAI50 endpoint) before deriving the binary endpoint based on the imputed score.

4.2.1.3 Composite Estimand - Non-Responder Imputation (NRI) analysis

This is the primary analysis for the ASAS40 primary efficacy endpoint.

In the NRI analysis, study participants will be classified as 'non-responders' (ie, as if they have not responded to the treatment) at all time points after the date of discontinuation of study treatment for any reason.

In addition, for all endpoints except the ASAS response for which the response is still calculable when there is only 1 component with missing data (see Section 4.2.1.1), study participants will also be classified as 'non-responders' at all timepoints they had missing data.

4.2.1.4 Modified Composite Estimand - Multiple Imputation (MI)

This section describes the algorithms to be implemented for the MI – Markov-Chain Monte Carlo (MCMC)/Monotone Regression procedures for non-composite binary and composite binary endpoints.

Unless mentioned otherwise, all study participants from the population set of interest (eg, the randomized set) should be included in the MI steps.

4.2.1.4.1 Non-composite binary endpoints (Modified Composite Estimand)

For non-composite binary endpoints, the MI method will be applied as follows based on raw values:

- Step 1: Any observation after date of discontinuation of study treatment for any reason will be set to missing.
- Step 2: Create a dataset, one for each treatment group, of study participants with observed values and those needing estimation by multiple imputation. For the imputation step, missing values will include non-monotone missing values (ie, intermittent missing values between completed assessments) and monotone missing values (ie, where all study participant data is missing after a given time point).
  - **Step 2a**: First, the intermittent missing values in each dataset will be imputed using the MCMC method with multiple chains, monotone missing data imputing pattern, and non-informative priors for all parameters. Unless specified differently, the first 200 iterations will not be used (the 'burn-in' option). A total of 100 sets of imputations will be

performed. The seed used for these imputations will be 2017 and all other multiple imputation procedures described in this SAP will use this same seed as well.

- **Step 2b**: If an imputed value falls outside of the pre-defined range of values for the given variable, the value will be truncated to be within the predefined range of values for the endpoint of interest. For example, the imputed value for BASDAI will be updated to 0 or 10 in the case of an imputed value less than 0 or greater than 10, respectively (see Table 4–2).

The resulting 100 imputed data sets will have a monotone missing data pattern and will be imputed using a method for monotone missingness:

- **Step 2c**: For monotone missing data, monotone regression will be used to impute missing data. A separate regression model is estimated for each variable with missing values (ie, measurement at each time point). Based on the resulting model, a new regression model is then drawn and is used to impute the missing values for the variable. Since the data set has a monotone missing data pattern, the process is repeated sequentially for variables with missing values. The procedure will be based on the 100 imputed datasets generated from the MCMC procedure and will be performed by Imputation.
- **Step 2d**: If an imputed value falls outside of the pre-defined range of values for the given variable, the value will be truncated to be within the predefined range of values for the endpoint of interest.

The SAS® PROC MI procedure will be used for the imputation.

In both cases, the imputation model will include, in the following order: prior TNFα inhibitor exposure, region, the value at Baseline and at each post-Baseline visit (prior to the week of interest).

The visits to include in the MCMC and monotone regression steps will be the same, and will correspond to all data available at the time of the analysis, as specified below:

- For dry-runs conducted prior to the Week 24 interim analysis, all weeks up to Week 24 will be included,
- For the Week 24 interim analysis, all weeks up to Week 24 will be included,
- For the Week 52 interim analysis and final analysis, all weeks up to Week 52 will be included.

The stratification factors (prior TNFα inhibitor exposure and/or region) may be dropped from the imputation model to facilitate model convergence if required (and this holds true for all imputation models). The post-Baseline values will need to be specified in chronological order in the imputation model so that the SAS® PROC MI imputes variables from left to right (ie, the Week 1 value will be first imputed using regression based on Baseline value, and then the Week 2 value will be imputed using regression based on Baseline and Week 1 values, etc.). Study participants with missing Baseline will also be included in the imputation step.

The resulting datasets for each treatment arm will be combined into one complete dataset based on each of the 100 imputations (containing 100 times the number of study participants analyzed).

The imputation model based on the MCMC method for intermittent missing values will only allow multivariate normal variables. Therefore, region which has 4 levels will be re-coded as indicator variable:

One indicator variable (region1) will be defined as 0 for regions other than North America and 1 for North America. Two more indicator variables (region2 and region3) will be defined similarly replacing North America with Eastern Europe and Western Europe respectively. In the VAR statement of the imputation model, the indicator variables will be ordered as follows: region1 region2 and region3. In the eventuality that two regions required to be pooled due to a low percentage of study participants randomized in one region (see Section 3.8), the indicator variables defined above will be revised accordingly.

  - One indicator variable will be defined for the prior TNFα inhibitor exposure stratification variable: will be coded as 1 for yes and 0 for No.

  To maintain consistency with the MCMC method for intermittent missing values, these indicator variables will also be used in the monotone regression step for monotone missing values and treated as continuous variables.

- Step 3: For each of the 100 complete imputed data set obtained from Step 2, the binary responder variable will be derived. In practice, the value at the week of interest (eg, Week 16) in the imputed data sets will be used to categorize the study participant as a responder or not. If study participants have a treatment discontinuation due to AE or lack of efficacy, then the endpoint will be changed to "non-response" for all visits after the IE date.
- Step 4:
  - For secondary efficacy endpoints: For each of the datasets obtained in Step 3, the responder rate will be analyzed using a logistic regression model with factors of treatment group, region, and prior TNFα inhibitor exposure as fixed effects. The suitability of including region and prior TNFα inhibitor exposure at Baseline as fixed effects will be assessed using goodness-of-fit tests (Deviance and Pearson's and Hosmer-Lemeshow) and added if appropriate and allows model convergence.
  - For other efficacy binary endpoints: At each timepoint, the mean proportion of the responders and 95% CIs will be calculated from the imputed datasets using SAS® PROC MIANALYZE.
- Step 5 (for secondary efficacy endpoints only): The results obtained from the 100 logistic regression analyses in step 4 (ie, the adjusted responder rates for each treatment group and the corresponding 95% CI, the odds ratio of the treatment comparison and corresponding 95% CI) will be combined for overall inference using Rubin's rules, which account for the uncertainty associated with the imputed values (Rubin, 1987). This will be done using SAS® PROC MIANALYZE. The method for determining the degrees of freedom when combining results will be the freedom default Rubin's method.

4.2.1.4.2 Composite binary Endpoint (Modified Composite Estimand)

For composite binary endpoints, the MCMC/monotone regression method described above in Steps 1 to 3 for handling intermittent and monotone missing data will be performed separately for each individual component variable. As mentioned above, study participants with missing

Baseline will also be included in the imputation step. Refer to Section 4.2.1.1 for details on the handling of missing and zero Baseline for ASAS endpoints.

Then, based on the multiply imputed data sets obtained for each component, the binary response will be derived as follows:

- The 100 data sets obtained for each component will be merged by imputation number and study participant number.
- For each of the 100 complete imputed data sets, the binary endpoint will be derived for each study participant/visit based on the component values. In the case of the ASAS40 and ASAS20 where only one component value (among the 4) is missing at a visit, the response will be derived without using the imputed value of the missing component.
- If study participants have a treatment discontinuation due to AE or lack of efficacy, then the response associated to the binary endpoint will be changed to "non-response" for all visits after the IE date.

Procedures described for Step 4 to Step 5 will then follow.

For the primary endpoint, the 'burn-in' option in PROC MI will be set to the default value (200 in this case). The convergence of the MCMC will be assessed graphically, and the results will be provided in a SAS® output.

In addition, for ASDAS-MI endpoint, the following actions should also be taken:

- in step 2b and 2d, any imputed CRP values <0.5 * the LLOQ (0.1) will be reset to 0.05 after each MCMC and monotone steps to account for potentially different LLOQs across individuals,
- when determining ASDAS-MI, ASDAS-CRP values <2 should be reset to 2 only after the monotone regression imputation step (i.e., after Step 2d) prior to calculation of the composite score.

4.2.1.5 Calculation of adjusted responder rates, odds ratio and CIs for binary endpoints.

Estimates of the adjusted responder rates for each treatment group and the associated standard errors are obtained from the logistic regression in Step 4 on the logit scale and as such are assumed to follow a normal distribution. These estimates will be combined using Rubin's rules and the combined estimates and associated SEs will be used to construct 95% CIs on the logit scale. The combined estimates and 95% CIs on the logit scale will be back-transformed using the inverse logit link function to obtain the adjusted responder rates (%) and associated 95% CIs.

Some key points to consider relative to the calculation of the odds ratios and corresponding confidence intervals. As the estimates of the odds ratios from the logistic regression model in Step 4 follow a lognormal distribution, a log transformation is needed to normalize these 100 odds ratio estimates since the procedures for combining results from multiple imputed datasets are based on the assumption that the statistics estimated from each imputed dataset are normally distributed. Therefore, the log of the odds ratio estimates from the logistic regression model are used when combining into a single inference (Step 5). Additionally, the standard errors (SE) for the odds ratios are transformed as follows:

$$SE = \frac{\log(UCL) - \log(LCL)}{2Z_{\alpha/2}}$$

where UCL and LCL are the upper and lower confidence limit, respectively, for the CI of the odds ratio from the logistic regression model, and $Z_{\alpha/2}$ is the relevant critical value from the standard normal distribution (1.96 for a 95% CI). The estimates of the log odds ratio for bimekizumab relative to placebo and the corresponding upper and lower confidence limits will be provided. The odds ratio will be then estimated by exponentiating the estimate of the log odds ratio. The odds ratio and the confidence limits of the odds ratio will be estimated as follows:

$$OR = \exp(Log\ odds\ ratio\ estimate\ )$$

$$LCL = OR * \exp(-SE * Z_{\alpha/2}\ )$$

$$UCL = OR * \exp(SE * Z_{\alpha/2}\ )$$

where OR is the back-transformed estimate of the odds ratio just described, SE is the standard error of the log odds ratio and $Z_{\alpha/2}$ is the relevant critical value from the standard normal distribution (1.96 for a 95% CI). For mean proportion of responders or responder rates, the LCL will be truncated at 0. These calculations will be done such that odds ratios and corresponding CIs are calculated for the odds ratio of bimekizumab versus placebo.

4.2.1.6 Reference-Based Multiple Imputation (MI)

Reference-based MI is an imputation model developed based on data from the placebo group only (Mallinckrodt, 2013). It will be the primary method of analysis for the secondary continuous endpoints which are part of the testing hierarchy, and will be also used for the components of the primary efficacy variable.

Reference-based multiple imputation assumes that the statistical behavior of the bimekizumab and placebo-treated study participants resembles that of the placebo-treated study participants in the study. All time points after discontinuation of the double-blind study treatment for both the bimekizumab and placebo groups will be considered missing. Multiple imputations will be used to replace missing outcomes for bimekizumab- and placebo-treated study participants who discontinued using multiple draws from the posterior predictive distribution estimated from the placebo arm, including intermittent missing data point. Missing Baseline values will however be imputed using data from all study participants. In the case of continuous endpoints, the procedure will be implemented as follows on the raw values:

- Step 1: Data will be processed sequentially, one timepoint (visit) at a time, by repeatedly calling SAS® PROC MI to impute missing outcome data at visits t=1,…, Week 16 (Week 16 being the time point of interest) using data from all study participants for imputation of missing baseline data and using data from the placebo-treated study participants arm only for imputation of missing post-baseline data.
  - **Step 1a**: Impute missing values at Baseline using an imputation model containing prior TNFα inhibitor exposure, region and the baseline visit based on all study participants.
  - **Step 1b** *(Initialization)*: Set t=1 (Baseline visit).
  - **Step 1c** *(Iteration*): Set t=t+1. Create a dataset combining records from bimekizumab- and placebo-treated study participants with columns for covariates (prior TNFα inhibitor exposure and region) and outcomes at visits 1 to t. Outcomes for bimekizumab-treated

study participants with missing values at visit t are set to observed or previously imputed values at visits 1 to t-1. Bimekizumab-treated study participant with observed values at visit t are not included. Outcomes for placebo-treated study participants are set to observed at visit t or observed or previously imputed values at visits 1 to t-1. The outcomes should be sorted in chronological order in the model.

- **Step 1d** *(Imputation):* Impute missing values for visit t using previous outcomes for visits 1 to t-1, prior TNFα inhibitor exposure, and region. Note that only placebo data will be used to estimate the imputation model since no outcome is available for bimekizumab-treated study participants at visit t. As a consequence, the input dataset should include all study participants from placebo but only study participants from the bimekizumab arm that have values at timepoint t missing.
- **Step 1e**: Repeat steps 1a-1e for all timepoints, 100 times with different seed values (seeds ranging from 201 to 300) to create 100 imputed complete datasets. Study participants whose missing values were imputed in the last PROC MI call will be included in the input dataset for the next PROC MI call. Standard rounding rules will be applied to the imputed values. If the MI procedure yields a value outside of the pre-defined range for the given variable, the value will be updated to be within the predefined range of values for the endpoint of interest. For example, the imputed value for BASDAI will be updated to 0 or 10 in the case of an imputed value less than 0 or greater than 10, respectively (see Table 4–2).

- Step 2: The analysis will be done on the change from Baseline, and the model will be an ANCOVA model with treatment group, region, and prior TNFα inhibitor exposure as fixed effects and the Baseline value as covariate.

  For generation of summary statistics, the 100 imputed datasets from Step 1 will be combined, and simple means and standard errors will be calculated using Rubin's rules (via SAS® PROC MIANALYZE). For calculation of other descriptive statistics (median, Q1, Q3, minimum and maximum), Rubin's rules do not apply. Multiple imputation estimates will be computed by calculating arithmetic means of the estimates from the multiple repetitions of the imputation algorithm. Thus, for median, Q1, Q3, minimum and maximum the following approach will apply:

  - The data will be summarized by treatment, visit and imputation and the summary statistics will be computed.
  - Results will be summarized by taking the mean value of each summary statistic at each visit across all imputations.
  - The number of decimal places will remain the same as the original for display purposes (ie, if the mean was presented to 1 decimal place, the mean of the means will also be presented to 1 decimal place).

- Step 3: The results obtained from the 100 ANCOVA analyses in Step 2 (i.e. the Least Square Means for the treatment difference and the 95% CI for the contrasts) will be combined with Rubin's rules.

4.2.1.7 Hypothetical Estimand - Multiple Imputation (MI)

For secondary continuous efficacy endpoints, the methods described above in Steps 1 to 5 of Section 4.2.1.4.1 for the non-composite binary endpoints will apply based on the values at the timepoint of interest, with the following differences:

- The intercurrent event is study treatment discontinuation due to any reason and the IE date is as defined in Section 3.4.
- Step 3 is not relevant for continuous endpoints.
- In Step 4 (**Excluding hs-CRP**), the 100 imputed datasets from Step 2 will be combined, and simple means and standard errors will be calculated using Rubin's rules (via SAS® PROC MIANALYZE). For calculation of other descriptive statistics (median, minimum and maximum), Rubin's rules do not apply. Multiple imputation estimates will be computed by calculating arithmetic means of the estimates from the multiple repetitions of the imputation algorithm. Thus, for median, minimum and maximum the following approach will apply:
  - The data will be summarized by treatment, visit and imputation and the summary statistics will be computed.
  - Results will be summarized by taking the mean value of each summary statistic at each visit across all imputations.
  - The number of decimal places will remain the same as the original for display purposes (ie, if the mean was presented to 1 decimal place, the mean of the means will also be presented to 1 decimal place).
- In Step 4 **for hs-CRP only**: The CRP data will be presented using the geometric mean, median, Q1, Q3, minimum and maximum. The change from Baseline will be expressed as the ratio to Baseline in the by-visit summaries. The following approach will be applied:
  - Any imputed CRP values <0.5*LLOQ (0.1) will be reset to 0.05
  - The natural logarithm of the absolute values will be calculated
  - The change from Baseline (based on logged values) will be calculated
  - The change from Baseline (based on logged values) will be summarized by treatment, visit and imputation
  - The datasets will be combined using PROC MIANALYZE in order to get the mean for the absolute values and change from Baseline (based on logged data) across imputations
  - The estimates of the mean for the absolute values and changes from Baseline will be back-transformed to obtain the geometric mean on the original scale. This back-transformation of the mean change from Baseline (based on logged data) gives the geometric mean ratio to Baseline (on the original scale).
  - For the median, Q1, Q3, minimum and maximum the procedure outlined above for the other endpoints will be followed (based on logged values). And the resulting summary statistics will finally be back-transformed to obtain the values on the original scale.

- In Step 4, the analysis will be done on the change from Baseline, and the model will be an ANCOVA model with treatment group, region, and prior TNFα inhibitor exposure as fixed effects and the Baseline value as covariate.
- In Step 5, the results obtained from the 100 ANCOVA analyses in Step 4 (ie, the Least Square Means for the treatment difference and the 95% CI for the contrasts) will be combined with Rubin's rules.

For other continuous efficacy endpoints, the following steps will apply:

- The same Steps 1 to Step 5 as the ones described for non-composite binary efficacy endpoints (but excluding Step 3) will be implemented.
- For ASDAS-CRP endpoint only, the following should also be considered:
  - after each MCMC and monotone steps to account for potentially different LLOQs across individuals: any imputed CRP values <0.5 * the LLOQ (0.1) will be reset to 0.05.
  - any imputed CRP values <2 will be rest to 2 only one time after monotone regression imputation step and immediately prior to calculation of the component score.
- The same Step 4 and Step 5 as those described above for secondary endpoints (but excluding the "Step 4 **for hs-CRP only**") will be implemented.

4.2.1.8 Tipping Point analysis

Tipping point analyses will also be implemented as a supportive analysis for the primary efficacy endpoint only.

The objective of the tipping point analyses is to evaluate the sensitivity of results to departures from the missing at random assumption and to identify the point at which departures cause results "tip" from statistically significant to statistically insignificant. As such, these tipping point analyses will only be performed if the primary efficacy analysis results in a statistically significant treatment effect ($p<0.05$).

In this sensitivity analysis, data for study participants after date of discontinuation of treatment for any reason (as defined in Section 3.4) will be changed to missing prior to imputation but will not be changed to non-response after imputation.

As a first step, the worst-case scenario will be evaluated for the primary endpoint (ASAS40 at Week 16). All missing primary endpoint values for study participants randomized to BKZ (where missing values are due to, e.g., discontinuation of treatment or missing components of the primary endpoint) will be imputed as non-responders, while all missing values for placebo-randomized study participants will be imputed as responders. While there is little justification for such an approach, it makes the most putative assumption possible against a bimekizumab treatment effect. After applying this imputation approach, a logistic regression model consistent with the one described for the primary analysis will be applied. If the p-value for the odds ratio of bimekizumab versus placebo is less than 0.05, then no further tipping point analyses are needed.

If the analysis based on the worst-case scenario results in a p-value greater than 0.05, then additional tipping point analyses will be performed to identify the point at which results switch from significant to insignificant. In this analysis, a shift parameter or delta adjustment is applied

to missing, and subsequently imputed values (where missing values include observations after the date of discontinuation of treatment and any other missing values). These delta adjustments will be implemented on the Week 16 imputed values only for each component of ASAS40, as follows:

1. Data after date of discontinuation of treatment for any reason will be set to missing.
2. The same MCMC method described in Section 4.2.1.4.1 Step 2a and 2b will be implemented for non-monotone (intermittent) missing pattern values, using the same imputation model. This will be based on 100 sets of imputations.
3. Based on the 100 datasets obtained in Step 2b, a monotone regression model will be applied (using the same imputation model as in Step 2a) as described in Section 4.2.1.4.1 Step 2. This will be based on 1 imputation.
4. Delta adjustments will be made to all Week 16 imputed values only, independently in each treatment group as described in detail below.
5. Delta adjusted imputed values will be truncated so that they are within the range of allowable values for each component.
6. Following the delta adjustments for the individual components, the composite binary endpoint (ASAS40) will then be derived based on the multiply delta-adjusted imputed data sets obtained for each component as described in Section 4.2.1.4.1.

Each of the 100 imputed datasets will then be analyzed using a logistic regression model with factors of treatment group, region, and prior TNFα inhibitor exposure as fixed effects.

7. The results obtained from the 100 logistic regression analyses in Step 6 will be combined for overall inference using Rubin's rules, and the results obtained for each shift parameter will be presented in a single table.
8. Step 3 to Step 7 will be repeated so that, at each iteration, missing values are adjusted with a larger delta than at the previous iteration. Depending on the results obtained, shift parameters with more granularity (eg, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9) may also be investigated. The process will go on until the p-value for the odds ratio between bimekizumab and placebo is no longer statistically significant (ie, ≥ 0.05). The odds ratio, 95% CI, and p-values obtained for each value of delta will be combined in one single table.

The delta adjustments performed here assume that study participants randomized to bimekizumab and who have missing data have a lower probability of response compared to study participants randomized to placebo with missing data. Therefore, for endpoints where high scores are associated with a more favorable outcome:

- A negative shift is applied to the imputed value for study participants randomized to bimekizumab in order to decrease the imputed value and reduce the likelihood of response.
- A positive shift is applied to the imputed value for study participants randomized to placebo in order to increase the imputed value and increase the likelihood of response.

For endpoints for which high scores are associated with a less favorable outcome:

- A positive shift is applied to the imputed value for study participants randomized to bimekizumab in order to increase the imputed value and reduce the likelihood of response.

- A negative shift is applied to the imputed value for study participants randomized to placebo in order to decrease the imputed value and increase the likelihood of response.

To start, imputed values within each continuous component, will be adjusted by the same integer value in each treatment arm. Imputed values will initially be adjusted by values of 1, 2, and 3. After these initial analyses, additional adjustment values may be tested to better identify the point at which results "tip". More robust primary analysis results will require larger adjustments to tip the results from significant to insignificant.

4.2.1.9 Treatment Policy Strategy

Another supportive analysis will be performed on the primary efficacy endpoint to address intercurrent events.

The treatment policy strategy will include all available data observed at the week of interest (Week 16) regardless of the occurrence of intercurrent events. This means the analysis includes on- and off-treatment values collected after study participants prematurely discontinued study treatment but agreed to remain on the study and continued to attend visits and provide assessments at those visits. Those observed values will be analyzed according to the study participant's randomized treatment. Study participants for whom efficacy data cannot be obtained at the week of interest, despite attempts to retain them in the study, will have their data imputed using MI under the assumption of MAR (ie, MI-MCMC/Monotone Regression).

In practice, the Step 1 for the non-composite binary endpoints described in Section 4.2.1.4.1 will not apply, and the final ASAS40 response will not be changed to "non-responder" at the end of the imputation step (ie, Step 4) for study participants having an intercurrent event.

4.2.1.10 Observed Cases analysis

The observed case analysis only applies where study participants are still on the original randomized treatment. Therefore, data collected after the date of study treatment discontinuation (for any reason) will not be considered in observed case summaries.

In addition, for all endpoints except the ASAS response for which the response is still calculable when there is only 1 component with missing data (see Section 4.2.1.1), all other missing data will also be excluded.

4.2.1.11 Last Observation Carried Forward

The LOCF analysis will only apply for the continuous secondary efficacy endpoints that are part of the testing hierarchy, if the reference-based multiple imputation analysis does not converge (even after having dropped the covariates).

Post-Baseline missing data will be imputed by carrying forward the last available observation before the missing visit. Baseline data can be carried forward.

If a study participant has a value that is missing at Baseline, then the study participant will not be included in the analysis.

4.2.1.12 Allowed ranges for continuous efficacy variables

The table below indicate the possible range of values for each continuous efficacy variable:

Table 4–2: Allowable ranges for continuous efficacy variables

Variable	Minimum	Maximum
BASDAI	0	10
BASFI	0	10
BASMI	0	10
Lateral spinal flexion	0.75	25
MASES	0	13
PGADA	0	10
Total NSP	0	10
SF-36 MCS[a]	-3.33	80.09
SF-36 PCS	5.02	79.78
SF-36 Bodily Pain	21.68	62
SF-36 General Health	18.95	66.5
SF-36 Mental Health	11.63	63.95
SF-36 Physical Functioning	19.26	57.54
SF-36 Role Emotional	14.39	56.17
SF-36 Role Physical	21.23	57.16
SF-36 Social Functioning	17.23	57.34
SF-36 Vitality	22.89	70.42
ASQoL	0	18

[a] SF-36 ranges are used for imputation purposes only. Any observed values outside the ranges based on norm-based scores will be reported as calculated in all tables and listings.

4.2.2 Handling of missing data for AE

For analyses of AEs, a complete date must be established to correctly identify the AE as occurring during treatment or not. For purposes of imputing missing components of partially reported start and stop dates for AEs, the algorithms listed below will be followed. In the event of ambiguity or incomplete data which makes it impossible to determine whether the AE was treatment emergent, the AE will be considered treatment emergent. If an imputed date occurs after a study participant's date of death, the imputed date will be set to the date of death.

Although the algorithms for treatment-emergence depend on the onset date, imputation rules are provided for stop date as well, as these may be needed for certain statistical analyses, such as an analyses of AE prevalence or AE duration.

Start and stop dates of AEs will be displayed as reported in the study participant data listings (ie, no imputed values will be displayed in data listings).

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 AE 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 AE onset, then use the date/time of first dose
- If only the month and year are specified, and the study participant did switch treatment during that month and year, then use the date of first switch treatment.
- If only the year is specified, and the year of first dose is not the same as the year of AE onset, and if the study participants did not switch from placebo to bimekizumab at Week 16, then use January 1 of the year of AE onset
- If only the year is specified, and the year of first dose is the same as the year of AE onset, then use the date/time of first dose
- If only the year is specified, and the year of first dose is the same as the year the study participant did switch treatment, then use the date of first switch treatment.
- If only the year is specified, and the year of first dose is the same as the year of the AE onset and the same as the year the study participant did switch treatment, then use the date of first dose.
- If the imputed stop date is prior to the imputed start date
  - For missing start day and start month:
    - If the year of start date is the same as the year of first dose and the imputed stop date is on or after the date of first dose then set the start date to the date of first dose
    - Otherwise set to the 1st January of the year of the start date
  - For missing start day only
    - If the month and year of the start date is the same as the month and year of first dose and the imputed stop date is on or after the date of first dose then set the start date to the date of first dose
    - If the month and year of the start date is the same as the month and year of first dose and the imputed stop date is before the date of first dose then set the start date to the 1st of that month
- If the AE onset date is completely unknown, then use the date of first dose

Imputation of Partial Stop 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 31st of that year
- If the AE is resolved and the stop date is completely unknown, then do not impute the stop date

Other imputation

In addition, the following will apply for presenting AE in summary tables:

- If the intensity of an AE is unknown, it will be considered as severe.
- If the relationship to study drug is missing, it will be considered as related.

For seriousness, no imputation rule will be applied for the interim analyses, but the worst-case approach will be applied for the final analysis (ie, the missing seriousness will be considered as serious if missing).

4.2.3 Handling of missing data for prior and concomitant medications

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. In the event of ambiguity or incomplete data which makes it impossible to determine whether a medication was concomitant or not, the medication will be considered as concomitant. Such imputations will only be performed for these classifications and calculations; in the listings all data will be shown as recorded on the electronic case report form (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, and the study participant did not switch treatment during that month and year, 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 this is leading to a start date after the stop date of then use the 1st of the month
- If only the month and year are specified, and the study participant did switch treatment during that month and year, then use the date of first switch treatment.
- If only the year is specified, and the year of first dose is not the same as the year of the start date and the study participant did not switch treatment during that year, 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 start date is completely unknown, and the stop date is unknown or not prior to the date of first dose, then use the date of first dose
- If the imputed stop date is prior to the imputed start date
  - If the year of start date is the same as the year of first dose and the stop date is after the date of first dose then set the start date to the date of first dose
  - Otherwise set to the 1st January of the year of the start date

Imputation of Partial Stop 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 stop date is completely unknown, do not impute the stop date

4.3 Interim analysis and data monitoring

4.3.1 Interim analyses

Two interim analyses will be conducted:

- The first interim analysis will be conducted after the planned number of randomized study participants have completed the Double-Blind Treatment Period and the 24 week assessments or have withdrawn from IMP or the study. The purpose of this analysis is to prepare a regulatory submission for a Marketing Authorization Application based on this analysis.
- The second interim analysis will be conducted after the planned number of study participants have completed Week 52 or have withdrawn from IMP or the study. The purpose of this second interim analysis is to perform a comprehensive evaluation of all available data for the study and to supplement the regulatory submission based on the Week 24 interim analysis.

The database for the Week 24 interim analysis will include all available data collected up to the day the last planned study participant completed the Week 24 Visit (even if this study participant has withdrawn from IMP). Study participants who discontinued the study prior to Week 24 will be included in this first interim analysis.

The Week 24 interim analysis will evaluate the primary, secondary and other efficacy variables, as well as the safety and PK/immunogenicity variable results. The disposition, demographics and other Baseline characteristics, protocol deviations and compliance will also be described.

The efficacy analyses will present data until Week 24, and the safety analyses will include all available data collected up to the day the last planned study participant completed the Week 24 Visit. This may include safety data beyond the Week 24 timepoint. Events on-going at the time of the interim analysis (eg, concomitant medications, or adverse events) will be analyzed as they are collected to date (ie, up to Week 24 of last participant for safety endpoints), and analyses based on these events will be revised at time of the second interim analysis and/or final analysis.

For this first interim analysis, a database freeze will be performed, and the treatment codes will be made available to the UCB personnel. An interim CSR report will be written to include efficacy, safety, PK, and ADAb analyses. NAb analyses are not planned for inclusion in the interim CSR report. The investigators and study participants will remain blinded to the assigned bimekizumab dosing regimen until the last study participant completes the study. A blinding maintenance plan will be written to evaluate the potential bias of the Maintenance Period and describe the process of interim results generation and dissemination. The plan will be finalized prior to the freeze of the database at the Week 24 interim analysis.

The Week 52 interim analysis will be conducted similarly as the first interim analysis: the database will include all available data collected up to the day the last study participant completed the Week 52 Visit (even if this study participant has withdrawn from IMP). Study participants who discontinued the study prior to Week 52 will be included in this second interim analysis. The efficacy analyses will present data until Week 52, and the safety analyses will include all available data collected up to the day the last study participant completed the Week 52 Visit.

No formal alterations to the further study conduct (eg, stopping rules, sample size re-estimation, or changes to eligibility criteria) are planned for the two interim analyses.

No separate SAP for the interim analyses will be provided. The TFL shells for the interim and final analyses will be provided in the same document.

The type of efficacy and safety analyses to be provided for the two interim analyses is detailed in Table 4–3.

Table 4–3: Summary of efficacy and safety analyses for each interim analysis

Variable Type of tables	Analysis Sets (study period)	Treatment groups	Interim Analysis 1	Interim Analysis 2
Primary and secondary efficacy variables				
Tables on primary efficacy variable	RS, FAS, PPS	• Placebo • BKZ 160mg Q4W	Yes	Yes, rerun of tables provided for interim analysis 1
Tables on secondary efficacy variables	RS	• Placebo • BKZ 160mg Q4W	Yes	Yes, rerun of tables provided for interim analysis 1
Other efficacy variables				
Tables/graph on time to ASAS response	RS (DB-period only)	• Placebo • BKZ 160mg Q4W	Yes, for all visits up to Week 16 for each participant	Yes, rerun of analyses provided for interim analysis 1
By-visit tables	RS	• Placebo • BKZ 160mg Q4W	Yes, for all visits up to Week 24 for each participant	Yes, for all visits up to Week 52 for each participant
Safety Variables				
Tables of adverse events	SS (DB-period only)	• Placebo • BKZ 160mg Q4W	Yes, for all visits up to Week 16 for each participant	No
	MS (MP only)	• BKZ 160mg Q4W Total	No	Yes, from Week 16 up to Week 52 for each participant
	SS (Overall period)	• BKZ 160mg Q4W Total	Yes, for all visits up to date of Week 24 visit of last planned participant	Yes, for all visits up to date of Week 52 visit of last planned participant
By-visit tables on laboratory or VS data	SS (Overall period)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W	Yes, for all visits up to date of Week 24 visit of last planned participant	Yes, for all visits up to date of Week 52 visit of last planned participant

Variable Type of tables	**Analysis Sets (study period)**	**Treatment groups**	**Interim Analysis 1**	**Interim Analysis 2**
Shift table or table on TEMA for laboratory data or Hy's Law cases	SS (DB-period only)	• Placebo • BKZ 160mg Q4W	Yes, for all visits up to Week 16 for each participant	No
	MS [a] (MP only)	• BKZ 160mg Q4W Total	No	Yes, from Week 16 up to Week 52
	SS [a] (Overall period)	• BKZ 160mg Q4W Total	Yes, for all visits up to date of Week 24 visit of last planned participant	Yes, for all visits up to date of Week 52 visit of last planned participant
Table on TEMA for VS data	SS (DB-period only)	• Placebo • BKZ 160mg Q4W	Yes, for all visits up to Week 16 for each participant	No
	MS [a] (MP only)	• BKZ 160mg Q4W Total	No	Yes, from Week 16 up to Week 52 for each participant
	SS [a] (Overall period)	• BKZ 160mg Q4W Total	Yes, for all visits up to date of Week 24 visit of last planned participant	Yes, for all visits up to date of Week 52 visit of last planned participant
Tables on ECG or tuberculosis data	SS (Overall period)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W	Yes, for all visits up to date of Week 24 visit of last planned participant	Yes, for all visits up to date of Week 52 visit of last planned participant
Tables on PHQ-9 data	SS (Overall period)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W	Yes, for all visits up to date of Week 24 visit of last planned participant	Yes, for all visits up to date of Week 52 visit of last planned participant
Tables on eC-SSRS data	SS (DB-period only)	• Placebo • BKZ 160mg Q4W	Yes, for all visits up to Week 16 for each participant	No

Variable Type of tables	**Analysis Sets (study period)**	**Treatment groups**	**Interim Analysis 1**	**Interim Analysis 2**
	MS (MP only)	• BKZ 160mg Q4W Total	No	Yes, from Week 16 up to Week 52 for each participant
	SS (Overall period)	• BKZ 160mg Q4W Total	Yes, for all visits up to date of Week 24 visit of last planned participant	Yes, for all visits up to date of Week 52 visit of last planned participant

[a] For placebo-randomized study participants, results are determined based on the Baseline Week 16 (See Section 3.3).
Study participants in the Placebo/ BKZ 160mg Q4W treatment group are taking BKZ 160 mg Q4W treatment after Week 16.

The final analysis will occur once last study participant has completed the SFU Visit. The final analysis will consist of a rerun of all listings and all safety summary analyses created for the Week 52 interim analysis but will include data occurring during the SFU period that were not available for the Week 52 interim analysis.

4.3.2 Data Monitoring Committee

An independent Data Monitoring Committee (DMC) will be formed to monitor the ongoing safety of the study through period review. The general scope of DMC activities is presented in the protocol, and the composition and operation of the DMC will be defined in the DMC charter. The presentation and analysis of data for the DMC meetings will be described within a separate DMC SAP.

4.4 Multicenter studies

The data from all centers will be pooled for the purposes of the analysis. Centers will be grouped in the geographic regions North America, Western Europe, Eastern Europe, and Asia.

No exploration of treatment by center interaction will be investigated.

4.5 Multiple comparisons/multiplicity

A fixed sequence testing procedure will be applied for the primary endpoint and the key secondary endpoints. The testing procedure will account for multiplicity and control the family-wise type I error rate at alpha=0.05 (2-sided).

For each test, on each binary efficacy endpoint, the null hypothesis is that the conditional odds ratio is equal to one (H_0: $OR_{T1T2} = 1$). The alternative hypothesis is that the conditional odds ratio is not equal to one (H_A: $OR_{T1T2} \neq 1$).

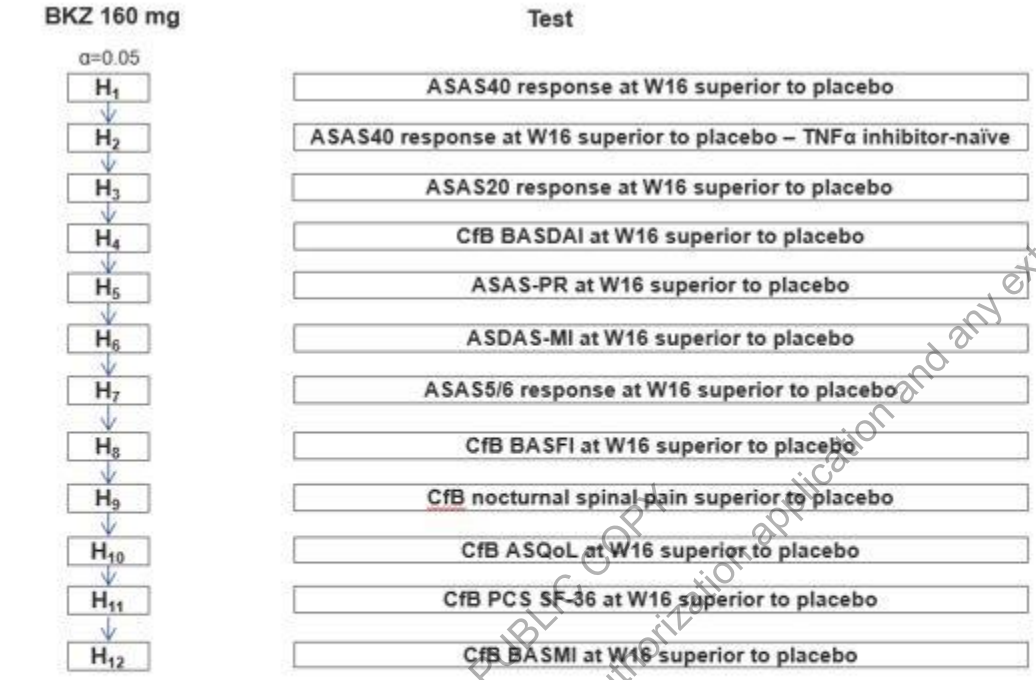
For each test, on each continuous efficacy endpoint, the null hypothesis is that there is no difference between treatment groups (H_0: $T_1 - T_2 = 0$). The alternative hypothesis is that there is a difference between treatment groups (H_A: $T1 - T2 \neq 0$).

T_1 refers to bimekizumab and T_2 to placebo.

According to this strategy, the statistical testing of an endpoint can be investigated only if the null hypothesis for the previous endpoint has been rejected (ie, if $p<0.05$).

Figure 4–1 shows the testing order for these endpoints.

Figure 4–1: Sequential testing procedure of primary and key secondary efficacy endpoints



ASAS40 (20)=Assessment of SpondyloArthritis International Society 40% (20%) response criteria; ASAS5/6=Assessment of SpondyloArthritis International Society 5 out of 6 response criteria; ASAS-PR=Assessment of SpondyloArthritis International Society partial remission; ASDAS-MI=Ankylosing Spondylitis Disease Activity Score major improvement; ASQoL=Ankylosing Spondylitis Quality of Life; BASDAI=Bath Ankylosing Spondylitis Disease Activity Index; BASFI=Bath Ankylosing Spondylitis Functional Index; BASMI=Bath Ankylosing Spondylitis Metrology Index; BKZ=bimekizumab; CfB=change from Baseline; H=hypothesis; PCS=physical component summary; SF-36=Short Form 36-item Health Survey; TNFα=tumor necrosis factor alpha; W=Week

4.6 Use of an efficacy subset of study participants

Analysis of the primary endpoint will be performed based on the RS and repeated for all study participants in the PPS, and for all study participants in the FAS as supportive analyses. The FAS analysis will evaluate whether there are differences in the efficacy analysis between randomized study participants and randomized study participants with a Baseline assessment, while the PPS analysis will evaluate the effect of IPDs and other possible exclusions on the analysis.

4.7 Active-control studies intended to show equivalence

Efficacy outcomes are based on superiority comparisons of the bimekizumab 160mg sc Q4W to placebo, this section is then not applicable for this study.

4.8 Examination of subgroups

Subgroups analyses will be performed for the primary endpoint ASAS40 and ASDAS-MI as listed in Table 4–4. In addition, ASAS40 will be analyzed based on the timing of participant

enrollment and timing of the Week 16 visit relative to the COVID-19 pandemic periods. The complete list of subgroups is listed in Table 4–4.

Furthermore, most of the analyses defined in this SAP will be repeated for the China total study population only (and respectively for the Japanese population only) in order to meet the China (and respectively the Japan) submission requirements. The details of the analyses to be produced for these submissions will be indicated in the TFL shell document. For the analysis based on MI, all study participants will be first included in the imputation model, and then the analysis step will be restricted to the subset of China (or respectively Japanese) subjects.

Table 4–4: Categories of variable for efficacy subgroup analyses

Subgroup	Categories
Age (years)	<45, ≥45
Gender	Male, Female
Race 1	Black, White, Other
Race 2	White, Asian Other
Region 1 [a]	Asia, Eastern Europe, North America, Western Europe
Disease duration (years)	<2, ≥2
Body mass index (BMI) (kg/m^2)	<18.5, ≥ 18.5 to <25, ≥ 25 to <30, ≥30
hs-CRP level	≤upper limit of normal (ULN) [b], >ULN
Prior TNFα inhibitor exposure [c]	Yes, No
csDMARDs	Yes, No

Subgroup	Categories
ASDAS status	<1.3 [inactive disease], 1.3 to ≤2.1 [low disease activity], >2.1 to ≤3.5 [high disease activity], >3.5 [very high disease activity]
HLA-B27 positivity [d]	Yes, No
Timing of study participant enrollment relative to COVID-19 pandemic periods as defined in Section 3.2.2.2	Enrolled prior to the COVID-19 pandemic, Enrolled during the COVID-19 pandemic, Enrolled after the COVID-19 pandemic
Timing of Week 16 Visit relative to the COVID-19 pandemic periods as defined in Section 3.2.2.2	Study participants who had the Week 16 Visit prior the COVID-19 pandemic, Study participants who had the Week 16 Visit during the COVID-19 pandemic, Study participants who had the Week 16 Visit after the COVID-19 pandemic

[a] The categories may be pooled as defined in Section 3.8.
[b] ULN refers to the central laboratory ULN definition (ie, 5mg/L).
[c] The actual TNF alpha stratum the study participant belongs to will be used for the subgroup analysis.
[d] HLA-B27 = human leukocyte antigen B27.

5 STUDY POPULATION CHARACTERISTICS

5.1 Study participant disposition

Study participant disposition will be summarized for all study participants screened overall, by region and by site. In this summary, the dates of first study participant screened and date of last study participant last visit, number of screened study participants, and the number of study participants included in each analysis set (ES, RS, SS, MS, FAS, PPS, PK-PPS, COVID-19 free set) will be presented.

The disposition of study participants into treatments groups and analysis sets (RS, FAS, SS, and PPS, COVID-19 free set) will also be summarized on the RS.

Reasons for screen failures (as collected on the Study Termination Screen Failure CRF page) will be summarized for all screened study participants who failed to be randomized.

The number and percentage of randomized study participants who completed the study, who completed the SFU, and who discontinued the study will be tabulated overall on the randomized set and by treatment group. The primary reason for study discontinuation as collected on the Study Termination CRF page will be provided in this table. The numbers and percentages of study participants entering and completing, or who remain ongoing for each of the study periods (16-Weeks Double-Blind Treatment Period, 36-Weeks Maintenance Period) will additionally be presented in this table. The number of study participants who completed the Double-Blind Treatment Period but discontinued prior to entering the Maintenance Period will be displayed in this table. The numbers and percentages of randomized study participants entering the extension study will also be presented in this table.

The disposition and study discontinuations reasons will additionally be provided on the RS up to Week 24, as well as on the maintenance set.

Study participants who prematurely discontinued study medication and are continuing to attend their scheduled visits through the end of each study period (DBP or MP) will also be considered as having completed the period. However, they will be labeled as 'completed while not on randomized treatment' in disposition and study discontinuation summaries. The primary reason for study medication discontinuation (respectively during the Double-Blind Treatment Period or Maintenance Period) as collected on the Study Medication termination CRF page will be summarized in separate tables.

The number of randomized study participants who discontinued the study due to AEs (during the Double-Blind Treatment Period or Maintenance Period) will be summarized for EudraCT reporting.

Finally, the number of study participants enrolled during each COVID-19 pandemic period, as well as the number of study participants still in the study during each COVID-19 pandemic period will be summarized in one table on the randomized set. The table will be done overall and by region. Refer to Section 3.2.2.2 for definition of the pandemic periods.

Study disposition and termination details will be listed for all study participants screened. Additional listings will be created on study discontinuation, study participant analysis sets and visit dates. A listing will also be provided for study participants who did not meet the study eligibility criteria. A by-study participant will be said to have completed the study if she/he had completed the last scheduled study visit, not including SFU visits.

5.2 Impact of COVID-19

A listing of visits affected by COVID-19 will be presented based on the enrolled set including the visit, date of visit, relationship to COVID-19, impact category, and a narrative (short description) of the event. These data will be presented by treatment group as well as by region. A supportive summary table will also be provided on the randomized set by treatment group. The table will be done overall and then repeated by country.

In addition, in order to assess the potential impact of COVID-19 on the collection and reporting of efficacy data, two summary tables on the RS will be presented to display missing data and data collected via an alternative modality (eg, phone interview, video call), respectively, for efficacy endpoints included in the testing hierarchy (Figure 4–1). For the purpose of these displays, missing data will be presented only for visits affected by COVID-19, as reported on the dedicated eCRF page. Missing data at other visits and for other reasons will not be included.

Some assessments must be conducted in person and cannot be assessed when a visit is completed remotely (eg, BASMI, TJC or SJC) and therefore these assessments will be missing for those visits.

In addition, for any missed visit or a visit conducted remotely, the CRP assessment will also be missing. Such assessments will be considered to be missing as a result of COVID-19. For these visits, it will therefore not be possible to assess ASDAS-CRP.

A listing will be presented showing each impacted visit, based on the Enrolled Set. This will include the planned visit, the visit date, and details of the assessments conducted remotely and/or those missing as a result of COVID-19.

These data will be presented in a summary table by treatment group for the RS.

For both the listing and the summary table, only visits at which efficacy assessments are scheduled will be included.

5.3 Protocol deviations

The definition of an IPD is given in Section 3.5.

A first summary displaying the number and percentage of study participants with an important protocol deviation will be provided on the RS by treatment group and overall. The summary will be overall (any IPD) and by type of deviation (inclusion criteria deviation, exclusion criteria deviation, withdrawal criteria deviation, prohibited concomitant medication, incorrect treatment or dose, treatment non-compliance, procedural non-compliance). The IPDs will also be classified as COVID-19 related or not. This summary will be presented on the Double-Blind Treatment Period and overall. A second summary displaying the number and percentage of study participants excluded from the PPS and PK-PPS will be provided on the Double-Blind Period, on the RS, by treatment group and overall. For the PPS exclusion, the table will provide the breakdown between study participants who are excluded due to IPD from study participants who are excluded because of a decrease in dosing or dosing frequency of axSpA background medication due to intolerance/ AE/ side-effects as permitted per study protocol with potential impact on primary efficacy endpoint.

Criteria for exclusion of study participants from the PPS or PK-PPS will be defined in a separate document.

A by-participant listing of important protocol deviations will be provided for all study participants in the RS.

A separate listing of all COVID-19 related PD will be presented based on the RS.

A listing of study participants excluded from PPS because of a decrease in dosing or dosing frequency of axSpA background medication (due to intolerance/ AE/ side-effects as permitted per study protocol) will be provided on the RS.

A listing of study participants excluded from COVID-19 free set and a listing of study participants excluded from PK-PPS will also be provided on the RS.

6 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Unless otherwise specified, all summaries will be based on the SS and repeated using the RS. If the SS and RS analysis sets are identical, the summaries will not be repeated. Some summaries will also be presented on the MS.

6.1 Demographics

The following demographic variables measured at Baseline will be summarized overall and by treatment group: age (years), gender, and country, race, ethnicity, height (cm), body weight (kg) and BMI (kg/m2). Age, body weight and BMI will be summarized as continuous variables and as categorical variables.

Any age values collected in the database it will be recalculated for the analysis. The recalculation should use the year of birth and the year of randomization. Age calculation for non-randomized study participants will not be done. For age, three sets of categories will be defined:

- ≤18, 19 to <65, ≥65 years (clinicaltrials.gov requirement),
- 18 to <65, 65 to <85, ≥85 years (EudraCT requirement),
- <45, ≥ 45 years

Even if BMI is available in the database, BMI (in kg/m2) will be recalculated during analysis based on height (in m) and weight (in kg) values collected in the database, and the calculated values will be used in the statistical analysis since they are considered more accurate. The formula for BMI (kg/m2) calculation is:

$$\text{BMI} = \frac{\text{Weight (kg)}}{(\text{Height (m)})^2}$$

BMI categories will be:

- <25, 25 to <30, ≥30 kg/m2
- <18.5, 18.5 to <25, 25 to <30, ≥30 kg/m2

For body weight, the following categories will be used:

- ≤100, >100 kg
- <70, 70 to <95, 95 to < 115, ≥115kg

In addition, the number and percentages of study participants in each stratum for each region stratification factor will be included in the demographic summary table, as per the information collected in the Stratification CRF page. The actual information on the prior use of TNFα inhibitor at Baseline will also be reported.

To understand the study population during the different periods of the COVID-19 pandemic, the reporting of Baseline demographics will be repeated by COVID-19 enrollment period (prior/during/post) on the SS. Refer to Section 3.2.2.2 for definition of the periods.

A separate frequency table will summarize the study participant's lifestyle on the SS only.

By-study participant listings on demographics and study participant's childbearing potential data will be provided for all screened study participants. Listing on lifestyle will be provided on the RS.

6.2 Baseline characteristics

6.2.1 Baseline efficacy variables

The ASAS components at Baseline (PGADA, total spinal pain NRS score, BASFI, mean of the BASDAI questions 5 and 6) will be summarized descriptively overall and by treatment group.

BASDAI spinal pain (question 2 only) and BASDAI total score will also be presented in this table, as well as hs-CRP (mg/L) and ASDAS-CRP. Hs-CRP and ASDAS-CRP will be summarized as continuous variables and as categorical variables and the same categories as the ones defined in Table 4–4 for the efficacy subgroup analyses will be used.

Total spinal pain will be derived from question 1 of the total and nocturnal spinal pain questionnaire.

To understand the study population during the different periods of the COVID-19 pandemic, the reporting of Baseline efficacy characteristics will be repeated by COVID-19 enrollment period (prior/during/post) on the SS.

6.2.2 Other baseline characteristics

The following categorical variables will also be summarized:

- HLA-B27 (positive, negative)
- Past anti-TNF therapy (yes, no)
- Any past biologic therapy (yes, no)
- Past biologic therapy other than TNF inhibitor (yes, no)
- Current nonsteroidal anti-inflammatory drug (NSAID) drug therapies (yes/no)
- Current number of NSAID drug therapies (0, 1, 2, ≥3)
- Current conventional synthetic disease-modifying antirheumatic drug (DMARD) (yes, no)
- Current conventional synthetic DMARD type (methotrexate, sulfasalazine, hydroxychloroquine, apremilast, leflunomide, other)
- Current oral corticosteroid use (yes, no)
- Current Analgesic/Opioid therapies (yes, no)
- Disease duration (<2, >= 2 years)

Additionally, a separate table containing a summary of variables related to ankylosing spondyloarthritis (AS) history will be prepared containing the following variables:

- Time since first diagnosis of AS (years) (continuous and in categories <2, ≥ 2)
- Time since first symptoms of AS (years)
- Age at first diagnosis of AS (years)
- Age at first symptoms of AS (years)

Time since first diagnosis of AS (years) will be calculated as:

$$\frac{\text{Randomization date} - \text{Date of diagnosis}}{365.25}$$

using the information collected in the AS Diagnosis CRF form. If the date of diagnosis is partial, it should be imputed to the most recent feasible date (ie, last day of the month if only day is missing, or the last day of the year if day and month are missing, if this imputation would lead to a date of diagnosis after screening date, the screening date will be used for imputation). If the date of randomization is missing, then the duration of disease will be derived using the date of screening.

Age at first diagnosis (years) will be calculated as:

$$\frac{\text{Date of first diagnosis} - \text{Date of birth}}{365.25}$$

Since only the year of birth may be entered into the eCRF, age at diagnosis will be calculated after imputing the date of birth to be 01 Jan XXXX.

Age at first symptoms (years) will be calculated as:

$$\frac{\text{Date of first symptoms} - \text{Date of birth}}{365.25}$$

Time since first symptoms should be calculated in the same manner as time since first diagnosis, using the information collected in the AS Diagnosis CRF form as well.

The percentage of study participants that meet each of three clinical criteria as well as the number and percentage of study participants that meet radiologic criterion will be summarized. For AS0011 eligibility, study participants must meet criterion defined below. The specific definition for these criteria is as follows:

- Clinical criterion a – Study participant had low back pain and stiffness for more than 3 months which improves with exercise, but is not relieved by rest
- Clinical criterion b – Study participant had limitation of motion of the lumbar spine in both the sagittal and frontal planes
- Clinical criterion c – Study participant had limitation of chest expansion relative to normal values corrected for age and sex
- Radiologic criterion – Study participant had sacroiliitis grade ≥2 bilaterally or sacroiliitis grade 3 to 4 unilaterally

To understand the study population during the different periods of the COVID-19 pandemic, the reporting of other baseline characteristics and AS history will be repeated by COVID-19 enrollment period (prior/during/post) on the SS.

Finally, a by participant-listing will be presented on all baseline characteristics for all screened study participants.

6.3 Medical history and concomitant diseases

Previous medical history is defined as conditions that have resolved prior to study entry.

Ongoing medical history are defined as conditions that are ongoing at the time of study entry.

Previous and on-going medical history will be summarized together by MedDRA System Organ Class (SOC) and Preferred Term (PT), by treatment and overall including the number and percentage of study participants with each condition. Axial spondyloarthritis history will not be included in this table as it will be presented separately (see Section 6.2). The table summaries will be ordered alphabetically for SOC and in terms of decreasing frequency for PT within SOC in the bimekizumab treatment group. In the event of ties, PT will be ordered alphabetically.

Pre-specified medical history coming from ‘the ‘Infection history’ CRF form will be summarized in a separate table.

The above summaries will be presented on the overall study treatment period.

Additionally, the extra-articular assessments (including uveitis, inflammatory bowel disease (IBD), psoriasis, peripheral arthritis, enthesitis (heel, non-heel), dactylitis) will be summarized at Screening and Baseline as collected in EDC. Uveitis and IBD assessments will also be summarized at post-Entry Visit (ie Visits after Visit 2). The summary will be presented on the SS for the overall period.

Medical history will be listed by treatment and study participant including the reported term, PT, and SOC for the RS. A glossary of all medical history conditions will be presented on the RS including the reported term, PT and SOC. Concomitant medical procedures will be listed on the RS.

6.4 Prior and concomitant medications

Prior medications include any medications that started prior to the start date of study medication.

Concomitant medications are medications taken at least one day in common with the study medication dosing period (defined as from first dose of study medication [including placebo] up to last dose of study medication + 28 days for the double-blind treatment period.

For study participants who discontinue early, the dosing period ends at the last study medication date + 28 days (one dosing interval). For study participants who complete the study as planned, the dosing period ends at the later of the following two dates:

- Last study medication date + 28 days
- The last scheduled visit date not including SFU

In the case of missing data, the classification of medications as prior or concomitant will be performed as described in Section 4.2.3. Imputations of missing data will be performed before calculation of relative study days.

The number and percentage of study participants in the SS taking prior medications will be summarized by treatment group, overall and by ATC class, presenting Anatomical Main Group (ATC Level 1), Therapeutic/Pharmacological Subgroup (ATC level 3), and PT. The table summary will be ordered alphabetically for the Anatomical Main Group and in terms of decreasing frequency for PT within Anatomical Main Group in the bimekizumab treatment group. In the event of ties, PT will be ordered alphabetically.

The number and percentage of study participants taking concomitant medications will be summarized similarly in a separate table. The summary will be presented on the SS for the Double-Blind Treatment period, as well as on the MS for the Maintenance Treatment Period.

In addition, separate summary tables for concomitant NSAIDs and for selected concomitant medications will be created for each treatment period

A by-study participant listing of all prior and concomitant medications as well as glossary of all medical medications will be provided on RS.

6.5 Rescue medication

Refer to the study protocol for the list of rescue medications.

The number and percentage of study participants who administered rescue medication will be summarized by ATC class, presenting ATC Level 1, ATC level 3, and PT. The table summary will be ordered alphabetically for the ATC class and in terms of decreasing frequency for PT

within ATC class in the bimekizumab treatment group. In the event of ties, PT will be ordered alphabetically. The same rule than the one defined in Section 6.4 will be used to define concomitant medications.

The summaries on rescue medication will be presented on the Maintenance Treatment Period, as well as on the overall period up to Week 20.

Rescue medication will be listed on RS.

7 MEASUREMENTS OF TREATMENT COMPLIANCE

Due to the method of administration of the treatments, compliance will be examined in terms of completed injections. Treatment compliance will be calculated as:

$$\frac{\textit{T}\text{otal number of actual (completed) injections}}{\text{Total number of expected injections}} \times 100\%$$

The total number of expected injections is derived relative to when the study participant finishes treatment. If a study participant completes treatment, four injections are expected during the Double-Blind Treatment Period and nine during the Maintenance Period. A total of thirteen injections are then expected (at Baseline, and at each study visits from Week 4 until Week 48). If a study participant discontinues early, then the number of expected injections is based on the time of early discontinuation relative to the dosing visits.

A summary of percent treatment compliance categorized as <75% and ≥75% will be provided by treatment group, as well as a by-study participant listing of treatment compliance.

Compliance will be summarized on the SS for the Double-Blind Period, on the MS for the Maintenance Period and on the SS for the overall period.

Injections missed due to COVID-19 will be summarized and tracked in the listings.

8 EFFICACY ANALYSES

8.1 Statistical analysis of the primary efficacy variable

8.1.1 Derivation of ASAS40 response

The rules described in Section 4.2.1.1 will be followed in the derivation of the ASAS40 response. The ASAS40 response at Week 16 is the primary efficacy variable. The ASAS40 response is defined as:

- an improvement of at least 40%, and
- absolute improvement of at least 2 units on a 0 to 10 numeric rating scale (NRS) in at least 3 of the 4 following domains:
  - Patient's Global Assessment of Disease Activity (PGADA)
  - Pain assessment (total spinal pain, question 1 from total and nocturnal spinal pain)
  - Physical function (measured by the Bath Ankylosing Spondylitis Functional Index (BASFI))
  - Inflammation (represented by the mean of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) questions 5 and 6) concerning morning stiffness intensity and duration)

- and no worsening at all in the remaining domain.

8.1.1.1 Derivation of the patient's global assessment of disease activity (PGADA)

Study participants will score the global assessment of their disease activity in answering the following "How active was your spondylitis on average during the last week?" using a NRS where 0 is "not active" and 10 is "very active".

8.1.1.2 Derivation of the pain assessment domain

The pain assessment domain will correspond to the total spinal pain and will be measured by the study participant in answering the following question "How much pain of your spine due to spondylitis do you have?". To answer the question, the study participant will consider the average amount of pain in the preceding week and use a NRS ranging from 0 to 10, where 0 represents "no pain" and 10 represents "most severe pain".

8.1.1.3 Derivation of the physical function domain of the Bath Ankylosing Spondylitis Functional Index (BASFI)

The BASFI contains 10 questions and will be measured by the study participant for assessing their physical functions during the past week. The first 8 questions evaluate activities related to functional anatomical limitations due to the course of AS. The final 2 questions evaluate the study participants' ability to cope with everyday life. An NRS ranging from 0 ("Easy") to 10 ("Impossible") will be used to answer the questions on the test. The arithmetic mean of the 10 scores gives the BASFI score, which is a value between 0 and 10, with lower scores indicating better physical function.

In case of missing answers to 1 or 2 of the single items within the BASFI questionnaire, the BASFI score will be calculated by imputing missing items with the mean of the completed items. Then, the BASFI score will be calculated as described above. If more than 2 of the items are missing, the BASFI score will be left missing.

8.1.1.4 Derivation of the inflammation domain (mean of questions 5 and 6 of the BASDAI)

The BASDAI consists of 6 horizontal NRSs and will be used to measure the disease activity of ankylosing spondylitis from the study participant's perspective. Each NRS contains 10 units to measure the severity of the 5 major symptoms: fatigue, spinal pain, peripheral joint pain and swelling, enthesitis, and morning stiffness (both severity and duration) over the last week.

The ASAS inflammation component is calculated as the average of the two scores relating to morning stiffness measurements (ie, question 5: "How would you describe the overall level of morning stiffness you have had from the time you wake up?" and question 6: "How long does your morning stiffness last from the time you wake up?"). The ASAS inflammation score ranges from 0 to 10.

If one of the two morning stiffness measurements is missing, the other one will be used as ASAS inflammation score.

If more than one measurement is missing, the ASAS inflammation score will be set to missing.

8.1.2 Primary analysis of the primary efficacy variable

The primary efficacy variable (ASAS40 response at Week 16) will be analyzed for all study participants in the RS.

The primary efficacy analysis will evaluate the composite estimand (NRI) that combines the clinically meaningful improvement from Baseline in ASAS40 response at Week 16 and the intercurrent event of not discontinuing early from study treatment for any reason prior to Week 16.

Note that only **permanent** discontinuations are considered as intercurrent events. This definition is applicable to all analyses defined in this SAP.

The following 4 attributes describe the composite estimand that will be used to define the treatment effect of interest for the primary efficacy analysis:

- Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP
- Study participant-level outcome = ASAS40 at Week 16
- Intercurrent event handling = An intercurrent event is defined as discontinuation of study treatment prior to Week 16. A composite strategy will be implemented in which a positive clinical outcome is defined as achieving ASAS40 at Week 16 and not discontinuing study treatment through Week 16.
- Population-level summary measure = Conditional odds ratio comparing bimekizumab to placebo.

Intercurrent events will then be acknowledged as an unfavorable outcome for the composite estimand in considering study participants with intercurrent events as 'non-responders' to the study treatment.

Consequently, if the date of an intercurrent event (as defined in Section 3.4) occurred prior to or at Week 16, study participants will be considered as "non-responders" at Week 16.

Additionally, missing data at Week 16 that are not preceded by an intercurrent event will be imputed as non-responders.

As outlined in Section 4.5, the statistical null hypothesis for the ASAS40 response at Week 16 is that there is no difference in the proportion of study participants with ASAS40 response between the bimekizumab treatment and placebo (ie, the conditional odds ratio for ASAS40 response in the bimekizumab treatment compared with placebo is equal to 1). The alternative hypothesis is that there is a difference between bimekizumab treatment and placebo.

A logistic regression model will be used to assess the treatment effect on ASAS40 response at Week 16. The model will include fixed effects for treatment, and prior TNFα inhibitor exposure and region as stratification factors. The suitability of including these variables in the model will be assessed using the Hosmer-Lemeshow goodness-of-fit test.

Also, as mentioned in Section 4.1, if the logistic regression model is unable to converge the stratification factors may be dropped to facilitate the model convergence.

The SAS® PROC LOGISTIC will be used to run the logistic regression.

The summary table results will present the number of responders, adjusted responder rates and associated 95% CI for bimekizumab and placebo, the adjusted odds ratio and 95% CI for the comparison bimekizumab versus placebo, and the p-value that the OR=1. The treatment comparison will be made using the 2-sided Wald test at a significance level of α=0.05.

8.1.3 Secondary analyses of the primary efficacy variable

Subgroup analyses will be performed on the primary efficacy variable on the RS.

The variables for subgroup analyses are defined in Section 4.8.

Unless otherwise notified, for each subgroup analysis, a logistic regression will be fitted involving the same terms from the primary analysis model and additional terms for the subgroup and the subgroup by treatment interaction. However, for the analysis by stratification variables (ie, prior TNFα exposure and region), the analysis model will contain the same terms from the primary analysis model plus the subgroup by treatment interaction only.

The same imputation method as the one used for the primary analysis (ie, NRI) will be used to handle missing data. The covariates will be provided in the same order as for the primary analysis model with the terms for subgroup and for subgroup by treatment interaction added at the end of the model statement.

For each subgroup category and each treatment group, the number of responders, and adjusted responder rate with associated 95% CI, and the adjusted OR (for the comparison bimekizumab and placebo) and the 95% CI will be provided. Results obtained for each subgroup will be presented in one single table. The ORs and associated 95% CIs for each subgroup category will also be displayed on a single forest-plot.

8.1.4 Supportive analyses of the primary efficacy variable

The following sections described the different supportive analyses planned for the primary efficacy endpoint. Unless otherwise specified:

- the same model as the one specified for the primary analysis will be used
- and the same statistic results as the ones produced for the primary efficacy analysis will be calculated

The adjusted odds ratios and 95% CI for the comparison of bimekizumab versus placebo obtained from the primary analysis and from the supportive analyses described in Section 8.1.4.1, Section 8.1.4.2, and Section 8.1.4.4 to 8.1.4.8 will be displayed on a forest-plot.

8.1.4.1 Analysis on the PPS

The primary analysis described in Section 8.1.2 will be repeated based on the PPS to evaluate the effect of important protocol deviations and protocol-permitted decrease in dosing or dosing frequency of axSpA background medication (due to intolerance/AE/side-effects) on the analysis.

8.1.4.2 Analysis on the FAS

The primary analysis described in Section 8.1.2 will be repeated on the FAS to evaluate whether there are differences in the efficacy analysis between randomized study participants and randomized study participants with a valid Baseline. This analysis will only be produced if the number of study participants in the RS and FAS differs.

8.1.4.3 Analysis on individual components of ASAS40

The change from Baseline in each individual component of the ASAS40 response (PGADA, total spinal pain score, BASFI, and BASDAI) will be analyzed at Week 16 using an analysis of covariance (ANCOVA) model. The analyses for the component variable BASFI will be described in Section 8.2.2 where analyses for secondary endpoints are described. For all four variables, the model will include treatment, prior TNFα inhibitor exposure and region as fixed effects and the respective Baseline values as covariate. These analyses will be done on the RS and will explore the effect of the signs and symptoms of the individual components on the composite endpoint.

The following 4 attributes describe the estimand that will be used to define the treatment effect of interest for each of the four components for this analysis:

- Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP
- Study participant level outcome = The change from Baseline in the ASAS component variable at Week 16
- Intercurrent event handling = An intercurrent event is defined as discontinuation of study treatment due to any reason prior to Week 16. This estimand targets the treatment difference in a scenario where withdrawal from study treatment does not occur, such that outcomes for study participants without an IE are as observed, and outcomes for study participants with an IE are treated as they had completed the randomized study treatment through Week 16 but on placebo. In that strategy, missing data and non missing data after IE (reset as missing) will be imputed using reference-based MI, in which the MI model is based on data from the placebo group.
- Population-level summary measure = the difference in the adjusted means between bimekizumab and placebo.

Intermittent missing data will be imputed using the reference-based MI method as well. See Section 4.2.1.6 for technical details on how the reference-based MI method will be implemented.

The SAS® PROC MIXED will be used to run the ANCOVA model.

For each ANCOVA model specified in this analysis, the following statistics will be presented:

a) For placebo and bimekizumab: the adjusted least-square means (LSM) and standard error (SE)

b) For the comparison between placebo and bimekizumab: the difference between the LSM, the associated 95% confidence intervals for the contrasts, and the corresponding p-value.

An additional table will present the descriptive statistics for each ASAS40 component variable at Week 16. Those statistics will be the mean, SE, median, minimum and maximum obtained after MI.

The following supportive and additional analyses described in Section 8.1.4.4 to Section 8.1.4.8 are planned for the primary efficacy endpoint.

8.1.4.4 Analysis using modified composite estimand - MI

A supportive analysis will evaluate the modified composite estimand (Section 4.2.1.4.2) that combines the clinically meaningful improvement from Baseline in ASAS40 response at Week 16 and the intercurrent event of not discontinuing early from study treatment due to AE or lack of efficacy prior to Week 16.

The same logistic regression model as the one used for the primary efficacy analysis will be used on the RS. The following 4 attributes describe the estimand that will be used to define the treatment effect of interest for this analysis:

- Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP
- Study participant-level outcome = ASAS40 at Week 16
- Intercurrent event handling = An intercurrent event is defined as discontinuation of study treatment due to AE (including death) or lack of efficacy, prior to Week 16. A composite strategy will be implemented in which a positive clinical outcome is defined as achieving ASAS40 at Week 16 and not discontinuing study treatment due to AE or lack of efficacy through Week 16.
- Population-level summary measure = Conditional odds ratio comparing bimekizumab to placebo.

Intercurrent events (Section 3.4) will be acknowledged as an unfavorable outcome for this modified composite estimand in considering study participants with intercurrent events as 'non-responders' to the study treatment.

Missing data at Week 16 that are not preceded by an intercurrent event will be imputed by standard MI techniques. Refer to Section 4.2.1.2 for further details.

8.1.4.5 Analysis using treatment policy strategy imputation for missing data

The treatment policy strategy will be investigated for addressing the intercurrent events. In this analysis, all available data at Week 16 will be considered regardless of the occurrence of intercurrent events. Missing data will be imputed by standard MI techniques. Refer to Section 4.2.1.9 for further details. This analysis will be based on the same logistic regression model as the one specified for the primary analysis where study participants are analyzed according to their randomized treatment even if they discontinued treatment prior to Week 16 and were no longer on the randomized study treatment when the assessment was performed at Week 16.

The following 4 attributes describe the estimand that will be used to define the treatment effect of interest for this analysis:

- Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP
- Study participant-level outcome = ASAS40 at Week 16
- Intercurrent event handling = There are no intercurrent events defined for this estimand. Discontinuation of study medication for any reason will be handled by using the data as observed.

- Population-level summary measure = Conditional odds ratio comparing ‘bimekizumab to ‘placebo.

8.1.4.6 Analysis on observed cases

An observed case (OC) analysis will additionally be conducted, where only observed data for study participants who are still on the randomized treatment at Week 16 are included. Study participants with missing data and all data after date of treatment discontinuation will be treated as missing and missing data will not be imputed. The same model than the one used for the primary efficacy analysis will be conducted on the RS. The following 4 attributes describe the estimand that will be used to define the treatment effect of interest for this analysis:

- Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP
- Study participant-level outcome = ASAS40 at Week 16
- Intercurrent event handling = A “while on randomized treatment” strategy will be implemented.
- Population-level summary measure = Conditional odds ratio comparing bimekizumab to placebo.

Refer to Section 4.2.1.10 for further details.

8.1.4.7 Tipping point analysis

The degree of the departure from the missing at random (MAR) assumption to overturn conclusions from the primary analysis will be investigated in a tipping point sensitivity analysis. This analysis will be performed only if the primary analysis is significant at α=0.05. In this analysis, it will be assumed that study participants who have missing data and are randomized to bimekizumab have a lower probability of response compared to study participants who have missing data and are randomized to placebo.

This analysis will generally follow the steps for multiple imputation described in Section 4.2.1.4.1. However, data for study participants after date of discontinuation of treatment will be changed to missing prior to imputation, but will not be changed to non-response after imputation, and all imputed values will be "adjusted" prior to calculating the binary response.

For the worst-case scenario, study participants who have missing ASAS40 response (after setting data after date of discontinuation of study treatment for any reason as defined in Section 3.4 to missing) and are randomized to bimekizumab are set to non-responders and study participants with missing data and randomized to placebo are set to responders. If the result tips from statistically significant to statistically insignificant, then the delta adjustment tipping point analysis will be conducted where missing and imputed values for each component of the ASAS40 endpoint are delta-adjusted. For both analyses, the same model as used for the primary efficacy analysis will be conducted on the RS. More robust primary analysis results will require larger adjustments to tip the results from significant to not significant.

Refer to Section 4.2.1.8 for more details on the methodology.

8.1.4.8 Analyses including COVID-19 impact

To assess the impact of the COVID-19 pandemic on the primary efficacy endpoint, the following sensitivity analyses will be conducted using the same analysis method as the primary analysis described in Section 8.1.2:

- On the COVID-19 free set,
- By timing of the Week 16 Visit relative to the start and end of the COVID-19 pandemic (ie, study participants will be categorized according to if their Week 16 Visit took place prior to, during or after the pandemic). See Section 3.2.2.2 for definition of pandemic periods.

For this last analysis, the same model than the one described in Section 8.1.3 will be used. The ORs and associated 95% CI results of this analysis will be displayed on a forest plot.

8.2 Statistical analysis of the secondary efficacy variables

8.2.1 Derivation of secondary efficacy variables

8.2.1.1 Other assessment in Axial SpondyloArthritis international Society response criteria (ASAS20, ASAS5/6, and ASAS-PR)

The ASAS20 response is defined as relative improvement of 20%, and an absolute improvement of at least 1 unit on a 0 to 10 NRS in at least 3 of the 4 following domains:

- Patient's Global Assessment of Disease Activity (PGADA)
- Pain assessment (total spinal pain, question 1 from total and nocturnal spinal pain)
- Function (represented by the Bath Ankylosing Spondylitis Functional Index (BASFI))
- Inflammation (represented by the mean of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) questions 5 and 6) concerning morning stiffness intensity and duration)

and absence of deterioration in the potential remaining domain. A deterioration will be defined as a relative worsening of at least 20% and an absolute worsening of at least 1 unit.

The ASAS 5/6 response is defined as at least 20% improvement in 5 of 6 domains, which includes, in addition to the 4 domains mentioned above for ASAS20, spinal mobility (ie, lateral spinal flexion component of the BASMI) and hs-CRP as more objective measures.

The ASAS-PR response is defined as a score of ≤2 units on a 0 to 10 unit scale in all 4 domains listed above for ASAS20.

ASAS40 response in the TNF alpha inhibitor-naïve study participants represents the ASAS40 response for study participants with "Prior TNFα inhibitor = No" (which are defined per actual information on prior use of TNFα inhibitor). For the supportive analysis based on MI, all randomized study participants will be included in the imputation model (since this analysis is a subgroup analysis of the primary endpoint), but the analysis step will only be done on the subgroup of TNF alpha inhibitor-naïve study participants.

8.2.1.2 Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

As mentioned in Section 8.1.1.4, the BASDAI consists of six 10-unit NRSs measuring the disease activity of AS over the past week on 5 major symptoms. To give each symptom equal weighting, the average of the 2 scores relating to morning stiffness is taken. The BASDAI is calculated as follows:

$$BASDAI = \frac{Q1 + Q2 + Q3 + Q4 + \left(\frac{Q5 + Q6}{2}\right)}{5}$$

where Q1 – Q6 are the six questions from the BASDAI questionnaire.

The BASDAI score is between 0 and 10, with lower scores indicating lower disease activity.

As mentioned in Section 8.1.1.4, if 1 of the 2 morning stiffness measurements (ie, questions Q5 and Q6) is missing, the other one will be used for the morning stiffness calculation.

If one major symptom of the BASDAI is missing, the total score of the remaining symptoms will be divided by the number of symptoms assessed. If more than one major symptom is missing, the total BASDAI score will be set to missing.

8.2.1.3 Ankylosing Spondylitis Disease Activity Score major improvement (ASDAS-MI)

The ASDAS-CRP consists of a number of assessments which are scored by the study participant and physician and multiplied by a proven formula (van der Heijde et al, 2009) as shown below:

- 0.121 x Back pain (BASDAI Question 2 result)
- 0.058 x Duration of morning stiffness (BASDAI Question 6 result)
- 0.110 x Patient's Global Assessment of Disease Activity (PGADA)
- 0.073 x Peripheral pain/swelling (BASDAI Question 3 result)
- 0.579 x (natural logarithm of the (hs-CRP [mg/L] + 1))

Back pain, PGADA, duration of morning stiffness, peripheral pain/swelling and fatigue are all assessed on a numerical scale (0 to 10 units) (Lukas et al, 2009). If the hs-CRP value is <2 mg/L, the constant value of 2 mg/L should be used in the calculation. The sum of these weighted components gives the ASDAS-CRP.

For analyses with imputation of missing data, the imputed value for ASDAS-CRP will be based on the imputed values for the individual components.

For observed case analyses, if any individual component score is missing, then the ASDAS-CRP will be set to missing.

Disease activity categories based on ASDAS-CRP are as follows:

- ASDAS-Inactive Disease (ASDAS-ID): ASDAS-CRP <1.3
- ASDAS-Low Disease (ASDAS-LD): ASDAS-CRP ≥1.3 to <2.1
- ASDAS-High Disease Activity (ASDAS-HD): ASDAS-CRP ≥2.1 to ≤3.5
- ASDAS-very High Disease Activity (ASDAS-vHD): ASDAS-CRP >3.5

The variables related to ASDAS improvement are defined as follows:

- ASDAS-Clinically Important Improvement (ASDAS-CII): ASDAS-CRP reduction (improvement) of ≥1.1 relative to Baseline
- ASDAS-Major Improvement (ASDAS-MI): ASDAS-CRP reduction (improvement) of ≥2.0 relative to Baseline

For both ASDAS disease category and ASDAS improvement variables, multiple imputation methods will be used to impute the ASDAS-CRP value and then dividing into the appropriate category or improvement groups.

8.2.1.4 Nocturnal spinal pain (NSP)

The pain in the spine at night due to AS will be measured by the study participant in answering question 2 of the total and nocturnal spine questionnaire "How much pain of your spine due to spondylitis do you have at night?". To answer the question, the study participant will consider the average amount of pain in the preceding week and use a NRS ranging from 0 to 10, where 0 represents "no pain" and 10 represents "most severe pain".

8.2.1.5 Ankylosing Spondylitis Quality of Life (ASQoL)

The ASQoL consists of 18 items which are scored by the study participant, each item with a score of 0=no or 1=yes. A score of "1" is given where the item is affirmed, indicating adverse quality of life. The ASQoL total score is based on the sum scores, ranging from 0 to 18, with higher scores indicating worse quality of life.

If 3 or fewer items are missing, the missing responses will be imputed with the mean of the available responses from that visit to calculate a total score. If more than 3 items are missing, the total score will be left missing.

8.2.1.6 Short-Form 36-item Health Survey (SF-36) Physical Component Summary (PCS)

The SF-36 (version 2, standard recall) is a 36-item generic health-related quality of life (HRQoL) instrument that uses a recall period of past 4 weeks. Items are grouped into the following 8 health domains: Physical Functioning (10 items), Role Physical (4 items), Bodily Pain (2 items), General Health (5 items), Vitality (4 items), Social Functioning (2 items), Role Emotional (3 items), Mental Health (5 items). The concepts represented by these domains contribute to physical, mental and social aspects of HRQoL. [redacted]

In addition to domain scores, the PCS and MCS scores are calculated from the 8 domains (excluding the Health Transition item). Component scores appreciate the impact of each domain on physical and mental health status (Maruish, 2011).

The two SF-36 component summaries (PCS and MCS) and the 8 domain scores are standardized with a mean of 50 and a standard deviation (SD) of 10 in the general US population (Maruish, 2011). An individual respondent's score that falls outside the T-score range of 45 to 55 should be considered outside the average range for the U.S. general population. When considering group-level data, a score below 47 should be considered indicative of impaired functioning within that health domain or dimension. Similar to individual respondent data, group mean scores 47 or greater should be considered average or above average as compared to the general US population.

For the calculation of the SF-36 domain scores and the two component summary scores, the scoring software QualityMetric's Health Outcomes™ will be used. The software uses updated 2009 U.S. population norms and applies a Full Missing Score Estimation method as follows:

- A health domain score (except the Physical Functioning domain) will be estimated provided that at least one non-missing response is available within that domain

- For the Physical Functioning domain item response, theory will be used to develop a model for estimates of the missing score
- Regression methods are then applied to estimate the PCS and MCS on the basis of the domain scores

8.2.1.7 Bath Ankylosing Spondylitis Disease Metrology Index (BASMI)

The BASMI characterizes the spinal mobility of study participants with AS consisting of 5 clinical measures to reflect study participant axial status: cervical rotation; tragus to wall distance; lateral spinal flexion; modified Schober test; intermalleolar distance. Each of the 5 movements is scored according to the linear BASMI definition given in Table 8–1. The mean of the 5 scores provides the BASMI score. The BASMI score ranges from 0 to 10. The higher the BASMI score the more severe the study participant's limitation of movement due to their AS.

Table 8–1: BASMI linear definition

Clinical Movement	S = 0 if:	S between 0 and 10:	S=10 if:
Lateral spinal flexion* (cm)	A ≥ 21.1	S = (21.1 cm – A) / 2.1 cm	A ≤ 0.1
Tragus-to-wall distance* (cm)	A ≤ 8	S = (A – 8 cm) / 3 cm	A ≥ 38
Lumbar Flexion* (modified Schober) (cm)	A ≥ 7.4	S = (7.4 cm – A) /0.7 cm	A ≤ 0.4
Intermalleolar distance (cm)	A ≥ 124.5	S = (124.5 cm – A) / 10 cm	A ≤ 24.5
Cervical rotation angle* (º)	A ≥ 89.3	S = (89.3° - A) / 8.5°	A ≤ 4.3

S = BASMI Score, A = assessment.

* For cervical rotation, tragus-to-wall distance, and lateral spinal flexion, the mean of the left and right measurements will be calculated, if both are available. Otherwise, the available measurement will be used. If multiple attempted measurements are obtained, calculate the mean of the attempts. The mean of the attempts should be calculated after averaging the left and right measurements, where applicable.

From van der Heijde et al. 2008.

For all clinical measures, values outside the ranges presented in Table 8–2 (Maksymowych 2006) will be flagged as invalid and treated as if they were missing values. Where two attempted measurements for a single clinical measure are entered at a given visit (ie, lateral lumbar flexion, cervical rotation, tragus-to-wall distance, and lumbar flexion) and only one attempted measurement is valid, the valid measurement will be used. If both attempts are invalid, the measure will be treated as missing.

Table 8–2: BASMI linear definition (pre-defined ranges)

Clinical Movement	Range*
Lateral spinal flexion (mean right/left)	0.75 – 25.00 cm
Tragus-to-wall distance (mean right/left)	9.00 – 37.50 cm
Lumbar flexion (modified Schober)	0.00 – 9.00 cm

Clinical Movement	Range*
Intermalleolar distance	13.00 – 160.00 cm
Cervical rotation (mean right/left)	0.00 – 94.00°

* From Maksymowych 2006

The order of events to derive the BASMI scores is as follows:

1. Discard individual assessment values (A) that are invalid (out of range according to the Table 8–2 above)
2. Average remaining valid left and right assessment values (A) by study participant, analysis visit, parameter and attempt, if more than one valid value is present
3. Average assessment values (A) across attempts, by study participant, analysis visit and parameter, if more than one attempt is present with valid values
4. From these averaged (A) values, calculate a single BASMI component score (S), by study participant, analysis visit and parameter
5. Re-map early termination visits according to the specifications given in Section 3.2.3, Re-mapping will be performed on a per-parameter basis prior to the calculation of the BASMI score (this method is adopted in the event that not all parameters are assessed at each visit)
6. If 1 or 2 BASMI components scores (S) are missing for a given study participant and analysis visit, carry forward the most recent non-missing value for that study participant and component to the current analysis visit; it is allowed to carry forward a screening or Baseline score value (S) to a post-Baseline analysis visit
7. If one value is available for each of the 5 component scores for a given study participant and analysis visit, calculate BASMI score as the average of the 5 component scores (S); otherwise set the BASMI score to missing for that study participant and analysis visit

8.2.1.8 Maastricht Ankylosing Spondylitis Enthesitis (MASES) index

The MASES index assesses enthesitis through 13 items (bilateral costochondral 1, costochondral 7, anterior superior iliac spine, posterior iliac spine, iliac crest and proximal insertion of the Achilles tendon sites, and the fifth lumbar vertebral body spinous process) (Heuft-Dorenbosch et al, 2003). Each item is scored as 0=yes or 1=no and then summed for a possible score of 0 to 13, with higher scores indicating worse enthesitis.

If 7 or more items are available, MASES will be derived by dividing the sum score by the number of available assessments and multiplying the result by 13. If less than 7 items are available, MASES will be treated as missing.

Analyses on MASES index will be restricted to the subset of study participants with enthesitis present at Baseline (defined as MASES index score>0).

Enthesitis-free state based on MASES index is a status defined as study participants having achieved a MASES index score of 0 at Week 16.

For the analysis based on multiple imputation, only randomized study participants with enthesitis at Baseline will be included in the imputation models.

8.2.2 Analysis of secondary efficacy variables

The secondary efficacy variables will be analyzed for all study participants in the RS as summarized in Table 4–1.

As indicated in Section 4.5, there are 11 key secondary endpoints included in the testing hierarchy that will be tested successively one after the other if the previous endpoint was found to be statistically significant.

Three secondary endpoints (change from Baseline in BASMI at Week 16, change from Baseline in MASES index at Week 16 in the subgroup of study participants with enthesitis at Baseline, and enthesitis-free state at Week 16 based on the MASES Index in the subgroup of study participants with enthesitis at Baseline) will not be part of the testing hierarchy and for this reason the p-values produced by the statistical models will be considered nominal since these endpoints are not controlled for multiplicity. The following analyses will be conducted for the Secondary Efficacy Endpoints:

- For the secondary binary endpoints:
  - Composite Estimand - NRI: The same composite estimand structure as the one defined in Section 8.1.2 for the primary efficacy analysis will be used. The same analysis model will be considered, and the analysis results will be presented similarly as for the primary efficacy analysis. The imputation strategy for handling missing data will be the same as for the primary endpoint, ie the NRI approach (see Section 4.2.1.1).
  - Modified Composite - MI: A similar modified composite estimand structure as the one defined in Section 8.1.4.4 for the primary efficacy analysis will be used. The same analysis model will be considered, and the analysis results will be presented similarly as for the primary efficacy analysis. The imputation strategy for handling missing data will be the MI-MCMC/Monotone regression for the composite and non-composite binary endpoints as detailed in Section 4.2.1.4.
  - Observed Case: A similar analysis as the one defined in Section 4.2.1.10 for the primary efficacy analysis will be used.
- For the secondary continuous endpoints:
  - Reference-Based Estimand - MI: The same hypothetical estimand (Section 4.2.1.6) structure as the one defined in Section 8.1.4.3 for the analysis on component variables for the primary efficacy endpoint will be used. The same analysis model and imputation strategy for handling missing data will also be considered. The analysis results will be presented similarly as for this analysis on the individual ASAS40 components.

    This estimand will only be implemented for the secondary continuous endpoints in the testing hierarchy.
  - Hypothetical Estimand - MI:

    The following 4 attributes describe the estimand that will be used to define the treatment effect of interest for this analysis:
    - Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP

- Study participant level outcome = The change from Baseline in the variable at Week 16
- Intercurrent event handling = An intercurrent event is defined as discontinuation of study treatment due to any reason prior to Week 16. A hypothetical strategy will be implemented in which the estimand targets the treatment difference in a scenario where intercurrent does not occur, such that outcomes for study participants without an intercurrent event are as observed, and outcomes for study participants with an intercurrent event are treated as though they had completed the randomized study treatment through Week 16. A multiple imputation (MI) strategy will be used to impute any missing data including observed data after the IE date which will be set to missing prior to running MI.
- Population-level summary measure = the difference in the adjusted means between bimekizumab and placebo.

The same analysis model as the one defined in Section 8.1.4.3 used for the analysis on the individual ASAS40 components will be considered. The analysis results will also be presented similarly.

– Observed Case: A similar analysis as the one defined in Section 4.2.1.10 for the primary efficacy analysis will be used.

Note that the ANCOVA analysis on the change from Baseline in BASFI will be considered a supportive analysis for the primary endpoint since BASFI is one of the components of the primary endpoint.

An overview table will combine the results of the primary analysis for the primary efficacy endpoint and the results of the primary analysis for each key secondary endpoint.

Subgroup analyses will also be conducted for ASDAS-MI endpoint using NRI. The variables of interest will be age, gender, race, region, ASDAS status and prior TNFα inhibitor exposure. Exact definition of the subgroups is provided in Table 4–4. The same methodology as the one used for the ASAS40 subgroup analyses will be followed. The results obtained for each subgroup will be combined in one single table as well as presented on a forest plot.

Finally, to assess the impact of the COVID-19 pandemic, the primary analysis of all secondary efficacy endpoints included in the testing hierarchy with be analyzed on the COVID-19 free set, using the Reference-Based Estimand.

8.2.3 Additional statistical analysis of selected secondary efficacy variables

For the "change from Baseline in MASES index at Week 16 in the subgroup of study participants with enthesitis at Baseline", and for the "enthesitis-free state at Week 16 based on the MASES Index in the subgroup of study participants with enthesitis at Baseline" secondary efficacy variables, additional statistical tests and calculation of inferential statistics will be performed. As these tests are not part of the multiplicity-controlled procedure, the associated p-values are considered nominal and are not controlled for multiplicity.

For the enthesitis-free state responder variable, the logistic regression model described in Section 8.1.2 will be used for analysis. The p-values will be based on the conditional logistic regression model to test general association. Missing values will be imputed using the NRI.

For the MASES continuous variable, the Hypothetical Estimand approach used for other continuous variables will be applied based on the ANCOVA model describe in Section 4.2.1.7 and Section 8.2.2.

The results of these inferential statistics will be presented in a single table summarizing the estimates (OR for the binary endpoint, LSM for the continuous endpoint) with their 95% CI and the associated p-values. The results will be pooled with the results obtained for selected other efficacy variables (see Section 8.3.3).

8.3 Analysis of other efficacy variables

The following sections provide the derivation of the other efficacy endpoints which have not been mentioned earlier in the SAP.

8.3.1 Derivation of other efficacy endpoints

8.3.1.1 Time to ASAS20/40 response

Time to a given response (ASAS20 or ASAS40) will be defined as the time in days from the start date of first treatment dose until the first date when the response (respectively ASAS20, or ASAS40) is achieved.

Missing ASAS response data will not be imputed. Instead, study participants with no observed response, will be censored at their last visit with ASAS40 response data.

As mentioned later in Section 8.3.2, the analysis will be performed on the Double-Blind Treatment Period. Study participants who discontinue IMP prior to achieving a response will be censored at the minimum date between the date of treatment discontinuation as defined in Section 3.4 and the Week 16 visit date. Study participants who reach the end of the Double-Blind Treatment Period visit (and are still on treatment) without achieving the given response will be censored at the date of the end of the Double-Blind Treatment Period visit (ie, Week 16 Visit date).

8.3.1.2 ASDAS status

The ASDAS status endpoint is derived as follows based on ASDAS-CRP endpoint previously defined in Section 8.2.1.1:

- Inactive disease: when ASDAS-CRP <1.3
- Low disease activity: when ASDAS-CRP ≥1.3 to <2.1
- High disease activity: when ASDAS-CRP ≥2.1 to ≤3.5
- Very high disease activity: when ASDAS-CRP >3.5

8.3.1.3 BASDAI50 response

A response variable called BASDAI50 is defined as an improvement of at least 50% in the BASDAI compared to Baseline. See Section 8.2.1.2 for details on the BASDAI endpoint.

8.3.1.4 Physician's Global Assessment of Disease Activity (PhGADA)

PhGADA is recorded by the investigator who will assess the overall status of the study participant with respect to their AS signs and symptoms and functional capacity using an NRS in which 0= "very good, asymptomatic and no limitations of normal activities" and 10= "very poor, very severe symptoms which are intolerable and inability to carry out normal activities".

8.3.1.5 High-sensitivity C-Reactive Protein (hs-CRP)

In the summary tables presenting hs-CRP values, the observed values and ratio to Baseline values will be displayed.

Hs-CRP values which are below the lower limit of quantification (LLOQ) should be set to the midpoint between 0 and the LLOQ. The listing will show that the value was below the LLOQ. The LLOQ will be set according to data received by the laboratory vendor. Specifically, any numeric value preceded by "<" will be reset to a value equal to 0.5 * the numeric value following the "<" for each individual.

Particular attention will be given to extreme CRP values. Any CRP values of ≥500mg/L will be set to missing prior to performing the MI procedure; the subsequently imputed CRP value will be included in the summary tables. Listings will display the original result.

8.3.1.6 European QoI-5Dimensions 3-Level (EQ-5D-3L)

The EQ-5D-3L health questionnaire provides a descriptive profile and a single index value for health status. The instrument is composed of a 5-item health status measure (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) and a visual analogue scale (VAS). Each of the 5 health states is divided into 3 levels (no problem, some or moderate problems, and extreme problems) and is scored as 1, 2, and 3, respectively. The 5 numerical responses are averaged to create the 'utility scores'.

The change from Baseline in each of the 5 responses of the EQ-5D-3L will be analyzed categorically by visit. Additionally, the change from Baseline in the utility score will be summarized by visit using the numerical scores.

The EQ-5D VAS records the respondent's self-rated health status on a vertical 20 cm visual analogue scale, graduated from 0 to 100 (0=worst imaginable health status, 100=best imaginable health status). The VAS will be evaluated by changes from Baseline and actual scores over time.

8.3.1.7 Short Form-36 (SF-36) Mental Component Summary (MCS) and the 8 domains scores (Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional, and Mental Health)

As described in Section 8.2.1.6, the SF-36 is a survey that measures 8 health domains rated by the study participants over the past four weeks. MCS is based on the aggregate of these 8 health concepts and all these 8 health domain scales are used to score the MCS. The MCS and each domain scores are standardized with a mean of 50 and a standard deviation of 10 in the general US population, and a higher score indicates better health status.

Results for MCS as well as for each domain scores will be reported.

8.3.1.8 Medical Outcomes Study (MOS) sleep scale

The MOS Sleep Scale will be used by the study participant to measure specific aspects of sleep. The frequency with which each sleep problem has been experienced during the previous 4 weeks is rated on a 5-point scale ranging from "none of the time" to "all of the time".

Out of the 12 sleep problems/items covered by the MOS sleep scale, only items relevant to derive the Sleep Disturbance and the Sleep Problems Index II subscales will be considered in this

study. Table 8–3 below describes which items are used for each sleep subscale of interest, and which items need to be reversed scored (these are indicated with an R in brackets).

Table 8–3: MOS sleep subscales

Subscale name	Number of items	Item scores
Sleep Disturbance	4	1, 3 (R), 7 (R), 8 (R)
Sleep Problems Index II	9	1, 3 (R), 4, 5 (R), 6 (R), 7 (R), 8 (R), 9 (R), 12

Note: (R) refers to a reversed item.

The subscale scores are created by averaging the respective rescaled item scores. Scales with at least 1 item answered can be used to generate a scale score. Items that are missing are not taken into account when calculating the scale scores. Scores represent the average for all items in the scale that the respondent answered.

Prior to averaging, each item score is transformed linearly to range from 0 to 100. Higher scores indicate more of the attribute implied by the scale name (eg, more sleep disturbance or more sleep problems).

QualityMetric Health OutcomesTM Scoring Software will be used to calculate the domains of Sleep disturbance and Sleep problems Index II.

8.3.1.9 Functional Assessment of Chronic Illness Therapy (FACIT)-Fatigue subscale

The FACIT-Fatigue is a 40-item measure that assesses self-reported fatigue and its impact upon daily activities and function within the past 7 days. The scale consists of 5 subscales (physical well-being, social/family well-being, emotional well-being, functional well-being, and fatigue). The fatigue subscale will be the only domain used in this study. It is composed of 13 items, all scored from 0 (Not at all) to 4 (Very much). The FACIT-Fatigue subscale score ranges from 0 to 52 with 0 being the worst possible score and 52 being the best possible score. To obtain a score from 0 to 52, all negatively worded questions have to be recoded, so that responses range from worst (0) to the best (4) outcome. This implies that the values to all questions (except for [redacted]

The FACIT-Fatigue subscale score is then obtained by summing up the responses of all 11 reversed items and responses of item 7 and 8 with equal weight.

In cases where some answers are missing, a total score is re-scaled from the score of the answered items, so long as more than 50% of the items (ie, at least 7 of 13) were answered.

The formula to derive the FACIT-Fatigue subscale score is then:

$$FACIT - Fatigue = 13\ x\ \frac{\sum Score\ of\ items}{number\ of\ item\ answered}$$

8.3.1.10 Work Productivity and Activity Impairment Questionnaire–Specific Health Problem (WPAI-SHP)

The WPAI-SHP is a study participant-reported questionnaire that assesses study participant's employment status, work absenteeism, work impairment while working, overall work, and daily activity impairment attributable to a specific health problem.

Five out of 6 items of the WPAI-SHP are regrouped into the 4 dimensions (absenteeism, presenteeism, work productivity and activity impairment) with scores expressed as percentage, where higher numbers indicate greater impairment and less productivity (ie, worse outcomes) as described in the WPAI-SHP scoring rules.

The scoring rules for the WPAI-SHP are as follows:

[redacted]

Question 1 is a Yes/No question.

[redacted]

Questions Q5 and Q6 are measured on a 0 to 10 scale (0= no effect, 10= complete interference).

In order to make data consistent and amenable to statistical analysis, the following counting rules will be applied to handle out of range and ambiguous answers of the WPAI-SHP. These rules will be applied prior to conducting any type of statistical analysis of the data. Due to the inter-relation between certain questions of the WPAI-SHP, the priority order for implementing these specific counting rules is as in the listed order below.

- Employment status:
  - If (Q1=missing) and (Q2>0 or Q3>0 or Q4>0), then Q1=YES.
- Hours missed due to Ankylosing spondylitis:
  - If Q1 = NO, then Q2 = missing.
  - If (Q2 =0 or missing) and Q1 = missing, then Q2=missing.
- Hours missed due to other reasons:
  - If Q1 = NO, then Q3 = missing.
  - If (Q3 =0 or missing) and Q1 = missing, then Q3=missing.
- Hours actually worked:

  - If Q1 = NO, then Q4 = missing.
  - If (Q4 =0 or missing) and Q1 = missing, then Q4=missing.
- Work productivity:
  - If Q1 = NO, then Q5 = missing.
  - If (Q5 =0 or missing) and Q1 = missing, then Q5=missing.
  - If Q4=0, then Q5 = missing.

In the listings, the original values will be kept and displayed.

After implementing the missing data rules above for questions Q1 to Q5, the following four scores will be derived:

- Percent work time missed due to problem from the target disease:

$$\frac{Q2\ hours}{Q2\ hours + Q4\ hours} * 100\ ,\ \text{Q2 hours} +\ \text{Q4 hours} \neq 0$$

- Percent impairment while working due to problem from the target disease:

$$\frac{Q5\ \text{score}}{10} * 100$$

- Percent overall work impairment due to problem from the target disease:

$$[\frac{Q2\ \text{hours}}{Q2\ hours + Q4\ hours} + \left(1 - \frac{Q2\ \text{hours}}{Q2\ hours + Q4\ hours}\right) * \frac{Q5\ \text{score}}{10}] * 100$$

- Percent activity impairment due to problem from the target disease:

$$\frac{Q6\ \text{score}}{10} * 100$$

WPAI scores are based on 1-item (presenteeism, activity impairment), 2-items (absenteeism) and multiple items (overall work productivity). A score cannot be calculated if there is a missing response to the corresponding item.

Because questions 2 to 6 are only relevant when the answer to question 1 is Yes, and because imputation of results is questionable without knowing the answer to question 1, only observed case analysis will be performed for this endpoint.

8.3.1.11 Tender joint count (TJC) and swollen joint count (SJC)

Each of the following 44 joints will be assessed for tenderness:

- Upper body (4) - bilateral sternoclavicular, and acromioclavicular joints.
- Upper extremity (26) - bilateral shoulders, elbows, wrists (includes radiocarpal, carpal and carpometacarpal bones considered as a single unit), metacarpophalangeal (I, II, III, IV, and V), thumb interphalangeal, and proximal interphalangeal (II, III, IV, and V).
- Lower extremity (14) - bilateral knees, ankles, and metatarsophalangeals (I, II, III, IV, and V).

All of these 44 joints will also be assessed for swelling.

Each of the joints can be graded as follows by the assessor:

- Permanently not assessable
- Temporarily not assessable
- Asymptomatic
- Tender only
- Swollen only
- Tender and swollen

Permanently not assessable joints will be considered missing for both tender and swollen joint counts at the visit at which the grading was recorded and all subsequent visits.

For the statistical analysis, swelling and tenderness will be each graded on a 2-point scale as described below in Table 8–4:

Table 8–4: Swelling and tenderness grading

Grade	Tenderness response (44)	Swelling response (44)
0	Not tender	None
1	Tenderness present	Detectable synovial thickening

The joint graded by the assessor will then have its grade mapped as follows:

- The "Tender only" joints will correspond to a tenderness response of 1, and a swollen response of 0.
- The "Swollen only" joints will correspond to a swollen response of 1, and a tenderness response of 0.
- The "Tender and swollen" joints will correspond to a swelling and tenderness response of 1.

In addition, injected joints will be counted as swollen and tender (ie, swelling and tenderness grading of 1) from the date of injection up to 1 year after injection.

Dactylitic digits will be identified as follows and the joints will be counted as follows:

- Fingers 2-5: Swelling Distal Interphalangeal Joint x (side), Swelling Proximal Interphalangeal Joint x (side), Swelling Metacarpophalangeal Joint x (side) – if they are in any of the gradings 'swollen only' or 'tender and swelling' or 'injected' then only 1 finger will be added to the swollen joint count
- Toes 1-5: Swelling Interphalangeal Joint x (side), Swelling Metatarsophalangeal Joint x (side) – if they were in any of gradings 'swollen' or 'tender and swollen' or 'injected' then only 1 toe will be added to the swollen joint count

The tender joint count and swollen joint count are weighted joint counts. If there are missing observations in the tender or swollen joint assessments (TJ and SJ, respectively), then the remaining observations will be assessed and weighted by the number of the assessed joints (AJ) as shown below:

$$TJC = n \times \frac{\sum_{i=1}^{n} TJ}{\sum_{i=1}^{n} AJ}$$

$$SJC = n \times \frac{\sum_{i=1}^{n} SJ}{\sum_{i=1}^{n} AJ}$$

Where n represents the number of total joints.

However, if a joint is missing at Baseline, then that joint will be set to missing throughout the study. If more than 50% of the planned tender joint assessments (ie, more than 22) or 50% of the swollen joint assessments (ie, more than 22) are missing at any post-Baseline visit, then no imputation will be done and the total TJC or SJC will be set to missing for that visit.

8.3.1.12 Ankylosing Spondylitis spine MRI-activity (ASspiMRI-a) (Berlin modification) score

The Berlin modification of the ASspiMRI-a is a scoring system for assessing the sacroiliac joint with a concentration on short-tau-inversion recovery (STIR) sequences without other fat saturation techniques. This scoring method quantifies active changes in 23 vertebral units (VU) of the spine (from C2 to S1). A VU is defined as the region between 2 virtual lines through the middle of each vertebra. Active inflammation is scored by grading the degree of bone marrow edema from 0 to 3 in 1 dimension on one or more consecutive slices that represent the highest level of inflammation in a particular VU. Each VU score is called a subscore. These subscores are combined into a total spine ASspiMRI-a score in the Berlin modification (total score), which can range from 0 to 69.

The ASspiMRI-a (Berlin modification) score will only be assessed for study participants in the sacroiliac joint and spine MRI assessment sub-study who have both a Baseline and at least one post-Baseline MRI of the spine available.

The following imputation rules will be used for calculating the total spine ASspiMRI-a (Berlin modification) score:

- If all subscores are not readable or poor (N) at a visit, the total will be considered missing for that visit and the study participant will not be included in the analysis.
- If one or more but not all subscores are ‘N’, then subscores = 'N' will be considered as 0 when computing the total score.

Note that a reviewer may consider an assessment as not readable and assign a “N” value if the reviewer feels unable to evaluate a scan or an area of a scan at a specific location. This could be related to artifacts, image quality or something else.

If 2 independent readers are used, the analysis will use the average of the scores from the 2 independent reviewers.

In cases where disagreement between the two readers is large, a third adjudication reviewer will be required. The spine MRI adjudication trigger will occur when the difference in change scores (ie, change from Baseline to post-Baseline timepoint) of the two primary readers differs by >=9. If there are 3 readings, the analysis will use the average of the 2 closest change scores.

As already indicated in Section 3.3, the efficacy analyses for the MRI substudy will be performed using two sets of study participants based on the date of the Baseline MRI relative to the first study medication administration: 1) all eligible substudy participants with an MRI any

time prior to the first study medication administration, and 2) eligible substudy participants with MRI performed any time from 3 weeks prior to the first dose of study drug visit up to the first dose of study drug visit.

For the Week 16 visit, a time window of +/-3 weeks before or after the scheduled visit may be considered appropriate for mapping MRI data to the given visit.

For the Week 52 visit, a time window of +/- 3 weeks before or after the scheduled visit may be considered appropriate for mapping MRI data to the given visit.

Only study participants with non-missing Baseline and at least one post-Baseline image will be included in this analysis. Therefore, no MI will be performed. The analyses will be conducted on the observed cases.

8.3.1.13 Spondyloarthritis Research Consortium of Canada (SPARCC) sacroiliac joint (SIJ) MRI score

The SPARCC is a scoring system for assessing the sacroiliac joint lesions based on STIR sequence. Each SI joint is divided into 4 quadrants (upper iliac, lower iliac, upper sacral, and lower sacral). The presence of increased signal on STIR in each of these 4 quadrants is scored on a dichotomous basis, where 1 = increased signal and 0 = normal signal. Joints that include a lesion exhibiting intense signal are each given an additional score of 1 per slice that demonstrated this feature. Similarly, each joint that included a lesion demonstrating continuous increased signal of depth greater than or equal to 1cm from the articular surface is also given an additional score of 1. The maximum score for a single coronal slice is 12. The scoring is repeated in each of 6 consecutive coronal slices. The total sacroiliac joint SPARCC score can range from 0 to 72.

Like the ASspiMRI-a (Berlin modification) score, the SPARCC score is only derived for study participants participating in the MRI sub-study who have both a Baseline and at least one post-Baseline MRI. The two sets of participants to be analyzed based on the date of the Baseline MRI relative to the first study medication administration are as described in Section 3.3, and the rules for handling poor or not readable images defined in Section 8.3.1.12 also apply. If 2 independent readers are used, the analysis will use the average of the scores from the 2 independent reviewers.

In cases where disagreement between the two readers is large, a third adjudication reviewer will be required. The SIJ MRI adjudication trigger will occur when the difference in change scores (ie, change from Baseline to post-Baseline timepoint) of the two primary readers differs by >=13. If there are 3 readings, the analysis will use the average of the 2 closest change scores. Only study participants with non-missing Baseline and post-Baseline data are included in this analysis. Therefore, no MI will be performed. The analyses will be conducted on the observed cases.

8.3.2 Analyses of other efficacy endpoints

The efficacy analyses described in this section are not part of the multiplicity-controlled testing procedure described in Section 4.5. Other efficacy endpoints will be analyzed as summarized in Table 4–1.

For the binary variables, the following two estimand structures will be defined:

- Composite Estimand - NRI:

  - Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP
  - Study participant-level outcome = The given variable and time point being summarized
  - Intercurrent event handling = An intercurrent event is defined as discontinuation of study treatment prior to the time point being summarized. A positive clinical outcome is defined as achieving the given variable at the specified time point and not having an intercurrent event date prior to the time point of interest.
  - Population-level summary measure = Unadjusted number and percent of responders

  Any missing data that are not preceded by an intercurrent event will be imputed based on NRI as well as any data after an intercurrent event, as described for the primary and secondary binary efficacy endpoints.

- Modified Composite Estimand – MI:
  - Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP
  - Study participant-level outcome = The given variable and time point being summarized
  - Intercurrent event handling = An intercurrent event is defined as discontinuation of study treatment due to AE or lack of efficacy, prior to the time point of interest. A composite strategy will be implemented in which a positive clinical outcome is defined as achieving the endpoint at the time point of interest and not discontinuing study treatment due to AE or lack of efficacy through the time point of interest.
  - Population-level summary measure = Mean proportion of responders and 95% confidence interval.

  An intercurrent event will be acknowledged as an unfavorable outcome for this modified composite estimand in considering study participants with intercurrent events as 'non-responders' to the study treatment.
  Missing data that are not preceded by an intercurrent event will be imputed by standard MI techniques. Refer to Section 4.2.1.4 for further details.

All binary other efficacy variables will also be summarized using OC approach as defined in Section 8.1.4.6.

For all these variables, the results based on the NRI will be combined with the ones obtained on OC and from the MI analysis in one single table.

In addition, a line plot by treatment group will be produced over time for some of the endpoints, that are part of the sequential testing procedure, based on NRI results, including the following binary variables: ASAS40 responder rate, ASAS40 responder rate in the TNF alpha inhibitor naïve study participants, ASAS20 responder rate, ASAS5/6 responder rate, ASAS-PR responder rate, and ASDAS-MI responder rate.

Continuous other efficacy variables that are components of the primary endpoint will be analyzed using the reference-based MI estimand and procedure described in Section 8.1.4.3.

For the continuous and categorical other efficacy variables (not including change from Baseline in WPAI-SHP, ASspiMRI-a (Berlin modification) score, and SPARCC score, and EQ-5D-3L Utility scores), the following estimand structure will be defined:

- Hypothetical Estimand (Section 4.2.1.7) – MI:
  - Population = Study participants enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP
  - Study participant-level outcome = The given variable and time point being summarized
  - Intercurrent event handling = An intercurrent event is defined as discontinuation of study treatment prior to the time point being summarized. This estimand targets the treatment difference in a scenario where the intercurrent event does not occur, such that outcomes for study participants without an intercurrent event are as observed at the given time point, and outcomes for study participants with an intercurrent event are treated as though they had completed the randomized study treatment through the time point being summarized
  - Population-level summary measure = Unadjusted mean.

  The same imputation strategy for handling missing data as the one used for the secondary continuous efficacy endpoints as described in Section 8.2.2 will be implemented. Missing data for ASDAS status endpoint will also be imputed using MI, and the imputation will be done on the components of ASDAS-CRP prior to calculating ASDAS-CRP and categorizing ASDAS status. The mean proportion of responders in the multiply imputed data sets (with the associated 95% CI based on SE) will be displayed for each ASDAS category.

In addition, a line plot by treatment group will be produced for some of the endpoints in the sequential testing hierarchy, displaying the mean change from Baseline for the following variables based on results obtained using MI: ASDAS-CRP, BASDAI, BASFI, Total Spinal Pain, Nocturnal Spinal Pain, PCS SF-36 and ASQoL. The plot will also be produced for ASDAS-CRP.

All other continuous and categorical variables will be summarized using descriptive statistics for Baseline, absolute values and change from Baseline at each visit (except for hs-CRP endpoint for which ratio to Baseline will be displayed). For other continuous variables whose value at Week 16 is part of the hierarchical testing procedure, descriptive statistics will also be provided using OC analysis. Categorical variables will be furthermore described using OC analysis.

The analyses on time to ASAS20/40 response will be restricted to the Double-Blind Treatment Period. Each time to ASAS20/40 response will be estimated and presented using the Kaplan-Meier product-limit method for each treatment group. The median time to response, including the 2-sided 95% CI, will be calculated for each treatment group. The Greenwood estimator will be used to estimate the variance of the survival. Comparisons of bimekizumab versus placebo will be analyzed using a log-rank test stratified by region and prior TNFα inhibitor exposure. The hazard ratio and 95% CI obtained from the Cox proportional hazards model as well as the p-value from the log-rank test will be displayed. The time points for calculating the number at risk is calculated in days but expressed in weeks, with each week corresponding to 7 days.

8.3.3 Additional statistical analysis of selected other efficacy variables

Although statistical testing is not planned for other efficacy variables per protocol, statistical testing and inferential statistics are of interest for selected endpoints. As these tests are not part of the multiplicity-controlled procedure, the associated p-values are considered nominal and should not be used for inference. Similarly as described in Section 8.2.3, statistical tests and calculation of inferential statistics will be performed for the following selected other efficacy endpoints:

- ASAS40 at Week 1, Week 2, Week 4, Week 8, Week 12
- ASAS40 in the TNF alpha inhibitor-naïve study participants at Week 1, Week 2, Week 4, Week 8, Week 12
- ASAS20 at Week 1, Week 2, Week 4, Week 8, Week 12
- PGADA at Week 1, Week 2, Week 4, Week 8, Week 12, Week 16
- Total spinal pain at Week 1, Week 2, Week 4, Week 8, Week 12, Week 16
- Nocturnal spinal pain at Week 1, Week 2, Week 4, Week 8, Week 12
- BASFI at Week 1, Week 2, Week 4, Week 8, Week 12
- Mean BASDAI Q5 and Q6 score at Week 1, Week 2, Week 4, Week 8, Week 12, Week 16
- BASDAI at Week 1, Week 2, Week 4, Week 8, Week 12,
- BASDAI50 at Week 1, Week 2, Week 4, Week 8, Week 12, Week 16
- CRP at Week 1, Week 2, Week 4, Week 8, Week 12, Week 16
- ASDAS-CRP at Week 16
- MASES at Week 16
- Enthesitis free state at Week 16
- TJC at Week 16
- SJC at Week 16
- FACIT-Fatigue at Week 16
- WPAI at Week 16
- ASspiMRI at Week 16
- SPARCC at Week 16

The same models as described in Section 8.2.3 will be run, and results will be presented in one single table together with the results for secondary endpoints.

In addition, the Week 16 results from the primary analysis for the primary endpoint ASAS40, and key secondary endpoints in the hierarchy (ASAS20, BASDAI, BASFI, nocturnal spinal pain) will also be included in this table.

9 PHARMACOKINETICS AND IMMUNOGENICITY

9.1 Pharmacokinetics

Only study participants treated with bimekizumab will be included in the PK analyses. For the study participants randomized to placebo group, Baseline will be the Week 16 pre-dose sample.

9.1.1 Definition of ADAb Sample Status (positive, negative, missing)

ADAb sample results will be provided as "Positive Screen" (PS) or "Negative Screen" (NS). For samples presenting anti-BKZ antibody levels that are PS, further confirmatory assay will be performed, and the result of which will be reported as either "Positive Immunodepletion" (PI) or "Negative Immunodepletion" (NI).

ADAb status for each sample will be derived as following:

- Sample values that are either NS, or PS and NI and where the BKZ concentration is less than the validated ADAb assay drug tolerance limit (200 µg/mL) will be defined as ADAb negative.
- Sample values that are either NS, or PS and NI and where the BKZ concentration exceeds the validated ADAb assay drug tolerance limit will be defined as inconclusive.
- Sample values that are PS and PI will be defined as ADAb positive (regardless of availability of a titer value).
- Missing or non-evaluable samples will be defined as missing.

PI samples will be titrated, and the ADAb titer (reciprocal dilution factor including MRD) reported. The PI samples will also be subject to a neutralizing assay to evaluate whether the anti-BKZ antibody neutralizes target binding of bimekizumab (IL17AA or IL17FF or both) in-vitro (as described in the next section).

9.1.2 Definition of Cumulative ADAb Status

The ADAb status (positive, negative or missing) will be considered in a cumulative manner at each time point.

A study participant will be counted positive from the first visit at which the study participant achieved a positive ADAb sample result to the end of the treatment period, regardless of any missing/inconclusive or negative ADAb sample result.

If a study participant has only negative ADAb samples or only one missing/inconclusive sample with all negative ADAb samples up to that timepoint, the study participant will be classified as negative. An exception remains for the Baseline Visit where only one sample could be available. If the sample is missing/inconclusive, then the sample will be classified as being negative for the cumulative ADAb status.

Otherwise, the study participant will be classified in the missing ADAb category.

9.1.3 PK Sample Inclusion and Exclusion Rules

PK samples collected at all visits except the SFU visit and meeting the following requirements will be included in summaries and corresponding figures. Samples not meeting the following requirements will be excluded from summaries and figures, and only listed:

- For PK samples associated with Week 2: include samples collected ≥1 week and ≤ 3 weeks after the first BKZ dose and before the subsequent BKZ dose.
- For PK samples associated with Week 52 or the ET visit: include samples collected >14 days and <42 days after the last/previous dose received.
- For all other visits: include samples collected >14 days after the preceding dose, <42 days after the preceding dose, and no later than 1 hour after the current visit dose including unscheduled assessments as described in Section 3.1.

For the SFU visit, all concentrations obtained at the SFU visit will be included in the summary tables but will not be included in the figures.

When multiple samples meeting the criteria above are associated with the same visit (either because multiple samples were collected or due to remapping or unscheduled visits), only results from one valid sample will be included in summaries and figures using the following rules. All others will be excluded and only listed:

- For PK samples with different dates, the sample that is closest to the target visit date will be included. For samples with different dates that are the same distance to the target visit date, the sample collected prior to IMP dosing will be included.
- For PK samples with the same date, the first sample as provided in the raw data will be included.

In addition, a PK sample could be excluded in the following situations:

- Sample collected after the study participants received more than one dose of BKZ at a single visit, or multiple doses separated by a few days,
- Sample collected after the study participants received less than a complete dose at a single visit,

In each of these 2 possible scenarios, the study participants will be flagged with an appropriate protocol deviation, and the sample to be excluded will be identified based on the review of the description of the deviation.

All excluded samples will be assigned one or more reasons for exclusion:

- Sample collected out of window relative to current dose
  - For PK samples associated with Week 2: includes samples collected <1 week or > 3 weeks after the first BKZ dose.
  - For PK samples associated with visits other than week 2, 52, ET, or SFU: includes samples >1 hour after the current visit dose.
- Sample collected out of window relative to previous dose
  - For PK samples associated with all visits except the Week 2 or SFU visit: includes samples collected <14 days or >42 days after the previous dose received.
- More than one sample obtained at the same visit
  - Includes all samples excluded due to multiple valid samples associated with a visit.

- Sample collected after the study participant received more than 160 mg of study drug in previous 4 week period
- Sample collected after the study participant received less than 160mg of study drug at previous visit

9.1.4 Summary of Analyses

PK variables will be summarized by treatment group (BKZ 160 mg Q4W and Placebo / BKZ 160 mg Q4W).

Bimekizumab plasma concentrations will be summarized for the above treatment groups, at each scheduled visit at which samples were taken, using the PK-PPS analysis set and based on observed values. No imputation will be used for missing samples. However, if plasma concentration measurements are below the level of quantification (BLQ), then for calculation of the derived statistics, the result will be set to ½ of the lower level of quantification (LLOQ) (ie, 1/2*0.250 = 0.125 ug/mL). Descriptive statistics including number of values, geometric mean, geometric mean 95% CI, geometric coefficient of variation, mean, standard deviation, median, minimum, and maximum if applicable will be calculated if at least 2/3 of the values of interest are above the LLOQ and n>=3. Otherwise, only number of values, median, minimum, and maximum will be presented.

In addition, the geometric mean of bimekizumab plasma concentration (with the 95% CI) will be plotted versus time on linear and semi-logarithmic scales. These figures will be generated overall for each treatment group.

The table summary and figures will be repeated by cumulative anti-bimekizumab antibody status as defined in Section 9.1.2, for the overall population. The figures will be produced separately for each treatment group.

All plasma concentration data will be reported in ug/mL in the tables, figures and listings.

Bimekizumab plasma concentrations will be listed for the SS. All concentrations will be listed as received prior to substitution of any BLQ values. The listing will include flags for concentrations that were excluded from the summary statistics and include the reason for exclusion as defined in Section 9.1.3.

If more than 10% of the PK concentration results have been excluded from the table summaries, the PK excluded results will be listed in a separate listing.

9.2 Immunology (Immunogenicity)

9.2.1 Anti-bimekizumab antibody

ADAb samples are analyzed when participants are on bimekizumab treatment. ADAb will be assessed using a three-tiered assay approach: screening assay, confirmatory assay and titration assay. Samples confirmed as positive within the confirmatory assay will be further evaluated for the presence of neutralizing anti-bimekizumab antibodies (NAb) specific to IL-17AA, IL-17FF or both (described in the next section).

Screening, confirmatory and titer cut points of the respective assays will be determined by the bioanalytical laboratory.

9.2.1.1 Definition of ADAb Sample Status (Positive, Negative, Missing)

This will be as defined in Section 9.1.1 above.

9.2.1.2 Definition of Overall ADAb Status (Positive, Negative, Missing)

A participant's Overall ADAb Status will be derived for the following periods:

- Overall ADAb Status in the Treatment Period: Including any visit during the treatment period (as defined in Section 3.2.2). Thus, this summary will exclude data obtained at the SFU visit and will include data obtained at Baseline. This summary will display the overall ADAb status up to Weeks 16, 24 and 52 by treatment group (BKZ 160 mg Q4W, Placebo / BKZ 160 mg Q4W).
- Overall ADAb Status including SFU: Including any visit during the study. Thus, this summary will include data obtained at both the SFU visit and at Baseline.

Study participants who rolled over to the OLE study will not have a SFU visit per study protocol. For these study participants, the 'overall ADAb status including SFU' will then be considered as being identical to the 'overall ADAb status up to Week 52'.

For each time period, the participant will be classified as:

- Positive if the study participant has at least one positive sample up to the time point of interest (regardless of having missing/inconclusive data).
- Negative if the study participant has all the samples negative or only one missing/inconclusive sample with negative ADAb samples up to the timepoint of interest.
- Missing if the study participant has missed more than one ADAb result (or have more than one inconclusive sample) and all other available ADAb samples are negative up to the time point of interest.

9.2.1.3 Definition of ADAb Categories

In addition, the anti-BKZ antibody status will be further classified as outlined below:

- **Pre ADAb negative – treatment emergent ADAb negative (Category 1):** Includes study participants who are negative at Baseline and antibody negative at all sampling points (including SFU), one post-Baseline missing/inconclusive sample is allowed for study participants with pre-ADAb negative sample. This group also includes study participants who have a missing/inconclusive pre-treatment sample (eg, either missing/inconclusive or insufficient volume) at Baseline with all post-Baseline samples as ADAb negative.
- **Pre ADAb negative – treatment emergent ADAb positive (Category 2)**: Includes study participants who are ADAb negative at Baseline and ADAb positive at any sampling point post treatment (including SFU). This group also includes study participants who have a missing pre-treatment sample (either missing or insufficient volume) at Baseline with one or more ADAb positive post-treatment samples.
- **Pre ADAb positive – treatment emergent reduced ADAb (Category 3):** Includes study participants who are ADAb positive at Baseline, and antibody negative at all sampling points post-treatment (including SFU).
- **Pre ADAb positive – treatment emergent unaffected ADAb positive (Category 4**): Includes study participants who are positive at Baseline and are positive at any sampling

point post-treatment (including SFU) and all post-Baseline samples have titer values < 2.07 times the Baseline titer (where 2.07 is Minimum Significant Ratio (MSR))

- **Pre ADAb positive – treatment emergent ADAb boosted positive (Category 5):** Includes study participants who are positive at Baseline and have at least one positive post-treatment sample (including SFU) that has titer value ≥ 2.07 times the Baseline titer (where 2.07 is Minimum Significant Ratio (MSR)

  Note: For any study participant who is positive at Baseline and positive at a post-Baseline time point, but for whom titers are not available to determine treatment unaffected or treatment boosted status, the study participant will be considered as treatment boosted (ie., Category 5), assuming no other samples are available.

- **Inconclusive (Category 6)**: Includes study participants who have a positive pre-treatment sample and some post-treatment samples are missing/inconclusive, while other post-treatment samples are ADAb negative.

- **Total treatment-emergent (Category 7** [Categories 2 and 5 combined]): Includes study participants who are pre ADAb negative – treatment emergent ADAb positive (Category 2) and pre ADAb positive – treatment boosted ADAb positive (Category 5).

- **Total prevalence of pre-ADAb positivity (Category 8** [Categories 3, 4, 5 and 6 combined]): Study participants that are tested ADAb positive at Baseline.

- **Missing (Category 9):** Includes study participants who have a negative or missing/inconclusive pre-treatment sample at Baseline and more than one post-treatment samples are missing/inconclusive, while other samples are ADAb negative. Includes also study participants who have missing/ inconclusive pre-treatment sample at Baseline, and only one missing post-baseline sample with no ADAb positive.

9.2.1.4 ADAb status groups for Efficacy Endpoint Summaries

For the Efficacy Endpoint by ADAb Summaries (described below), the following ADAb-efficacy status groups will be defined:

- ADAb positive - Defined as study participants having at least 2 ADAb positive samples up to the time point of interest (i.e., excluding Baseline, excluding SFU) regardless of other ADAb negative samples and/or missing or inconclusive samples

- ADAb negative - Defined as study participants for whom either (1) all samples to the time point of interest are ADAb negative and there are no missing or inconclusive samples, (2) only 1 sample is ADAb positive and all other to the time point of interest (including Baseline) are ADAb negative or missing/inconclusive or 3) only one sample is missing/inconclusive and the remaining ADAb samples are negative.

- Missing - Defined as study participants who do not fulfil the criteria for one of the 2 groups listed above.

ADAb data from Baseline up to the timepoint of interest will be used to derive the ADAb-efficacy status groups.

9.2.1.5 ADAb Sample Inclusion and Exclusion Rules

The following rules will be implemented for by-visit ADAb summaries and apply to all visits except Week 52, ET and SFU:

- From Baseline through Week 24 Visit, if the ADAb sample is collected within ±14 days relative to the visit date at which BKZ was administered (or ±14 days from a scheduled visit at which BKZ dosing was not performed), the ADAb result for that sample will be associated with the scheduled visit and summarized accordingly.
- After the Week 24 Visit, if the ADAb sample is collected within ±21 days relative to the visit date at which BKZ was administered (or ±21 days from a scheduled visit at which BKZ dosing was not performed), the ADAb result for that sample will be associated with the scheduled visit and summarized accordingly.
- The two rules described above will include unscheduled assessments as described in Section 3.1 if a dose was administered at the unscheduled visit.

Samples collected outside the relevant windows described above will be excluded from the ADAb by-visit summaries and will be listed only.

For samples associated with the Week 52, ET or SFU visits, all ADAb data obtained will be included in the by-visit summaries.

All other summaries of ADAb status (as described in the previous section) will use all available (scheduled and unscheduled) data.

The reason for exclusion will be assigned as one or more of the following:

- Sample collected out of window relative to current dose (or visit)
- More than one sample obtained at the same visit

9.2.2 Summary of Analyses

9.2.2.1 Immunogenicity Summaries

All the following analyses will be prepared on the Safety Set population by treatment group (BKZ 160 mg Q4W, Placebo / BKZ 160 mg Q4W), unless notified differently:

- Summary table displaying the number and percentage of study participants with ADAb sample status (positive, negative, total of positive and negative, missing) at each visit, overall ADAb status in the treatment period, and overall ADAb status including SFU as defined above. If there are ≥95% of study participants included in the non-missing groups, the missing group will not be displayed on the table.
- The above summary table displaying the number and percentage of study participants in each of the ADAb status (positive, negative, total of positive and negative, missing) will be repeated by region. The overall ADAb Status in the Treatment Period will be presented for up to and including Weeks 16, 24, 52 and SFU.
- A summary table of the number (%) of study participants with the first occurrence of any ADAb positivity or treatment-emergent positivity during the study will be produced. This summary will include the following categories:
  - Any ADAb+: ADAb positive sample regardless of category during the treatment period

- Category 2: Pre ADAb negative – treatment-emergent ADAb positive
- Category 5: Pre ADAb positive – treatment-boosted ADAb positive

The table will summarize the number and percentage of study participants with first occurrence of any ADAb+ sample, and also study participants who are either treatment-emergent ADAb positive or treatment-boosted ADAb positive for the first time at the specified time point in the study and will include the cumulative number and percentage of study participants with any ADAb+ sample and treatment-emergent ADAb positive results at each time point.

Treatment-emergent ADAb+ in this table will be based on ADAb category 7 (total treatment-emergent ADAb).

In the event that ≥ 10% of study participants are classified as category 5, the summaries will be further split by the following categories:

- Category 2: Pre ADAb negative – treatment-emergent ADAb positive
- Category 5: Pre ADAb positive – treatment-boosted ADAb positive

- A box-plot of the ADAb titer by time of occurrence of ADAb positivity will be created for BKZ 160mg Q4W group, as well as for the placebo/BKZ 160mg Q4W group. The ADAb titer results will be presented on the log-scale. The time of occurrence is defined as the time in weeks from Baseline until the visit of interest when a sample is ADAb positive. Study participants who don't have any positive ADAb samples post-Baseline will be excluded from this analysis.
- A summary table displaying the number and percentage of study participants in each of the 9 ADAb categories as defined above in Section 9.2.1.3 by treatment group.
- Figure summarizing the time to achieving any ADAb positivity on a cumulative basis will be presented by treatment group. Similarly, figure representing the time to achieving treatment-emergent ADAb positivity by treatment group will be plotted. Study participants will be considered to have an event at the time point at which an ADAb+ result (or treatment-emergent ADAb positivity) is first achieved. Treatment-emergent ADAb+ in this plot will be based on ADAb category 7 (total treatment-emergent ADAb).

  In the event that >= 10% of study participants are classified as category 5, the lines will be further split by the following categories (thus the plot will include 3 lines per treatment group):

  - Category 2: Pre ADAb negative – treatment-emergent ADAb positive
  - Category 5: Pre ADAb positive – treatment-boosted ADAb positive

9.2.2.2 Efficacy Endpoint by ADAb Summaries

Several figures, tables, and listings will be created:

- Figures will summarize efficacy response (ASAS40 and ASAS20 responders based on NRI) versus ADAb titer quartiles. The x-axis will display the ADAb titer quartiles at Week 16 (categorized as negative, Q1, Q2, Q3 and Q4) and the y-axis will display percentage of ASAS40 (or ASAS20) responders at Week 16.

The plots will therefore display the percentage of ASAS40 (or ASAS20) responders as a function of the number of study participants within each ADAb titer categories. Study participants with negative ADAb results at Week 16 will be included in the 'negative' category on the x-axis. This plot will include data from study participants only randomized to BKZ 160 mg Q4W.

For each efficacy endpoint (ASAS40 and ASAS20 responders), the figure will be repeated using Week 24 data and then using Week 52 data, These figures will include all study study participants receiving BKZ and will be split by treatment group (BKZ 160 mg Q4W throughout, Placebo/ BKZ160 mg Q4W starting at Week 16). Thus, 2 figures per efficacy variable and time point will be presented.

- Two figures will summarize efficacy response (ASAS40 responders based on NRI) over time for each treatment group (Placebo/BKZ 160 mg QW4 (starting at Week 16) and BKZ throughout) by the ADAb-efficacy status groups defined above in Section 9.2.1.4 (3 lines per plot).

  If ≥95% of study participants are included in the non-missing groups, the missing group will not be displayed on the figure.

  The first two figures will present data up to Week 24. Two similar figures will present data up to Week 52. These four figures will then be repeated for the ASAS20 endpoint. For each figure, the ADAb-efficacy status groups will be calculated based on data up to the timepoint of interest.

  An additional figure for BKZ 160 mg Q4W group only will be generated for ASAS40 and ASAS20 with data up to Week 16.

  The percent of study participants in each of the ADAb-efficacy status groups considered for efficacy sub-group analyses will be summarized.

- The ASAS20 and ASAS40 responder rates over time by ADAb-efficacy status groups defined above in Section 9.2.1.4 up to Week 16 will be presented in a tabular format with ASAS20 and ASAS40 results presented side by side (the analysis will be presented for BKZ 160 mg Q4W only). The analysis will be repeated by ADAb-efficacy status groups up to Week 24, up to Week 52, and up to SFU Visit and presented by treatment group (BKZ 160 mg Q4W and placebo / BKZ 160 mg Q4W).

- Individual plots of bimekizumab concentrations, ADAb titer and ASAS40 response (based on NRI) will be created. All three endpoints will be plotted on the Y-axes by visit (x-axis) for the full treatment period, including SFU where a study participant has not progressed into the OLE. Plots should be labeled and grouped into the 7 pre-defined ADAb categories (described above in Section 9.2.1.3). For the Week 24 interim analysis, the 9 ADAb categories will be defined based on data up to Week 24.

- Spaghetti plots of ADAb titer (y-axis) will be reported on the log-scale by visit (x-axis) by treatment group, for all study participant in the following ADAb categories:
  - Category 2: Pre ADAb negative – treatment-emergent ADAb positive
  - Category 5: Pre ADAb positive – treatment-boosted ADAb positive

Plots will be presented using a semi-logarithmic scale for the ADAb titers (ADAb negative samples will therefore be excluded from the plot).

- All individual study participant-level ADAb results will be listed including the screening assay, confirmatory assay, ADAb sample status and titers if applicable. Note, that titer results will only be available if the confirmatory assay is positive. The listing will also include flags and reasons for exclusion for ADAb measurements that were excluded from the by-visit summaries, as well as information on whether the BKZ concentration exceed the ADAb assay drug tolerance limit.
- Finally, if more than 10% of the ADAb results have been excluded from the table summaries, the ADAb excluded results will be listed in the same listing as the one mentioned in Section 9.1.4 for excluded PK results.

9.2.3 Neutralizing anti-bimekizumab antibodies

NAb will be assessed using IL-17AA- and IL-17FF-specific assay methods, respectively. NAb results will be positive or negative to each specific NAb assay or both.

Any NAb results derived from samples with drug concentrations higher than the drug tolerance limits of the NAb assays (100µg/mL) will be labeled 'inconclusive'. All inconclusive results will be regarded as missing.

Participants will be assigned an overall NAb classification, inclusive of Baseline and post-Baseline results from NAb assay:

- NAb negative: No NAb positive samples for IL-17AA and IL-17FF at Baseline or post-Baseline. This group will also include study participants who have only 1 missing sample and all other available samples during the period of interest are negative. Study participants who are NAb negative will be further classified as follows:
  - ADAb positive / NAb negative: ADAb positive study participants who are 1) NAb negative for all available ADAb positive samples or 2) with only one missing NAb sample and all other evaluated ADAb positive samples are NAb negative.
  - ADAb negative: if the study participant has all the samples as ADAb negative or only one missing/inconclusive sample with all other available samples as negative ADAb. Note that ADAb negative samples are not participants to the neutralizing assay.
- NAb positive: One or more positive samples (IL-17AA positive, IL-17FF positive, or both) at Baseline or post-Baseline (regardless of missing samples) Study participants who are NAb positive will be further classified as follows:
  - Positive for IL-17AA only: one or more positive samples for IL-17AA at Baseline or post-Baseline. No positive samples for IL-17FF
  - Positive for IL-17FF only: one or more positive samples for IL-17 FF at Baseline or post-Baseline. No positive samples for IL-17AA
  - Positive for both IL-17AA and IL-17FF: one or more positive samples for both IL-17AA and IL-17FF at Baseline or post-Baseline

    Drug tolerance limit for both the IL-17AA-specific NAb method and the IL-17FF-specific NAb method is 100 µg/mL.

- NAb Missing: more than one relevant NAb samples are missing/inconclusive and other available NAb samples during the period of interest are negative, eg, missing or insufficient sample left for NAb testing.

Summary tables and a data listing will be produced to summarize the NAb status overall in the study.

The listing will be sorted by treatment group, participant identifier, and visit and, will summarize the following information for each participant assessed for NAbs:

- Visit
- Study week since first bimekizumab dose
- laboratory sampling date and time
- Time since previous dose (weeks)
- The corresponding bimekizumab plasma concentration level at each visit (ug/mL)
- Anti-bimekizumab anti-body titer at each visit
- IL-17AA NAb status and corresponding IL-17AA signal/negative control result
- IL-17FF NAb status and corresponding IL-17FF signal/negative control result

The table will provide the following overall summary statistics by treatment group (based on the total number of participants):

- The number and percentage of participants confirmed as anti-bimekizumab antibody positive and anti-bimekizumab antibody negative at any visit.
- Number and percentage of participants who are NAb positive, NAb negative and missing at any visit.

The NAb summary table will be repeated with percentages calculated on a denominator based on the number of ADAb positive participants at any visit in the study.

The table summaries will display the overall NAb categories derived based on data from Baseline up to Weeks 16, 24, 52 and SFU by dose group (BKZ 160 mg Q4W and placebo / BKZ 160 mg Q4W).

For the analysis of NAb effect on pharmacokinetics, efficacy and safety of bimekizumab, the data from AS0011 may be pooled with other phase 3 studies for an integrated analysis and may be reported elsewhere.

10 SAFETY ANALYSES

Safety variables will be analyzed on the SS, as well as on the MS for analyses covering the Maintenance Period (MP) only.

Summaries for the MP will also include the SFU period for study participants that do not roll enter the OLE or discontinue early in the MP. The SFU period for study participants that discontinue in the DBTP will only be summarized in the Overall Period tables.

10.1 Extent of exposure

Two summary tables will be created to describe the extent of exposure.

The first table consists of a descriptive summary of study medication duration (in days), total study medication duration in (subject-years), and total time at risk (in subject-years). The information will be presented overall (ie, for the entire treatment period (Double-Blind Treatment Period (DBTP) and Maintenance Period (MP) combined)) as well as by treatment periods.

Study medication duration in subject-years is obtained as study medication in days divided by 365.25. Similarly, total time at risk in subject-years is obtained as total time at risk in days divided by 365.25.

A second summary will be created on the cumulative study medication duration for the overall period, as well as by treatment periods.

- The following time durations categories will be considered for the overall period: >=0 weeks, >=16 weeks, >=24 weeks, >=48 weeks.
- The following time durations categories will be considered for the Double-Blind Treatment Period:>=0 weeks, >=4 weeks, >=8 weeks, >=12 weeks.
- The following time durations categories will be considered for the Maintenance Period:>=0 weeks, >=16 weeks, >=24 weeks, >=36 weeks.

A by-study participant listing of date of first and last dose and the duration of exposure will be performed.

The study medication duration (in days) and the time at risk (in days) are defined depending on the study period of interest. See Section 10.1.1 to Section 10.1.5 for further details.

Throughout this section, date of last clinical contact for each study participant is defined as the maximum of [last visit date including SFU visit, last AE start date (including imputed AE start dates), date of study termination or completion, last date of study drug administration].

10.1.1 Duration of exposure and time at risk in the Double-Blind Treatment Period (DBTP)

- The study medication duration during the DBTP will be calculated as follows:

$$Date\ of\ last\ dose\ (DBTP) - Date\ of\ first\ dose\ (DBTP) + 28$$

(this formula also applies for study participants who discontinue/lost to follow-up during the DBTP).

  - If the date of last dose of the DBTP + 28 days extends to a date beyond the date of first dose of the Maintenance period, then this calculation reverts to:

$$Date\ of\ first\ dose\ (MP) - Date\ of\ first\ dose\ (DBTP)\ + 1$$

  Note: If the date of last dose + 28 extends to a date beyond the final visit date (not including SFU), then this calculation reverts to:

$$Date\ of\ last\ visit\ (not\ including\ SFU) -\ Date\ of\ first\ dose\ + 1$$

  - For study participants who die during the DBTP, this calculation reverts to:

$$Date\ of\ death - Date\ of\ first\ dose\ (DBTP) + 1$$

- The time at risk during the DBTP will be calculated as follows:

- For study participants who complete the final visit of the DBTP and continue to the Maintenance Period:

$$Date\ of\ first\ dose\ (MP) - Date\ of\ first\ dose\ (DBTP) + 1$$

- For study participants who discontinue on or prior to the final visit of the DBTP (ie, Week 16), use the minimum of the following:
  - Total number of days in the DBTP:
    - For those that discontinue prior to the Week 16 visit, use 112 days,
    - For those that discontinue at the Week 16 visit, use the total number of days based on Week 16 visit date, which can be a maximum of 115 days due to visit windows
  - $Date\ of\ last\ clinical\ contact - Date\ of\ first\ dose\ (DBTP) + 1$, last clinical contact can be defined as last scheduled site visit.
- For study participants who die during the DBTP, then this calculation reverts to:

$$Date\ of\ death - Date\ of\ first\ dose\ (DBTP) + 1$$

10.1.2 Duration of exposure and time at risk in the Maintenance Period

- The study medication duration during the Maintenance Period will be calculated as follows:

$$Date\ of\ last\ dose\ (MP) - Date\ of\ first\ dose\ (MP) + 28$$

- If date of last dose in the Maintenance Period + 28 days extends to a date beyond the final visit date of the Maintenance Period (not including SFU), then this calculation reverts to:

$$Date\ of\ last\ visit\ (MP - not\ including\ SFU) - Date\ of\ first\ dose\ (MP) + 1$$

- For study participants who die during the Maintenance Period, then this calculation reverts to:

$$Date\ of\ death - Date\ of\ first\ dose\ (MP) + 1$$

- For study participants who discontinue/lost to follow-up during the Maintenance Period, then this calculation reverts to:

$$Date\ of\ last\ dose\ (MP) - Date\ of\ first\ dose\ (MP) + 28$$

- The time at risk during the Maintenance Period will be calculated as follows:
  - For study participants who complete the Maintenance Period as planned and continue into an open-label study (and, therefore, do not have the SFU visit in the feeder study):

$$Date\ of\ last\ visit\ (MP) - Date\ of\ first\ dose\ (MP) + 1$$

  - For study participants who die during the Maintenance Period, then this calculation reverts to:

$$Date\ of\ death - Date\ of\ first\ dose\ (MP) + 1$$

  - For all other study participants, use the minimum of the following:
    - $Date\ of\ last\ dose\ (MP) - Date\ of\ first\ dose\ (MP) + 140 + 1$
    - $Date\ of\ last\ clinical\ contact - Date\ of\ first\ dose\ (MP) + 1$

This last group could include study participants who discontinue the Maintenance Period early, study participants who complete the Maintenance Period as scheduled but choose not to continue into an open-label study, or study participants who are ongoing in the SFU period at the time of the interim data snapshot.

Note that the Week 24 visit (ie, Visit 10) will be the last visit for the Week 24 interim analysis, and that the Week 52 visit (ie, Visit 17) will be the last visit for the Week 52 interim analysis.

10.1.3 Duration of exposure and time at risk in the combined Double-Blind Treatment and Maintenance Periods

- The study medication duration during the combined treatment period will be calculated as follows:
  - For study participants who do not switch study treatments per study design:

$$Date\ of\ last\ dose - Date\ of\ first\ dose + 28$$

    - If the date of last dose + 28 days extends to a date beyond the date of the final visit date (not including SFU), then this calculation reverts to:

$$Date\ of\ last\ visit\ (not\ including\ SFU) - Date\ of\ first\ dose + 1$$

    - For study participants who die, then this calculation reverts to:

$$Date\ of\ death - Date\ of\ first\ dose + 1$$

  - For study participants who switch study treatments per study design (ie, study participants initially randomized to placebo who are re-allocated to bimekizumab treatment):
    - During the DBTP (attributed to initially randomized treatment):

$$Date\ of\ last\ dose\ (DBTP) - Date\ of\ first\ dose\ (DBTP) + 28$$

      Note: If date of last dose in the Initial Period + 28 days extends to a date beyond the date of first dose in the Maintenance Period, then this calculation reverts to:

$$Date\ of\ first\ dose\ (MP) - Date\ of\ first\ dose\ (DBTP) + 1$$

    - During the Maintenance Period (attributed to the treatment initiated in the Maintenance Period), the study medication duration algorithm specified for the Maintenance Period in Section 10.1.2 applies.
- Definitions for time at risk during the combined treatment period will be calculated as follows:
  - For study participants who do not switch study treatments per study design:
    - For study participants who complete the Maintenance Period as planned and continue into an open-label study (and, therefore, do not have the SFU visit in the feeder study):

$$Date\ of\ last\ visit - Date\ of\ first\ dose + 1$$

    - For study participants who die prior to the final visit:

$$Date\ of\ death - Date\ of\ first\ dose + 1$$

- For all other study participants, use the minimum of the following:
  - $Date\ of\ last\ dose - Date\ of\ first\ dose + 140 + 1$
  - $Date\ of\ last\ clinical\ contact - Date\ of\ first\ dose + 1$

  This last group could include study participants who discontinue early, study participants who complete the Maintenance Period as scheduled but choose not to continue into an open-label study, or study participants who are ongoing in the SFU period at the time of the data snapshot.

– For study participants who switch study treatments per study design:
  - During the DBTP (attributed to initially randomized treatment):

    $$Date\ of\ first\ dose\ (MP) - 1 - Date\ of\ first\ dose\ (DPTB) + 1$$

    This assumes that any study participant in this category has completed the DBTP and doses [with a new study treatment] in the Maintenance Period.
  - During the Maintenance Period (attributed to the treatment initiated in the Maintenance Period), the time at risk algorithm specified for the Maintenance Period in Section 10.1.2 applies.

Note that the Week 24 visit (ie, Visit 10) will be the last visit for the Week 24 interim analysis, and Week 52 visit (ie, Visit 17) will be the last visit for the Week 52 interim analysis.

In addition, Week 20 visit will correspond to the last dose of the Week 24 interim analysis. and the Week 48 visit will correspond to the last dose of the Week 52 interim analysis.

10.1.4 Duration of exposure by COVID-19 pandemic period

- If the pandemic end date is known, then the study medication exposure duration will be calculated as follows:

$$date\ of\ last\ exposure\ date\ on\ or\ after\ pandemic\ start\ date - date\ of\ first\ exposure\ on\ or\ before\ pandemic\ end\ date + 28$$

- If the pandemic end date is unknown, then the study medication exposure duration will be calculated as follows:
  – If the first exposure date is on or after the pandemic start date, then:
    - For study participants who did not die:

      date of last exposure - date of first BKZ dose + 28

    Note: If the last exposure date + 28 extends to a date beyond the final visit date (not including SFU), then this calculation reverts to:

    date of last visit (not including SFU) – date of first BKZ dose +1
    - For study participants who die on or after the pandemic start date:

      date of death - date of first BKZ dose + 1

- If the first exposure date is before the pandemic start date and the study drug period is overlapping with the pandemic (ie, the study participant started or ended the treatment during the pandemic) then:
  - For study participants who did not die:

date of last exposure - 11Mar2020 + 28

Note: If the last exposure date + 28 extends to a date beyond the final visit date (not including SFU), then this calculation reverts to:

date of last visit (not including SFU) – 11Mar2020 +1

  - For study participants who die:

date of death - 11Mar2020 + 1

10.1.5 Time at risk in subject-years by COVID-19 pandemic period

In this section, "study drug" refers to BKZ treatment.

10.1.5.1 Time at risk in the pre-COVID-19 pandemic period

Only study participants starting study drug during the "pre-COVID-19 pandemic period" will have a time at risk calculated during that period. The following rules will be applied:

- For study participants who did not discontinue medication in the "pre-COVID-19 pandemic period" (ie, for study participants with date of last dose of study drug on or after 11Mar2020):

10Mar2020 – date of first dose + 1

- For study participants who have died on or before 10Mar2020:

date of death – date of first dose + 1

Note: If the date of death is later than 140 days following the last dose of BKZ, this rule does not apply, and the duration of exposure will follow the rule below.

- For study participants who discontinued medication in the "pre-COVID-19 pandemic period" (ie, for study participants with date of last dose of study drug before 11Mar2020):
  - If the earliest date between (date of last dose + 140) and (date of last clinical contact) is on or after 11Mar2020, then:

10Mar2020 – date of first dose + 1

  - If the earliest date between (date of last dose + 140) and (date of last clinical contact) is before 11Mar2020, then use the minimum of:
    - date of last dose + 140 – date of first dose +1
    - date of last clinical contact – date of first dose +1

10.1.5.2 Time at risk in the COVID-19 pandemic period

This period is applicable for all study participants in the SS who are still considered at risk on the 11MAR2020 (ie, those who have their first dose is on or after 11MAR2020, or for those for whom either dosing is continuing, or for whom the 140 day SFU period after last dose has not

been completed [in the case of premature treatment discontinuation or for those who do not roll over to the OLE]). The following rules will be applied:

- For study participants who started IMP prior to 11MAR2020 and did not discontinue medication in the "during COVID-19 pandemic period" (ie, for study participants with the last dose of BKZ 160mg is on or after the date of the end of the pandemic +1):

  date of the end of pandemic – 11Mar2020 + 1

- For study participants who started IMP on or after 11MAR2020 and did not discontinue medication in the "during COVID-19 pandemic period" (ie, for study participants with the last dose of BKZ 160mg is on or after the date of the end of the pandemic +1):

  date of the end of pandemic – date of first dose + 1

- For study participants who have died on or after 11Mar2020, and on or before the pandemic end date:

  date of death – maximum (11MAR2020, date of first dose) + 1

  Note: If the date of death is later than 140 days following the last dose of BKZ, this rule does not apply.

- For study participants who discontinued medication:
  - in the "pre-COVID-19 pandemic period" (ie, for study participants with date of last dose of study drug before 11Mar2020), and if the earliest date between (date of last dose + 140) and (date of last clinical contact) is on or after 11Mar2020, then use minimum between:
    - date of last dose + 140 – 11Mar2020 +1
    - date of last clinical contact – 11Mar2020 + 1
  - in the "during COVID-19 pandemic period" (ie, for study participants with date of last dose of study drug on or after 11Mar2020, and on or before the pandemic end date),
    - For study participants who started IMP prior to 11MAR2020, and if the earliest date between (date of last dose + 140) and (date of last clinical contact) is after the end of pandemic, then use:

      date of end of pandemic – 11MAR2020 + 1

    - For study participants who started IMP on or after 11MAR2020, and if the earliest date between (date of last dose + 140) and (date of last clinical contact) is after the end of pandemic, then use:

      date of end of pandemic – date of first dose + 1

    - Else time at risk will be calculated as minimum between:
      - date of last dose date + 140 – maximum (11MAR2020, date of first dose) + 1
      - date of last clinical contact – maximum (11MAR2020, date of first dose) + 1

10.1.5.3 Time at risk in the post-COVID-19 pandemic period

This period is applicable for all study participants in the SS who are still considered at risk on the day after the end date of the pandemic (ie, those for whom either dosing is continuing or for

whom the 140 day SFU period after last dose has not been completed [in the case of premature treatment discontinuation or for those who do not roll over to the OLE]). The following rules will be applied:

- For study participants who did not discontinue medication in the "during COVID-19 pandemic period" (ie, for study participants with date of last dose of study drug after the pandemic period end date), then use minimum between:
  - date *of last dose + 140 – date of the end of pandemic*
  - date *of last clinical contact – date of the end of pandemic*
- For study participants who have died after the *pandemic end date*:

  date of death – date of the end of pandemic

  Note: If the date of death is later than 140 days following the last dose of BKZ, this rule does not apply.
- For study participants who discontinued medication in the "during COVID-19 pandemic period" (ie, for study participants with date of last dose of study drug on or before the pandemic period end date), and if the earliest date between (date of last dose + 140) and (date of last clinical contact) is after the pandemic period end date, then use minimum between:
  - date of last dose + 140 – date of the end of pandemic
  - date of last clinical contact – date of the end of pandemic

To assess the impact of the COVID-19 pandemic, the study medication duration and time at risk will be reported by COVID-19 pandemic period.

10.2 Adverse events

Adverse events with start date prior to first administration of study medication are defined as pre-treatment AEs. No summaries of these events will typically be produced although they will be included in listings.

Adverse events that started more than 140 days after the last administration of study medication are defined as post-treatment AEs. Such events will not be included in any tabulated summaries but will be listed.

Treatment-emergent AEs (TEAEs) are defined as those AEs that have a start date on or following the first dose of study treatment through the final dose of study treatment + 140 days (covering the 20-week SFU period).

All summaries will be provided by treatment group. The treatment group for each study participant will be defined as the treatment received at the onset of the AE. If the onset of the AE occurs on the day where the study participant switches treatment from the Double-Blind Treatment Period to the Maintenance Period and receives a new treatment, it will be attributed to the originally randomized study treatment. The only exception to this is if the AE fulfills the criteria for an anaphylactic reaction. See Section 3.7 for further details.

AEs will be presented as "number of study participants (percentage of study participants) [number of events]". In this style of output, "[number of events]" will include all cases of an AE including repeat occurrences in individual study participants, while "number of study participants" will count each study participant only once.

Adverse event summaries by SOC, HLT and PT will be ordered alphabetically for SOC and HLT within SOC and in terms of decreasing frequency for PT within HLT in the bimekizumab treatment group and then similarly by decreasing frequency in the placebo group.

AE summaries by PT will be ordered in terms of decreasing frequency for PT within the bimekizumab treatment group, and in the event of ties, PT will be sorted alphabetically.

Study participant time at risk represents the time a study participant is at risk for having an AE. The definitions for study participant time at risk (in days) are outlined in Section 10.1. These definitions will be used for exposure-adjusted AE summaries.

Selected AE summaries will include the exposure adjusted incidence rate (EAIR) with associated 95% confidence interval and the exposure adjusted event rate (EAER).

The EAIR is defined as the number of study participants (k) with a specific AE adjusted for the exposure and will be scaled to 100 subject-years:

$$\text{EAIR} = 100 \times \frac{k}{\sum_{i=1}^{n}(T_{Exp(i)})}$$

where n is the total number of study participants in the respective treatment group, and $T_{Exp(i)}$ is the time of exposure for each study participant defined as:

- For study participants with the specific AE of interest: years since first dose of the study treatment to the first occurrence for the AE of interest at the level of coding evaluated (see Section 10.1)
- For study participants without the AE: the time at risk in years (see Section 10.1)

Exact Poisson 95% confidence intervals for incidence rates are calculated using the relationship between the Poisson and the Chi-square distribution (Ulm, 1990; Fay and Feuer, 1997):

$$LCL = \frac{\chi^2_{2n,\alpha/2}}{2}$$

$$UCL = \frac{\chi^2_{2(n+1),1-\alpha/2}}{2}$$

where n is the number of study participants with a specific AE for the incidence rate of interest and is the basis for the number of the degrees of freedom for the chi-square quantile for the upper tail probability χ^2.

The EAER will be the number of AEs including repeat occurrences in individual study participants divided by the total time at risk scaled to 100 subject-years and calculated using:

$$\text{EAER} = 100 \times \frac{N_{AE}}{\sum_{i=1}^{n}(T_{Risk(i)})}$$

where N_{AE} is the total number of AEs, n is the total number of study participants and $T_{Risk(i)}$ is the time at risk in years for each study participant.

No confidence interval will be computed for EAER.

In the following sections, all tables based on "100 subject-years" will include EAIR (with 95% confidence interval) and EAER.

As previously indicated in Table 4–3, the AE summary tables described in Section 10.2.1 and Section 10.2.2 will be presented:

- on the SS for the overall period, presenting the BKZ 160 mg Total treatment column only.
- on the SS for the Double-Blind Treatment Period for the Week 24 interim analysis only, presenting the Placebo and BKZ 160mg treatment columns.
- and on the MS for the Maintenance Period for the Week 52 interim analysis only, presenting the BKZ 160 mg Total treatment column only.

10.2.1 Standard AE summaries

The following tabular summaries will be produced:

- Incidence of TEAEs – Overview.
- Incidence of TEAEs per 100 subject-years by SOC, HLT, and PT.
- Incidence of Serious TEAEs per 100 subject-years by SOC, HLT, and PT.
- Incidence of TEAEs Leading to Study Discontinuation per 100 subject-years by SOC, HLT, and PT.
- Incidence of TEAEs Leading to Study Drug Discontinuation per 100 subject-years by SOC, HLT, and PT
- Incidence of TEAEs Leading to Death by SOC, HLT, and PT.
- Incidence of TEAEs by Relationship by SOC, HLT, and PT.
- Incidence of TEAEs by Intensity, SOC, HLT, and PT.
- Incidence of TEAEs by decreasing frequency of PT.
- Incidence of TEAEs Above Reporting Threshold of 5% by SOC and PT.
- Incidence of Non-Serious TEAEs Above Reporting Threshold of 5% by Relationship SOC and PT.
- Incidence of Related TEAES by SOC, HLT, and PT
- Incidence of Related TEAES Above Reporting Threshold of 5% by SOC and PT.
- Incidence of TEAEs per 100 subject-years by SOC, HLT, and PT and by Time of Onset Relative to Anti-bimekizumab Antibody Status. The table will include columns for the following:
  - TEAEs starting before the first ADAb positive result (includes ADAb Categories 2 and 5) where TEAEs have occurred before the following events:

    a) the first positive ADAb result for study participants in category 2 and

    b) the first post-Baseline boosted ADAb titer result for study participants with titer results or the first post-Baseline positive ADAb result for study participants with positive ADAb at Baseline with no other samples with titer available for study participants in category 5.

- TEAEs starting on the same date or after the first ADAb positive result (includes ADAb Categories 2, 3, 4, 5 and 6) where TEAEs have occurred on or after the following events:

  a) the first positive ADAb results for study participants in categories 2, 3, 4 and 6, and

  b) the first post-Baseline boosted ADAb titer result for study participants with titer results and the first post-Baseline positive ADAb result for study participants with positive ADAb at Baseline with no other samples with titer available for study participants in category 5.

- TEAEs for study participants who are ADAb negative at all time points (includes ADAb Category 1).

Note: For TEAE by onset relative to ADAb positivity status, all available ADAb data at the time of IA cut-off will be utilized to derive the study participant-level ADAb status categories.

The following summary tables will be produced to support results disclosure activities

- Incidence of Serious TEAEs by Relationship SOC, HLT, and PT
- Incidence of TEAEs Leading to Death by Relationship by SOC, HLT, and PT
- Incidence of Non-Serious TEAEs Above Reporting Threshold of 5% (in any treatment group, and in any study period) by SOC and PT

By-study participant listings on all AEs and all deaths will be provided on the enrolled set. Serious AEs and AEs leading to study discontinuation will also be listed on the safety set. AEs occurring on the day of first study drug administration will also be listed on the randomized set. For study participants randomized to Placebo, this will include AEs occurring at first administration of Bimekizumab at Week 16. A glossary listing will also be created for all TEAEs.

10.2.2 Other safety topics of interest summaries

The safety topics of interest (or AEs with special monitoring) for the bimekizumab program are:

- Infections (serious, fungal, opportunistic and tuberculosis)
- Malignancies
- Major adverse cardiac events (MACE)
- Neutropenia
- Suicidal Ideation and Behavior
- Inflammatory bowel disease
- Hypersensitivity (including anaphylactic reactions)
- Hepatic events and drug-induced liver injury (DILI)

The analyses produced for the TEAE of special monitoring are based on the specifications described in the Bimekizumab safety topic of interest document version from 26-Mar-2020.

Separate tables will be created to summarize each of the above AEs (except when notified) per 100 subject-years by system organ class, high level term, and PT.

Infections

The summary table for fungal infection will include all TEAEs (serious and non-serious) which code to the High Level Group Term "Fungal infectious disorders".

Opportunistic infections (including tuberculosis) will be summarized in a stand-alone table and will include all opportunistic infection TEAEs identified as per the UCB-defined search criteria which were adjudicated as opportunistic infections (refer to Appendix 2 for further details on the process for identifying opportunistic infections).

No dedicated table will be produced for serious infections, as there will be identified based on MedDRA classification (SOC "Infections and infestations") using the "Any SAE" table specified in Section 10.2.1.

Fungal infections and opportunistic infections will be listed separately.

Malignancies

Two tables will be created on malignancies:

- one table will display all events in the criteria standardized MedDRA Query (SMQ)= "Malignant or unspecified tumors (SMQ)",
- and another table will display all events in the criteria SMQ= "Malignant tumors (SMQ)".

The SMQ search will include all TEAEs which code to a PT included in the Scope=Narrow group within each SMQ.

The output tables will include two different overall incidence rows:

- The first overall incidence row will summarize "Any Malignancy" and this row will summarize the incidence of all AEs flagged for inclusion in the table (using the appropriate SMQ depending on the table), regardless of the HLT it codes to.
- The second overall incidence row will summarize "Any Malignancy excluding non melanomic skin cancers HLT" and this row will summarize the incidence of AEs flagged for inclusion in the table (using the appropriate SMQ depending on the table), excluding those which code to an HLT of "skin neoplasms malignant and unspecified (excl melanoma)".

All malignancy events will be listed.

Cardiac events

The table on cardiac events will summarize events considered as MACE by the external cardiovascular event adjudication committee.

Extended MACE events will be presented in a separate table and listing. All events which are classified by the adjudication committee as any of the following events types will be considered an extended MACE event:

- Non-Fatal Myocardial Infarction (MI)
- Non-Fatal Stroke: hemorrhagic
- Non-Fatal Stroke: ischemic

- Non-Fatal Stroke: embolic
- Non-Fatal Stroke: undeterminable
- Hospitalization or ER for Unstable Angina with urgent revascularization
- Hospitalization for Heart Failure
- Coronary Revascularization Procedures (eg, percutaneous coronary intervention, coronary artery bypass grafting)
- Urgent Revascularization Procedures (ie, due to symptoms of brain ischemia or pending infarction)
- Death due to Myocardial Infarction (MI
- Death due to Stroke
- Sudden Cardiac Death
- Other CV Death (eg, heart failure, pulmonary embolism, cardiovascular procedure-related)
- Cardiovascular: Undetermined Cause of Death (ie, cause of death unknown)

A separate table and listing will present the adjudicated cardiovascular events by type. For each cardiovascular event type (24 total), the individual PTs which fall within each event type will be summarized.

Additionally, a listing of all events identified for potential review by the cardiovascular event adjudication committee will be produced. This listing will indicate whether each event was escalated to the committee for formal review/adjudication.

Neutropenia

The table on neutropenia will be based on the following PTs (regardless of seriousness):

- Autoimmune neutropenia
- Band neutrophil count decreased
- Cyclic neutropenia
- Febrile neutropenia
- Idiopathic neutropenia
- Neutropenia
- Neutropenic infection
- Neutropenic sepsis
- Neutrophil count decreased

All neutropenia events will be listed.

Suicidal Ideation and Behavior

An external neuropsychiatric adjudication committee will evaluate potential neuropsychiatric events and determine whether any of those events were associated with suicidal ideation and behavior (SIB). If an event is adjudicated as SIB, further information will be provided. A table

and a listing for SIB events as determined by the adjudication committee will be produced as well as a listing for all events escalated to the committee but which were not adjudicated as a SIB event.

A separate table will present the adjudicated neuropsychiatric events by type. For each neuropsychiatric event type (6 total), the individual PTs which fall within each event type will be summarized. It will include events adjudicated as SIB and events adjudicated as non-suicidal. Note that the event type Suicidal ideation can be classified as either SIB or non-suicidal.

Additionally, a listing of all events identified for potential review by the neuropsychiatric adjudication committee will be produced. This listing will indicate whether each event was escalated to the committee for formal review/adjudication.

Inflammatory bowel disease

An external inflammatory bowel disease (IBD) adjudication committee will evaluate potential IBD events and will classify each one as follows:

- Event Type Code 1: Possible IBD – Crohn's Disease
- Event Type Code 2: Probable IBD – Crohn's Disease
- Event Type Code 3: Definite IBD – Crohn's Disease
- Event Type Code 4: Possible IBD – Ulcerative Colitis
- Event Type Code 5: Probable IBD – Ulcerative Colitis
- Event Type Code 6: Definite IBD – Ulcerative Colitis
- Event Type Code 7: Possible IBD – Unclassified
- Event Type Code 8: Probable IBD – Unclassified
- Event Type Code 9: Definite IBD – Unclassified
- Event Type Code 10: Symptoms not consistent with IBD
- Event Type Code 11: Possible Inflammatory Bowel Disease - Microscopic Colitis
- Event Type Code 12: Probable Inflammatory Bowel Disease - Microscopic Colitis
- Event Type Code 13: Definite Inflammatory Bowel Disease - Microscopic Colitis
- Event Type Code 14: Possible Inflammatory Bowel Disease – no further differentiation possible
- Event Type Code 15: Probable Inflammatory Bowel Disease – no further differentiation possible
- Event Type Code 16: Definite Inflammatory Bowel Disease – no further differentiation possible
- Event Type Code 99: Not enough information to adjudicate

A table for adjudicated definite IBD events (event type codes 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15 and 16) as determined by the adjudication committee will be produced. It will summarize events determined by the adjudication committee as definite IBD (event type codes 3, 6, 9, 13, and 16), probable IBD (event type codes 2, 5, 8, 12, and 15) and possible IBD (event type codes

1, 4, 7, 11, and 14). Definite and probable IBD will also be aggregated and summarized. This table will be stratified by study participants with or without a previous medical history of IBD. Previous medical history of IBD will be determined using the information recorded on the Extra-Articular Assessment at Screening CRF page ("Does subject have a history of IBD").

A separate table will present the adjudicated gastrointestinal events by type. For each gastrointestinal event type (17 total), the individual PTs which fall within each event type will be summarized. It will include events determined by the adjudication committee as definite IBD probable IBD and possible IBD. It will also include events determined as Symptoms not consistent with IBD (event type code 10) and Not enough information to adjudicated (event type code 99).

A listing of all events identified for potential review by the IBD adjudication committee will be produced. This listing will indicate whether each event was escalated to the committee for formal review/adjudication.

A separate table and listing will present the adjudicated IBD events by type. For each IBD event type (event type codes 1 through 16 and 99; 17 total), the individual PTs which fall within each event type will be listed.

A third listing will present the individual diagnostic criteria met for each adjudicated IBD event.

Hypersensitivity (including anaphylactic reactions)

The MedDRA anaphylaxis Algorithm described in Appendix 1 will be used to summarize acute anaphylactic events (reported on the same day as when an injection was administered or one day after). An AE glossary table will also be produced to summarize the MedDRA coding for these events. The glossary table will include the following fields: reported term, PT, LLT, HLT, and SOC.

A separate table will be prepared to summarize hypersensitivity events, identified using the SMQ "Hypersensitivity (SMQ)". All TEAEs which code to a PT included in the Scope=Narrow search will be included in this table. The analysis will be repeated for serious hypersensitivity reactions. An AE glossary table will also be produced to summarize the MedDRA coding for these events. The glossary table will include the following fields: reported term, PT, LLT, HLT, and SOC.

No separate table will be needed for injection site reactions that will be evaluated based on the table on any TEAE by looking under the HLTs "Administration site reactions NEC" and "Injection site reactions".

Separate listings will be created for anaphylaxis and hypersensitivity events.

An additional analysis will report the serious hypersensitivity events.

Hepatic events and drug induced liver injury

The table on hepatic events will be created based on the SMQ of "Drug related hepatic disorders - comprehensive search (SMQ)" excluding the two sub-SMQs "Liver neoplasms, benign (incl cysts and polyps) (SMQ)" and "Liver neoplasms, malignant and unspecified (SMQ)". The SMQ search should include all events which code to a PT included in the Scope=Broad and/or Scope=Narrow, regardless of whether they have been judged as related to study medication or not by the investigator. Hy's Law cases will be reported separately in a liver function test table (see Section 10.3)

Hepatic events (including DILI) will also be listed.

Finally, the incidence of each safety topics of interest occurring in the Double-Blind Treatment Period will be reported in one summary table. The risk difference (bimekizumab minus placebo) associated to each safety event will be calculated, together with the associated 95% confidence interval. These results will also be displayed graphically. For each safety topic of interest, the risk difference (RD) will be derived as follows:

$$RD = \frac{\text{Number of study participants treated with bimekizumab having the safety event}}{\text{Total number of study participants treated with bimekizumab}} - \frac{\text{Number of study participants treated with placebo having the safety event}}{\text{Total number of study participants treated with placebo}}$$

The lower and upper confidence interval bounds for RD will be calculated as follows:

$$RD - Z_{\alpha/2} * SE\ (RD), RD + Z_{\alpha/2} * SE\ (RD)$$

Where $Z_{\alpha/2}$ is the critical value from the standard normal distribution (1.96 for a 95% CI).

10.2.3 Extra-Articular Manifestations

A summary table will be presented on study participants having TEAE uveitis. The table will include all events defined by one of the following preferred terms: "Autoimmune uveitis", "Iridocyclitis", "Iritis", "Uveitis". The TEAE will be summarized per 100 subject-years by system organ class, high level term, and PT. The analysis will be stratified by study participants with or without a previous medical history of uveitis at Baseline. Study participants with uveitis at Baseline will be identified as the ones having documented an uveitis in the" Extra-articular Assessments form at Baseline" form ("Does subject have a history of uveitis?").

The analysis will be presented on the overall period as well as on the Double-Blind period.

10.2.4 Impact of COVID-19

To assess the impact of the COVID-19 pandemic on safety, additional summaries and listings will be presented.

For reporting purposes, AEs will be assigned to either Pre-COVID-19, COVID-19 or post-COVID-19 pandemic period by comparing the AE start date (based on imputed date) to the COVID-19 pandemic period start and end dates (Section 3.2.2.2) (AE allocated to a period if it starts during the period):

The following categories of TEAE will be summarized by MedDRA SOC, HLT and PT, including EAIR and EAER:

- All TEAEs by time of onset relative to COVID-19 pandemic period (prior/during/post) by region and overall
- All serious TEAEs by time of onset relative to COVID-19 pandemic period (prior/during/post) by region and overall
- All TEAEs leading to study discontinuation by time of onset relative to COVID-19 pandemic period (prior/during/post) by region and overall
- All TEAEs leading to permanent withdrawal of study medication by time of onset relative to COVID-19 pandemic period (prior/during/post) by region and overall

- All COVID-19 related TEAEs by treatment group, region and overall

COVID-19 related TEAEs will be identified by searching for Preferred Terms "Corona virus infection" and "Coronavirus test positive". These will include confirmed or suspected COVID-19 infections. All the above table summaries will be produced on the overall period, except for the summary on the COVID-19 related TEAEs that will be produced for all treatment periods.

A separate listing of all COVID-19 related AEs will be presented, where COVID-related AEs are identified as described above.

In addition, all listing of AEs will include a column for COVID-19 relatedness (as defined above) and the time of onset of each AE relative to the COVID-19 pandemic.

For the purpose of calculating EAIR and EAER by COVID-19 pandemic period, the calculation of exposure time at risk presented in Section 10.1.5 will be used. An individual study participant may therefore be counted in the denominator for several COVID-19 pandemic periods dependent on whether the study participant is still considered at risk on the COVID-19 pandemic start and stop dates. In this case time at risk will be calculated separately for each period.

10.3 Clinical laboratory evaluations

The list of the laboratory measurements/variables that will be evaluated in the study is specified in Table 10–1 per laboratory category.

Table 10–1: Laboratory measurements

Hematology	**Biochemistry**	**Urinalysis**
Basophils	Calcium	pH
Eosinophils	Chloride	Albumin (protein)
Lymphocytes	Magnesium	Glucose
Monocytes	Potassium	Blood
Neutrophils	Sodium	Leukocyte esterase
Hematocrit	Glucose	Nitrite
Hemoglobin	BUN	Urine dipstick for pregnancy testing
MCH	Creatinine	Urine drug screen
MCHC	AST	
MCV	ALT	
Platelet count	ALP	
RBC count	GGT	
WBC count	Total bilirubin	
	LDH	
	Total cholesterol	
	Serum pregnancy testing	

Hematology	Biochemistry	Urinalysis
	hs-CRP	
Hematology	Biochemistry	Urinalysis
Basophils	Calcium	pH

ALP=alkaline phosphatase; ALT=alanine aminotransferase; AST=aspartate aminotransferase; BUN=blood urea nitrogen; hs-CRP=high sensitivity C-reactive protein; ET=early termination; GGT=gamma-glutamyltransferase; HDL=high density lipoprotein, IMP=investigational medicinal product; LDH=lactate dehydrogenase; MCH=mean corpuscular hemoglobin; MCHC=mean corpuscular hemoglobin concentration; MCV=mean corpuscular volume; RBC=red blood cell; SFU=Safety Follow-Up; WBC=white blood cell.

The laboratory analyses will mainly be prepared using data from the central laboratory, except for determining treatment-emergent abnormal values, potential drug-induced liver injury or potential Hy's Law cases. In the latter cases, both central and local laboratory results can be used if the local laboratory values meet specific criteria detailed thereafter:

- The local laboratory value should not be considered without a provided local laboratory range in the database on the local laboratory CRF page,
- The local laboratory value should have been reviewed and checked,
- Only local laboratory results coming from the same assessment and laboratory at the exact same time point can be used for the determination of potential drug-induced liver injury or potential Hy's law criteria,
- Local laboratory values which are not reported in standard international (SI) units will need to be converted in SI units prior to inclusion in the statistical analyses,
- An indicator will be added in the table summaries and study participant listings when local laboratory contributed to the displayed results.

All laboratory summaries will be presented in SI units and will be based on observed case values. In the case where laboratory values are below the lower limit of quantification, then these will be set to the midpoint between 0 and the lower limit of quantification for the purpose of summarizing the data.

The following two summary statistic tables will be created for all laboratory continuous variables (unless otherwise notified). These two tables will be provided on the SS for the overall period, presenting the Placebo/BKZ 160mg Q4W and the BKZ 160mg treatment columns. Baseline will correspond to the original Baseline Day 1.

- A summary of the absolute and change from Baseline values in each laboratory variable by treatment group and visit.
- A summary of the number and percentage of study participants with a given common terminology criteria for adverse events (CTCAE) grade (0, 1, 2, 3 or 4) based on the minimum/maximum post-Baseline value, by laboratory variable and treatment group.

  This analysis will be restricted to the hematology and biochemistry laboratory parameters only, and will include the following parameters: hemoglobin, platelet count, WBC count, lymphocytes, neutrophils, creatinine, sodium, potassium, calcium, magnesium, total cholesterol, total bilirubin, AST, AST and GGT.

The hs-CRP parameter will not be included in the summary statistics on absolute and change from Baseline values because the results are already planned to be presented as part of an efficacy analysis (see Section 8.3.1.5).

As indicated in Section 3.1, for tables where data are summarized by visit only values occurring at scheduled visits will be included.

A shift table of the number and percentage of study participants with a given CTCAE grade 0, 1, 2, 3 or 4 values (as applicable) at Baseline to minimum/maximum post-Baseline CTCAE grade, by laboratory variable and treatment group will additionally be created for the hematology and biochemistry laboratory parameters only. The analysis will be performed for the following parameters: hemoglobin, platelet count, WBC count, lymphocytes, neutrophils, creatinine, sodium, potassium, calcium, magnesium, total cholesterol, total bilirubin, AST, AST and GGT.

In addition, a summary table of the number and percentage of study participants experiencing markedly abnormal values by laboratory variable, treatment group and visit will be created for biochemistry, hematology and liver function test values.

Markedly abnormal values are defined as those with a severity of Grade 3 and above based on the CTCAE criteria Version 4.03. Definitions of markedly abnormal values using the Grade 3 cut points are given in Table 10–2, Table 10–3 and Table 10–4 in order to allow for a more thorough review of elevated LFTs.

All values observed at any time while on treatment (ie, occurring at scheduled and unscheduled visits) will be included in this table summarizing markedly abnormal values. For this summary, Baseline values and values observed more than 140 days after the last administration of study medication are not considered.

Table 10–2: Definitions of markedly abnormal biochemistry values

Parameter name	Unit	Criteria	Abnormal designation
Creatinine	mmol/L	>3.0 x Baseline or >3.0 x ULN	AH
Glucose	mmol/L	<2.2 >13.9	AL AH
Calcium	mmol/L	>3.1 <1.75	AH AL
Magnesium	mmol/L	>1.23 <0.4	AH AL
Potassium	mmol/L	>6.0 <3.0	AH AL
Sodium	mmol/L	>155 <130	AH AL
Cholesterol	mmol/L	>10.34	AH

AH=abnormal high; AL=abnormal low; dL=deciliter; L=liter; mg=milligram; mmol=millimoles; ULN=upper limit of normal.

Table 10–3: Definitions of markedly abnormal hematology values

Parameter name	Unit	Criteria	Abnormal designation
Hemoglobin	g/L	<80 >40 above ULN	AL AH
Lymphocytes Absolute	10^9/L	<0.5 >20.0	AL AH
Neutrophils Absolute	10^9/L	<1.0	AL
Platelets	10^9/L	<50	AL
WBC/ Leukocytes	10^9/L	<2.0 >100	AL AH

AH=abnormal high; AL=abnormal low; dL=deciliter; L=liter; mg=milligram; mmol=millimoles; ULN=upper limit of normal.

Table 10–4: Definitions of markedly abnormal liver function values

Paramter name	Unit	Criteria	Abnormal designation
Alkaline Phosphatase	U/L	>5.0 x ULN	AH
ALT	U/L	>5.0 x ULN	AH
AST	U/L	>5.0 x ULN	AH
Total Bilirubin	umol/L	>3.0 x ULN	AH
GGT	U/L	>5.0 x ULN	AH

ALT=alanine aminotransferase; AST=aspartate aminotransferase; GGT=gamma-glutamyltransferase

Another table will list the count and percentage of study participants meeting the below criteria at any time during the study while on treatment:

- AST: >2xULN, >3xULN, >5xULN, >10xULN, >20xULN
- ALT: >2xULN, >3xULN, >5xULN, >10xULN, >20xULN
- AST or ALT: >2xULN, >3xULN, >5xULN, >10xULN, >20xULN
- Bilirubin: >1xULN, >1.5xULN
- ALP: >2xULN

In addition, a table will be produced to summarize potential Hy's Law cases. The following two definitions will be used in that table:

- [AST ≥3xULN or ALT ≥3xULN] and Total Bilirubin ≥2xULN
- [AST ≥3xULN or ALT ≥3xULN] and Total Bilirubin ≥2xULN in the absence of ALP ≥2xULN

Neutrophil values for study participants with at least one treatment emergent markedly abnormal neutrophil value will be plotted over time.

A spaghetti plot will be created for ALT values for study participants with at least one treatment emergent markedly abnormal ALT value. The plot will also be created for AST values.

Results for the shift tables, the graphs and summary tables on markedly abnormal values and the Hy's Law cases tables will be presented as follows:

- on the SS for the Double-Blind Treatment Period for the Week 24 interim analysis only, presenting the Placebo and BKZ 160mg treatment columns,
- on the MS for the Maintenance Period for the Week 52 interim analysis only, presenting the BKZ 160 mg Total treatment column only,
- on the SS for the overall period, presenting the BKZ 160 mg Total treatment column only.

For the two last analyses presenting data occurring during the Maintenance Period, Baseline Week 16 will be used as Baseline for study participants randomized to Placebo that switched to bimekizumab at Week 16.

A by-study participant listing of all laboratory data (excluding urinalysis results) will be provided. In this listing, values that are out of normal range will be flagged (abnormal values that

are below the lower limit of the reference range flagged as 'L' (low) and abnormal values that are above the upper limit of the reference range flagged as 'H' (high)).

In addition, the laboratory results classified as Grade 3 or Grade 4 will be listed separately.

Urinalysis laboratory results will be listed separately, as well as all additional laboratory samples that may have been collected for a study participant.

An additional listing will be created for study participants with suspected hepatic events.

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

10.4.1 Vital signs, weight and physical findings

The following results will be tabulated for the vital sign variables (systolic blood pressure (mmHg), diastolic blood pressure (mmHg), and pulse rate (beats/min)), by treatment group and visit:

- A summary of the absolute and change from Baseline value

The table will be done on the SS, will cover the overall period, and present the Placebo/BKZ 160mg Q4W and the BKZ 160mg treatment columns.

- A summary of the number and percentage of study participants experiencing at least one markedly abnormal value for a vital sign variable as defined in Table 10–5 below.

The table will be presented:

  - on the SS for the Double-Blind Treatment Period for the Week 24 interim analysis only, presenting the Placebo and BKZ 160mg treatment columns.
  - on the MS for the Maintenance Period for the Week 52 interim analysis only, presenting the BKZ 160 mg Total treatment column only.
  - and on the SS for the overall period, presenting the BKZ 160 mg Total treatment column only.

For the two last analyses presenting TEMA data occurring during the Maintenance Period, Baseline Week 16 will be used as Baseline for study participants randomized to Placebo that switched to bimekizumab at Week 16.

Table 10–5: Definitions of markedly abnormal blood pressure values

Parameter (unit)	Markedly abnormal low	Markedly abnormal high
Systolic blood pressure (mmHg)	<90 and a decrease from Baseline of ≥20	>180 and an increase from Baseline of ≥20
Diastolic blood pressure (mmHg)	<50 and a decrease from Baseline of ≥15	>105 and an increase from Baseline of ≥15

At Visit 2, vital signs are planned to be assessed three times (prior to the study treatment administration, 30 minutes after the study treatment administration and 1 hour after the study treatment administration). Baseline will be defined as the last available value up to and including the assessment collected at Visit 2 prior to first dose.

At Week 16, vital signs collected before IMP belong to the double-blind treatment period, while vital signs collected after IMP belong to the maintenance period.

A by-study participant listing of all vital sign results (including body temperature) with the details of abnormities (when applicable) will be provided (abnormal values will be flagged as "L" or "H" accordingly).

Similarly, physical examination findings together with the details of abnormalities (when applicable) will be listed by treatment group, study participant and visit.

Body weight will also be listed.

10.4.2 Electrocardiograms

Since it is only planned to assess electrocardiogram (ECG) at the Screening Visit, the Baseline value will be derived based on the Screening assessment.

The date and time of the ECG will be recorded in the eCRF together with the Investigator interpretation. A summary of the number and percentage of study participants with normal, abnormal not clinically significant, and abnormal clinically significant ECG results will be presented for Baseline and post-Baseline timepoints.

A summary of the absolute and change from Baseline values in each ECG variable will also be created at each post-Baseline/Screening timepoints by treatment group. The following ECG variables will be summarized: heart rate (bpm), RR interval (ms), PR interval (ms), QRS duration (ms), QT interval (ms), QTcF interval (ms) and QTcB interval (ms).

QTc outliers are defined as QTcF (or QTcB) values following dosing that are greater than 450 ms or are increases from Baseline greater than 30 ms. QTcF and QtcB outliers will be summarized in a dedicated table using the following categories:

- Values >450 ms, >480 ms, >500 ms
- Increase from Baseline of >30 ms, increase from Baseline of >60 ms, increase from Baseline of >90 ms
- Values >450 ms and increases of >30 ms. Values >500 ms and increases of >60 ms

In this table, the number and percentage of study participants who meet the ECG outlier criteria at any assessment post first dose will be summarized.

These three table summaries will cover the overall period on the SS and will present the Placebo/BKZ 160mg Q4W and the BKZ 160mg treatment columns. Baseline Day 1 will serve as Baseline.

A by-study participant listing of all 12-lead ECG data will be provided where QTcF and QTcB outliers will be highlighted.

10.4.3 Other safety variables

Suicidal ideation and behavior

Suicidal ideation and behavior will be assessed using the electronic Columbia-Suicide Severity Rating Scale (eC-SSRS) questionnaire. The questionnaire is an assessment tool that evaluates suicidal ideation and behavior that may occur during the study.

The eC-SSRS defines 5 subtypes of suicidal ideation and behavior in addition to self-injurious behavior with no suicidal intent.

Suicidal ideation is defined as an event in any of the following 5 categories:

- Category 1 - Wish to be dead
- Category 2 - Non-specific active suicidal thoughts
- Category 3 - Active suicidal ideation with any methods (not plan), without intent to act
- Category 4 - Active suicidal ideation with some intent to act, without specific plan
- Category 5 - Active suicidal ideation with specific plan and intent

Suicidal behavior is defined as an event in any of the following 4 categories:

- Category 6 - Preparatory acts or behavior
- Category 7 - Aborted attempt
- Category 8 - Interrupted attempt
- Category 9 - Actual attempt

Suicidal behavior or ideation is defined as an event in any of the above 9 categories.

Self-injurious behavior without suicidal intent is defined as an event in the category non-suicidal self-injurious injuries.

The incidence of study participants with suicidal ideation, suicidal behavior, suicidal behavior or ideation, and self-injurious behavior without suicidal intent will be summarized:

- on the SS for the Double-Blind Treatment Period for the Week 24 interim analysis only, presenting the Placebo and BKZ 160mg treatment columns,
- on the MS for the Maintenance Period for the Week 52 interim analysis only, presenting the BKZ 160 mg Total treatment column only,
- and on the SS for the overall period, presenting the BKZ 160 mg Total treatment column only.

A by-study participant listing of the eC-SSRS questionnaire data will be provided, as well as a subset listing presenting eC-SSRS data for study participants having at least one positive response.

The visit name associated to the eC-SSRS data will not be transferred by the vendor since the way the device is designed would lead to incorrect naming of the visits. The visit date collected in EDC will rather be used to assign the correct visit to the eC-SSRS data, (in merging based on study participant number and date) as follows:

- If the eC-SSRS assessment date is matching an eCRF scheduled visit date, then the eC-SSRS visit identification will be the one corresponding to eCRF visit date.
- If the eCRF visit date directly matches the date of the eC-SSRS assessment, then the eC-SSRS assessment will be considered in the SDTM datasets as being done on an unscheduled visit.

Among the C-SSRS assessments previously considered in the SDTM datasets as being done on an unscheduled visit, if the assessment is matching an eCRF visit date within a window of +/- 3 days, then the eC-SSRS assessment will be considered as acceptable in the ADaM datasets and will be considered as being done on that eCRF scheduled Visit. Otherwise, the eC-SSRS assessment will stay recorded in an unscheduled visit.

The following rules will be implemented in case of visit with artificially duplicated questionnaires caused by the technical limitations of the data capture system, like time-out settings with a visit resetting trigger

- If there is a visit with one (or more) incomplete questionnaires and at least one complete questionnaire without discrepancies, then the incomplete questionnaire will be excluded from the ADaM dataset.
- If there is a visit with one (or more) incomplete questionnaires and at least one complete questionnaire **with** discrepancies, then the incomplete questionnaire will be excluded from the ADaM dataset.
- If there is a visit with one (or more) full duplicates without discrepancies, then the unscheduled repetition of the first completed questionnaire will be excluded from the ADaM dataset.
- If there is a visit with one (or more) full duplicates with discrepancies, then the "wrong" questionnaire with discrepancies will be excluded from the ADaM dataset.
- If there is a visit with only a single incomplete questionnaire, the incomplete questionnaire will be included in the ADaM dataset and used for the TFL generation like a normal visit.

Patient Health Questionnaire-9 (PHQ-9)

The PHQ-9 is a multipurpose instrument for screening, diagnosing, monitoring, and measuring the severity of study participant's depression. The PHQ-9 score is based on nine questions assessing the depression over the last two weeks. Each criteria is scored from 0 (not at all) to 3 (nearly every day). The total score for the nine questions is added up and provide the PHQ-9 score and can range from 0 to 27 with higher scores indicating worse state. If one of the nine criteria is missing, the PHQ-9 score will be set to missing.

The following depression states are defined based on the PHQ-9 score:

- A score of 5-9 is considered to be minimal symptoms of depression.
- A score of 10 to 14 is considered minor depression, dysthymia, or mild major depression.
- A score of 15 to 19 is considered to indicate moderately severe major depression.
- A score >=20 is considered to be severe major depression.

The percentage of study participants with PHQ-9 scores below 5, between 5 and 9, between 10 and 14, between 15 and 19, and greater than or equal 20 will be summarized by visit and treatment group, as well as the change from Baseline over time. These two analyses will cover the overall period on the SS and will present the Placebo/BKZ 160mg Q4W and the BKZ 160mg treatment columns.

Tuberculosis

The laboratory test results (negative, positive or indeterminate) from the TB assessment performed by interferon gamma release assay will be summarized by treatment group (Placebo/BKZ 160mg Q4W and the BKZ 160mg) for Screening and Week 44 timepoints on the SS, as well as listed.

The results from the 'Evaluation of signs and symptoms of tuberculosis' questionnaire data will be also listed.

Questionnaires

Study participants for who the wrong user role was referenced by the site, at time of completion of study participant facing questionnaires will be listed on the randomized set.

11 REFERENCES

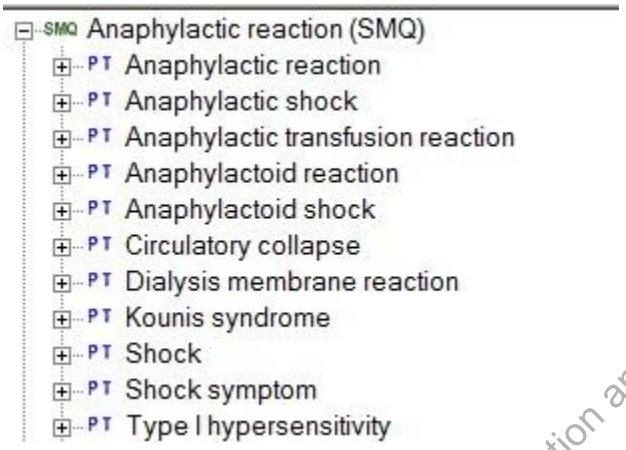
1. Maksymowych WP, Mallon C, Richardson R, et al. Development and Validation of the Edmonton Ankylosing Spondylitis Metrology Index. Arthritis Rheum. 2006;55:575-82.
2. Heuft-Dorenbosch L, Spoorenberg A, van Tubergen A, Landewé R, van der Tempel H, Mielants H, et al. Assessment of enthesitis in ankylosing spondylitis. Ann Rheum Dis. 2003;62:127-32.
3. Mallinckrodt CH, Lin Q, Lipkovich I, Molenberghs G. A structured approach to choosing estimands and estimators in longitudinal clinical trials. Pharm Stat. 2012;11:456-61.
4. Mallinckrodt CH. Preventing and Treating Missing Data in Longitudinal Clinical Trials: A Practical Guide. New York, NY: Cambridge University Press; 2013.
5. O'Kelly, Michael, and Bohdana Ratitch. Clinical trials with missing data: A guide for practitioners. John Wiley & Sons, 2014
6. Rubin, Donald B. "The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm." Journal of the American Statistical Association 82, no. 398 (1987): 543-546.
7. Ulm, K., A simple method to calculate the confidence interval of a standardized mortality ratio. American Journal of Epidemiology, 1990; 131(12), 373-375

12 APPENDICES

APPENDIX 1. MEDDRA ALGORITHMIC APPROACH TO ANAPHYLAXIS

The SMQ Anaphylactic reaction consists of three parts:

- A narrow search containing PTs that represent core anaphylactic reaction terms (Category A core anaphylactic terms)



- A **broad search** that contains additional terms that are added to those included in the narrow search. These additional terms are signs and symptoms possibly indicative of anaphylactic reaction and categorized in B, C or D
  - Cat B

- Acute respiratory failure
- Asthma
- Bronchial oedema
- Bronchospasm
- Cardio-respiratory distress
- Chest discomfort
- Choking
- Choking sensation
- Circumoral oedema
- Cough
- Cyanosis
- Dyspnoea
- Hyperventilation
- Irregular breathing
- Laryngeal dyspnoea
- Laryngeal oedema
- Laryngospasm
- Laryngotracheal oedema

– Cat C

- Allergic oedema
- Angioedema
- Erythema
- Eye oedema
- Eye pruritus
- Eye swelling
- Eyelid oedema
- Face oedema
- Flushing
- Generalised erythema
- Injection site urticaria
- Lip oedema
- Lip swelling
- Nodular rash
- Ocular hyperaemia
- Oedema
- Periorbital oedema

- Pruritus
- Pruritus allergic
- Pruritus generalised
- Rash
- Rash erythematous
- Rash generalised
- Rash pruritic
- Skin swelling
- Swelling
- Swelling face
- Urticaria
- Urticaria papular

– Cat D

- Blood pressure decreased
- Blood pressure diastolic decreased
- Blood pressure systolic decreased
- Cardiac arrest
- Cardio-respiratory arrest
- Cardiovascular insufficiency
- Diastolic hypotension
- Hypotension

- An **algorithmic approach** which combines a number of anaphylactic reaction symptoms in order to increase specificity. A case must include either **(on the same day):**
  - A narrow term or a term from Category A;
  - A term from Category B - (Upper Airway/Respiratory) AND a term from Category C - (Angioedema/Urticaria/Pruritus/Flush);
  - A term from Category D - (Cardiovascular/Hypotension) AND [a term from Category B - (Upper Airway/Respiratory) OR a term from Category C - (Angioedema/Urticaria/ Pruritus/Flush)]
- Hypersensitivity events will be identified using the “Hypersensitivity (SMQ)”. All TEAEs which code to a PT included in the Scope=Narrow search will be included.

APPENDIX 2. UCB-DEFINED SEARCH CRITERIA FOR IDENTIFYING OPPORTUNISTIC INFECTIONS

Opportunistic infections are identified in two steps:

Step 1: Refer to column B of the spreadsheet (Opportunistic infections MedDRA v 19.xlsx) which identifies the Preferred Terms (PTs) to be classified as opportunistic infections using either a single 'x' or a double 'xx'.

TEAEs which code to a PT flagged with a single 'x' need to also be serious to be considered an opportunistic infection.

All TEAEs which code to a PT flagged with a double 'xx' are considered to be an opportunistic infection, regardless of seriousness.

Step 2: Refer to column C of the spreadsheet which identifies the PTs that need to be evaluated on a case-by-case basis by the study physician to determine whether or not it is an opportunistic infection. If Column C has a single 'x', then the corresponding preferred term should be flagged for case-by-case review by the study physician.

13 AMENDMENT(S) TO THE STATISTICAL ANALYSIS PLAN (SAP)

13.1 AMENDMENT 1

Rationale for the amendment

The major purpose of this SAP amendment was to implement changes in response to Protocol Amendment 4 (16FEB2021) and discussions and feedback provided at meetings between UCB and PAREXEL technical teams or for clarifications. The main changes are rules for handling missing data and guidelines on the implementation of multiple imputation and latest guidelines from the BKZ AE of special monitoring convention document.

The protocol amendments derived changes include clarification of the 12-item MOS Sleep Scale is rated using the 5-point scale, the increase of the threshold for subgroup analyses by disease duration from 2 to 5 years.

The SAP amendments can be summarized as follows:

- Re-classification of Change from Baseline in Patient Health Questionnaire-9 (PHQ-9) variable from other efficacy variable to other safety variable
- New rules for handling of missing data for the primary and key secondary efficacy variables
- Use of latest guidelines from the BKZ AE of special monitoring convention document
- Clarification of the 12-item MOS Sleep Scale is rated using the 5-point scale
- The increase of the threshold for subgroup analyses by disease duration from 2 to 5 years
- Clarification on use and multiple imputation approaches in supportive analyses
  - Observed cases
  - NRI
- Imputation of partial start dates
- Clarification of outputs for China and Japan submissions
- Update in the pharmacokinetic and immunogenicity analyses

Modifications and changes

Global changes:

The following changes were made throughout the SAP:

- The use of "subject" was updated to "study participant".
- The list of abbreviations was updated accordingly.
- Minor spelling, editorial, and formatting changes were made throughout the document.

Specific changes

In addition to the global changes, the following specific changes have been made (formats as missing spaces or redundant spaces are not listed, typos):

13.1.1 Efficacy variables (Section 2.2.1)

Other efficacy variable moved to the safety section:

- Change from Baseline in Patient Health Questionnaire-9 (PHQ-9)

Other efficacy variables: Additional detail provided as below:

- Change from Baseline in Ankylosing Spondylitis spine MRI-activity (ASspiMRI-a) in the Berlin modification score (Note that this endpoint will be analyzed for study participants in the MRI substudy only)
- Change from Baseline in sacroiliac joint Spondyloarthritis Research Consortium of Canada (SPARCC) score (Note that this endpoint will be analyzed for study participants in the MRI substudy only)

13.1.2 Immunological variables (Section 2.2.4)

New variable:

- The neutralizing antidrug antibody (NAb) status

13.1.3 Safety variables (Section 2.2.5)

- Secondary safety variables now focused on treatment emergent events

New other safety variables:

- Change from Baseline in Patient Health Questionnaire-9 (PHQ-9)

13.1.4 General presentation of summaries and analyses (Section 3.1)

New rules:

- ADaM variables AVAL, CHG or PCHG are rounded to 12 decimal places prior to the comparison to the threshold
- The summary statistics for hs-CRP should also contain arithmetic mean.

13.1.5 Study periods (Section 3.2.2)

New updated study periods definition for the classification of :

- Safety data (Screening period, 16-Week DBTP, 36-Week MP, SFU period)
- Efficacy data (Screening period, 16-Week DBTP, 36-Week MP)
- COVID-19 pandemic periods (Prior/during/post)

13.1.6 Definition of baseline variables (Section 3.3)

New updated baseline definition:

- When no treatment information is available, the randomization date will be used as reference in place of the first day of study drug administration
- Baseline value for HLA-B27
- Baseline value for MRI substudy

13.1.7 Treatment Discontinuation Date and Intercurrent event date (Section 3.4)

- New definition of treatment discontinuation date and intercurrent event date

13.1.8 Protocol deviations (Section 3.5, Section 5.3)

New category of protocol deviation:

- Exclusion from PPS
- COVID-19 related PD

13.1.9 Analysis sets (Section 3.6)

Additional study participants to be excluded from the PPS:

- Study participant excluded from the FAS

New analysis sets:

- Immunogenicity Safety Set
- COVID-19 Free Set

13.1.10 Treatment assignment and treatment groups (Section 3.7)

New category of anaphylactic reaction:

- Events, that fulfil the hypersensitivity reaction criteria

Efficacy analysis will be based:

- on the RS as randomized, for the Double-Blind Treatment Period (displaying data under Placebo and BKZ 160mg Q4W columns),
- on the RS as randomized/assigned treatment for the overall period (displaying data under Placebo/BKZ and BKZ 160mg Q4W columns).

13.1.11 Changes to protocol-defined analyses (Section 3.10)

New changes from the protocol:

- The continuous secondary efficacy endpoints which are part of the sequential testing procedure, as well as the components of the primary ASAS40 endpoint, will be primarily analyzed in implementing a reference-based imputation method.

13.1.12 Changes related to COVID-19 (Section 3.11, Section 5.2)

Additional analyses including analyses by COVID-19 pandemic periods:

- Study participant disposition (Section 5.1)
- Details of impacted visits and effects on collection and reporting of efficacy data (Section 5.3)
- Protocol deviations (Section 5.3)
- Exposure (Section 10.1.4 and Section 10.1.5)
- Adverse events (Section 10.2)

13.1.13 Handling of missing data for efficacy analyses (Section 4.2.1)

- New updated table 4.1 to summarize the different approaches to handle missing data for efficacy endpoints

Change in missing data handling approaches for efficacy endpoints :

- MI - MCMC/ Logistic Regression method no longer used.

Change of primary analysis method for variables in the testing hierarchy

- Primary analysis for binary variables performed under the composite estimand approach (NRI) instead of MI.
- Primary analysis for continuous endpoints to be analyzed using the referenced-based multiple imputation method.

Change of supportive analysis method for variables in the testing hierarchy

- For binary variables, supportive analyses now include Modified Composite estimand approach (MI) (where IE are defined as study treatment discontinuation due to AE or Lack of efficacy)
- For continuous variables (including ACR component at Week16), supportive analyses now performed under the Hypothetical estimand approach (MI).

13.1.14 Handling of missing data for AE (Section 4.2.2)

New rules of imputation of partial start dates:

- When the imputed stop date is prior to the imputed start date
  - For missing start day and start month
  - For missing start day only

13.1.15 Handling of missing data for prior and concomitant medications (Section 4.2.3)

New rules of imputation of partial start dates:

- When the imputed stop date is prior to the imputed start date

13.1.16 Examination of subgroups (Section 4.8)

New categories of variable for efficacy subgroup analyses:

- csDMARDs (Yes/No)
- Timing of participant enrollment relative to COVID-19 pandemic periods as defined in Section 3.2.2.2 (Enrolled prior/during/after to the COVID-19 pandemic)
- Timing of Week 16 Visit relative to the COVID-19 pandemic periods as defined in Section 3.2.2.2 (Study participants who had the Week 16 Visit prior/during/after the COVID-19 pandemic)

13.1.17 Study participant disposition (Section 5.1)

New statistics:

- The number of study participants who completed the DBTP but discontinued prior to entering the MP
- The numbers and percentages of randomized study participants entering the extension study
- The disposition and study discontinuations reasons on the RS and MS up to Week 24
- The number of study participants enrolled during each COVID-19 pandemic period, as well as the number of study participants still in the study during each COVID-19 pandemic period on the RS (overall and by region).

13.1.18 Demographics (Section 6.1)

New categories:

- BMI: <18.5, 18.5 to <25, 25 to <30, ≥30 kg/m2
- Body weight: ≤100, >100 kg
- Body weight: <70, 70 to <95, 95 to < 115, ≥115kg

13.1.19 Baseline characteristics (Section 6.2)

Additional parameter for baseline efficacy variables:

- ASDAS-CRP

Additional parameters for other baseline characteristics:

- Any past biologic therapy (yes, no)
- Current nonsteroidal anti-inflammatory drug (NSAID) drug therapies (yes/no)
- Current oral corticosteroid use (yes, no)
- Current Analgesic/Opioid therapies (yes, no)
- Disease duration (<5, >= 5 years)

13.1.20 Medical history and concomitant diseases (Section 6.3)

New summarized tables:

- Pre-specified medical history coming from 'the 'Infection history' CRF form
- The extra-articular assessments at Screening and Baseline
- Uveitis and IBD assessments at post-Entry Visit

13.1.21 Statistical analysis of the primary efficacy variables (Section 8.1)

Updated the definition of intercurrent event:

- An intercurrent event is defined as discontinuation of study treatment due to **any reason** prior to Week 16.

Additional details:

- The adjusted OR (for the comparison bimekizumab and placebo) and the 95% CI will be provided for secondary analyses of primary efficacy variable
- Reference-based MI method is used for analysis on individual components of the ASAS

- A supportive analysis using modified composite estimand added
- The worst-case scenario added for the tipping points analysis
- Analyses including COVID-19 impact added

13.1.22 Statistical analysis of the secondary efficacy variables (Section 8.2)

Additional details added for derivation:

- BASDAI: if 1 of the 2 morning stiffness measurements (ie, questions Q5 and Q6) is missing, the other one will be used for the morning stiffness calculation
- Disease activity categories based on ASDAS-CRP added
- Re-classification of Change from Baseline in Patient Health Questionnaire-9 (PHQ-9) variable from other efficacy variable to other safety variable
- SF-3 component summaries (PCS and MCS scores) and normal range
- Updated BASMI linear definition (table 8.1, table 8.2)
- The order of events to derive the BASMI scores
- MASES: only randomized study participants with enthesitis at Baseline will be included in the multiple imputation models

Change of analysis of secondary effect variables

- Secondary binary endpoints performed under the composite estimand approach (NRI)
- Secondary continuous endpoints performed under the referenced-based MI for primary analyses and hypothetical estimand MI method for supportive analyses in the testing hierarchy
- New section for describing additional statistical analyses of selected secondary efficacy variables

13.1.23 Analysis of other efficacy variables (Section 8.3)

Additional details added for derivation:

- Time to a given response (ASAS20 or ASAS40) will be defined as the time in days from the start date of first treatment dose
- Extra rules for calculation of hs-CRP
- Formula to derive the FACIT-Fatigue subscale score
- WPAI-SHP: derivation of additional four scores after implementing the missing data rules for questions Q1-Q5
- Extra rules for calculation of ASspiMRI-a score and SPARCC score

New rules for handling of missing data for the analysis of other efficacy endpoints

Additional analyses for other efficacy endpoints:

- Addition of some extra analyses of efficacy endpoint with statistical testing

13.1.24 Pharmacokinetics (Section 9.1)

Additional details added for:

- Definition of ADAb Sample Status
- Definition of Cumulative ADAb Status
- PK Sample Inclusion and Exclusion Rules
- Summary of Analyses

13.1.25 Immunology (Section 9.2)

Additional details added for:

- Definition of Overall ADAb status
- Definition of ADAb Categories
- ADAb status groups for Efficacy Endpoint Summaries
- ADAb Sample Inclusion and Exclusion Rules
- Detailed immunology analyses
- Neutralizing anti-bimekizumab antibodies

13.1.26 Extent of exposure (Section 10.1)

- Now also calculated by COVID-19 pandemic period

13.1.27 Adverse events (Section 10.2)

- Use of latest guidelines from the BKZ AE of special monitoring convention document

New tables/listings for:

- The adjudicated neuropsychiatric events by type
- The adjudicated IBD events by type
- The incidence of each safety topics of interest occurring in the DBTP
- Extra-Articular Manifestations
- Impact of COVID-19

Additional details provided:

- Hypersensitivity now includes injection site reactions in addition to anaphylaxis
- Calculation of the risk difference with the upper and lower confidence interval

13.1.28 Clinical laboratory evaluations (Section 10.3)

Changes the definitions of markedly abnormal values for:

- Creatinine
- Glucose
- Cholesterol

New figures for:

- Neutrophil values for study participants with at least one treatment emergent markedly abnormal neutrophil value over time
- ALT (or AST) values for study participants with at least one treatment emergent markedly abnormal value

13.1.29 Other safety variables (Section 10.4.3)

New added variables:

- Patient Health Questionnaire-9 (PHQ-9)
- Questionnaires

13.2 AMENDMENT 2

Rationale for the amendment

The major purpose of this SAP amendment was to fix formatting issues in the SAP document and to add clarifications on how to analyze specific data.

13.2.1 Summary of Analyses (Section 9.1.4)

Removal of PK analysis by region subgroup.

13.2.2 Definition of ADAb categories (Section 9.2.1.3)

Update on the ADAb category 9 definition.

13.2.3 Handling of missing data for efficacy analyses (Section 4.2.1)

Change in missing data handling approaches for efficacy endpoints :

- LOCF will be used for the secondary continuous endpoints in case of convergence issue of the Reference-Based Multiple Imputation analyses.
- Update in the rule described in Step 1c for the reference-based imputation

New section added:

- Section 4.2.1.11 for Last Observation Carried Forward
- Section 4.2.1.12 for allowed ranges for continuous efficacy variables

13.2.4 Clinical laboratory evaluations (Section 10.3)

Change in the use of the local laboratory results.

13.2.5 Other safety variables (Section 10.4.3)

New rules for visits with incomplete eC-SSRS questionnaire or duplicated questionnaires.

13.2.6 Baseline characteristics (Section 6.2)

Change in the cut-off category for disease duration and time since first diagnosis of AS, from 5 to 2 years.

13.2.7 Examination of subgroups (Section 4.8)

Change in the cut-off category for disease duration, from 5 to 2 years.

13.2.8 Definition of baseline variables (Section 3.3)

Updated baseline definition for component scores.

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.

Approval Signatures

Name: AS0011-sap-amend-2

Version: 3.0

Document Number: CLIN-000144514

Title: AS0011-Statistical Analysis Plan

Approved Date: 15 Nov 2021

Document Approvals	
Approval Verdict: Approved	Name: [redacted] Capacity: Clinical Date of Signature: 15-Nov-2021 15:23:44 GMT+0000
Approval Verdict: Approved	Name: [redacted] Capacity: Document Author Date of Signature: 15-Nov-2021 15:30:04 GMT+0000
Approval Verdict: Approved	Name: [redacted] Capacity: Clinical Date of Signature: 15-Nov-2021 15:31:24 GMT+0000