STATISTICAL ANALYSIS PLAN

Study: PA0010

Product: Bimekizumab

A PHASE 3, MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, ACTIVE REFERENCE (ADALIMUMAB) STUDY EVALUATING THE EFFICACY AND SAFETY OF BIMEKIZUMAB IN THE TREATMENT OF SUBJECTS WITH ACTIVE PSORIATIC ARTHRITIS

SAP/Amendment Number	Date
Final SAP	27 MAY 2019
Amendment 1	06 JUL 2021
Amendment 2	22 OCT 2021
Amendment 3	06 MAY 2022

Confidentiality Statement

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACR	American College of Rheumatology
ACR20,50,70	American College of Rheumatology 20, 50, 70% response criteria
ADA	adalimumab
ADAb	antidrug antibody
ADR	adverse drug reaction
AE	adverse event
AMP	Active Medication Periods
AMS	Active Medication Set
ANCOVA	analysis of covariance
ATC	anatomic therapeutic class
ATP	Active Treatment-Blind Period
ATS	Active Treatment-Blind Set
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
bDMARD	biologic disease modifying antirheumatic drug
BKZ	bimekizumab
BLQ	below the limit of quantification
BMI	body mass index
BP	blood pressure
BSA	body surface area
CASPAR	Classification Criteria for Psoriatic Arthritis
CCP	cyclic citrullinated peptide
cDMARD	conventional disease modifying antirheumatic drug

CI	confidence interval
COVID-19	coronavirus disease 2019
CRF	case report form
CRP	C-reactive protein
CV	coefficient of variation
DAPSA	Disease Activity Index for Psoriatic Arthritis
DAS28[CRP]	Disease Activity Score-28 based on C-reactive protein
DBP	Double-Blind Treatment Period
DILI	drug-induced liver injury
DMARDs	disease modifying antirheumatic drugs
DMC	data monitoring committee
eC-SSRS	electronic Columbia Suicide Severity Rating Scale
ECG	electrocardiogram
EQ-5D-3L	Euro-Quality of Life 5-Dimensions 3 Level version
ES	erosion score
ET	Early Termination
EAER	exposure adjusted event rate
EAIR	exposure adjusted incidence rate
FACIT	Functional Assessment of Chronic Illness Therapy
FAS	Full Analysis Set
FDA	Food and Drug Administration
HAQ-DI	Health Assessment Questionnaire - Disability Index
HLA-B27	human leukocyte antigen B27

HLT	high level term
hs-CRP	high sensitivity C-reactive protein
ICH	International Council for Harmonisation
IE	intercurrent event
IGA	Investigator's Global Assessment
IMP	investigational medicinal product
IPD	important protocol deviation
IWRS	interactive web response system
JSN	joint space narrowing
KM	Kaplan-Meier
LCL	lower confidence limit
LDI	Leeds Dactylitis Index
LEI	Leeds Enthesitis Index
LLOQ	lower limit of quantification
LLT	lowest level term
LN	natural logarithm
LSM	least square means
MACE	major adverse cardiac events
MAR	missing at random
MCMC	Markov-Chain Monte Carlo
MCS	Mental Component Summary
MDA	Minimal Disease Activity
MedDRA	Medical Dictionary for Regulatory Activities

MI	multiple imputation
mNAPSI	modified Nail Psoriasis Severity Index
NAb	neutralizing anti-bimekizumab antibodies
NRI	non-responder imputation
NSAID	nonsteroidal anti-inflammatory drug
OC	observed case
OLE	open label extension
OR	odds-ratio
PASDAS	Psoriatic Arthritis Disease Activity Score
PASI	Psoriasis Area and Severity Index
PASI75, PASI90, PASI100	Psoriasis Area and Severity Index 75%, 90%, 100%
PASS	Patient acceptable symptom state
PCS	Physical Component Summary
PD	Protocol Deviation
PGA-Arthritis	Patient's Global Assessment of Arthritis
PGA-PsA	Patient's Global Assessment of Psoriatic Arthritis
PhGA-Arthritis	Physician's Global Assessment of Arthritis
PhGA-PsA	Physician's Global Assessment of Psoriatic Arthritis
PHQ-9	Patient Health Questionnaire-9
PK-PPS	Pharmacokinetics Per-Protocol Set
PPS	Per-Protocol Set
PsA	psoriatic arthritis
PsAID-12	Psoriatic Arthritis Impact of Disease-12

PsAQoL	Psoriatic Arthritis Quality of Life
PsARC	Psoriatic Arthritis Response Criteria
PSO	psoriasis
PT	preferred term
PtAAP	Patient's Assessment of Arthritis Pain
Q2W	every 2 weeks
Q4W	every 4 weeks
RAS	Radiographic Set
RS	Randomized Set
SAE	serious adverse event
SAP	Statistical Analysis Plan
SC	subcutaneous(ly)
SD	standard deviation
SFU	Safety Follow-up
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
TB	tuberculosis
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event

TFLs	tables, figures, and listings
TJC	tender joint count
TNF	tumor necrosis factor
TNFα	tumor necrosis factor alpha
UCL	upper confidence limit
ULN	upper limit of normal
VAS	Visual Analog Scale
vdHmTSS	van der Heijde modified Total Sharp Score
VLDA	Very Low Disease Activity
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 interim (Week 24 and Week 52 analyses) and the final statistical analyses for study PA0010. It also defines the summary tables, figures, and listings (TFLs) to be generated for the Clinical Study Report.

The SAP is based on Protocol amendment 2 (22 February 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 clinical efficacy of bimekizumab (BKZ) administered subcutaneously (sc) every 4 weeks (Q4W) for 16 weeks compared with placebo in the treatment of subjects with active psoriatic arthritis (PsA), as assessed by the American College of Rheumatology 50% response criteria (ACR50).

2.1.2 Secondary objectives

The secondary objectives are as follows:

- To assess the efficacy of BKZ compared with placebo
- To assess the safety and tolerability of BKZ
- To assess the impact of BKZ on patient-reported quality of life
- To assess the impact of BKZ on skin psoriasis (PSO) in the subgroup of affected subjects at Baseline
- To assess the impact of BKZ on functional improvement
- To assess the impact of BKZ on radiographic changes in the hands and feet
- To assess the impact of BKZ on extra-articular disease manifestations (dactylitis, enthesitis).

2.1.3 Other objectives

Other objectives are as follows:

- To descriptively assess the efficacy of BKZ with reference to adalimumab (ADA)
- To assess the immunogenicity of BKZ
- To assess the impact of BKZ treatment on axial disease

- To assess the maintenance of treatment effect
- To assess nail PSO in the subgroup of affected subjects at Baseline
- To explore the exposure response relationship of BKZ
- To assess the effect of BKZ on gene and protein expression, and explore the relationship between genomic, genetic, and proteomic biomarkers and disease biology, drug treatment and inflammatory and immune responses (from consenting subjects who agree to participate in the biomarker substudy)
- To assess the impact of BKZ on social life and work productivity.

2.2 Study variables

2.2.1 Efficacy variables

2.2.1.1 Primary efficacy variable

The primary efficacy variable is the ACR50 response at Week 16.

2.2.1.2 Secondary efficacy variables

The secondary efficacy variables are as follows:

- Change from Baseline in Health Assessment Questionnaire - Disability Index (HAQ-DI) at Week 16
- Reduction of 90% from Baseline in Psoriasis Area and Severity index (PASI90 response) at Week 4 and Week 16 in the subgroup of subjects with PSO involving at least 3% of Body Surface Area (BSA) at Baseline
- Change from Baseline in the Short Form 36-item Health Survey (SF-36) Physical Component Summary (PCS) at Week 16
- Minimal Disease Activity (MDA) response at Week 16
- Change from Baseline in Van der Heijde modified Total Sharp Score (vdHmTSS) at Week 16 on subjects with elevated high sensitivity C-reactive protein (hs-CRP) and/or at least 1 bone erosion at Baseline
- Enthesitis-free state based on the Leeds Enthesitis Index (LEI) at Week 16 in the subgroup of subjects with enthesitis at Baseline in the pooled population of PA0010 and PA0011
- Dactylitis-free state based on the Leeds Dactylitis Index (LDI) at Week 16 in the subgroup of subjects with dactylitis at Baseline in the pooled population of PA0010 and PA0011
- Change from Baseline in vdHmTSS at Week 16
- ACR20 response at Week 16
- ACR70 response at Week 16
- Proportion of subjects with an Investigator Global Assessment (IGA) score of 0 (clear) or 1 (almost clear) and at least a 2-grade reduction from Baseline at Week 4 and Week 16 in the subset of subjects with psoriatic skin lesions at Baseline

- Change from Baseline in the Patient's Assessment of Arthritis Pain (PtAAP) at Week 16
- Enthesitis-free state based on the Spondyloarthritis Research Consortium of Canada (SPARCC) index at Week 16 in the subgroup of subjects with enthesitis at Baseline
- Change from Baseline in Psoriatic Arthritis Impact of Disease-12 (PsAID-12) total score at Week 16.

2.2.1.3 Other efficacy variables

Change from Baseline variables evaluated during the Double-Blind Treatment Period (DBP) are relative to the Baseline (first dose) visit. For subjects randomized to placebo who switch to BKZ 160mg Q4W at the Week 16 visit, change from Baseline variables during the Active Treatment-Blind Period (ATP) may be evaluated relative to both the Baseline (first dose) visit and to the Week 16 visit (first dose of BKZ).

For simplicity, "change from Baseline" is used below for all such variables. More details will be provided below in Section 3.3.

The other efficacy variables listed below will be summarized for all timepoints except for the previously defined primary and secondary efficacy variables which will be also summarized separately at Week 16 (and at Week 4 and 16 for PASI90).

The following other efficacy variables will be assessed (all timepoints not specified in Section 2.2.1.1 and 2.2.1.2 are exploratory):

- Time to ACR20, ACR50, and ACR70 response from Baseline (Day 1)
- ACR20, ACR50, and ACR70 response
- PASI75, PASI90, and PASI100 response in the subgroup of subjects with PSO involving at least 3% BSA at Baseline
- Composite endpoint composed of ACR50 and PASI90 in subjects with PSO involving at least 3% BSA at Baseline
- Composite endpoint composed of ACR50 and PASI100 in subjects with PSO involving at least 3% BSA at Baseline
- Proportion of Psoriatic Arthritis Response Criteria (PsARC) responders
- Psoriatic Arthritis Disease Activity Score (PASDAS) categories
- Change from Baseline in the PASDAS
- Proportion of ACR50 responders at Week 16 and maintaining response at Week 52
- MDA response
- Very Low Disease Activity (VLDA) response
- Proportion of subjects with an IGA score of 0 (clear) or 1 (almost clear) and at least a 2-grade reduction from Baseline in the subset of subjects with psoriatic skin lesions at Baseline
- Disease Activity Index for Psoriatic Arthritis (DAPSA) score categories

- Change from Baseline in DAPSA score
- Change from Baseline in the Disease Activity Score-28 based on C-reactive protein (DAS28[CRP])
- Change from Baseline in all individual ACR core components:
  - Swollen Joint Count (SJC)
  - Tender Joint Count (TJC)
  - HAQ-DI
  - PtAAP
  - Physician's Global Assessment of Psoriatic Arthritis (PhGA-PsA)
  - Patient's Global Assessment of Psoriatic Arthritis (PGA-PsA)
  - hs-CRP
- Change from Baseline in vdHmTSS total score and in erosion and joint space narrowing (subscores) in subjects with elevated hs-CRP and/or with at least 1 bone erosion at Baseline
- Change from Baseline in vdHmTSS total score and in erosion and joint space narrowing (subscores) in the overall population
- Proportion of subjects with no radiographic joint damage progression (change from Baseline vdHmTSS of ≤0.5)
- Dactylitis-free state based on the LDI in the subgroup of subjects with dactylitis at Baseline in the pooled population of PA0010 and PA0011, as well as for PA0010 alone
- Enthesitis-free state based on the LEI in the subgroup of subjects with enthesitis at Baseline in the pooled population of PA0010 and PA0011, as well as for PA0010 alone
- Enthesitis-free state based on the SPARCC index in the subgroup of subjects with enthesitis at Baseline in the pooled population of PA0010 and PA0011, as well as for PA0010 alone
- Proportion of subjects with a decrease of HAQ-DI from Baseline of at least 0.35 in the subgroup of subjects with Baseline HAQ-DI≥0.35
- Change from Baseline in the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) in the subgroup of subjects with axial involvement defined by a score of ≥4 at Baseline
- Change from Baseline in the modified Nail Psoriasis Severity Index (mNAPSI) score in the subgroup of subjects with psoriatic nail disease at Baseline (mNAPSI score>0)
- Change from Baseline in the LEI in the subgroup of subjects with enthesitis at Baseline in the pooled population of PA0010 and PA0011, as well as for PA0010 alone
- Change from Baseline in the SPARCC index in the subgroup of subjects with enthesitis at Baseline in the pooled population of PA0010 and PA0011, as well as for PA0010 alone
- Change from Baseline in the LDI in the subgroup of subjects with dactylitis at Baseline in the pooled population of PA0010 and PA0011, as well as for PA0010 alone

- Change from Baseline in PsAID-12 total score, as well as the individual domain scores
- Proportion of subjects achieving PsAID-12 total score ≤4
- Proportion of PsAID-12 responders (decrease from Baseline in PsAID-12 total score ≥3) in subjects with PsAID-12 total score>3 at Baseline
- Change from Baseline in the Psoriatic Arthritis Quality of Life (PsAQoL) total score
- Change from Baseline in the SF-36 PCS and Mental Component Summary (MCS), as well as the 8 domain scores (Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional, and Mental Health)
- Change from Baseline in Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) subscale score
- Proportion of FACIT-Fatigue subscale responders (subjects with a minimum clinically important difference for FACIT-Fatigue subscale score, defined as an increase of ≥4) in subjects with FACIT-Fatigue subscale score ≤48 at Baseline
- Change from Baseline in Work Productivity and Activity Impairment Questionnaire-Specific Health Problem (WPAI-SHP) v2.0 adapted to PsA scores
- Responses to the EuroQol-5 Dimensions-3 Level (EQ-5D-3L)
- Change from Baseline in EQ-5D-3L Visual Analog Scale (VAS) scores
- Change from Baseline in Physician's Global Assessment of Arthritis (PhGA-Arthritis)
- Change from Baseline in Patient's Global Assessment of Arthritis (PGA-Arthritis).

2.2.2 Pharmacokinetic variable

The pharmacokinetic (PK) variable is the plasma concentration of BKZ. Subjects are asked to provide blood samples for these measurements at Baseline, Week 2, 4, 6, 8, 12, 16, 20, 24, 36, 52, end of treatment and SFU visit.

2.2.3 Pharmacogenomic variables

Genomic, genetic, epigenetic, proteins, and metabolite biomarkers may be measured to evaluate the relationship with response to treatment with BKZ, psoriatic arthritis disease biology, and inflammatory and immune response processes. The nature and format of these tentative substudy analyses will be determined when the results of the main study are made available.

The candidate exploratory variables are the blood or blood derivative (eg, serum), concentrations of cytokines and chemokines of relevance to IL-17A/F signaling pathway and psoriatic arthritis biology. Additional variables may include but will not be limited to serum complement concentrations. Where local regulations permit, additional blood samples will be collected at Baseline and at Week 16. They may be used to allow for potential exploratory analyses of ribonucleic acid, proteins, lipids, and metabolite biomarkers relevant to disease biology and progression, response to therapy, the inflammatory and immune response processes, and cardiovascular risk in PsA.

A specific SAP will be written to describe the analysis methods for those variables. The nature and format of these analyses will be determined later.

2.2.4 Immunological variables

The immunological variables allow evaluation of immunogenicity as following:

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

These 3 variables will be assessed at Baseline, Week 4, 8, 12, 16, 20, 24, 36, 52, end of treatment and at SFU visit.

2.2.5 Safety variables

2.2.5.1 Secondary safety variables

Main safety variables to be assessed are as follows:

- Incidence of treatment-emergent adverse events (TEAEs)
- Incidence of treatment-emergent serious adverse events (TESAEs)
- 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 the 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

2.3.1 Study description

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, noninferential active reference study to evaluate the efficacy and safety of BKZ in subjects with active PsA who are biologic disease-modifying antirheumatic drug (bDMARD) naive. To be eligible to participate in this study, subjects must be adults with a diagnosis of active PsA based on the Classification Criteria for Psoriatic Arthritis (CASPAR) and have disease with TJC ≥3 and SJC ≥3. In addition, subjects should not have been previously exposed to any bDMARD for PsA or PSO to be eligible for the study.

Approximately 840 subjects will be randomly assigned in a 3:2:1 ratio, stratified by region (North America, Western Europe, Eastern Europe, and Asia/Australia) and bone erosion (0, ≥1) to 1 of the 3 following blinded treatment groups:

- Bimekizumab 160mg sc Q4W (420 subjects),
- Placebo (280 subjects),
- Adalimumab sc 40mg every 2 weeks (Q2W) (140 subjects).

It is planned to enroll a minimum of 45% of subjects who have elevated hs-CRP (≥6mg/L) and/or who have at least 1 bone erosion at Screening.

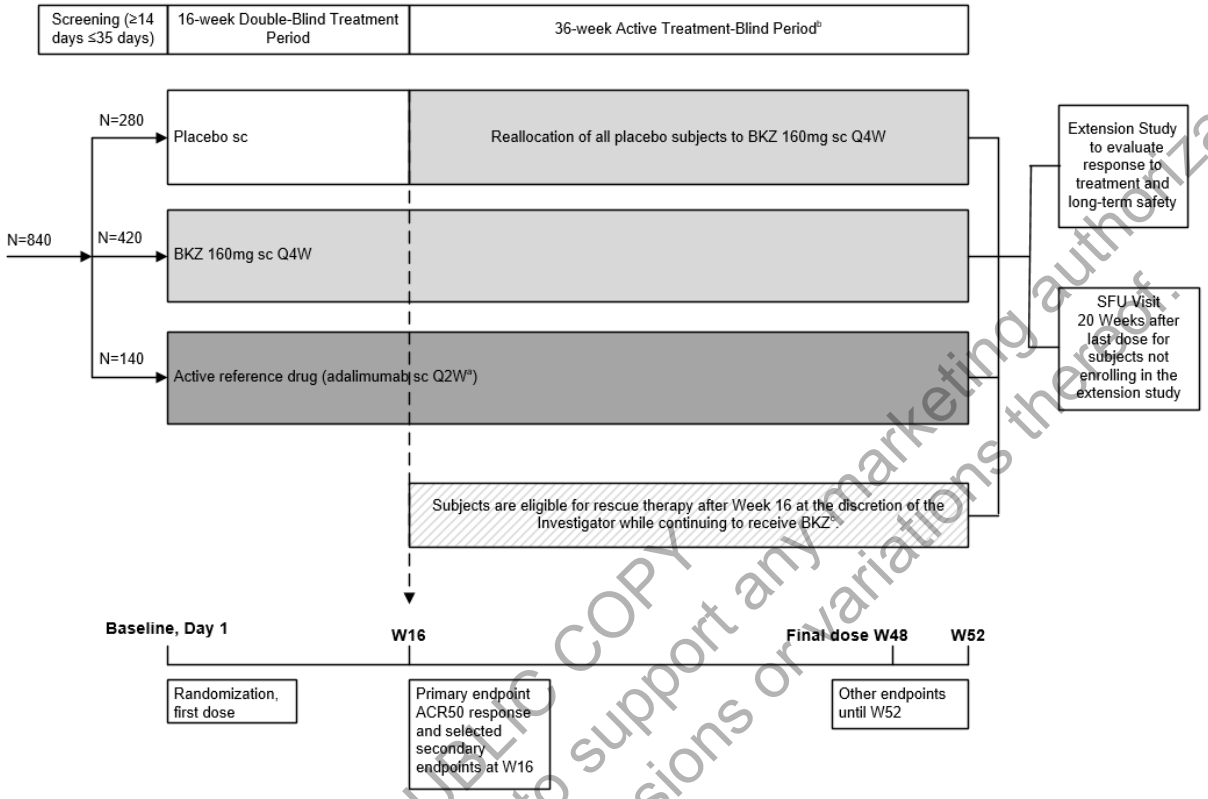
The overall study design consists of:

- a Screening Period (≥14 days to ≤35 days)
- a 16-week placebo controlled Double-Blind Treatment Period (DBP)
- a 36-week Active Treatment-Blind Period (ATP) (Week 16 to week 52)
- a Safety Follow Up (SFU) visit, 20 weeks after the final dose of IMP (for subjects not entering the open label extension (OLE) study or who discontinue early, including those withdrawn from IMP, as applicable).

The maximum study duration per subject will be up to 73 weeks.

The study schematic diagram for the study is presented in .

Figure 2–1: Study schematic diagram



ACR50=American College of Rheumatology 50% response criteria, BKZ=bimekizumab, PsA=psoriatic arthritis; Q2W=every 2 weeks, Q4W=every 4 weeks, sc=subcutaneous, SFU=Safety Follow-Up, W=week

[a] Adalimumab dosing is performed every 2 weeks from Baseline to Week 50.

[b] After 16 weeks of double-blind treatment, subjects will enter the Active Treatment-Blind Period. All subjects randomized to placebo will be reallocated to receive BKZ 160mg Q4W. Subjects randomized to BKZ 160mg Q4W will continue to receive their originally randomized dose. Subjects randomized to active reference (ADA) will continue with their active treatment.

[c] Permitted rescue therapy is defined in Protocol Section 5.4.

2.3.2 Study Periods

2.3.2.1 Screening Period

Subjects will enter a Screening Period (starting with a Screening visit - Visit 1) that will last for a minimum duration of 14 days and a maximum duration of 35 days and will involve obtaining laboratory data and verifying that the doses of nonsteroidal anti-inflammatory drugs (NSAIDs), or permitted DMARDs, if used to treat PsA, are stable. The Screening Period will also enable washout of any medications not permitted for use during the study.

During the Screening Period, X-rays of hands and feet will be assessed for bone erosion and used for the stratification of subjects at Baseline (Day 1).

This period starts at time of the informed consent date (Screening visit -Visit 1) and ends the day prior to the day of the first dose of study drug (Baseline visit – Visit 2).

2.3.2.2 Double-Blind Treatment Period

After the Screening Period, subjects will attend a Baseline visit (Visit 2, Day 1).

At the end of the Baseline visit, subjects will be randomized according to a 3:2:1 ratio (stratified by region and bone erosion [0, ≥1] at Screening) to receive 1 of the 3 following blinded treatments during the 16 weeks of the DBP:

- Bimekizumab 160mg sc Q4W
- Placebo
- Adalimumab 40mg sc Q2W.

The first administration of IMP will be delivered at the Baseline visit (Visit 2, Day 1).

During the 16 weeks of the DBP, IMP administration will be performed by unblinded study personnel every 2 weeks (allowed visit window between 12 to 16 days) from Day 1 through Week 16 (Visits 2 to 10).

For the BKZ group, active treatment will be administered every 4 weeks and placebo in between, to maintain the blind and still have an injection every 2 weeks.

Visit 10 (Week 16) ends the DBP, and the last assessments for that period are collected during that visit.

This period starts on the day of first dose administration of study drug (Visit 2) and ends at Week 16 (Visit 10) or at the Early Termination visit (ET Visit) for subjects withdrawn from the study before Week 16. This period should last 16 weeks maximum.

2.3.2.3 Active Treatment-Blind Period

After the 16-week DBP, subjects will enter the 36-week ATP. At Week 16 (Visit 10), subjects will be allocated to treatment regimens as follows:

- Subjects in the BKZ 160mg sc Q4W group will continue to receive BKZ 160mg sc Q4W
- Subjects in the placebo group will be reallocated to BKZ 160mg sc Q4WSubjects in the ADA 40mg sc Q2W group will continue to receive ADA 40mg sc Q2W.

As for the DBP, BKZ and ADA will be administered sc by unblinded study personnel at the clinical site. The first dose in the ATP will be administered at Week 16 after all assessments have been completed.

During the 36 weeks of the ATP, IMP administration will be performed every 2 weeks (allowed visit window between 11 and 17 days) at Study visits (from Visit 10 – Week 16 through Visit 27 - Week 50).

Visit 28 (Week 52) ends the ATP, and the last assessments for that period are collected during that visit.

Subjects who withdraw early from the study will undergo the ET visit assessments and will enter the SFU Period, as applicable. Subjects who withdraw from IMP during the ATP will be encouraged to return for all remaining scheduled visits up to Week 52 and the SFU visit (20 weeks after their final dose of IMP).

At the completion of the ATP, investigators should discuss treatment options with the subject. Subjects who complete PA0010 (ie who performed all visits up to Week 52 and not having permanently discontinued IMP and who did not meet any of the withdrawal criteria) will be given the opportunity to enter the OLE study at Week 52.

The ATP starts on the day of dose administration at Week 16 and ends at Week 52 (Visit 28) or at the Early Termination visit (ET Visit) for subjects withdrawn from the study after Week 16 and before Week 52. This period should last 36 weeks maximum.

2.3.2.4 Safety Follow-up

All subjects who complete the study and do not enter the OLE study, or who discontinue early, including those withdrawn from IMP, will have a SFU visit 20 weeks after their final dose of IMP, as applicable.

2.4 Determination of sample size

2.4.1 Power calculations for the primary endpoint

The sample size assumptions for BKZ versus placebo are based on the ACR50 response data from the Phase 2b BKZ study in subjects with moderate-to-severe PsA (PA0008). The median ACR50 responses of the top 3 dose groups (BKZ 160mg, 320mg, and 320mg [initial dose] plus 160mg) at Week 12 in the Tumor necrosis factor-naïve (TNF-naïve) population are conservatively assumed for the Week 16 endpoint. The observed median ACR50 response rate of the top 3 BKZ doses in the TNF-naïve population in study PA0008 was 43.8%.

The placebo ACR50 response at Week 16 is based on the TNFα-naïve population in PA0008 (6.1% at Week 12, subjects with available measurement (n)=33); Mease et al, 2015), FUTURE 2 study in the subgroup of TNFα-naïve subjects (15.9%, n=63; McInnes et al, 2015), FUTURE 3 study (11.8%, n=93; Nash and Mease, 2018), and FUTURE 5 study in a mixed tumor necrosis factor alpha (TNFα) exposure population (8.1% at Week 16, n=332; Mease et al, 2018). Therefore, the estimated ACR50 response at Week 16 in the placebo group is assumed to be 16%.

The sample size for showing statistical superiority of BKZ vs placebo was calculated using a 2-sided 2-sample Chi-square test with continuity correction (Fleiss et al, 1980). Assuming 420 subjects in the BKZ group and 280 subjects in the placebo group, the test for detecting statistical superiority of BKZ 160mg Q4W vs placebo based on ACR50 response at Week 16 has >99% power to detect a true treatment difference of 27.8% (odds ratio (OR) 4.09).

2.4.2 Power calculations for secondary endpoints

The assumptions for power calculations of the secondary endpoints included in the hierarchy, and for which supporting data exists, are based on the interim results of the Phase 2b BKZ study PA0008 and the FUTURE 1, FUTURE 2, FUTURE 5, and SPIRIT P1 studies. 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).

For the PASI90 response at Week 16, the BKZ 160mg Q4W treatment response at Week 12 is 45% (n=20) in the TNFα-naïve population of PA0008 at Week 12 and 46% (n=28) at Week 12

in the mixed TNFα exposure population. The placebo PASI90 response at Week 16 is based on the subgroup analyses of TNFα therapy-naïve subjects in PA0008 (9.1% at Week 12; n=22), FUTURE 2 study (9.7% at Week 24; n=63), and SPIRIT P1 study (1.5% at Week 12; n=67). Therefore, a placebo PASI90 response rate of 10% is assumed. With those assumptions, the study has >99% power to detect a true treatment difference at an assumed 60% of subjects with BSA ≥3% of the planned sample size.

For change from Baseline in HAQ-DI at Week 16, the between treatment differences of the change from Baseline in HAQ-DI of in the mixed TNFα exposure population of the PA0008 at Week 12 were used. The BKZ 160mg Q4W treatment group mean change from Baseline in HAQ-DI at Week 12 of -0.37, Standard deviation (SD)=0.47 (n=41) versus placebo mean change from Baseline in HAQ-DI at Week 12 of -0.13, SD=0.50 (n=42). With those assumptions, the study has >99% power to detect a true treatment difference of -0.24 at the planned sample size.

For change from Baseline in PCS of SF-36 at Week 16, the between treatment differences of the change from Baseline in PCS of SF-36 of the TNFα-naïve population of FUTURE 2 study (Kavanaugh et al, 2016) at Week 24 were used. The BKZ 160mg Q4W treatment group assumes the secukinumab 150mg treatment group mean change from Baseline in SF-36 PCS at Week 24 of 7.91, SD 7.38 (n=63) versus placebo mean change from Baseline in SF-36 PCS at Week 24 of 2.08, SD 9.51 (n=63). With those assumptions, the study has >99% power to detect a true treatment difference of 5.83 at the planned sample size.

For MDA at Week 16, the BKZ 160mg Q4W treatment response at Week 12 in PA0008 is 46.3% (N=41) in the mixed TNFα exposure population of PA0008. In the placebo group, 14.3% of subjects achieved MDA at Week 12 (n=42). With those assumptions, the study has >99% power to detect a true treatment difference of 32% (OR=5.17) at the planned sample size.

For change from Baseline in vdHmTSS at Week 16, the between treatment differences of the change from Baseline in vdHmTSS at Week 16 of the SPIRIT P1 study in a TNFα-naive population (Mease et al, 2017) were used. The SPIRIT P1 study ixekizumab 80mg sc Q2W treatment group least square mean (LSM) change from Baseline at Week 16 0.06 and SD 0.720; and LSM change from Baseline 0.36 and SD 0.710 for placebo are assumed. With those assumptions, the study has a 95% power to detect a true treatment difference of -0.30, assuming 45% of subjects with elevated hs-CRP and/or bone erosion, in the planned sample size.

There is uncertainty in the final percentage of subjects with elevated hs-CRP and/or bone erosion at Baseline that will be recruited to the study and the variability in the subgroup. To demonstrate the sensitivity of the sample size calculation for this study, Table 2–1 shows the power function for a fixed total sample size of 840 subjects in the Randomized Set (RS) population for this secondary efficacy analysis, varying the percentage of subjects with elevated hs-CRP and/or bone erosion at Baseline and the treatment difference and estimated standard deviation on the change from Baseline in vdHmTSS at Week 16.

Table 2–1: Treatment effect elevated hs-CRP and/or bone erosion at Baseline for secondary efficacy analyses

Subjects with elevated hs-CRP and/or bone erosion at Baseline	Treatment difference/standard deviation	
	Treatment effect = -0.3 SD = 0.72 vs. 0.71	**Treatment effect = -0.35 SD = 0.81 vs. 0.93**
20%	67	62
25%	76	72
30%	84	79
35%	89	85
40%	92	89
45%	95	92
50%	96	95

hs-CRP=high-sensitivity C-reactive protein ; SD=standard deviation

For dactylitis-free state at Week 16, dactylitis resolution data of the FUTURE 5 study (Mease et al, 2018) at Week 16 were used to estimate the BKZ 160mg Q4W treatment and placebo groups. The treatment effect estimate from the secukinumab 150mg treatment group at Week 16 of 56% (n=103) is assumed versus a placebo group estimate of 32% (n=124). With those assumptions and having 11% of subjects with Baseline dactylitis, the pooled PA0010/PA0011 studies have a 66% power to detect a true treatment difference of 24% (OR= 2.71).

For enthesis-free state at Week 16, enthesis resolution data from the FUTURE 5 study at Week 16 were used to estimate the BKZ 160mg Q4W treatment and placebo groups. In FUTURE 5 The treatment effect estimate from the secukinumab 150mg treatment group at Week 16 was 55% (n=141) and the placebo group estimate was 36% (n=192). Assuming a BKZ response of 55% and a placebo rate of 39% and having 25% of subjects with Baseline enthesis the pooled PA0010/PA0011 the studies have a 69% power to detect a true treatment difference of 16% (odds ratio 1.91). Treatment difference in change from Baseline in vdHmTSS at Week 16 in the overall population is estimated using the data from the FUTURE 5 study in a mixed TNFα exposure population. The BKZ 160mg Q4W group treatment difference compared to placebo is estimated to be the same as the difference in change from Baseline in vdHmTSS at Week 16 between the secukinumab 150mg group and placebo which is 0.42, assuming a SD of 1.13 for BKZ and SD=2.44 for placebo. With these assumptions the study has 76% power to detect a treatment difference of 0.42 in the planned sample size.

3 DATA ANALYSIS CONSIDERATIONS

3.1 General presentation of summaries and analyses

Statistical analysis and generation of TFLs and statistical outputs will be performed using SAS® Version 9.3 or higher unless otherwise specified. 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 subjects in each category will be presented. Unless otherwise noted, the denominator for percentages will be based on the number of subjects included in the respective analysis set. Subjects with missing data can generally be accounted for using one of the 2 following approaches:

- Percentages will be summarized based on all subjects in the analysis set and a "Missing" category (corresponding to subjects with missing data for the variable being summarized) will be included as the last row in the list of categories being summarized. This approach will be used when reporting demography, Baseline characteristics, some NAb tables and some shift tables from Baseline for laboratory data.
- Percentages will be based only on those subjects with observed data for the variable being summarized. As the denominator may be different from the number of subjects in the analysis set being considered, the denominator should be displayed in the table. The general format for displaying this will be "n/Nsub (%)".

Unless otherwise noted, all percentages will be displayed to 1 decimal place. No percentage will be displayed for zero counts, and no decimal will be presented when the percentage is 100%.

For continuous parameters, descriptive statistics will include number of subjects 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 (for the value and the ratio to Baseline). The geometric CV (%) will be calculated using the following formula:

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

Where SD_{ln} represents the SD of the ln-transformed hs-CRP value.

For BKZ PK concentrations, geometric mean, geometric CV, 95% confidence intervals (CIs) for geometric mean (assuming log-normally distributed data) will be calculated if at least ⅔ of the values of interest are above the lower limit of quantification (LLOQ). 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 (Q1 and Q3 when applicable) will use 1 additional decimal place compared to the original data (for a derived score, number of decimals of the original data considered is the one obtained where deriving the score in the absence of missing data)
- CV (%) will be presented with 1 decimal place
- Minimum and maximum will have the same number of decimal places as the original value.

If the number of decimal places reported in the original 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.

The table below provides the number of decimals to be reported for each continuous efficacy endpoint:

Table 3–1: List of decimals for continuous efficacy endpoints

Continuous efficacy endpoint	Decimal place used for minimum and maximum	Decimal places used for Mean, SD (or SE) and median
BASDAI	1	2
DAPSA	1	2
DAS28[CRP]	2	3
EQ-5D-3L dimension scores	0	1
EQ-5D-3L (VAS)	0	1
FACIT-Fatigue subscale score	1	2
HAQ-DI	3	4
hs-CRP	2	3
LDI	1	2
LEI	1	2
mNAPSI	0	1
PASDAS	1	2
PASI	1	2
PGA-Arthritis /PhGA-Arthritis	0	1
PGA-PsA / PhGA-PsA	0	1
PsAID-12	1	2
PsAQoL	1	2
PsARC	1	2
PtAAP	0	1
SF-36	2	3
SPARCC	1	2
TJC / SJC	1	2
vdHmTSS, Erosion score Joint space narrowing score	1	2
WPAI-SHP	1	2

BASDAI=Bath Ankylosing Spondylitis Disease Activity Index; DAPSA=disease activity index for psoriatic arthritis; DAS28[CRP]=disease activity score-28 based on C-reactive protein; EQ-5D-3L=Euro-Quality of Life 5-Dimensions 3 Level version; FACIT-Fatigue=Functional assessment of chronic illness therapy - Fatigue; HAQ-DI=Health assessment questionnaire – Disability index; hs-CRP=High sensitivity C-reactive protein; LDI=Leeds Dactylitis Index; LEI=Leeds Enthesitis Index; mNAPSI=Modified Nail Psoriasis Severity Index; PASDAS= Psoriatic Arthritis Disease Activity Score; PASI=Psoriasis Area and Severity Index; PGA-Arthritis=Patient's global assessment of Arthritis; PGA-PsA=Patient's Global assessment of Psoriatic Arthritis; PhGA-Arthritis=Physician's global assessment of Arthritis; PhGA-PsA=Physician's Global assessment of Psoriatic Arthritis; PHQ-9=Patient Health Questionnaire-9; PSAID-12=Psoriatic Arthritis impact of disease-12; PsAQoL=Psoriatic Arthritis Quality of Life; PsARC= Psoriatic Arthritis Response Criteria; PtAAP=Patient's assessment of Arthritis Pain; SF-36=Short-Form 36-item Health Survey; SJC= Swollen Joint Count; SPARCC= Spondyloarthritis Research Consortium of Canada ; TJC=Tender Joint Count; vdHmTSS=Van der Heijde modified Total Sharp Score; WPAI-SHP=work productivity and activity impairment questionnaire – specific health problem

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.

Unless stated otherwise, statistical tests of efficacy variables will be presented as 2-sided p-values rounded to 3 decimal places. P-values less than 0.001 will be presented as "<0.001" and p-values greater than 0.999 will be presented as ">0.999." Statistical comparisons will be performed by 2-sided statistical tests at the 0.05 level of significance. Least square means (LSM), OR and corresponding CI will be presented using 3 decimals.

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 subject from a responder to a non-responder (and vice versa) if these values occur on 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, temporary variables are created (ie, for absolute values, change and percentage change from Baseline) during programming which are rounded to 12 decimal places prior to comparison to a specific threshold in the derivation of a response parameter. This does not imply inherent rounding on the analysis variables for absolute value, change or percentage of change which are retained unrounded in the final analysis dataset. Thus, rounding is applied exclusively during the derivation of new response parameters or critical value variables, and the rounded values are created on a temporary basis only.

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 Table 4–5) 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 this SAP, the order of treatment groups presented in tables from left to right will be:

- Placebo
- Bimekizumab 160mg Q4W

- Adalimumab 40mg Q2W.

Selected tables may also include columns for all subjects (regardless of study treatment) and/or all subjects on BKZ (regardless of whether they shifted from placebo after the DBP or not).

The abbreviation for bimekizumab is BKZ, and ADA for adalimumab and will be used in table and listing headers. No abbreviation will be used for Placebo.

In the TFLs, subjects randomized to BKZ 160mg Q4W will be labeled as "BKZ 160mg Q4W" , subjects randomized to ADA 40mg Q2W will be labeled as "ADA 40mg Q2W" and subjects randomized to Placebo will be labeled as "Placebo / BKZ 160mg Q4W" (or "Placebo" when the TFLs report data from the DBP only).

For the ATP, all subjects on BKZ 160mg Q4W (that includes the subjects switching from Placebo at Week 16) will be labeled as "BKZ 160mg Q4W Total" in the TFLs.

Per protocol, visit windows of ±2 days from the first dose at all visits through Week 16 and ±3 days after Week 16 are permissible. 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.

For the medical imaging collection date (at Week 16 and 52), the allowed visit window will be extended to ±14 days around the scheduled visit date.

All by-visit tables will contain nominal (ie scheduled) visits only. Unscheduled visits will not be mapped to scheduled visits except for some assessments that may occur within the above visit 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 unless data are present for that scheduled visit. Only data collected/mapped at the scheduled visits will be included in the analysis. The only deviation may concern unscheduled vendor data collected (Section 10.5).

Assessments collected at unscheduled and repeated visits will only be listed, except in the case of determining treatment-emergent markedly abnormal criteria for laboratory and vital sign parameters where all post-Baseline (scheduled and unscheduled) values will be used.

For by-visit tables summarizing efficacy data, the SFU visit will not be included but SFU efficacy data will be listed. By-visit tables summarizing safety data will however include the SFU visit. For PK, immunological and immunogenicity analyses, SFU data will be part of the data reported in some tables (Section 9).

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 group, subject 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 non-randomized subjects, the non-randomized subjects will be shown first in the listing, ordered by subject number. All listings will include repeated and unscheduled measurements. Such measurements will appear in chronological order together with the scheduled visits (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 Relative day

The relative day will be included in some listings. 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 occurred on or after the date of first IMP administration (but before the last IMP administration), then the following calculation is used: Relative Day=Date of event - Date of first IMP administration + 1.
- If the event occurred before the date of first IMP administration, then the following calculation is used: Relative Day=Date of event - Date of first IMP administration. Relative days that occur before the date of first IMP administration should be preceded by a "-".
- If the event occurred after the date of last IMP administration, then the following calculation is used: Relative Day=Date of event - Date of last IMP administration. Relative days that occur after the date of last IMP administration should be preceded by a "+".

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

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

For subjects in the placebo group that switch to BKZ at Week 16, for events occurring after Week 16, an additional relative day will be calculated using the same rules as above but based on the date of first injection of BKZ. 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 before the date of first IMP administration. A complete date must be established to correctly identify the AEs. Section 4.2.3 describes imputation rules in case of missing data for AEs.

3.2.2 Mapping of assessments performed at Early Discontinuation Visit

Study assessments at an early termination visit where visit date matches the visit date of a scheduled visit will be summarized at the scheduled visit with the same visit date. Premature study termination visit assessments that do not have a scheduled visit with a matching date will be assigned a visit based on the protocol-defined visit windows (±2 days up to and including Week 16, and ±3 days from Week 20 up to and including Week 52). For early discontinuation visits that fall between protocol defined visit windows and not within them, the later of the 2 visits is assigned.

If there is an existing scheduled site visit in the window, then the assessments at the ET visit will be mapped to the next scheduled site visit.

The ADAb assessments, PK and the vdHmTSS are exceptions to this rule:

- ADAb levels, PK from an early termination visit will be mapped to the next visit where they are measured.

- VdHmTSS from an early termination visit will be mapped to the Week 16 assessment (if the early termination is before Week 16) or to the Week 52 assessment (if the early termination is between Week 16 and Week 52).

3.3 Definition of Baseline values

A Baseline value for a subject is defined as the latest measurement for that subject up to and including the day of first administration of IMP (Visit 2 by default), unless otherwise stated.

By default, for randomized subjects for whom no start date of treatment is available, the Baseline value will be considered as the last available value prior to the randomization date.

If a Baseline assessment is taken on the same day as first administration of IMP, 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 administration of IMP. 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 IMP 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 IMP. One exception to this rule is plasma concentration. If Baseline plasma concentration is measured at a time after the first administration of IMP, then it should not be eligible to be considered as a Baseline plasma concentration. Such cases should be discussed with the quantitative clinical pharmacologist.

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 IMP, then the Baseline value is treated as missing.

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

When the time of first dose is derived, it should be based on the first injection of IMP, regardless of whether it is an active treatment.

Chest x-ray performed within 3 months before the Screening visit is considered as Baseline measurement.

For subjects who switch treatment from placebo to BKZ at Week 16 (Visit 10), an additional Baseline value (called Week 16 Baseline) will be defined for efficacy variables. The Week 16 value will be considered as the Baseline value regardless of whether the Placebo subjects switched to BKZ on a later visit.

The variables listed below will be analyzed relative to the Week 16 Baseline value for the ATP, for timepoints after Week 16:

- PASI90 response
- ACR50 response.

3.4 Study treatment discontinuation and intercurrent event

The concept of intercurrent event (IE) is one of the estimand attributes.

Unless otherwise noted, treatment discontinuation due to any reason will be considered an IE. Since there is no treatment discontinuation date collected, the treatment discontinuation date is defined as:

- Treatment end date + 31 days.

3.5 Protocol deviations

An important protocol deviation (IPD) is defined as a protocol deviation (PD) likely to have a meaningful impact on the study conduct, the primary efficacy outcome, the key safety or the PK outcomes for an individual subject. The IPDs will be identified and classified by the deviation types defined in the appropriate protocol-specific document. All PDs will be reviewed as part of the ongoing data cleaning and data evaluation process and IPD will be identified and documented prior to unblinding to confirm exclusion from Per Protocol Set (PPS) and the Pharmacokinetics Per Protocol Set (PK-PPS).

An IPD will not always necessitate the removal of a subject from the PPS. Depending on the type of PD, the subject may remain in the PPS. The IPD document developed by the clinical study team should clearly state which PD will result in a removal from the PPS.

In addition to IPDs resulting in the exclusion from the PPS, subjects who reduce the dose or dosing frequency of certain medications due to intolerance, AE, side-effects or receive new prohibited medication for AE will be removed from the PPS. While this is not an IPD, as it is allowed per protocol for safety reasons, these subjects will be removed from the PPS as this non-PD could have an effect on the primary efficacy outcome in the same way as flagged IPDs resulting in exclusion from the 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 DBP, ie only subjects with IPD (or the non-PD above) observed prior to or at Week 16 (Visit 10) will be excluded from the PPS. Subjects with IPD after Week 16 will not be excluded from the PPS.

A specific category of PD as a consequence of the Coronavirus Disease 2019 (COVID-19) pandemic (called COVID-19 related PD) will be assessed based especially on the information collected on a dedicated electronic case report form (eCRF) page (and other sources).

3.6 Analysis sets

3.6.1 Enrolled Set

The Enrolled Set consists of all subjects who have given informed consent.

Subject dispositions will be summarized on the Enrolled Set.

3.6.2 Randomized Set

The Randomized Set (RS) consists of enrolled subjects that have been randomized.

Demographic tables, primary, secondary (except VdHmTSS) and other efficacy variables will be presented on the RS.

3.6.3 Safety Set

The Safety Set (SS) consists of all subjects who received at least 1 dose of the IMP.

Demographic tables, study treatment compliance, exposure and safety variables will be presented on the SS.

3.6.4 Full Analysis Set

The Full Analysis Set (FAS) consists of all randomized subjects who received at least 1 dose of IMP and have valid measurements of all the components of the primary efficacy variable at Baseline.

Supportive analysis of the primary efficacy variable will be performed in the FAS.

3.6.5 Active Medication Set

The Active Medication Set (AMS) will consist of all subjects who have received at least 1 dose of active IMP (BKZ or ADA). The AMS will cover the analysis of data collected during the Active Medication Periods (AMP), ie:

- The ATP for subjects randomized to placebo,
- The DBP and the ATP for subjects randomized to BKZ or ADA.

The AMS will be used for summaries of safety during the AMP.

The anti-BKZ antibody will also be analyzed on the AMS for subjects receiving BKZ.

In this analysis set, the order of the treatment groups presented in tables from left to right will be:

- BKZ Total (all BKZ treated subjects)
- ADA 40mg Q2W.

For the analysis of the compliance, study medication and time at risk, the order of the treatment groups to be presented in tables from left to right will be:

- Placebo / BKZ 160mg Q4W
- BKZ 160mg Q4W
- BKZ 160mg Q4W Total (all BKZ treated subjects)
- ADA 40mg Q2W.

3.6.6 Active Treatment-Blind Set

The Active Treatment-Blind Set (ATS) consists of all subjects who received at least 1 dose of active treatment (BKZ or ADA) during the ATP (Week 16 and after).

Disposition, demographics, and Baseline characteristics will be reported on the ATS. The ATS will also be used to report data from the ATP such as study treatment compliance and exposure, adverse events, treatment-emergent markedly abnormal (TEMA) data for vital signs and laboratory data and selected efficacy analyses.

In this analysis set, the order of treatment groups presented in tables from left to right will be:

- Placebo / BKZ 160mg Q4W
- BKZ 160mg Q4W

- ADA 40mg Q2W.

For safety data, the first 2 groups will also be summarized as “BKZ 160mg Q4W Total”.

3.6.7 Per Protocol Set

The Per-Protocol Set (PPS) will consist of all subjects in the RS who had no IPD or non-PD related to prohibited medications affecting the primary efficacy variable (only IPD/non-PD related to prohibited medications observed prior to Week 16 are considered for exclusion from the PPS). The deviations will be predefined and subjects with deviations will be evaluated during ongoing data cleaning and data evaluation meetings prior to unblinding of the data (Section 3.5). Exclusion from the FAS will be considered as an IPD that also results in exclusion from the PPS.

Supportive analysis of the primary efficacy variable will be performed on the PPS.

3.6.8 Pharmacokinetic Per-Protocol Set

The Pharmacokinetic Per-Protocol Set (PK-PPS) consists of all randomized subjects who received at least one dose of the BKZ and provided at least one quantifiable plasma concentration post-dose (ie: after first BKZ dose) without IPD that would affect the concentration.

PK variables will be analyzed for all subjects in the PK-PPS.

3.6.9 Radiographic Set

The Radiographic Set (RAS) will consist of all subjects in the RS who received at least 1 dose of IMP and have a valid radiographic image of the hands and feet (with an assessment performed by at least the 2 reviewers) at Screening.

The analysis of vdHmTSS will be performed on the RAS.

3.6.10 COVID-19-free Set

The COVID-19-free Set will consist of all subjects in the RS who had no COVID-19 impact up to the primary efficacy endpoint. This will be defined as subjects (up to Week 16):

- not having a COVID-19 related IPD
- not having an impact based on the COVID-19 eCRF
- not having an AE related to COVID-19 (Section 10.2.6)
- not discontinuing due to COVID-19.

The disposition data, the primary efficacy endpoint and the secondary efficacy endpoints included in the testing hierarchy will be analyzed on the COVID-19-free Set.

3.7 Treatment assignment and treatment groups

It is expected that subjects receive treatment as randomized and hence safety analyses will be based on the SS, as randomized. However, after unblinding, study treatment received might be in some cases different from the randomized treatment. If subjects randomized to placebo received BKZ at any time during the DBP, then for safety analyses these subjects will be reallocated to the BKZ treatment group. If subjects randomized to placebo received ADA at any time during the DBP (but no BKZ), then for safety analyses these subjects will be reallocated to the ADA

treatment group. Subjects randomized to BKZ will only be reallocated to the placebo treatment group if they never received BKZ. Subjects randomized to ADA will only be reallocated to the placebo treatment group if they never received ADA.

Efficacy analyses will be performed according to randomization and not actual treatment received. Data collected after study treatment discontinuation will be listed only and not reported in table except in the context of the treatment policy analysis performed on the primary efficacy endpoint (Section 8.1.4.5).

3.8 Center pooling strategy

The 4 geographic regions considered for the study are the ones used for randomization stratification:

- North America (Canada, USA)
- Western Europe (Belgium, France, Germany, Italy, Spain, United Kingdom)
- Eastern Europe (Czech Republic, Hungary, Poland, Russia)
- Asia (Australia, Japan) identified as "Asia" in all analyses.

The above regions will be the ones used when considering the region as factor in the efficacy 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 March-2021. Medical procedures will not be coded.

To maintain consistency across studies in the BKZ 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:

- An additional supportive analysis for primary endpoint based on the analysis of the individual components of ACR will be performed using the Reference-Based imputation method.
- The main analyses of the secondary continuous variables included in the testing hierarchy will be performed using the Reference-Based imputation method.
- Proportion of subjects with a decrease of HAQ-DI from Baseline of at least 0.35 (HAQ-DI responders) in those subjects with HAQ-DI ≥0.35 instead of >0.35.
- An additional subgroup based on the combination of concomitant methotrexate (MTX) and Baseline conventional disease modifying antirheumatic drug (cDMARD).
- Additional subgroup analyses will be performed on HAQ-DI responders at Week 16.
- Proportion of PsAID-12 responders (decrease from Baseline in PsAID-12 total score ≥3) in subjects with PsAID-12 total score ≥3 at Baseline instead of >3 at Baseline

- The subgroup analysis on BASDAI will be performed on the categories: <4 vs. ≥4 rather than ≤4 vs. >4.
- The analysis of the proportion of responding subjects at Week 52 among those who responded at Week 16 will be extended to all post Week 16 visits (and not only at Week 52).
- The analysis of the proportion of FACIT-Fatigue subscale responders (subjects with a minimum clinically important difference for FACIT-Fatigue subscale score defined as an increase of ≥4) will be performed in subjects with FACIT-Fatigue subscale score ≤48 at Baseline.
- The time to ACR20/50/70 will be exclusively analyzed using observed cases data.
- The AMS is an analysis set that has been added for the analysis of safety variables.
- The ATS is an analysis set that has been added for the analysis of data collected during the ATP.

3.11 Changes related to COVID-19

The impact of the COVID-19 pandemic on study procedures/conduct as well as the efficacy and safety endpoints will be investigated, and additional analysis outputs will be provided as appropriate. These additional analyses were not planned as part of the original 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 (pre/during/post) as defined below:

- Pre-COVID-19 pandemic period: period prior to COVID-19 pandemic start date defined as 11-Mar-2020
- COVID-19 pandemic period: period from 11-Mar-2020 though the COVID-19 pandemic end date which is currently not defined at the time of approval of the SAP
- Post-COVID-19 pandemic period: period after the declaration of the end of the pandemic.

The additional analyses are described in the following sections:

- Subject disposition (Section 5.1)
- Demographics and other baseline characteristics (Section 6)
- Details of impacted visits and effects on collection and reporting of efficacy data (Section 5.2)
- Protocol deviations (Section 5.3)
- Efficacy analyses related to the hierarchy endpoints (Section 8.1.4.8 and 8.2.3)
- Adverse events (Section 10.2.6).

4 STATISTICAL/ANALYTICAL ISSUES

4.1 Adjustments for covariates

The primary efficacy analysis will investigate the treatment effect, adjusting on the 2 randomization stratification variables:

- Region:
  - North America, Western Europe, Eastern Europe, Asia

  or
  - North America, Eastern Europe, Western Europe + Asia if the percentage of randomized subjects is < 10% in either of the Asia or Western Europe regions
- Bone erosion (0, ≥1) at Baseline.

When adjusting on these 2 variables, a statistical model might not converge (ie: the likelihood maximization algorithm fails to converge). In that case, the statistical model will be run after dropping successively region and bone erosion at Baseline.

If a subject is stratified in the incorrect stratum (ie: the stratum recorded in the Interactive voice or web Response System differs from the actual stratum the subject belongs to), the actual stratum will be used for the analysis.

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

4.2 Handling of dropouts or missing data

4.2.1 Strategy for handling missing data for efficacy analyses

Different approaches will be used to handle missing data including how the IE will be considered (Section 3.4).

In this section, 3 terms will be used to define binary and continuous endpoints:

- Non-composite binary endpoint: Binary endpoint derived based on one continuous measurement (eg, PASI90)
- Composite binary endpoint: Binary endpoint derived based on several continuous measurements (eg, ACR50)
- Composite continuous endpoint: Continuous endpoint derived based on several continuous measurements (eg, PASDAS).

Below is a table summarizing for all continuous endpoints (excluding composite continuous endpoints) observed ranges and whether they are represented by an integer value. These endpoints will be analyzed and/or used in the derivation of composite or non-composite endpoints:

Table 4–1: List of non-composite continuous efficacy endpoints

Efficacy endpoint	Minimum	Maximum	Integer value
BASDAI	0	10	
EQ-5D-3L dimension scores	1	3	Yes
EQ-5D-3L (VAS)	0	100	Yes
FACIT-Fatigue subscale score	0	52	
HAQ-DI	0	3	
hs-CRP	LLOQ/2 where LLOQ=0.10 mg/L	No maximum	
IGA	0	4	Yes
LDI	0	No maximum	
LEI	0	6	
mNAPSI	0	13	Yes
PASI	0	72	
PGA-Arthritis /PhGA-Arthritis	0	100	Yes
PGA-PsA / PhGA-PsA	0	100	Yes
PsAID-12	0	10	
PsAQoL	0	20	
PtAAP	0	100	Yes
SF-36[a] Physical Functioning (PF) Role Physical (RP) Bodily Pain (BP) General Health (GH) Vitaly (VT) Social Functioning (SF) Role Emotional (RE) Mental Health (MH) Physical component summary (PCS) Mental component summary (MCS)	 19.26 21.23 21.68 18.95 22.89 17.23 14.39 11.63 5.02 -3.33	 57.54 57.16 62.00 66.50 70.42 57.34 56.17 63.95 79.78 80.09	
SPARCC	0	16	
Tender Dactylitis Count	0	20	Yes
TJC / SJC	0	68/66	
vdHmTSS Erosion score Joint space narrowing score	0	528 320 208	

Efficacy endpoint	Minimum	Maximum	Integer value
WPAI-SHP	0	100	

BASDAI=Bath Ankylosing Spondylitis Disease Activity Index; EQ-5D-3L=Euro-Quality of Life 5-Dimensions 3 Level version; FACIT=Functional Assessment of Chronic Illness Therapy; HAQ-DI=Health Assessment Questionnaire – Disability index; hs-CRP=High sensitivity C-reactive protein; IGA=Investigator Global Assessment; LDI=Leeds Dactylitis Index; LEI=Leeds Enthesitis Index; LLOQ=lower limit of quantification; mNAPSI=Modified Nail Psoriasis Severity Index; PASI=Psoriasis Area and Severity Index; PGA-Arthritis=Patient's Global Assessment of Arthritis; PGA-PsA=Patient's Global Assessment of Psoriatic Arthritis; PhGA-Arthritis=Physician's Global Assessment of Arthritis; PhGA-PsA=Physician's Global Assessment of Psoriatic Arthritis; PHQ-9=Patient Health Questionnaire-9; PSAID-12=Psoriatic Arthritis Impact of Disease-12; PsAQoL=Psoriatic Arthritis Quality of Life; PtAAP=Patient's Assessment of Arthritis Pain; SF-36=Short-Form 36-Item Health Survey; SJC= swollen joint count; SPARCC= Spondyloarthritis Research Consortium of Canada ; TJC=tender joint count; vdHmTSS=van der Heijde Modified Total Sharp Score; WPAI-SHP=Work Productivity and Activity Impairment Questionnaire – Specific Health Problem

[a] minimum and maximum of the norm-based scores – Normal range for SF-36 scores is 0-100

4.2.1.1 Primary analysis of binary endpoints that are part of the testing hierarchy (primary and secondary endpoints)

All binary endpoints (composite & non-composite) are based on continuous component variables.

The primary analysis of binary efficacy endpoints that are part of the testing hierarchy is considered under an estimand framework in which missing data (due to study treatment discontinuation) is considered indicative of failed treatment and imputed to non-response. Further missing data for subjects on treatment, or data observed while not on treatment is also imputed to non-response.

This composite estimand approach to handling missing data which is the primary analysis method for binary efficacy endpoints is similar to the non-responder imputation (NRI) method.

4.2.1.2 Primary analysis of continuous endpoints that are part of the testing hierarchy (secondary endpoints)

For continuous efficacy endpoints that are part of the testing hierarchy, the primary analysis method is based on the hypothetical estimand approach as follows:

- If subjects have missing data regardless of whether the missing data is preceded by an IE, then missing data will be imputed based on the MI-Markov-Chain Monte Carlo (MI-MCMC) /Reference-Based imputation method (Section 4.2.2.3).
- If subjects have non missing data after IE, then such data will be set to missing prior to running MI.

4.2.1.3 Supportive analyses of primary efficacy endpoint

For the primary endpoint (which is a composite binary endpoint) several supportive analyses assuming different missing data mechanisms will be conducted:

- The modified composite estimand approach: the IE is changed from all treatment discontinuation to treatment discontinuation due to AE or lack of efficacy. The same as the primary method will be used with the difference that only data after discontinuation due to AE or lack of efficacy is set to non-response. All other imputed data will be used.

- For the analysis of individual components of ACR, under the hypothetical estimand approach using the reference-based imputation methods (Section 4.2.2.3).
- The tipping point approach implemented within the MI framework, but only if the primary endpoint analysis result is statistically significant at alpha=0.05. (Section 4.2.2.4).
- The treatment policy strategy (Section 4.2.2.5).
- The observed case (OC) analysis that will only include observed data while on treatment. For subjects who had treatment discontinuation, only observed data up to treatment discontinuation date will be analyzed. For visits with missing data without treatment discontinuation, the data will remain missing. For observed data after treatment discontinuation, the corresponding visit data and subsequent visit data will be treated as missing.

4.2.1.4 Supportive analyses of endpoints that are part of the testing hierarchy (secondary endpoints)

The following supportive analysis methods will be performed for the secondary efficacy variables that are part of the testing hierarchy:

For binary endpoints:

- The modified composite estimand approach: the IE is changed from all treatment discontinuation to discontinuation of treatment due to AE or lack of efficacy. For composite or non-composite binary endpoints, the standard MI approach will be implemented (similarly as in Section 4.2.1.3) on the raw continuous score(s) (eg: PASI for the PASI90 endpoint) before deriving the binary endpoint based on the imputed score(s).
- The OC analysis.

For continuous endpoints:

- The hypothetical estimand approach:
  - If subjects have missing data regardless of whether the missing data is preceded by an IE, then missing data will be imputed based on the MI-MCMC/Monotone regression method (Section 4.2.2.2).
  - If subjects have non missing data after IE, then such data will be set to missing prior to running MI.
- The OC analysis.

4.2.1.5 Analyses of secondary endpoints that are not part of the testing hierarchy and other efficacy endpoints

The other efficacy endpoints and the secondary efficacy endpoints that are not part of the testing hierarchy will be analyzed as the secondary efficacy endpoints (with the exclusion of the MI-MCMC/referenced based imputation method for continuous variables) but with no designated priority.

For a combined binary endpoint based on the combination of a non-composite and a composite binary endpoint (as with the ACR50 and PASI90 combined endpoint), each binary endpoint will be derived independently, before deriving the combined endpoint.

4.2.1.6 Summary table of missing data handling approaches for efficacy analyses

The below table summarizes which missing data handling approaches will be used for each type of efficacy endpoint.

Table 4–2: Missing data handling approaches for efficacy endpoints

Efficacy Endpoint Priority	Variable Type	Missing data handling approach						
		Composite estimand (NRI)	Hypothetical estimand (RB-MI)	Modified Composite estimand (MI)	Hypothetical estimand (MI)	OC	Tipping Point Analysis (MI)	Treatment Policy (MI)
Primary endpoint	Composite Binary	P		S[a]		S	S[a]	S[a]
Secondary endpoints included in the testing hierarchy	Binary	P		S[a]		S		
	Continuous		P		S	S		
Other Secondary endpoints (endpoints not included in the testing hierarchy)	Binary	X		X[a]		X		
	Continuous				X	X		
Other endpoints	Binary	X		X[a,b]		X		
	Continuous		X[d]		X[b]	X		
	Categorical	X[c]		X[c]		X		

P=Primary analysis method, RB=Referenced Based, MI=multiple imputation, NRI=non-responder imputation, OC=observed cases, S=Supportive analysis 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 are imputed as non-response, and other missing data are also imputed as non-response.
Note : Referenced Based imputation refers to the approach in which it is assumed that the statistical behavior of the bimekizumab and placebo-treated subjects after discontinuing study medication becomes that of the placebo-treated subjects that remain in the study.
Note: Modified Composite Estimand (MI) refers to the approach in which data preceded by the intercurrent event of study treatment discontinuation due to AE or lack of efficacy are imputed as non-response (or worst category for categorical variable), and other missing data are imputed via a multiple imputation model.
Note: Hypothetical Estimand (MI) refers to the approach where outcomes for subjects without an intercurrent event of study treatment discontinuation are as observed, and outcomes for subjects with the intercurrent event 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 complete imputed dataset.
[b] MI will be used to provide point estimates at each timepoint only
[c] Missing data to be imputed to the worst category
[d] for ACR components analysis at Week 16

4.2.2 Methods for handling missing data for efficacy analyses

The sections below describe the method to be used for missing efficacy data handling.

For a specific endpoint, if analyzed subjects in a one or several treatment group(s) have no missing data, then analysis of observed cases will be performed in the concerned treatment group(s).

4.2.2.1 Non-responder imputation

For binary endpoints, the non-responder imputation (NRI) analysis, also described above as the composite estimand, will consider the following subjects as "non-responders" for the timepoint of interest:

- Subjects with missing data at the time point of interest. In the case of ACR50, non-response is imputed if ACR50 cannot be derived (ie: derived as "missing") based on available ACR components (Section 8.1.1)
- Subjects who remained in the study at the timepoint of interest but discontinuing study treatment before the timepoint of interest (data after IE)
- Subjects with missing Baseline value (for composite or non-composite binary variables based on change(s) from Baseline of continuous endpoint(s)).

4.2.2.2 MI – MCMC/Monotone Regression

In instances where MI is used, the missing value is replaced by a set of plausible values, where each value is a Bayesian draw from the conditional distribution of the missing data given the observed data. Intermittent missing data will be imputed using the MCMC method, followed by regression for monotone missing data (called "MI-MCMC/Monotone regression" method in this SAP). The MI procedures planned for efficacy analyses are based on an assumption of data missing at random (MAR).

This section describes the algorithms to be implemented for the MI – MCMC/Monotone Regression procedures for non-composite binary endpoints, composite binary endpoints, and continuous endpoints. These descriptions focus on the MI procedure itself and do not specifically account for dealing with IE (Section 4.2.1).

Non-composite binary endpoints

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

- Step 1 - Imputation of missing data using MI
  Create datasets, one for each treatment group, of subjects with observed values and missing values (needing estimation by MI). For the imputation step, missing values will be separated into 2 categories: intermittent missing values (ie, missing values for a given subject that has available data before and after the missing timepoint, including missing value at Baseline) and monotone missing values (ie, where all subject data is missing after a given time point). Datasets should be designed in a horizontal structure meaning that each subject should be presented in a single observation, with a set of values, one for each scheduled visit where the endpoint is scheduled to be collected according to the protocol (this exclude the unscheduled visits). For the Week 24 analysis, this will include visits up to Visit 14 (Week 24) and for the Week 52 analysis, visits up to Visit 28 (Week 52). Datasets should also be sorted by subject number before proceeding with the MI process.

For the intermittent missing values, the missing value in each dataset will be filled in 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 MI procedures described in this SAP will use this same seed as well. The procedure will be performed for each treatment group separately. The resulting 100 imputed datasets will have a monotone missing pattern and will be imputed using a method for monotone missingness.

For monotone missing data, one 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 dataset 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 for each treatment group separately by imputation. The SAS® PROC MI procedure will be used for the imputation.

In both cases, the imputation model will include the randomization strata as stratification variables (region and bone erosion at Baseline), the value at Baseline and at each post-Baseline visit (up to the week of interest). The imputation model based on the MCMC method will only allow joint multivariate normal variables. For the MCMC method (when imputing intermittent missing values), randomization strata will be re-coded as indicator variables and will always be specified in the following order: North America region (1 for North America, 0 otherwise), Western Europe region (1 for Western Europe,0 otherwise), Eastern Europe region (1 for Eastern Europe, 0 otherwise), For each imputation model, Bone erosion at Baseline (0 if bone erosion at Baseline =0 ; 1 if bone erosion at Baseline≥1).

Bone erosions will be read for all visits at both at the Week 24 analysis and the Week 52 / final analyses to ensure that unnecessary bias is not introduced. Given there may be differences at Baseline for a subject between the two sets of reads, the following rules will be applied for the MI process:

- For the Week 24 analysis, the first set of reads will be used.
- For the Week 52 and final analyses, Week 16 specific outputs will continue to use the first set of reads. Outputs involving over time data up to Week 52 will use the second set of reads.

For each imputation model, the randomization strata (region and bone erosion at Baseline) may be dropped to facilitate convergence if required as described below: If the MI fails to converge when adjusting on the stratification variables specified above, or if the percentage of randomized subjects is < 10% in either of the Asia or Western Europe regions, then the Western Europe and Asia region will be combined and the binary variables for region that will be left in the MI will then be for North America and Eastern Europe regions.

If after doing that, the MI still fails to converge then all region factors will be removed from the MI. if the MI still does not converge after dropping the regions, the remaining stratification variable (bone erosion at Baseline) will also be removed.

If a variable is dropped in order to allow convergence for one model in the study, that variable does not have to be dropped from other models in the study of the model converges without dropping the variable. That is, model convergence should be evaluated for each efficacy table independently.

The post-Baseline values will need to be specified in chronological order after the Baseline value in the imputation model so that the SAS® PROC MI imputes variables from left to right (ie, Week 2 value will be first imputed using regression based on Baseline value, and then Week 4 value will be imputed using regression based on Baseline and Week 2 value, etc.). The resulting datasets for each treatment arm will be combined into one complete dataset containing 100 times the number of subjects analyzed.

When imputing missing hs-CRP values, values below LLOQ (0.10 mg/L) will be replaced by the midpoint value between 0 and LLOQ prior to running MI.

- Step 2: If an imputed value falls outside of the range for the given variable (as listed in Table 4–1) the value will be updated to be within the predefined range. For example, the imputed value for PASI will be updated to 0 or 72 in the case of an imputed value < 0 or > 72, respectively. For endpoints that can take only integer values (eg: IGA as listed in Table 4–1), the imputed values will be rounded to the closest integer after each SAS® PROC MI.
  Note: For hs-CRP, the lower limit used the midpoint value between 0 and LLOQ. For some parameters, there is not a fixed upper limit to the imputed value (ie: LDI).
- Step 3: On the dataset 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 datasets will be used to categorize the subject as a responder or not. If subjects have an IE, then the endpoint at all subsequent visits will be set to "non-response". This will be performed:
  - When date of visit is available: by comparing the IE date vs. the date of visit
  - When a visit date is missing in case of fully imputed data at a visit: by comparing the next visit number after the visit where the last study medication occurs vs. the visit number where data are imputed.
- Step 4: At each timepoint the (unadjusted) proportion of responders will be calculated by treatment group from the imputed datasets using SAS® PROC FREQ.
- Step 5 (For primary and secondary efficacy endpoints only): For each value of the imputation number from 1 to 100, the adjusted proportion of responders will be analyzed using a logistic regression model with a fixed effect for treatment. The suitability of including region, and bone erosion at Baseline as fixed effects will be assessed using goodness-of-fit tests (Deviance and Pearson's and Hosmer-Lemeshow) and added if appropriate, if it allows model convergence. Covariates kept in the modelling should also be the same as the ones previously used for the MI. Comparisons will be made using 2-sided Wald test at a significance level of 5%. The results obtained from the 100 logistic regression analyses (ie, the adjusted proportion of responders for each treatment group, the OR and the difference of proportions for the BKZ-placebo 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.

Composite binary endpoints

For composite binary endpoints, the MI method will be applied as follow:

- Step 1: The MCMC/monotone regression approach described above in Step 1 for non-composite binary endpoints will be performed separately for each individual component variable.
- Step 2: The same rounding rules described in step 2 above for non-composite binary endpoints will be performed for each component.
- Step 3: Based on the multiply imputed datasets obtained for each component, the binary response will be derived as follows:
  - The dataset obtained for each component will be merged by imputation number and subject number.
  - On the dataset obtained, the binary endpoint will be derived for each subject/visit based on the component values.
  - If subjects have an IE, then the endpoint at all subsequent visits (from the day after the IE date, whether the data were observed or not) will be set to "non-response".
- Step 4 and 5: Same as step 4 and 5 above for non-composite binary endpoints.

For the primary endpoint, the 'burn-in' option in SAS® PROC MI may be set higher than the default value. The goodness-of-fit tests to assess convergence of the MCMC will be provided in a SAS® output.

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 (SEs) are obtained from the logistic regression in Step 5 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 need to be considered relative to the calculation of the OR and corresponding CIs. As the estimates of the odds ratios from the logistic regression models obtained in Step 5 follow a lognormal distribution, a log transformation is needed to normalize these 100 odds ratio estimates. This is because the procedures for combining results from multiple imputed datasets assume that the statistics estimated from each imputed dataset are normally distributed. Therefore, the log of the OR estimates from the logistic regression model are used when combining into a single inference (Step 5). Additionally, the SE for the OR 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 BKZ relative to placebo and the corresponding upper and LCLs will be provided. The OR will be then estimated by exponentiating the estimate of the log odds ratio. The OR and the confidence limits of the OR 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 SE 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). These calculations will be done such that OR and corresponding CIs are calculated for the OR of BKZ vs. placebo.

In addition to calculating the OR, associated CIs, and p-values for the comparisons of BKZ vs. placebo, the estimated proportion of responders (ie, estimated responder rate) and the difference in the proportion of responders between BKZ and placebo will be estimated, along with 2-sided 95% CIs. The creation of the estimates of this difference will be completed using the process detailed below:

- Use the logistic regression model to calculate:

  LSM estimates of the log odds of BKZ (G_B) and placebo (G_P), as well as their corresponding standard errors (S_B and S_P, respectively).

  Standard error of the LSM estimate of the log odds ratio (S_R).

- Compute estimates for predicted proportions using the following transformations:

  $P_B = \exp(G_B)/(1 + \exp(G_B))$

  $P_P = \exp(G_P)/(1 + \exp(G_P))$

  The difference in proportions is then given by:

  $D = P_B - P_P$

- Estimate the standard error of D by:

  $S_D = \text{sqrt}\ [P_B^2(1-P_B)^2S_B^2 + P_P^2(1-P_P)^2S_P^2 + P_B(1-P_B)P_P(1-P_P)S_R^2 - P_B(1-P_B)P_P(1-P_P)(S_B^2 + S_P^2])$

The MI/MCMC monotone regression method, as previously outlined, will be used to account for missing values. The calculation steps described above will be based on the results provided from the logistic regression model of the multiple imputed datasets. An estimate for the difference in proportions, D, and corresponding standard error, SD, will be computed for each of these datasets. The results from these analyses will be combined into a single estimate of the difference in predicted proportion of response and a 2-sided 95% CI interval using SAS® PROC MIANALYZE.

Continuous efficacy endpoints

For continuous efficacy endpoints, the MI method will be applied as follows based on the values at the timepoint(s) of interest:

- Step 1: The same imputation procedure as the described in step 1 for the non-composite binary endpoints will be used.
- Step 2: The same rounding rules described in step 2 above for non-composite binary endpoints will be performed.
- Step 3: On the dataset obtained with the imputation number from 1 to 100; change from Baseline will be derived. Simple means and SEs will be calculated using Rubin's rules (via SAS® PROC MIANALYZE) for each timepoints of interest. For calculation of other descriptive statistics such as the median, min and max, Rubin's rules do not apply. MI estimates will be computed by simply averaging the estimates from the multiple repetitions of the imputation algorithm. For CRP data (that will be presented using the geometric mean and corresponding 95% CI, arithmetic mean, median, Q1, Q3, minimum and maximum). The change from Baseline will be expressed as the ratio to Baseline (value at the visit divided by the value at Baseline) in the by-visit summaries. The following approach will be applied:
  - Following the MI procedure, the ratio to Baseline will be calculated for any of the imputed values
  - The natural logarithm of the absolute values and of the ratios to Baseline will be calculated
  - The logged values will be summarized by treatment, visit and imputation
  - The datasets will be combined using SAS® PROC MIANALYZE in order to get the mean from the absolute values and ratios to Baseline across imputations
  - The estimates of the mean will be back transformed to obtain the geometric mean on the original scale
  - For the median, Q1, Q3, minimum and maximum the procedure outlined above for the other endpoints will be followed.
- Step 4 (for primary and secondary efficacy variables): For each value of the imputation number from 1 to 100, the change from Baseline will be analyzed using and ANCOVA model with treatment group, region and bone erosion as fixed effects and the Baseline value as covariate. The results obtained from the 100 ANCOVA analyses (ie: the LSM for the treatment difference and the 95% CI for the contrasts) will be combined using Rubin's rules (via SAS® PROC MIANALYZE).

For composite continuous endpoints, the MI will be applied on each component the same way as for composite binary endpoint. The datasets obtained for each component will be merged by imputation number and subject number. On the dataset obtained, the endpoint will be derived for each subject/visit based on the component values. The endpoint obtained will be analyzed as above.

Ordinal efficacy endpoints

For ordinal efficacy endpoints, the MI method will be applied as for the continuous efficacy endpoints with the following exceptions:

- After the imputation of the intermittent and monotone missing data, the rounding rules will be performed to align the data with the possible responses for the respective endpoint (eg: If the variable can take the value 1, 2 or 3, the imputed values will be rounded to the closest possible value).
- If the ordinal efficacy endpoint is used to derive a binary endpoint, the same analysis strategy as described above for the binary endpoints will be used.
- If the ordinal efficacy endpoint is not used to derive a binary endpoint, the unadjusted proportions of subject by categories will be presented.
- If subjects have an IE, then the ordinal endpoint at all subsequent visits (from the day after the IE date, whether the data were observed or not) will be set to the value used for the worst category.

4.2.2.3 MI – MCMC/Reference-based imputation

For the primary analysis of the secondary continuous efficacy variables included in the testing hierarchy and the individual components of the ACR response, the referenced-based multiple imputation method will be used. In this imputation method, the missing data will be imputed based on data from the placebo group only (Mallinckrodt, 2013). This analysis will be performed only for the subjects randomized to Placebo and BKZ.

Reference-based MI assumes that the statistical behavior of the BKZ and placebo randomized subjects after discontinuing study medication becomes that of the placebo-treated subjects. All time points after study treatment discontinuation for BKZ and placebo groups will be considered missing. MIs are used to replace missing outcomes for BKZ and placebo-treated subjects who discontinued using multiple draws from the posterior predictive distribution estimated from the placebo arm.

Note : SAS 9.4 may be used for this analysis where equivalent steps to those outlined below will be used to perform the imputation.

In the case of continuous endpoints, the procedure will be implemented as follows on the raw values:

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 (Baseline), …, Week 16 (Week 16 being the time point of interest) using data from the placebo-treated subjects only.

a. Initialization. Set t=1 (Baseline visit). Create a dataset combining all records from BKZ and placebo subjects with columns for covariates (bone erosion and region) and outcome at Baseline. Impute missing values at Baseline using bone erosion at Baseline, and region. Note that both placebo and BKZ data will be used to estimate the imputation model.

b. *Iteration*. Set t=t+1. Create a dataset combining records from BKZ subjects with missing data at visit t and all placebo subjects with columns for covariates (bone erosion and region) and

outcomes at visits 1 to t. In this dataset, outcomes for all BKZ randomized subjects are missing at visit t and observed or previously imputed values at visits 1 to t-1. Outcomes for placebo-treated subjects are observed or missing 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.

c. *Imputation.* Impute missing values for visit t using previous outcomes for visits 1 to t-1, bone erosion at Baseline, and region. Note that only placebo data will be used to estimate the imputation model since no outcome is available for BKZ-treated subjects at visit t. As a consequence, the input dataset should include all subjects from Placebo but only subjects from the BKZ arm that have values at timepoint t missing.

d. Repeat steps 2a-2d for all timepoints after Baseline up to Week16, 100 times with different seed values (seeds ranging from 201 to 300) to create 100 imputed complete datasets. Subjects whose missing values were imputed in the last SAS® PROC MI call will be included in the input dataset for the next SAS® 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 TJC will be updated to 0 or 68 in the case of an imputed value less than 0 or greater than 68, respectively.

2. The analysis will be done on the change from Baseline, and the model will be an ANCOVA model with treatment group, region, and bone erosion classification as fixed effects and the Baseline value as covariate.

   For generation of summary statistics, the 100 imputed datasets 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. MI 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).

3. The results obtained from the 100 ANCOVA analyses in Step 4 (i.e. LSM for the treatment difference and the 95% CI for the contrasts) will be combined with Rubin's rules.

4.2.2.4 Tipping Point analysis

Tipping point analyses will also be implemented as a supportive analysis for the primary efficacy endpoint. This analysis will be performed only for the subjects randomized to Placebo and BKZ.

The objective of the tipping point analyses is to identify assumptions about the missing data under which the conclusions from the main analysis change (ie: under which there is no longer

evidence of a treatment effect). These tipping point analyses will only be performed if the primary efficacy analysis results in a statistically significant treatment effect ($p<0.05$).

As a first step, the worst-case scenario will be evaluated. Specifically, all BKZ randomized subjects with a missing ACR50 at Week 16 (or non-missing ACR50 after study treatment discontinuation) will be imputed as non-responders, while all placebo-randomized subjects with a missing ACR50 at Week 16 will be imputed as responders. While there is little justification for such an approach, it makes the most putative assumption possible against a BKZ 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 OR of BKZ versus placebo is < 0.05, then no further tipping point analyses are needed.

If this analysis based on the worst-case scenario results in a p-value > 0.05, then additional tipping point analyses will be performed. In practice, it implies different delta adjustments will be made to the assumed responses on the missing data (where missing values include observations after the IE date and any other missing values) in each treatment group independently with various degrees of plausibility with the goal to find for each treatment group the "tipping point" that will significantly reverse the primary result which yielded a p-value < 0.05. These delta adjustments will be done on each component of ACR50 as follows:

- Step 1: The same MCMC method described in Section 4.2.2.2 will be implemented for non-monotone (intermittent) missing pattern values, using the same imputation model. This will be based on 100 imputations.
- Step 2: Based on the datasets obtained in Step 1, a monotone regression model will be applied (using the same imputation model as in Step 1) and as described in Section 4.2.2.2. This will be based on 1 imputation.
- Step 3: Imputed values will be shifted using a delta adjustment independently in each treatment group.

  Once defined, the same delta adjustment value will be applied on the imputed values for the Week 16 visit only. The selected scenario will assume that subjects randomized to BKZ and who have missing data have a lower probability of response compared to subjects randomized to placebo with missing data.

  – For ACR components for which high scores are associated with a less favorable outcome, it will mean that:
    - A positive shift is applied to the imputed value for subjects randomized to BKZ to increase the imputed value.
    - A negative shift is applied to the imputed value for subjects randomized to placebo to decrease the imputed value.

  A set of possible values will be first pre-defined for the delta parameter as below:

  – For all ACR component, the value of the initial delta parameter will be equivalent to a specific percentage of the possible range of each component (i.e. 5%). For hs-CRP, the range will be based on the log-transformed values at Week 16. The delta parameters for each endpoint are listed in the table below:

Table 4–3: Tipping point analysis: Delta parameter for each ACR component

ACR component[1]	Range	Delta
SJC	0-66	3.3
TJC	0-68	3.4
HAQ-DI	0-3	0.15
PtAAP, PhGA-PsA, PGA-PsA	0-100	5
hs-CRP[2]	Observed range of the $\log_e$ transformed values for all BKZ and Placebo subjects (at Week 16)	5% of the observed range

[1]The shifted imputed value should not exceed the range for the ACR component.
[2]The delta adjustment will be applied on the $\log_e$ transformed imputed and observed values which will then be exponentiated prior to deriving endpoint.

- Step 4: Standard rounding rules will be applied to the imputed values. If the SAS® PROC MI yields a value outside of the range for the given component, the value will be updated after the imputation has been performed to be within the predefined range.
- Step 5: Repeat Steps 1 to 4 for each ACR component. The composite binary endpoint (ACR50) will then be derived based on the multiply shifted imputed datasets obtained for each component.
- Step 6: Additionally, as the primary endpoint is derived using NRI, subjects randomized to BKZ with missing data (where missing values include observations after the IE date and any other missing values) should be set to non-response, after applying the delta adjustment outlined in Step 3 above. This ensures subjects randomized to BKZ do not have a higher probability of response in the tipping point analyses compared to the primary analysis (i.e. a subjects randomized to BKZ who is non-responder in the primary analysis cannot become a responder in the tipping point analyses).
- Step 7: In the data obtained, for each value of the imputation number, ACR50 will be analyzed using a logistic regression model with factors of treatment group, regions, and bone erosion at Baseline as fixed effects.
- Step 8: 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.
- Step 9: Step 3 to 8 will be repeated so that, at each iteration, missing values are adjusted with a larger shift than at the previous iteration. Depending on the results obtained, delta parameters with more granularity (eg, 2 times, 3 times the initial delta, etc.) will be investigated. The process will go on until the p-value for the OR between BKZ and placebo is no longer statistically significant (ie ≥ 0.05). The OR, 95% CI, and p-values obtained for each value of delta will be combined in one single table.

4.2.2.5 Treatment Policy Strategy

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

The treatment policy strategy will include all available data observed at the week of interest (Week 16) regardless of the occurrence of IE. This means the analysis includes on- and off-treatment values collected after subjects 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 subject's randomized treatment. Subjects 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-MCMC / monotone regression (Section 4.2.2.2) based on non missing data collected up to Week 16.

4.2.3 Handling of missing data for AEs

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.

Although the algorithms for treatment-emergence depend on the start 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 subject data listings (ie, no imputed values will be displayed in data listings).

Imputation of partial AE start dates

- If only the month and year are specified:
  - if the month and year of first dose of IMP is not the same as the month and year of the start date, then use the 1st of the month
  - if the month and year of first dose of IMP is the same as the month and year of the start date, then use the date of first dose of IMP
  - if the subject did switch treatment during that month and year from placebo to BKZ at Week 16, then use the date of treatment switch.
- If only the year is specified:
  - if the year of first dose is not the same as the year of the start date, and if the subjects did not switch from placebo to BKZ at Week 16 during that year then use January 1 of the year of the start date
  - if the year of first dose is not the same as the year of the start date, and if the subjects did switch from placebo to BKZ at Week 16 during that year then use the date of treatment switch
  - if the year of first dose is the same as the year of the start date and the same as the year the subject did switch from placebo to BKZ at Week 16, then use the date of first dose.
- If the date is completely unknown:
  - use the date of first dose of IMP if the stop date is unknown or not prior to the onset 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 31st of that year
- If the stop date is completely unknown, do not impute the stop date.

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.

In the event of ambiguity or incomplete data which makes it impossible to determine whether an AE is treatment-emergent, the AE will be considered as treatment emergent.

Other imputations

In addition, the following imputations will be applied when 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

No imputation rule will be applied for missing seriousness.

4.2.4 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 eCRF.

Imputation of partial start dates

- If only the month and year are specified:
  - if the month and year of first dose of IMP is not the same as the month and year of the start date and the subject did not switch from placebo to BKZ at Week 16 during that month and year, then use the 1st of the month.

  - if the month and year of first dose of IMP is the same as the month and year of the start, then use the date of first dose of IMP.
  - if the subject did switch from placebo to BKZ at Week 16 during that month and year, then use the date of treatment switch.
- If only the year is specified:
  - If the year of first dose of IMP is not the same as the year of the start date and the subject did not switch from placebo to BKZ at Week 16 during that year, then use the 1st of January of the year of the start date.
  - If the year of first dose of IMP is the same as the year of the start date, then use the date of first dose of IMP.
  - if the year of first dose of IMP is the same as the year the subject did switch treatment from placebo to BKZ at Week 16, then use the date of treatment switch.
  - if the year of first dose of IMP is the same as the year of the start date and the same as the year the subject did switch from placebo to BKZ at Week 16, then use the date of first dose of IMP.
- If the start date is completely unknown:
  - if the stop date is unknown or not prior to the date of first dose of IMP, then use the date of first dose of IMP.
  - if the stop date is prior to the date of first dose of IMP, then set the start date to the 1st of January of the year of the end 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 31st of that year.
- If the stop date is completely unknown, do not impute the stop date. There will be no imputation of any other missing data.

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.

In the event of ambiguity or incomplete data which makes it impossible to determine whether a medication was concomitant, the medication will be considered as concomitant.

4.3 Planned analyses and data monitoring

4.3.1 Interim analyses & final analysis

Two interim analyses will be performed:

- Interim analysis 1: Week 24 analysis.

The Week 24 analysis is the first interim analysis after all randomized subjects have completed the DBP and the ATP up to Week 24 or have withdrawn from the IMP or the study before Week 24. The purpose of this analysis is to perform a comprehensive evaluation of all available data for the 3 treatment arms to prepare a regulatory submission for a marketing authorization application based on this analysis. This interim analysis will cover the analysis of all efficacy and non-safety by visit tables on data collected up to Week 24 and the analysis of safety in tables on data collected up to the cut-off date (defined as Week 24 visit) for the last subject (all available data will be reported in listings). The type of efficacy and safety tables provided for the first interim analysis is detailed in Table 4–4.

For this analysis, the database will be cleaned and locked, and the treatment codes will be made available to the independent unblinded UCB and PAREXEL personnel. The database will include all data available at the time of the lock including available SFU data (for subjects having completed the SFU Period). An interim report will be written. The investigators and the subjects will remain blinded until the final analysis is completed. A blinding plan will be written to ensure blinding of investigators and subjects, define blinded and unblinded teams at UCB and PAREXEL and describe the process of generation and dissemination of the results of this first interim analysis. This plan will be finalized prior to the lock of the database for this analysis.

- Interim analysis 2: Week 52 analysis.

The Week 52 analysis is the second interim analysis after all subjects have completed Week 52 or have withdrawn from the IMP or the study. The purpose of this analysis is to demonstrate the long-term efficacy and safety of BKZ, perform a comprehensive evaluation of all available data for the study and to supplement the regulatory submission based on the Week 24 interim analysis. This analysis will comprise the results of the 16 weeks of the DBP, the 36-weeks of the ATP and the available SFU data at the time of this analysis (for subjects having completed the SFU Period). The type of efficacy and safety tables provided for the second interim analysis is detailed in Table 4–4.

No formal alterations to the further study conduct (eg, stopping rules, sample size re-estimation, or changes to eligibility criteria) are planned for the 2 interim analyses.

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

The below table summarize the type of efficacy and safety analyses to be performed for the 2 interim analyses.

Table 4–4: Type of efficacy and safety tables provided for the interim analyses

Variable **Type of tables**	**Period or Timepoint (Analysis set)**	**Treatment groups**	**Provided for Interim Analysis 1 (Week 24 analysis)**	**Provided for Interim Analysis 2 (Week 52 analysis)**
Tables of primary efficacy variable	Week 16 (RS FAS PPS)	• Placebo • BKZ 160mg Q4W • ADA 40mg Q2W	Yes	Yes, rerun of tables provided for interim analysis 1
Tables of secondary efficacy variables.	Week 4 and 16 (RS*)	• Placebo • BKZ 160mg Q4W • ADA 40mg Q2W	Yes	Yes, rerun of tables provided for interim analysis 1
Tables / Graphs of time to ACR response	DBP (RS)	• Placebo • BKZ 160mg Q4W • ADA 40mg Q2W	Yes, subjects not achieving response at Week 16 are censored at Week 16.	Yes, rerun of tables provided for interim analysis 1
	Overall Study Period (RS)	• Placebo/BKZ 160 mg Q4W • BKZ 160mg Q4W • ADA 40mg Q2W	Yes, subjects not achieving response at Week 24 are censored at Week 24.	Yes, subjects not achieving response at Week 52 are censored at Week 52.
By-visit tables of other efficacy endpoints	Overall Study Period (RS*)	• Placebo / BKZ 160 mg Q4W • BKZ 160mg Q4W • ADA 40mg Q2W	Yes, from Baseline to Week 24	Yes, from Week 0 until Week 52.
Extra by-visit tables for ACR50 and PASI90 using Baseline Week 16	ATP (ATS)	• Placebo / BKZ 160mg Q4W	Yes, from Week 20 until Week 24	Yes, from Week 16 until Week 52.
Extent of exposure and time at risk	DBP (SS)	• Placebo • BKZ 160mg Q4W • ADA 40mg Q2W	Yes	Yes
	ATP (ATS)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W • BKZ 160mg Q4W Total • ADA 40mg Q2W	Yes, until Week 24	Yes
	AMP (AMS)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W • BKZ 160mg Q4W Total • ADA 40mg Q2W	Yes, until Week 24	Yes
Cumulative study medication duration	Overall Study Period (SS)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W • ADA 40mg Q2W	Yes, until Week 24	Yes

Variable **Type of tables**	**Period or Timepoint (Analysis set)**	**Treatment groups**	**Provided for Interim Analysis 1 (Week 24 analysis)**	**Provided for Interim Analysis 2 (Week 52 analysis)**
Tables of AEs or Tables of AEs of Special Interest or Tables of Other Safety Topics of Interest.	DBP (SS)	• Placebo • BKZ 160mg Q4W • ADA 40mg Q2W	Yes	Yes
	ATP (ATS)	• Placebo / BKZ 160 mg Q4W • BKZ 160 mg Q4W • BKZ 160mg Q4W Total • ADA 40mg Q2W	Yes, until cut-off date for this analysis	Yes
	AMP (AMS)	• BKZ 160mg Q4W Total • ADA 40mg Q2W	Yes, until cut-off date for this analysis	Yes, for treatment-emergent events until Week 52 and including available SFU data
By-visit and shift tables of laboratory data or vital signs	Overall Study Period (SS)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W • ADA 40mg Q2W	Yes, until cut-off date for this analysis	Yes, for all visits until Week 52
Tables of markedly abnormal laboratory values or vital signs	DBP (SS)	• Placebo • BKZ 160mg Q4W • ADA 40mg Q2W	Yes	Yes
	ATP (ATS)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W • BKZ 160mg Q4W Total • ADA 40mg Q2W	Yes, until cut-off date for this analysis	Yes
	AMP (AMS)	• BKZ 160mg Q4W Total • ADA 40mg Q2W	Yes	Yes
By-visit tables of ECG, table of ECG abnormalities	Overall study period (SS)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W • ADA 40mg Q2W	Yes, until cut-off date for this analysis.	Yes, for all visits until Week 52.
Table of eC-SSRS events.	Overall Study Period (SS)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W • ADA 40mg Q2W	Yes, until cut-off date for this analysis	Yes
Table of tuberculosis testing.	Screening and W48 (SS)	• Placebo / BKZ 160mg Q4W • BKZ 160mg Q4W • ADA 40mg Q2W	Yes, Screening and Week 48 values (if observed prior to cut-off date).	Yes, Screening and Week 48 value.

ACR= American College of Rheumatology, ADA=Adalimumab, AE=adverse events, ATP=Active Treatment-Blind Period, ATS=Active Treatment-Blind Set, AMP=Active Medication Periods, AMS=Active Medication Set, BKZ=Bimekizumab, DBP=Double-Blind Treatment Period, eC-SSRS=electronic Columbia Suicide Severity Rating Scale, ECG=electrocardiogram, FAS=Full Analysis Set, PASI= Psoriasis Area and Severity Index, RS=Randomized Set, SS=Safety Set
*For subjects in the Radiographic Set for vdHmTSS

Note: Subjects in the Placebo/ BKZ 160mg Q4W treatment group are taking BKZ 160 mg Q4W treatment after Week 16.

- The final analysis.

The final analysis will consist of a rerun of all analyses provided during the preceding Week 52 analysis. This includes new SFU data that were not available for the Week 52 analysis. If there is no SFU data ongoing, the final analysis will be identical to the Week 52 analysis.

Prior to the final analysis, additional data cuts may be prepared following regulatory requests or for publication purposes.

4.3.2 Data Monitoring Committee

A Data Monitoring Committee (DMC) will be reviewing safety data on an ongoing basis. The DMC membership includes experienced clinicians and a statistician, all of whom have expertise in clinical studies. Further details are specified in the DMC Charter. The presentation and analysis of data for the DMC meetings will be described in a separate DMC SAP.

4.4 Multicenter studies

In general, the data from all centers will be pooled for the purposes of the analysis. However, the effect of center (using a pooling of centers by region) on results will be evaluated as mentioned in Section 3.8.

4.5 Multiple comparisons/multiplicity

A step-down closed testing procedure will be applied for the primary endpoint and some 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 BKZ 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 (i.e., if $p<0.05$). Table 4–5 shows the testing order for these endpoints.

Table 4–5: Sequential testing procedure of primary/secondary efficacy endpoints (fixed sequence testing procedure) (All efficacy endpoints at Week 16)

BKZ 160mg Q4W	
H_1	ACR50 Response superior to Placebo
H_2	Change from Baseline HAQ-DI superior to Placebo

BKZ 160mg Q4W	
H_3	PASI90 Response superior to Placebo
H_4	Change from Baseline SF-36 PCS superior to Placebo
H_5	MDA superior to Placebo
H_6	Change from Baseline vdHmTSS* superior to Placebo on subjects with elevated hs-CRP and/or with at least one bone erosion (hs-CRP ≥6mg/L and/or erosion-positive)
H_7	Enthesitis-free state superior to Placebo (based on pooled PA0010 and PA0011 W16 data)
H_8	Dactylitis-free state superior to Placebo (based on pooled PA0010 and PA0011 W16 data)
H_9	Change from Baseline vdHmTSS** superior to Placebo

ACR=American College of Rheumatology; BKZ=bimekizumab; CfB=change from Baseline; H=hypothesis; HAQ-DI=Health Assessment Questionnaire - Disability Index; hs-CRP=high sensitivity C-reactive protein; LEI=Leeds Enthesitis Index; MDA=Minimal Disease Activity; PASI=Psoriasis Area and Severity Index. PCS=Physical Component Summary; Q4W=every 4 weeks; SF-36=Short-Form 36-item Health Survey; vdHmTSS=Van der Heijde modified Total Sharp Score.
* It is planned to enroll a minimum of 45% of subjects who are positive for elevated hs-CRP (hs-CRP ≥6mg/L) and/or who have at least 1 bone erosion at Screening.
**Based on the overall population

4.6 Use of an efficacy subset of subjects

The RS will be the primary analysis set for efficacy analyses, but analyses will also be repeated on the FAS and the PPS for the primary efficacy endpoint. The FAS analysis will evaluate whether there are differences in the efficacy analysis between randomized subjects and randomized subjects with a Baseline assessment, while the PPS analysis will evaluate the effect of IPD 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

Subgroup analyses will be performed on the variables below. They are all assessed at Baseline except concomitant cDMARDs, concomitant MTX and ADAb status which will be assessed during the 16-week DBP. Subgroup analyses will be performed on the ACR50 response, the PASI90 response and the HAQ-DI response (subjects with a decrease of HAQ-DI from Baseline of at least 0.35) at Week 16. ADAb status will also be used for subgroup analysis for the PK endpoints. The variables for subgroup analyses will be:

- Age (<45 years of age, ≥45 years of age)
- Gender (male, female)
- Disease duration (<1 year, ≥1 year)

- Region (eg, North America, Western Europe, Eastern Europe, Asia. Western Europe will be combined with Asia if any of these 2 regions have less than 10% of the randomized subjects)
- Race (White, Black and Other)
- Body weight at Baseline (≤100kg, >100kg)
- Bone erosion (≥1) at Baseline (Yes, No)
- hs-CRP at Baseline (<6mg/L, ≥6mg/L)
- Bone erosion (≥1) and/or hs-CRP ≥6mg/L at Baseline (Yes, No)
- Prior cDMARDs (0, 1, ≥2) (taken prior to Baseline,). Such medications will be classified using the adjudicated spreadsheet.
- Concomitantly receiving cDMARDs versus no concomitant cDMARDs
- Concomitantly receiving MTX versus no concomitant MTX
- Concomitantly receiving MTX at Baseline vs. other cDMARDs at Baseline (MTX at Baseline, no MTX at Baseline and cDMARDs at Baseline, no MTX at Baseline and no cDMARDs at Baseline)
- PSO affected BSA at Baseline (<3%, ≥3% to 10%, >10%)
- BASDAI at Baseline (<4, ≥4)
- ADAb status (positive, negative) (Section 9.2) (for the BKZ 160mg Q4W group only)
- Human leukocyte antigen B27 (HLA-B27) (positive, negative).

In the context of submission to Japanese health authorities, a subset of tables and figures will be provided in the subgroup of subjects randomized in Japan(when MI is used, data for subjects randomized in Japan will be extracted from the MI datasets generated for the full population).

5 STUDY POPULATION CHARACTERISTICS

5.1 Subject disposition

Subject disposition will be summarized for all subjects screened overall, by region and by site. In this summary, the dates of first subject screened and dates of last subject last visit (end of the study), number of screened subjects, and the number of subjects included in each analysis set (RS, SS, FAS, AMS, ATS, PPS, RAS, PK-PPS, COVID-19-free Set) will be presented.

The disposition of subjects into treatment groups and analysis sets will also be summarized on the RS.

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

The number and percentage of randomized subjects who completed the study, completed the study not on randomized treatment, who discontinued IMP overall and by reason and who discontinued the study, with the primary reason for study discontinuation (as collected on the Study Termination CRF page), by study treatment period (DBP and ATP) will be tabulated

overall and by treatment group. The numbers and percentages of randomized subjects entering the SFU Period will additionally be presented. The numbers and percentages of these subjects who either complete the SFU visit or not will additionally be presented.

The numbers and percentages of randomized subjects entering the Open Label Extension (OLE) study and those not entering the OLE study will be presented. Of the randomized subjects not entering the OLE study.

A subject will be said to have completed the study if she/he had completed the last scheduled study visit, not including SFU visit (the SFU visit is typically 140 days after the last dose of IMP, which is roughly equivalent to 5 half-lives of BKZ).

A subject will be considered to have completed a period if she/he completes the last scheduled study visit for that period.

Subjects who previously discontinued IMP and are continuing for all scheduled visits through the end of each period (DBP or ATP) will also be considered as completing the period. A separate summary of subjects who discontinue IMP will be provided.

Finally, the number of randomized subjects who discontinued the study due to AEs (during the DBP and the ATP separately) by type of AEs (serious fatal, non-fatal, other) will be summarized.

Study disposition and termination details will be listed for all screened subjects. Additional listings will be created on study discontinuation, subject analysis sets and visit dates (both actual visit and visit remapped for analysis). A listing will also be provided for subjects who did not meet the study eligibility criteria.

To assess subject disposition (entry and periods in the study) during the COVID-19 pandemic, subject disposition will be assessed by period of the COVID-19 pandemic (pre – during) (Section 3.11), by comparing the dates of visits (or events) to the dates of the COVID-19 pandemic period.

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, dates of visit, relationship to COVID-19, impact category and a narrative (short description) of the event. These data will be summarized by treatment group and overall, for randomized subjects.

A summary of study visits by COVID-19 pandemic period (pre – during) will be presented for subjects enrolled prior or during the pandemic.

In addition, in order to assess the potential impact of COVID-19 on the collection and reporting of efficacy data, a separate summary based on the RS will be presented to display missing data as well as data collected via an alternative modality (eg: phone, video call) for efficacy endpoints included in the hierarchy (Table 4–5). For 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.

For visits conducted remotely, it was not possible to assess some endpoints (eg, TJC, SJC, LDI, LEI, PASI, etc.) and therefore these assessments will be missing for those visits. In addition, for any missed visit or visit conducted remotely, the CRP assessment will also be missing. Such

assessments will be considered missing because of the COVID-19 pandemic. For these visits, it will therefore not be possible to assess composite endpoints like ACR response. For this summary table based on the RS, 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 summary displaying the number and percentage of subjects with an IPD will be provided by treatment group separately for the DBP on the RS and on the ATP on the ATS. This will include a summary of subjects excluded from the PPS (DBP only) and the PK-PPS. The summary will be provided overall 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, COVID-19 related IPD) as well as the number and percentage of subjects excluded from the PPS due to reason other than PD.

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

A by-subject listing of IPDs will be provided for all subjects in the RS.

A by-subject listing of COVID-19 related PDs will be provided for all subjects in the RS.

A by-subject listing of subjects excluded from the PPS for reasons other than IPD will be provided.

6 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Unless otherwise specified, all summaries will be based on the SS and repeated on the RS (unless the SS and RS analysis sets are identical) and on the ATS.

6.1 Demographics

The following demographic variables measured at Baseline will be summarized overall and by treatment group: age (years), gender, country, geographic region, country, race, ethnicity, height (cm), body weight (kg) and Body Mass Index (BMI) (kg/m^2).

Age, body weight and BMI will be summarized as continuous variables and as categorical variables based on the categories specified below.

For age, 3 sets categories will be defined:

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

For body weight, categories will be: ≤100 kg, >100 kg.

BMI (in kg/m^2) will be recalculated during analysis based on height (in meter) and weight (in kg). The formula for BMI (kg/m^2) calculation is:

$$BMI = \frac{Body\ weight\ (kg)}{(Height(m))^2}$$

For BMI, categories will be: <25, 25 to <30, 30 to <35, ≥35 kg/m^2.

A separate frequency table will summarize the subject's lifestyle on the SS.

By-subject listings on demographics, female subject's childbearing potential data will be provided for all screened subjects. Listing on lifestyle will be provided on the RS.

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 pandemic period (pre or during as determined by the Baseline visit date) on the SS.

6.2 Baseline characteristics

6.2.1 Baseline values for components of the primary efficacy variable

Baseline values for the 7 components of the primary efficacy variable will be summarized by treatment group and overall. The following variables will be summarized:

- TJC
- SJC
- PGA-PsA
- PhGA-PsA
- PtAAP
- HAQ-DI
- hs-CRP (mg/L).

For hs-CRP, SJC and TJC, the following rule will be applied: if the Baseline value is missing, then the most recent value (taken at previous unscheduled visit or at Screening) will be considered. If the value is still missing, the Baseline value will be considered missing.

The hs-CRP variable will be summarized using descriptive statistics and in classes (≥6mg/L, <6mg/L).

Efficacy variables not listed above will not be presented in the Baseline characteristics section. They will be presented in by-visit tables in the context of the analysis of secondary and other efficacy variables.

To understand the study population during the different periods of the COVID-19 pandemic, the reporting of Baseline characteristics will be repeated by COVID-19 pandemic period (pre or during as determined by the Baseline visit date) on the SS.

6.2.2 Other Baseline characteristics

The following variables will be summarized by treatment group and overall:

- Bone erosion (0, ≥1) – based on IXRS randomization strata
- Bone erosion (0, ≥1) – based on actual randomization strata (as per imaging data)
- Bone erosion (≥1) and/or hs-CRP ≥6mg/L at Screening (Yes, No). To determine this, bone erosion will be derived from imaging data and corresponds to actual randomization strata
- % of BSA affected by PSO (<3%, ≥3% to ≤10%, >10%)
- PASI score at Baseline (for subjects with PSO involving at least 3% of BSA at Baseline). This variable will be summarized as a continuous variable and in classes (<10, 10 to 20,>20)
- BASDAI (<4, ≥4)
- Rheumatoid factor (positive, negative). A rheumatoid factor value <30IU/mL is defined as negative and a value ≥30IU/mL is defined as positive
- Anti-cyclic citrullinated peptide (CCP) antibodies (positive, negative). Anti-CCP antibodies negative is a value <20U/mL and a value ≥20U/mL is defined as positive
- Nail psoriasis (yes, no)
- Dactylitis ("yes" if LDI score>0, "no" if LDI score=0, missing)
- Enthesitis based on LEI ("yes" if LEI score>0, "no" if LEI score=0, missing)
- Enthesitis based on SPARCC ("yes" if SPARCC>0, "no" if SPARCC=0, missing)
- Prior NSAID (yes, no)
- NSAID at Baseline (yes, no)
- Past cDMARDs therapy (yes, no)
- Prior cDMARDs (0, 1, ≥2) (Section 4.8)
- cDMARDs at Baseline (yes, no)
- HLA-B27 (positive, negative)
- Methotrexate at Baseline (yes, no) (Section 4.8)
- Prior oral corticosteroids (yes, no)
- Oral corticosteroids at Baseline (yes, no).

To understand the study population during the different periods of the COVID-19 pandemic, the reporting of other Baseline characteristics will be repeated by COVID-19 period (pre or during as determined by the Baseline visit date) on the SS.

6.2.3 History of psoriatic arthritis

The following PsA history variables will be summarized overall and by treatment group:

- The time since first diagnosis of PsA (years). Time since first diagnosis will be summarized as a continuous variable and as categorical variables based on following categories: <1, 1-<2, ≥2 years
- The time since first diagnosis of PSO (years), calculated as :

$$\frac{Date\ of\ informed\ consent - Date\ of\ diagnosis}{365.25}$$

  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).
- The age at first diagnosis date of PsA (years), calculated as:

$$\frac{Date\ of\ PsA\ diagnosis - Date\ of\ birth}{365.25}$$

  If the date of PsA diagnosis is partial, it should be imputed to the most recent feasible date (ie: the last day of the month if only day is missing, or the last day of the year if day and month are missing. Partial date should not be imputed later than the subject's screening date). For subjects enrolled at German sites, only the year of birth may be entered into the eCRF for this study. For these subjects age will be calculated after imputing their date of birth to be on January 1st.
- The PsA subtype (polyarticular – symmetric arthritis, oligoarticular – asymmetric arthritis, distal interphalangeal, joint predominant, spondylitis predominant, arthritis mutilans)

The data on history of psoriasis and psoriatic arthritis will be listed for all screened subjects.

6.3 Medical history and concomitant diseases

Medical history conditions are defined as conditions that have resolved prior to study entry.

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

- Medical history and ongoing medical conditions will be summarized on the SS by MedDRA System Organ Class (SOC) and Preferred Term (PT), by treatment and overall including the number and percentage of subjects with each condition. The denominator of the percentages will be the number of subjects in the population considered. The tables will be ordered alphabetically for SOC and in terms of decreasing frequency for PT within SOC in the BKZ treatment group. In the event of ties, PT will be ordered alphabetically.

Medical history and ongoing medical conditions will be listed by treatment and subject including the reported term, PT, and SOC on the RS. The start date (month and year only) and the end date (or ongoing if applicable) will be included in the listing. Partial dates will not be imputed. A glossary of all medical history conditions will be presented 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 date of the first IMP administration.

Concomitant medications are medications taken at least one day in common with the IMP dosing period.

The IMP dosing period corresponding to the combined DBP and ATP will be calculated as follows:

The study medication dosing period start date is defined as the date of first dose.

The study medication dosing period stop date is defined as follows:

- For subjects who are ongoing at the time of the clinical data cut (not including participants who are in SFU) use date of last clinical contact
- For subjects who died prior to last visit (not including those who died during SFU), use the minimum of the following:
  - Date of death
  - Date of last dose of any study medication + 28 days
- For subjects who complete the study as planned, the dosing period ends at the later of the following 2 dates:
  - Date of last administration of IMP + 14 days
  - The last scheduled visit date not including SFU
- For all other subjects, use the maximum of the following:
  - Date of last dose of any study medication + 28 days
  - Date of last visit (not including SFU).

In the case of missing data, the classification of medications as prior or concomitant will be performed as described in Section 4.2.4. Imputations of missing data will be performed before calculation of relative study days.

The number and percentage of subjects taking prior medications (excluding past psoriasis medications) will be summarized on the SS by treatment group, overall and by ATC class, presenting Anatomical Main Group (ATC Level 1), Pharmacological Subgroup (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 BKZ treatment group. In the event of ties, PT will be ordered alphabetically.

The number and percentage of subjects taking concomitant medications will be summarized similarly on the SS by treatment group (for the Week 24 analysis on the period from Baseline to Week 24, for the Week 52 analysis on the overall study treatment period).

A by-subject listing of all prior (on the RS) and concomitant medications (on the RS) as well as glossary of all medical medications will be provided.

6.5 Prohibited medications and permitted rescue therapy after Week 16

Prohibited or restricted medications are defined in the protocol (section 7.8.2). Prohibited medications will be listed on the SS.

Some permitted rescue therapies are allowed after Week 16 as defined in Section 5.4 of the protocol.

The number and percentage of subjects who used permitted rescue medications after Week 16 will be summarized by ATC class, presenting ATC Level 1, ATC level 3, and PT. This table will be produced on the ATS.

Permitted rescue therapies will be listed on the SS.

7 MEASUREMENTS OF TREATMENT COMPLIANCE

Treatment compliance summaries will be provided:

- On the SS for the DBP, by randomized treatment and for all subjects.
- On the ATS for the ATP, by randomized treatment, for all BKZ treated subjects and for all subjects
- On the AMS for the AMP, by randomized treatment, for all BKZ treated subjects and for all subjects.

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

$$\frac{Total\ number\ of\ actual\ (completed)\ injections}{Total\ number\ of\ expected\ injections} x\,100\ \%$$

The total number of expected injections is derived relative to when the subject finishes treatment:

- If a subject completes the DBP, 8 injections are expected (every 2 weeks from Baseline to Week 16)
- If a subject completes treatment over the entire study, 26 injections are expected (every 2 weeks from Baseline to Week 50)
- If a subject discontinues early, then the number of expected injections is based on the time of early discontinuation relative to the dosing visits. This number will be calculated by determining the number of expected dosing visits prior to the date of early discontinuation. If the early discontinuation coincides with an expected dosing visit, this visit will used to determine the number of expected injections.

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

8 EFFICACY ANALYSES

8.1 Statistical analysis of the primary efficacy variable

8.1.1 Derivation of ACR50 response

The ACR50 response at Week 16 is the primary efficacy variable.

This is a composite endpoint that represents at least 50% improvement from Baseline at Week 16 for each of the following:

- TJC (based on 68 joints),
- SJC (based on 66 joints),
- At least 3 of the 5 remaining core set measures:
  - PGA-PsA
  - PhGA-PsA
  - PtAAP
  - HAQ-DI
  - hs-CRP.

Prior to deriving ACR50, the percent of change for each component will be calculated as:

$$\text{Percent improvement from Baseline} = 100 \times \frac{\text{Baseline Value} - \text{Week 16 value}}{\text{Baseline Value}}$$

As the ACR response is based on 7 different component scores, it is necessary to consider various data scenarios that could impact the calculation of ACR response. The rules described here are applicable in the context of the calculation of ACR 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 ACR response)

The following rules will be applied prior to invoking any imputation analysis at the variable level:

- If a subject has a component value that is equal to 0 at Baseline and the post-Baseline value is ≥ 0, then the percent improvement for that component will be treated as 0 for purposes of ACR response calculations.
- If a subject has a component value that is missing at Baseline, then the percentage of improvement for that component will be treated as missing for purposes of ACR response calculations.

Observed data will be used to calculate ACR response where possible. In case of partial missing data where an observed response may be calculated, imputed data will not change the result.

To address possible missing data in ACR components, ACR50 will be derived as below:

- Positive response on ACR50:

  - Improvement at Week 16 of at least 50% from Baseline on SJC and TJC and at least 3 out of the 5 remaining ACR components (regardless of whether the 2 remaining components are missing or not).

- Negative response on ACR50:
  - No improvement at Week 16 of at least 50% from Baseline on SJC (regardless of whether the remaining components are missing or not).
  - No improvement at Week 16 of at least 50% from Baseline on TJC (regardless of whether the remaining components are missing or not).
  - No improvement for at least 3 out of the 5 following ACR components (PGA-PsA, PhGA-PsA, PtAAP, HAQ-DI, hs-CRP) regardless of whether the remaining components are missing or not.
- Missing ACR50 response:
  - For all other situations that are not specified above.

For the OC analysis (Section 8.1.4.6): if a given visit date falls after the treatment discontinuation date, the endpoint at this visit and all subsequent visits (whether the data were observed or not) will be derived but reported as missing.

For the MI analysis (Section 8.1.4.3): If a given visit date falls after the treatment discontinuation date, the endpoint at this visit and all subsequent visits (whether the data were observed or not) will be imputed. Responses falling after an IE (defined as treatment discontinuation due to AE or lack of efficacy for the modified composite estimand) will be set to non-response.

8.1.1.1 Derivation of tender joint count (TJC) and swollen joint count (SJC)

Tender and swollen joints are assessed based on the 68 and 66 joints for tenderness and swelling, respectively. The 68/66 joints that are assessed are:

- Upper body (6): bilateral temporomandibular, sternoclavicular, and acromioclavicular joints
- Upper extremity (34): bilateral shoulders, elbows, wrists (includes radiocarpal, carpal and carpometacarpal joints considered as 1 unit), metacarpophalangeals (I, II, III, IV, V), thumb interphalangeals, proximal interphalangeals (II, III, IV, V), distal interphalangeals (II, III, IV, V)
- Lower extremity (28): bilateral hips, knees, ankles, tarsi (includes subtalar, transverse tarsal, and tarsometastarsal joints considered as 1 unit), metarsalphalangeals (I, II, III, IV, V), great toe interphalangeals, and proximal interphalangeals (II, III, IV, V).

The assessments on swelling exclude the hips.

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 non-evaluable 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. The asymptomatic joints will be included in the joint count analysis with scores of 0.

For the statistical analysis, swelling and tenderness are each graded on a 2-point scale as described below:

Table 8–1: Swelling and Tenderness grading

Present	Tenderness response (68)	Swelling response (66)
No (0)	Not tender	None
Yes (1)	Positive response to questioning (tender), spontaneous response elicited (tender and winced) or withdrawal by subject on examination (tender, winced, and withdrew).	Detectable synovial thickening with or without loss of bony contours, or bulging synovial proliferation with or without cystic characteristics.

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 response of 1) from and including the date of injection up to 52 weeks after injection. Injected joints are allowed after Week 16 as part of rescue medications.

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 (TJC) and swollen joint count (SJC) 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.

If a joint is missing at Baseline, then that joint is set to missing throughout the study. If more than 50% of the planned tender joint assessments (ie more than 34) or 50% of the planned swollen joint assessments (ie more than 33) are missing at the time of a given assessment any post-Baseline visit, then no imputation will be done and the total TJC or SJC will be set to missing for that visit.

8.1.1.2 Patient and physician reported clinical outcomes on psoriatic arthritis

The physician's global assessment of psoriatic arthritis (PhGA-PsA), the patient's global assessment of psoriatic arthritis (PGA-PsA) and the patient's assessment of arthritis pain (PtAAP) are clinical outcomes based on visual analog scales (VAS) ranged from 0 to 100 (in mm).

- PhGA-PsA: The investigator will assess the overall status of the subjects with respect to their PsA signs and symptoms using a numerical rating scale where 0 is "very good, asymptomatic and no limitation of normal activities" and 100 is "very poor, very severe symptoms which are intolerable and inability to carry out all normal activities". This assessment will be based through the following question "Considering all the ways the disease affects your patient, mark a vertical line on the scale for how well his or her condition is today."
- PGA-PsA: Subjects will assess the impact of PsA in answering the following question "Considering all the ways your psoriatic arthritis affects you, please mark a vertical line on the scale below to show how well you are doing today." The subjects should be asked to consider all aspects of their disease (including joint and skin components) in their response to this question. Subjects will score PGA-PsA in using a VAS where 0 is "very good, no symptoms" and 100 is "very poor, severe symptoms."
- PtAAP: Subjects will assess their arthritis pain through the PtAAP using a VAS where 0 is "no pain" and 100 is "most severe pain."

8.1.1.3 Derivation of Health Assessment Questionnaire-Disability index (HAQ-DI) score

The HAQ-DI contains 20 items that measure the degree of difficulty experienced in the following 8 categories of the daily living activities: dressing and grooming (2 items), arising (2 items), eating (3 items), walking (2 items), hygiene (3 items), reach (2 items), grip (3 items), and common daily activities (3 items).

For each question/item, subjects should indicate the level of difficulty (ranged from 0 to 3) in the past week as follows:

- 0: 'without any difficulty'

- 1: ‘with some difficulty’
- 2: ‘with much difficulty’
- 3: ‘unable to do’

Each category is given a score by taking the maximum value for each question.

For each question related to a specific daily living activity, subjects can specify whether an aid or devices is usually used.

If a category score equals 0 or 1, but a device related to that category is used, or help from another person is provided for that category, then the category score is increased to 2.

If a category score already equals 2, and a device related to that category is used, or help from another person is provided for that category, the score for that category remains 2.

Table 8–2 details how each aid and device is associated with the category scores. Aid or device considered as ‘other’ than the ones listed will not be considered in the analysis.

Table 8–2: Aid and device associated with HAQ-DI category

[redacted]	**Will be associated with category score**
[redacted]	Walking
[redacted]	Dressing and grooming
[redacted]	Eating
[redacted]	Arising
[redacted]	Hygiene
[redacted]	Hygiene
[redacted]	Grip
[redacted]	Hygiene
[redacted]	Reach
[redacted]	Hygiene

If all questions within a given category are unanswered, no score will be provided for that category (this rule applies even if aids and devices are non-missing).

The HAQ-DI score will be calculated by dividing the sum of the highest score in each category (0 to 24) by the number of categories with at least 1 question answered. If fewer than 6 categories have responses, no HAQ-DI score will be calculated (ie HAQ-DI score will be considered as missing).

The HAQ-DI score ranges from 0 to 3. A lower HAQ-DI score indicates an improvement in function.

8.1.1.4 High sensitivity C-reactive protein levels (hs-CRP)

High sensitivity C-reactive protein values which are below the lower limit of quantification (LLOQ) should be set to the midpoint between 0 and the LLOQ prior to the analysis. Listing will show values below the LLOQ. The LLOQ for hs-CRP is 0.10 mg/L.

8.1.2 Primary analysis of the primary efficacy variable

The primary efficacy variable will be analyzed for all subjects in the RS.

The primary endpoint is the ACR50 response at Week 16. The primary efficacy analysis will evaluate the composite estimand in the RS. The composite estimand combines the clinically meaningful improvement from Baseline in ACR50 response and not discontinuing study treatment early.

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

- Population = Subjects enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP.
- Subject-level outcome = ACR50 at Week 16.
- IE handling = An IE 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 ACR50 at Week 16 and not discontinuing study treatment through Week 16.
- Population-level summary measure = Conditional OR comparing BKZ to placebo.

Missing data at Week 16 that are not preceded by an IE, and any data after an IE will be imputed as non-responders. This results in a more traditional NRI approach and will be labeled as such throughout the SAP.

The statistical hypothesis for the ACR50 response at Week 16 is that the conditional OR for ACR50 response in the BKZ treatment compared to placebo treatment is equal to 1.

A logistic regression model will be used to assess the treatment effect on ACR50 response at Week 16. The model will include a fixed effect for treatment. The suitability of including randomization stratification variables (bone erosion at Baseline and region) will be assessed using Pearson and Deviance and The Hosmer-Lemeshow Goodness-of-Fit Tests (Hosmer and Lemeshow, 2000).

In the context of the Week 24 analysis, actual bone erosion at Baseline as a covariate will be based on data from the first set of reads. For Week 52 analysis, data from the second set of reads will be used.

P-values below 0.05 would lead to a reconsideration of the model to be used and if the logistic regression model is unable to converge the stratification variables will be dropped (Section 4.1). In that case, all supportive analyses as well as imputation models conducted for the primary endpoint will disregard these factors from their models.

The country-specific analyses performed on subjects randomized in Japan will not consider the region factor as a covariate for the modelling.

Considering that the number of subjects randomized in Japan is low (less than 10% of the Randomized Set), statistical models might not converge. If a model (Logistic Model or Mixed model) is not converging, all related adjusted statistics and p-value will not be presented : “NE” for “Not Evaluable” will be displayed instead.

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

The summary table results will present the adjusted responder rates and the associated 95% CIs for the 3 treatment groups, the adjusted OR and the corresponding 95% CI for the comparison of BKZ versus placebo, the p-value that the OR=1 and the difference of response rate between BKZ and placebo and associated 95% CI. Comparisons of BKZ vs. placebo will be made using the 2-sided Wald test at a significance level of α=0.05 (ie, H_1 in Table 4–5).

Any use of prohibited or rescue medications through Week 16 would constitute an important protocol deviation which would be accounted for when the sensitivity analysis based on the PPS is performed (Section 8.1.4.1).

8.1.3 Subgroup 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.

In the context of the Week 24 analysis, actual bone erosion at Baseline used as a covariate and/or a subgroup will be based on data from the first set of reads. For the Week 52/final analyses, data from the second set of reads will be used.

For each subgroup analysis (except for the analysis by region, by bone erosion at Baseline, by hs-CRP value combined with bone erosion at Baseline and by ADAb status), a logistic regression will be fitted involving the same terms that were retained when running the primary analysis model, plus a term for the subgroup and the subgroup by treatment interaction as detailed below :

- Fixed effect for treatment
- Bone erosion at Baseline (0 if bone erosion at Baseline=0; 1 if bone erosion at Baseline≥1)
- Region
- Subgroup
- Subgroup by treatment interaction.

For subgroup analyses by bone erosion, the terms that will be retained will be:

- Fixed effect for treatment
- Bone erosion at Baseline
- Region
- Bone erosion at Baseline by treatment interaction.

For subgroup analyses by region, the terms that will be retained will be:

- Fixed effect for treatment

- Bone erosion at Baseline
- Region
- Region by treatment interaction.

For subgroup analyses by bone erosion (≥1) and/or hs-CRP ≥6mg/L at Baseline (Yes, No), the terms that will be retained will be:

- Fixed effect for treatment
- Bone erosion (≥1) and/or hs-CRP ≥6mg/L at Baseline (Yes, No)
- Region
- Bone erosion (≥1) and/or hs-CRP ≥6mg/L at Baseline by treatment interaction.

For subgroup analyses by ADAb status, the terms that will be retained will be:

- Bone erosion at Baseline
- Region
- ADAb Status (Positive, Negative).

For all subgroup analyses, the term for the subgroups will coded as below:

- For Yes/No or positive /negative variables, the category "yes" or "positive" will be coded "1".
- For continuous variables in 2 classes, the highest category will be coded "1".
- For continuous variables in 3 classes (or more), the coding categories will start from 1 for the lowest value.
- For gender, "male" will be the category coded "1".

The same estimand approach as the one used for the primary analysis will be used to handle missing data (NRI).

For each subgroup category and each of the 3 treatment groups (for the BKZ group only for the subgroup analysis on the ADAb status), the mean proportion of responders on imputed datasets, the adjusted responder rates with associated 95% CI, will be provided. The adjusted OR (for the comparison of BKZ vs. placebo), the corresponding 95% CI, the difference in response rates between BKZ and placebo and the corresponding 95% CI will also be presented (except for the subgroup analysis on the ADAb status). The results obtained for each subgroup will be presented in one single table.

The observed rates of responses by subgroup categories will be also provided along with the results obtained on imputed data.

For each subgroup category, the ORs and the associated 95% CIs will also be displayed on a single forest-plot.

In addition, the mean proportion of ACR50 responders at Week 16 separated by the 6 pre-defined ADAb subcategories (as defined in Section 9.2) will be presented graphically for the BKZ randomized subjects.

8.1.4 Supportive analyses of the primary efficacy variable

The following supportive analyses for the primary efficacy variable will be conducted on the RS (except analyses described in Section 8.1.4.1 , 8.1.4.2 and 8.1.4.8). For the below analyses, unless specified:

- The same model as run for the primary analysis will be used.
- The same statistics will be provided: the mean proportion of responders (on imputed datasets when applicable), the adjusted responder rates for each treatment group, the adjusted ORs for the comparison BKZ versus placebo and its 95% CI , the p-value for the comparison between BKZ and placebo and the difference of response rates between BKZ and placebo and its corresponding 95% CI.

The adjusted OR and 95% CI for the comparison of BKZ vs. placebo obtained for the primary analysis and the supportive analyses described below (except for Section 8.1.4.4) will be displayed in a forest plot.

8.1.4.1 Analysis on the PPS

The analysis described in Section 8.1.2 will be repeated based on the PPS to evaluate the effect of IPD on the analysis.

8.1.4.2 Analysis on the FAS

The analyses described in Section 8.1.2 will be repeated for all subjects in the FAS to evaluate the consistency between the RS and the more restrictive FAS. This analysis will only be performed if the number of subjects in RS and FAS are different.

8.1.4.3 Analysis using a modified composite estimand where intercurrent events are defined as discontinuation due to AE or lack of efficacy

The modified composite estimand is detailed below:

- Population = Subjects enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP.
- Subject-level outcome = ACR50 at Week 16.
- IE handling = An IE is defined as discontinuation due to AE or lack of efficacy. A composite strategy will be implemented in which a positive clinical outcome is defined as achieving ACR50 at Week 16 and not discontinuing study treatment due to AE or lack of efficacy through Week 16.
- Population-level summary measure = Conditional OR comparing BKZ to placebo.

Any missing data at Week 16 that is not preceded by an IE will be imputed based on a predefined MI model (MI/MCMC - Monotone Regression) (Section 4.2.2).

The following categories of subjects will enter the MI/MCMC – Monotone regression process for each ACR component:

- Subjects with all post-baseline values missing
- Subjects withdrawn from the study treatment (data collected prior to the treatment withdrawal will be retained) - Those subjects will have values imputed for missing assessments and non missing assessment falling after study treatment discontinuation (falling after or before the study withdrawal)
- Subjects with missing value at Baseline. MI will be applied for each of the components separately before deriving the response.

In the case of partial missing data, ACR will be derived as in Section 8.1.1.

If an IE occurred prior to Week 16, the subjects will be considered as "non-responders".

In MI, the missing value is replaced by a set of plausible values, where each value is a Bayesian draw from the conditional distribution of the missing data given the observed data.

8.1.4.4 Analysis of individual components of the ACR

The primary comparison (BKZ vs. placebo) will also be repeated for all individual components of the ACR response to explore the effect of the signs and symptoms of the individual components on the composite endpoint. Since all ACR components are continuous variables (eg, change from Baseline in TJC), an analysis of covariance (ANCOVA) with treatment, region, and bone erosion at Baseline as fixed effects and the Baseline value as covariate will be used for the analysis. Missing data imputation described in Section 4.2.2.2 and 4.2.2.3 will be successively applied for each component.

The following 4 attributes describe the hypothetical estimand that will be used to define the treatment effect of interest for each of the 7 ACR components:

- Population = Subjects meeting the protocol-specified inclusion/exclusion criteria and randomized to IMP.
- Subject level outcome = Change from Baseline in ACR components at Week 16.
- IE handling =An IE is defined as discontinuation of study treatment 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 subjects without an IE are as observed, and outcomes for subjects with an IE are treated successively according to the two following hypothetical strategies:
  - (a) as they had completed treatment through Week 16. In that strategy, missing data and non missing data after IE (reset as missing) will be imputed using MI.
  - (b) 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 BKZ 160mg Q4W and placebo.

For the results obtained using the MI method, the SAS® PROC MIXED will be used to run the ANCOVA models.

The inferential statistics obtained from each model will be displayed in a single table. For all variables, the following statistics will be presented:

- The LSM and SE by treatment group
- For the comparison between BKZ and placebo: the difference between the LSM, the associated 95% CI for the contrasts, and the corresponding p-value.

An additional table will present the descriptive statistics for the 3 treatment groups for each ACR component variable at Week 16. Those statistics will be the mean, SE, median, minimum and maximum obtained after MI.

8.1.4.5 Analyses using treatment policy strategy imputation for missing data

The treatment policy strategy for addressing IE will be considered. This will be based on an analysis in which all available data at Week 16 regardless of the occurrence of IE will be kept in the analysis. This analysis will use the same models specified for the primary analyses where subjects are analyzed according to their randomized treatment, even if they discontinued study treatment for any reason prior to Week 16 and were no longer on the randomized study treatment when the assessment was performed at Week 16.

In the context of this analysis, in addition to the statistics provided in the context of MI, the number of observed ACR50 responses regardless of the occurrence of IE will be also reported.

The estimand for the treatment policy strategy analysis is detailed below:

- Population = Subjects enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP.
- Subject-level outcome = ACR50 at Week 16.
- IE handling = Discontinuation of study medication for any reason will be handled by using the data as observed irrespective of the IE occurring.
- Population-level summary measure = Conditional OR comparing 'BKZ and other medication as needed' to 'placebo and other medication as needed'.

8.1.4.6 Analysis on Observed cases

An OC analysis will additionally be conducted, where only observed data for subjects who are still on the randomized treatment at Week 16 are included. Subjects with missing data at Week 16 or who have prematurely discontinued study treatment prior to Week 16 will be treated as missing and missing data will not be replaced.

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

- Population = Subjects enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP.
- Subject-level outcome = ACR50 at Week 16 while on randomized treatment.

- IE handling = A ”while on randomized treatment” strategy will be implemented.
- Population-level summary measure = Conditional OR comparing BKZ to placebo.

The same analysis model as used for the primary efficacy analysis will be conducted on the RS.

8.1.4.7 Tipping point analysis

The degree of the departure from the MAR assumption to overturn conclusions from the primary analysis will be investigated in a tipping point analysis. This analysis will be performed on the monotone missing data and only if the primary analysis is significant at α=0.05. Intermittent missing data will be imputed using the MCMC method. In this analysis, it will be assumed that subjects who have missing data and are randomized to BKZ have a lower probability of response compared to subjects who have missing data and are randomized to placebo.

The analysis will be primarily conducted on the binary response endpoint itself in considering the worst-case scenario that subjects who have missing ACR50 response will be set as non-responders if they are randomized to BKZ and as responders if they are randomized to Placebo. If the result tips, then the tipping point analysis will be reconducted on the ACR50 endpoint after having delta-adjusted each individual component of the ACR50 endpoint. For both analyses, the same model as used for the primary efficacy analysis will be conducted on the RS. If the delta parameter value for which the study conclusion under MAR is reversed is plausible, the conclusion under MAR will then be questionable. Refer to Section 4.2.2.4 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, 2 sensitivity analyses will be conducted using the same analysis method as the primary analysis:

- on the COVID-19-free Set
- by COVID-19 period (pre or during as determined by the Week 16 date).

8.2 Statistical analysis of secondary efficacy variables

8.2.1 Derivation of secondary efficacy variables

The primary and supportive analysis methods of secondary efficacy variables for which the derivation method is provided below are described in Section 8.2.2 and 8.2.3. For the other secondary variables (which are not part of the testing hierarchy), there is no priority designated between methods.

8.2.1.1 HAQ-DI: Change from Baseline at Week 16

The derivation for HAQ-DI is described in Section 8.1.1.3 (The analysis of this variable is already covered in Section 8.1.4.3).

Change from Baseline at Week 16 in HAQ-DI is the 2nd endpoint in the sequential testing hierarchy (Table 4–5).

8.2.1.2 Psoriasis Area Severity Index 90 (PASI90) at Week 16 (in subjects with PSO involving at least 3% of BSA at Baseline)

PASI scoring of psoriatic plaques is based on 3 criteria:

- redness (R)
- thickness (T)
- scaliness (S).

Severity is rated for each index (R, S, T) on a 0-4 scale (0=none, 1=slight, 2=moderate, 3=marked, and 4=very marked).

The body is divided into 4 areas comprising:

- head (h)
- upper extremities (u)
- trunk (t)
- lower extremities (l).

In each of these areas, the fraction of total surface area affected is graded on a 0-6 scale (where 0=0% [clear], 1=>0% to <10%, 2=10% to <30%, 3=30% to <50%, 4=50% to <70%, 5=70% to <90%, and 6=90% to 100%).

The various body regions are weighted to reflect their respective proportion of BSA. The composite PASI score is then calculated by multiplying the sum of the individual-severity scores for each area by the weighted area-of-involvement score for that respective area, and then summing the 4 resulting quantities as follows:

$$\begin{aligned} PASI = & (0.1 \times (R_h + T_h + S_h) \times A_h) \\ & + (0.2 \times (R_u + T_u + S_u) \times A_u) \\ & + (0.3 \times (R_t + T_t + S_t) \times A_t) \\ & + (0.4 \times (R_l + T_l + S_l) \times A_l) \end{aligned}$$

Where:

R_h, R_u, R_t, R_l = redness score of plaques on the head, upper extremities, trunk, and lower extremities, scored 0-4 respectively

T_h, T_u, T_t, T_l = thickness score of plaques on the head, upper extremities, trunk, and lower extremities, scored 0-4 respectively

S_h, S_u, S_t, S_l = scaliness score of plaques on the head, upper extremities, trunk, and lower extremities, scored 0-4 respectively

A_h, A_u, A_t, A_l = numerical value translation of % area of psoriatic involvement score for the head, upper extremities, trunk, and lower extremities, scored 0-6 respectively.

The highest potential PASI score is 72 for severe disease and the lowest is 0 for no psoriasis lesions. PASI scores are treated as continuous.

If a subject is missing 1 or 2 severity measurements for a certain region, the average of the remaining non-missing severity measurement(s) within that region will be utilized to substitute for the missing severity measurement(s) in that region. If the area of affected skin and/or all severity measurements for up to 2 regions are missing, then the missing (R+T+S) x A for a region will be substituted by the average of the available (R+T+S) x A. Otherwise, the PASI will be set to missing.

The percent improvement in PASI scores from Baseline will be computed as:

$$\text{Percent improvement from Baseline} = 100 \times \frac{\text{Baseline PASI} - \text{PostBaseline PASI}}{\text{Baseline PASI}}$$

If a subject has experienced an improvement, this measure will be positive. If a subject has experienced a worsening in their condition, this measure will be negative.

PASI90 is a binary variable equal to 1 if the percentage improvement from Baseline in the PASI scores is ≥ 90% and 0 if the percentage improvement from Baseline is < 90%.

PASI90 is the 3rd endpoint in the sequential testing hierarchy (Table 4–5).

8.2.1.3 SF-36 PCS: Change from Baseline at Week 16

The SF-36 (Version 2, standard recall) is a 36-item generic health-related quality of life instrument that uses a recall period of 4 weeks. Items are grouped into 8 domains as follows:

- 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).

[redacted] The concepts represented by these domains contribute to physical, mental, and social aspects of health-related quality of life.

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

The norm-based T-scores for the 2 SF-36 component summary (PCS and MCS) and the 8 domains 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 US 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 of 47 or greater should be considered average or above average as compared to the general US population. Higher scores indicate a better health status.

For the calculation of the SF-36 norm-based T-scores for the 8 domains and the PCS and MCS, the scoring software Optum's PRO CoRE (version 1.4) 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 1 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 will then be applied to estimate the PCS and MCS on the basis of the available domains.

Change from Baseline at Week 16 in SF-36 PCS score is the 4th endpoint in the sequential testing hierarchy (Table 4–5).

8.2.1.4 MDA at Week 16

MDA is a state of disease activity deemed a useful target of treatment by both the subject and physician, given current treatment possibilities and limitations. The criteria that need to be met are based on key outcome measures in PsA and are shown below:

- Tender joint count ≤1
- Swollen joint count ≤1
- PASI ≤1(for subjects with PSO involving at least 3% of BSA at Baseline) or BSA ≤3
- PtAAP (VAS) ≤15 mm
- PGA-PsA (VAS) ≤20 mm
- HAQ-DI ≤0.5
- Tender enthesial points ≤1 (LEI≤1).

A subject has achieved MDA if 5 or more of the 7 above criteria are fulfilled.

The enthesitis instrument of the study (ie: LEI) will be used to determine whether a subject has ≤1 tender enthesial point.

The following rule will be applied for subjects with BSA <3 at Baseline: Subjects with BSA <3 at Baseline will always meet the criteria PASI ≤1 or BSA ≤3 except in the cases where a BSA score >3 is observed.

MDA at Week 16 is the 5th endpoint in the sequential testing hierarchy (Table 4–5).

8.2.1.5 Van der Heijde modified Total Sharp Score (vdHmTSS): Change from Baseline at Week 16

The degree of joint damage is to be assessed using the vdHmTSS as used in the evaluation of PsA (Van der Heijde et al, 2005). This methodology quantifies the extent of joint erosions (52 joints) and joint space narrowing (52 joints), with higher scores representing greater damage.

Radiographs of the hands and feet are to be taken using standardized imaging methodology, which is defined in the Hand and Foot Radiography Manual that will be provided as part of the Study Manual.

All enrolled subjects are required to have radiographs taken of both hands and both feet (a single posteroanterior view of each hand and a single dorsoplantar of each foot). Radiographs are to be read centrally and independently by 2 experienced readers. If the 2 primary readers disagree, adjudication will be performed by a third reviewer. Readers will be blind to treatment assignment and time course of the films. Additional details on radiograph reading, assessment, and adjudication should be given in the imaging charter.

The scoring methodology is outlined below.

Scoring of Erosions:

Bone erosions in the hands and feet will be scored in the following locations according to the following grading scheme:

Erosion Scoring Locations:

- 20 locations in each hand and wrist (left and right): 40 locations overall:
  - 4 Distal inter-phalangeal joints (2-5)
  - 4 Proximal Inter-Phalangeal Joints (2-5)
  - 5 Metacarpo-Phalangeal Joints (1-5)
  - Inter-Phalangeal Joint of the thumb
  - Proximal first Metacarpal Bone
  - Radius Bone
  - Ulnar Bone
  - Trapezium and Trapezoid (as one unit; multangular)
  - Navicular Bone
  - Lunate Bone.
- 6 joints in each foot (left and right), 12 locations overall for the 2 feet:
  - 5 Metatarso-phalangeal joints (1-5)
  - Inter-phalangeal joint of the first toe.

Erosion Grading Scheme:

- Scores will be based on the number and size of discrete erosions in each location as follows:
  - 0=Normal (no erosion)
  - 1=Discrete small erosion
  - 2=Either 2 discrete erosions or 1 larger erosion that do not cross the joint middle line

- 3=Combination of the above or 1 large erosion that cross the joint middle line of the bone
- 4=Combination of above adding up to 4
- 5=Combination of above adding up to 5 or more.

The maximum score for erosion is 5 per joint in the hand and 10 per joint in the feet.

A score of "N" will be assigned in the case that the location is non-evaluable due to advanced subluxation, interfering osteoarthritis, or poor radiographic depiction. A score of "S" will be assigned in the case that the location had joint replacement, amputation, or other surgery. Exported score will be from 0 to 5 for hand and 0 to 10 for foot. Score of "S" and "N" will not be exported as a score but in an additional text field ("Reason").

Scoring of Joint Space Narrowing:

Joint space narrowing (JSN) in the hands and feet will be scored in the following locations according to the following grading scheme:

Joint Space Narrowing Scoring Locations:

- 20 locations in each hand and wrist (40 locations overall):
  - 4 Distal inter-phalangeal joints (2-5)
  - 5 Metacarpo-phalangeal joints (1-5)
  - 4 Proximal inter-phalangeal joints (2-5)
  - Interphalangeal Joint of thumb (IP)
  - 3 Carpo-metacarpal joints (3-5)
  - Radio-carpal joint
  - Multangular-navicular joint
  - Capitate-navicular-lunate joint.
- 6 locations in each foot (12 locations overall):
  - 5 Metatarso-phalangeal joints
  - Inter-phalangeal joint of first toe.

Joint Space Narrowing Grading Scheme:

- The joints mentioned above will be scored using a 5-point scale from 0 to 4 as follows:
  - 0=normal
  - 1=asymmetrical or minimal narrowing up to a maximum of 25%
  - 2=definite narrowing with loss of up to 50% of the normal space
  - 3=definite narrowing with loss of 50–99% of the normal space or subluxation
  - 4=absence of a joint space, presumptive evidence of ankylosis, or complete luxation.

A score of ”N” will be assigned in the case that the location is non-evaluable due to advanced subluxation, interfering osteoarthritis, or poor radiographic depiction. A score of “S” will be assigned in the case that the location had joint replacement, amputation, or other surgery. Exported score will be from 0 to 5 for hand and 0 to 10 for foot. Score of “S” and “N” will not be exported as a score but in an additional text field (“Reason”).

Total Scores:

The individual joint scores will be summed separately to create a total erosion score (ES) and a total JSN score for the hands/wrists.

- The maximum total ES for both hands/wrists is 200 (20 locations × 5 maximum per joint × 2).
- The maximum total JSN score for both hands/wrists is 160 (20 locations × 4 maximum per joint × 2.

In the feet as for the hands/wrists the individual joint scores will be summed separately to create a total ES and a total JSN score for the feet.

- The maximum total ES for both feet is 120 (12 locations × 10 per joint)
- The maximum total JSN score for both feet is 48 (6 locations × 4 maximum per joint × 2).

Thus, the maximum possible scores are:

- 320 for erosions
- 208 for joint space narrowing
- 528 for the total score.

The total score is the sum of the ES and JSN score.

Radiographic assessments should be conducted using a standardized imaging methodology as defined in an imaging charter. Subjects reaching Week 52 will have 2 sets of reads for Baseline and Week 16 as x-rays for these subjects are read more than once in different sessions to reduce within subject variability by each of the independent reviewers.

Bone erosions will be read for all visits at both at the Week 24 analysis and the Week 52/final analyses to ensure that unnecessary bias is not introduced. Given there may be differences at a visit for the two sets of reads, the following rules will be applied for the derivation:

- For the analysis involving data up to Week 16 only (Week 24 analysis), the first set of reads will be used for all subjects.
- For the analyses involving data up to Week 52, the second set of reads will be used for all subjects.

For the analysis, the average of the scores (ES, JSN score and total score) from the 2 independent reviewers will be utilized. They will be calculated as:

- (Score from Reviewer 1 + Score from Reviewer 2)/2

When a third reviewer will perform adjudication, the following average will be used for the scores:

- (Sum of the adjudicated score + score closest to the adjudicated score)/2

If the 2 scores are equidistant to the adjudicated score, then the highest score will be taken.

For the assessment of an individual reviewer, the following imputation rules will be applied:

- If at least 1 score for a particular joint used to calculate JSN score or ES at a visit is available, the remaining score will be imputed:
  - For Baseline, missing scores for a particular joint will be set to 0
  - At post-Baseline visits, missing score for a particular joint (ie, the assigned score is "N" or "S"), the joint score from the previous visit will be carried forward to compute the total score.
- If the total score is missing at a post-Baseline visit (ie, only "N" or "S" is recorded for all joints) or if JSN or ES score is missing then the missing total score will be imputed using MI imputation.

For the change from Baseline analyses, subjects who withdraw early and have radiographs taken at their early withdrawal visit should be included in the analysis. In such cases, one or more imputation methods may be applied for the subjects with less than 2 radiographs. In addition to any such imputation approach that is selected, an analysis excluding subjects with less than 2 radiographs from the analysis should be performed.

In the testing hierarchy (Table 4–5), 2 endpoints are defined based on the vdHmTSS:

- Change at Week 16 from Baseline in vdHmTSS for subjects with elevated hs-CRP and/or with at least one bone erosion at Baseline (planned to be approximately 45% of the overall population). This is the 6th endpoint in the sequential testing hierarchy.
- Change at Week 16 from Baseline in vdHmTSS for all subjects. This is the 9th endpoint in the sequential testing hierarchy.

The 2 endpoints will be analyzed in the radiographic set.

At Week 16, imaging data collected outside ±14 days of the expected visit date will be excluded from the analysis.

For the Week 24 interim analysis only, in the primary analysis of these 2 endpoints, missing data at Week 16 will be imputed based on value at Screening (as there is no value collected in between) using MI-MCMC / Monotone regression. Since VdHmTSS is only collected at Baseline prior to Week 16, only the second MI step to address monotone missing pattern and generating 100 imputations will be processed (Section 4.2.2.2).

8.2.1.6 Enthesitis-free state at Week 16 in the subgroup of subjects with enthesitis at Baseline

The Leeds Enthesitis Index (LEI) consists of 6 items, 3 for the right part and 3 for the left part of the body. The body parts are Lateral epicondyles of the humerus, Medial femoral condyle, and

Achilles tendons. Each item is scored for tenderness with either 1 or 0 (tenderness, no tenderness, respectively). The total score ranges from 0 (no enthesitis) to 6 (severe enthesitis).

Note that LEI is collected for all subjects regardless of their Baseline score.

In case 4 items are missing, the LEI will be set to missing. If at least 3 items are available, the available items will be assessed and scored and then weighted by the number of the assessed items (AI) as shown in the formula below:

$$LEI = n \times \frac{\sum_{i=1}^{n} score}{\sum_{i=1}^{n} AI}$$

Note that n here refers to the number of total items.

Enthesitis-free state is based on a LEI response of 0 at Week 16, for subjects with enthesitis at Baseline (LEI>0).

Enthesitis-free state at Week 16 (pooled between PA0010 and PA0011) is the 7th endpoint in the sequential testing hierarchy (Table 4–5) and will be based on pooled data from PA0010 and PA0011 (for subjects in the BKZ160mg Q4W and the Placebo groups). The analysis will also be repeated using PA0010 data alone and the p-value for this analysis will only be considered nominal.

8.2.1.7 Dactylitis-free state at Week 16 in the subgroup of subjects with dactylitis at Baseline

The LDI measures the percent difference between the circumference of the affected digit and the circumference of the digit on the opposite hand or foot (referred to as the contralateral digit), where circumference is measured in millimeters. A minimum difference of 10% and the assessment of the investigator that the digit was affected by dactylitis will be used to define a dactylitis digit.

The percent difference between circumferences will be multiplied by a tenderness score (0 for non-tender, 1 for tender). The score using the binary tenderness score is referred to as the LDI basic. The results from each digit with dactylitis will then be summed to produce a final LDI score.

Table 8–3 is a table of normative values, which will be used to provide the comparison, if matching digits are thought to be involved.

Table 8–3: Normative values for LDI

	Digit	Men	Women
Hand	Thumb	70	58
	Index	63	54
	Middle	63	54
	Ring	59	50
	Little	52	44
Foot	Great toe	82	72

	Digit	Men	Women
	Second	52	46
	Middle	50	44
	Fourth	50	44
	Little	52	45

The following rules will be applied for the LDI calculation in case of unclear data.

- Circumferences of 0 will be considered as missing. Circumferences <15 will be assumed to be in cm instead of mm and will be multiplied by 10 before being used in summaries/analyses.
- If both digits of a given pair are recorded as affected, then each digit will be compared to the normative value (shown in the above table). Note that if both comparisons result in a difference ≥10%, then both digits will contribute to the final LDI score.
- If the circumference of the affected digit is smaller than the unaffected digit, then the LDI will be calculated by comparing the smaller digit to the normative value.
- If a digit is recorded as affected and the circumference of the contralateral digit is missing, the normative value will be used for comparison with the affected digit.
- If a digit is selected but recorded as 'not affected' but has circumference and contralateral circumferences collected, a missing tenderness score will be considered as 1.
- If the investigator did not complete the LDI questionnaire at Baseline, it will be assumed that no digits were affected (ie, a LDI score of 0 will be assigned).

Dactylitis-free state is based on a Leeds Dactylitis Index (LDI) at Week 16 of 0, for subjects with dactylitis at Baseline (LDI>0).

Dactylitis-free state at Week 16 (pooled between PA0010 and PA0011) is the 8th endpoint in the sequential testing hierarchy (Table 4–5) and will be based on pooled data from PA0010 and PA0011 (for subjects in the BKZ160mg Q4W and the Placebo groups). The analysis will also be repeated using PA0010 data alone and the p-value will only be considered nominal.

A tender dactylitis count will be calculated as the sum of the tenderness scores for all digits and will be used for the PASDAS score calculation (Section 8.3.1.6).

8.2.1.8 Psoriasis Area Severity index 90 (PASI90) at Week 4 (in subjects with PSO involving at least 3% of BSA at Baseline)

This variable will be derived as described in Section 8.2.1.2.

8.2.1.9 ACR20 and ACR70 response at Week 16

The ACR20 and ACR70 responses are defined based on respectively 20% and 70% or greater improvement from Baseline in the same measures than the primary endpoint (Section 8.1.1).

8.2.1.10 Proportion of subjects with an Investigator Global Assessment (IGA) score of 0 (clear) or 1 (almost clear) AND at least a 2-grade reduction from Baseline at Week 4 and Week 16 in the subset of subjects with psoriatic skin lesions at Baseline

A static IGA for PSO will be used to assess disease severity in all subjects during the study. The Investigator will assess the overall severity of PSO using the following 5-point scale presented in Table 8–4.

Table 8–4: Five-point IGA

Score	Short Descriptor	Detailed Descriptor
0	Clear	No signs of PSO; post-inflammatory hyperpigmentation may be present
1	Almost Clear	No thickening; normal to pink coloration; no to minimal focal scaling
2	Mild	Just detectable to mild thickening; pink to light red coloration; predominantly fine scaling
3	Moderate	Clearly distinguishable to moderate thickening; dull to bright red, moderate scaling
4	Severe	Severe thickening with hard edges; bright to deep dark red coloration severe/coarse scaling covering almost all or all lesions

IGA=Investigator's Global Assessment; PSO=psoriasis

The IGA response is defined as clear [0] or almost clear [1] assessment with at least a 2-category improvement from Baseline, meaning that this parameter will be only evaluable for subjects with psoriatic skin lesions (IGA score ≥2) at Baseline.

Note that per protocol and device setting, only subjects with BSA ≥ 3% at Baseline will have IGA assessed at post-Baseline visits.

For some subjects with BSA ≥ 3% at Baseline, IGA data have not been captured at Week 4 and Week 16 visits when BSA<3% at those visits. To address that, an additional supportive analysis will be performed in which these subjects will be considered as IGA responders at the respective visit.

8.2.1.11 Change from Baseline in the Patient's Assessment of Arthritis Pain (PtAAP) at Week 16

The change from Baseline at Week 16 in PtAAP will be analyzed as described in Section 8.1.4.4.

8.2.1.12 Enthesitis-Free state based on Spondyloarthritis Research Consortium of Canada (SPARCC) index at Week 16

The SPARCC index measures the severity of enthesitis through the assessment of 16 sites, 8 for the right part and 8 for the left part of the body: the greater trochanter (right/left), quadriceps tendon insertion into the patella (right/left), patellar ligament insertion into the patella and tibial tuberosity (right/left), achilles tendon insertion (right/left), plantar fascia insertion (right/left), medial and lateral epicondyles (right/left), and the supraspinatus insertion (right/left). Tenderness on examination is recorded as either present (1) or absent (0) for each of the 16 sites, for an overall score range of 0 (no enthesitis) to 16 (severe enthesitis).

In case 9 or more are missing, the SPARCC index will be set to missing. If at least 8 items are available, the available items will be assessed and scored and then weighted by the number of the assessed items as shown in the formula below:

$$SPARCC\ index = n \times \frac{\sum_{i=1}^{n} score}{\sum_{i=1}^{n} AI}$$

where *n* refers to the number of total items and AI the number of assessed items.

The analysis of enthesitis-free state based on SPARCC index at Week 16 from Baseline will be restricted to the subset of subjects with enthesitis at Baseline (defined as SPARCC>0).

8.2.1.13 Change from Baseline in PsAID-12 total score at Week 16

The PsAID questionnaires measure the impact of disease on the subject. The impact of PsA is measured by weighting 12 different domains of health to derive a weighted summary score. The domains are pain, fatigue, skin problems, work and/or leisure activities, functional capacity, discomfort, sleep disturbance, coping, anxiety/fear/uncertainty, embarrassment and/or shame, social participation, and depression. The PsAID score is based on the relative importance of each of the domains. Each domain is answered based on a numeric rating scale ranging from 0 to 10. The equation for the PsAID-12 score is as follows:

$$\begin{aligned} PsAID - 12 = \frac{1}{20}(&3 \times Pain + 2 \times Fatigue + 2 \times Skin\ Problems + 2 \times Work/Leisure \\ &+ 2 \times Functional\ Capacity + 2 \times Discomfort + 2 \times Sleep\ Disturbance \\ &+ Coping + Anxiety/Fear + Embarrassment/Shame \\ &+ Social\ Participation + Depression) \end{aligned}$$

The total score also ranges from 0 to 10 with higher scores indicating a worse impact of the disease. A score below 4 out of 10 is considered a subject-acceptable status. A change of 3 or more points is considered relevant absolute change.

If a value (from the 12 domains) is missing, the missing value will be imputed in taking the mean value of the remaining (non-missing) domains, and the imputed value will be used to calculate the PsAID with the formula above. If 2 or more of the domains are missing, the PsAID score will be set to missing.

8.2.2 Primary analysis of secondary efficacy variables

The secondary efficacy variables will be analyzed for all subjects in the RS by treatment group (except for vdHmTSS for which the analysis will be performed on the RAS).

As indicated in Table 4–5, there are 8 secondary endpoints included in the testing hierarchy that will be tested in a chronological order until one is failed to be statistically significant.

For the other secondary endpoints (secondary endpoints which are not part of the testing hierarchy), the p-values produced by the statistical models will be considered nominal since these endpoints are not controlled for multiplicity.

For the secondary composite and non-composite binary endpoints, the same estimand structure (composite estimand) as the one defined for the primary efficacy analysis of the primary efficacy endpoint (Section 8.1.2) will be used. The NRI approach for handling missing data and the same analysis model will be considered, and the analysis results will be presented similarly, with the exceptions that the analyses performed on pooled data from PA0010 and PA0011 subjects will consider the region and the study-id factor as possible covariates for the modelling.

The statistical hypothesis for the binary variables at Week 16 is that the conditional odds ratio in the BKZ group compared with placebo treatment group is equal to 1.

For the secondary continuous endpoints, the analysis will evaluate the hypothetical estimand as defined below:

- Population = Subjects enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP.
- Subject level outcome = variable as stated in Section 2.2.1.2
- IE handling = An IE is defined as discontinuation of study treatment prior to Week 16. A hypothetical strategy for addressing IE will be implemented. This estimand targets the treatment difference in a scenario where withdrawal from study treatment does not occur, such that outcomes for subjects without an IE are as observed, and outcomes for subjects with an IE are treated as though they had completed treatment through Week 16 but on Placebo. A MI strategy will be used to impute any missing data and observed data after IE which will be set to missing prior to running MI. Such data 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 BKZ 160mg Q4W and placebo.

The same analysis model and imputation strategy for handling missing data as in Section 8.1.4.4 will also be considered. The analysis results will be presented similarly as for this analysis on the individual ACR components.

For vdHmTSS, a similar estimand structure as the one defined above for continuous endpoints will be used with the following exception for population:

- When analyzing change at Week 16 from Baseline in vdHmTSS for subjects with elevated hs-CRP or with at least one bone erosion at Baseline: Population=Subjects enrolled according to the protocol-specified inclusion/exclusion criteria, randomized to IMP, who have a valid radiographic image of the hands and feet at Baseline in the subgroup of patients with elevated CRP and with at least one bone erosion at Baseline.

- When analyzing change at Week 16 from Baseline in vdHmTSS for all subjects: Population=Subjects enrolled according to the protocol-specified inclusion/exclusion criteria, randomized to IMP, who have a valid radiographic image of the hands and feet at Baseline.

In the case of MDA as the response is based on 7 different component scores, it is necessary to consider various data scenarios that could impact the calculation of response. The rules described here are applicable in the context of the calculation of MDA 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 MDA response).

The following rules will be applied to complete the derivation of MDA response based on the composite estimand definition:

- If a given visit has been preceded by an IE (treatment discontinuation):
  - The endpoint at all subsequent visits (whether the data were observed or not) will be set to "non-response" as the subject has not met the criteria for response.
- If a given visit has not been preceded by an IE:
  - If a subject is satisfying at least 5 of the 7 MDA criteria, the subject will be considered as MDA responder
  - If a subject is not satisfying at least 3 MDA criteria, the subject will be considered as MDA non-responder
  - In all other cases, the NRI approach will be applied.

An overview table will combine the results of the primary analysis for the primary and secondary efficacy endpoints included in the testing hierarchy.

For the country specific analyses performed on subjects randomized in Japan, the region factor will not be considered as a covariate for the modelling.

Considering that the number of subjects randomized in Japan is low (less of 10% of the Randomized Set), statistical models might not converge. If a model (Logistic Model or Mixed model) is not converging, all related adjusted statistics and p-value will not be presented, "NE" for "Not Evaluable" will be displayed instead.

8.2.3 Supportive analyses of the secondary efficacy variables

As mentioned in Table 4–2, the supportive analyses for all secondary variables will be performed using the modified composite estimand for binary variables (using MI-MCMC/monotone regression as in Section 8.1.4.3), the hypothetical estimand for continuous variables (using MI-MCMC/monotone regression as in Section 8.1.4.4) and OC.

For PASI90, the MI will be run on the PASI score on subjects involving at least 3% of BSA at Baseline.

For the MDA (and VLDA) MI analysis, subjects with BSA <3 at Baseline will always meet the criteria PASI ≤1 or BSA ≤3 except in the cases where a BSA score >3 is observed. Subjects involving at least 3% of BSA at Baseline will have their BSA values imputed using MI when deriving MDA (VLDA).

For enthesitis and dactylitis free-state based on the pooled data of PA0010 and PA0011 subjects (in the BKZ160mg Q4W and the Placebo group), the MI will be performed on the pooled dataset (using data from Baseline, Week 4, 8, 12 and 16) the same way as described in Section 4.2.2.2:

- with the following covariates in that order for each SAS® PROC MI: region (North America, Western Europe and Eastern Europe) and study-id ("1" for PA0011 and "0" for PA0010)
- For MI on LDI : on subjects with LDI>0 at Baseline
- For MI on LEI : on all subjects (but the results will be reported for subjects with LEI>0 at Baseline).

In addition, subgroup analyses will be performed on the PASI90 response at Week 16 (in subjects with PSO involving at least 3% of BSA at Baseline) (Refer to Section 4.8 for the list of subgroups of interest). The same analysis method used for the subgroup analysis of the primary endpoint will be performed, and the results will be presented similarly.

Enthesitis and dactylitis free state will be analyzed as supportive analyses in the PA0010 data only (successively using the NRI, the modified composite estimand and the OC strategies) .

To assess the impact of the COVID-19 pandemic, the analysis of secondary efficacy endpoints included in the hierarchy with be repeated on the COVID-19-Free Set (on subjects from the Radiographic Set included in the COVID-19-Free-Set for the analysis of VdHmTSS) using the primary analysis method.

8.3 Statistical analysis of other efficacy variables

8.3.1 Derivation of other efficacy variables

8.3.1.1 Time to ACR20, ACR50, ACR70 response from Baseline

Time to a given response will be defined as the length in days from Baseline until the first date when the response is achieved.

Time from Baseline to censoring will be considered for the following subjects:

- Subjects who discontinue study treatment prior to achieving a response will be censored at the date of study treatment discontinuation.
- Subjects in the placebo group who fail to achieve a response prior to Week 16 visit will be censored at the date of the Week 16 visit.
- Subjects who reach the end of treatment without achieving the given response will be censored at the date of the end of treatment visit (for the Week 24 analysis, subjects who reached Week 24 without achieving the given response will be censored at that visit).

8.3.1.2 ACR20, ACR50 and ACR70

Method for derivation of ACR50 can be found in Section 8.1.1. ACR20 and ACR70 will be derived similarly except that they will be based on a 20% and on 70% improvement from Baseline.

ACR20, ACR50 and ACR70 responses will be analyzed at Week 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 44, 52.

For the placebo group, ACR50 response will be presented both based on the Day 1 Baseline and on the Week 16 Baseline (on the ATS, for the ATP, for subjects switching to BKZ at Week 16).

8.3.1.3 PASI75, PASI90 and PASI100 response (in subjects with PSO involving at least 3% of BSA at Baseline)

Method for derivation of PASI score can be found in Section 8.2.1.2.

PASI75, PASI90, PASI100 are binary variables equal to 1 if the percentage improvement from Baseline in the PASI scores is ≥ 75%, ≥90%, 100% and 0 if the percentage improvement from Baseline is < 75%, 90%, 100%.

PASI75, PASI90, PASI100 as well as the PASI score and the change from Baseline for subjects with PSO involving at least 3% of BSA at Baseline will be analyzed at Week 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 52.

For the placebo group, PASI90 responses will be determined both based on the Day 0 Baseline and on Week 16 Baseline (on the ATS, for the ATP, for subjects switching to BKZ at Week 16).

8.3.1.4 Composite endpoints composed of ACR50 and PASI90 and ACR50 and PASI100 responses (subjects with PSO involving at least 3% of BSA at Baseline)

The composite endpoint composed of ACR50 and PASI90 is a binary variable equal to 1 for subjects responding to ACR50 (ACR50=1) and PASI90 (PASI90=1).

The composite endpoint will be equal to 0 if ACR50=0 or PASI90=0 regardless of whether any of the 2 endpoints is missing.

The same rules as above will be applicable for the composite endpoint on ACR50 and PASI100.

The composite endpoints will be missing if both endpoints are missing or if any of the 2 is missing and the other one equal to 1.

Each endpoint (ACR50 and PASI90 or PASI100) will be imputed separately before deriving the composite endpoint.

The composite endpoints for subjects with PSO involving at least 3% BSA at Baseline will be analyzed at Week 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 52.

8.3.1.5 Proportion of Psoriatic Arthritis Response Criteria (PsARC) responders

The PsARC is based on the TJC (68 joints) and SJC (66 joints), the PGA-PsA and PhGA-PsA. The PsARC response is defined as improvement from Baseline in at least 2 of the 4 measures (TJC, SJC, PGA-PsA, PhGA-PsA) 1 of which must be TJC or SJC and no worsening from Baseline in any of the 4 measures:

- Improvement for TJC and SJC is defined as a reduction of ≥30%. Improvement for PGA-PsA and PhGA-PsA is defined as a reduction of the 100-point VAS of ≥20 points
- Worsening for TJC and SJC is defined as an increase of ≥30%. Worsening for PGA-PsA and PhGA-PsA is defined as an increase of the 100-point VAS of ≥20.

PsARC response will be analyzed at Week 2, 4, 8, 12, 16, 20, 24, 36, 52.

8.3.1.6 Psoriatic Arthritis Disease Activity Score (PASDAS) categories and change from baseline in PASDAS

The PASDAS is a composite score that includes patient and physician global scores of skin and joint disease, SJC, TJC, LEI, tender dactylitis count (sum of tenderness scores), the physical component of the SF-36 Health Survey, and level of hs-CRP. PASDAS is calculated using the following equation:

$$PASDAS = \Big(\big(0.18 \times \sqrt{\text{Physician global VAS}}\big) + \big(0.159 \times \sqrt{\text{Patient global VAS}}\big) - \big(0.253 \times \sqrt{\text{SF36 PCS}}\big) + \big(0.101 \times \text{LN}(\text{SJC} + 1)\big) + \big(0.048 \times \text{LN}(\text{TJC} + 1)\big) + \big(0.23 \times \text{LN}(\text{LEI} + 1)\big) + \big(0.377 \times \text{LN}(\text{Tender dactylitis count} + 1)\big) + (0.102 \text{ x LN}(\text{hs} - \text{CRP} + 1) + 2)\Big) * 1.5$$

Note: LN=natural logarithm

If Tender Dactylitis count is missing because LDI=0 at Baseline, Tender Dactylitis count will be replaced by 0 in the formula.

If for any other reason a PASDAS component is missing, PASDAS will be set to missing.

The following categories are used to define the level of disease activity:

- Remission: PASDAS ≤1.9
- Low disease activity: PASDAS >1.9 to <3.2
- Moderate disease activity: PASDAS 3.2 to <5.4
- High disease activity: PASDAS ≥5.4.

The PASDAS categories and the change from Baseline in PASDAS score will be analyzed at Week 4, 16, 24, 36, 52.

8.3.1.7 ACR50 responders at Week 16 and maintaining response at post-Week 16 visits up to Week 52

For subjects who have an observed ACR50 response at Week 16, the ACR50 responders at Week 52 will be presented.

This summary will be presented on the ATS by randomized treatment for responder subjects at Week 16.

ACR50 responses for subjects responding at Week 16 will be analyzed at Week 20, 24, 28, 32, 36, 44, 52.

8.3.1.8 MDA response

Method for derivation of the MDA response can be found in Section 8.2.1.4.

MDA response will be analyzed at Week 4, 8, 12, 16, 20, 24, 36, 52.

8.3.1.9 Very Low disease activity (VLDA) response

Like MDA, VLDA is a state of disease activity deemed a useful target of treatment by both the subject and physician, given current treatment possibilities and limitations.

A subject is considered as having VLDA if all 7 of the criteria used for MDA and listed in Section 8.2.1.4 are fulfilled.

If any of the 7 MDA criteria is not met, then the subject will be considered as non-responder.

VLDA response will be analyzed at Week 4, 8, 12, 16, 20, 24, 36, 52.

8.3.1.10 Proportion of Subjects with an IGA score of 0 (clear) or 1 (almost clear) AND at least a 2-grade reduction from Baseline in the subset of subjects with psoriatic skin lesions at Baseline

Method for derivation of this variable can be found in Section 8.2.1.10.

The proportion of subjects with an IGA score of 0 (clear) or 1 (almost clear) AND at least a 2-grade reduction from Baseline in the subset of subjects with psoriatic skin lesions at Baseline will be analyzed at Week 4, 8, 12, 16, 20, 24, 36, 52.

For some subjects with BSA ≥ 3% at Baseline, IGA data have not been captured at some post-Baseline visits when BSA<3% at those visits. To address that, an additional supportive analysis will be performed in which these subjects will be considered as IGA responders at the respective visit.

In the context of this analysis, observed data for % of BSA affected by PSO will be also analyzed by visit using descriptive statistics.

8.3.1.11 Disease Activity Index for Psoriatic Arthritis (DAPSA) score, state and change from Baseline in DAPSA score

Disease Activity Index for Psoriatic Arthritis (DAPSA) is a composite score of patient global and pain VAS, TJC, SJC and hs-CRP (mg/L) that incorporates a pattern of peripheral arthritis that often is encountered in PsA. DAPSA score will be calculated as follows:

$$DAPSA = SJC + TJC + PGA\text{–}Arthristis \quad + PtAAP + hs\text{–}CRP$$

where:

- SJC range from 0 to 66
- TJC ranges from 0 to 68
- PGA-Arthritis represents the patient's global assessment of arthritis ranging from 0 to 10 ; 0=best, 10=worst (using the unrounded VAS value that ranges from 0 to 100 mm and dividing this value by 10) (Section 8.3.1.28)
- PtAAP ranging from 0 to 10, 0=best, 10=worst, (unrounded, by dividing the original value by 10)
- hs-CRP in mg/L (no upper limit applied).

DAPSA values will be categorized in the following disease activity states:

- Remission (REM) (DAPSA≤4)
- Low Disease Activity (LDA) (DAPSA >4 and ≤14)
- Medium Disease Activity (MDA) (DAPSA >14 and ≤28)
- High Disease Activity (HDA) (DAPSA>28).

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

For the OC analyses, if any individual component score is missing, the DAPSA score will be set to missing.

DAPSA state, DAPSA scores and changes in scores from Baseline will be analyzed at Week 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 44, 52.

8.3.1.12 Change from Baseline in the Disease Activity Score-28 based on C-reactive protein (DAS28[CRP])

The components for DAS28[CRP] include:

- the TJC and SJC based on 28 joints (Section 8.1.1.1):
  - Upper extremity (26) - bilateral shoulders, elbows, wrists (includes radiocarpal, and carpal and carpometacarpal bones considered as a single unit), metacarpophalangeal joints I, II, III, IV, and V, thumb interphalangeals, proximal interphalangeal II, III, IV, and V.
  - Lower extremity (2) - knees.
- the PGA-Arthritis (ranging from 0 to 10 by dividing the original value by 10)
- the hs-CRP (in mg/L).

The DAS28[CRP] is calculated as follows:

$$DAS28[CRP] = 0.56 \times \sqrt{TJC} + 0.28 \times \sqrt{SJC} + 0.014 \times \text{PGA–Arthritis} + 0.36 \times \ln(hs\text{–}CRP + 1) + 0.96$$

For analyses with imputation of missing data, the imputed value for DAS28[CRP] will be based on the imputed values for the individual components.

For OC analyses, if any individual component score is missing, the DAS28[CRP] score will be set to missing.

The changes from Baseline in DAS28[CRP] will be analyzed at Week 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 44, 52.

8.3.1.13 Change from Baseline in all individual ACR core components

Individual ACR components (as listed in Section 8.1.1) are:

- SJC
- TJC
- HAQ-DI

- PtAAP
- PhGA-PsA
- PGA-PsA
- hs-CRP.

The change from Baseline (ratio from Baseline for hs-CRP) for each of the ACR components will be summarized at Week 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40 (for SJC, TJC and hs-CRP), 44, 48 (for hs-CRP only), 52.

8.3.1.14 Change from Baseline in vdHmTSS total score and in erosion and joint space narrowing (sub-scores)

Scoring of vdHmTSS can be found in Section 8.2.1.5.

The change from Baseline in vdHmTSS total score, erosion sub-score and joint space narrowing sub-score will be analyzed at Week 16, 52, on the radiographic set (successively on the overall population , on subjects with elevated hs-CRP and/or with at least one bone erosion at Baseline and on subjects without elevated hs-CRP and without bone erosion at Baseline).

At Week 16 and 52, imaging data collected outside ±14 days around the expected visit date will be excluded from the analysis.

8.3.1.15 Percentage of subjects with no radiographic joint damage progression (Change from Baseline vdHmTSS of ≤0.5)

The higher the vdHmTSS score is the more radiographic damage the Subject will have.

The absence of joint damage progression (ie, stabilization or improvement in radiographic damage from Baseline) is then considered is the change of score from Baseline is ≤0.5.

The percentage of subjects with no joint damage progression will be analyzed at Week 16 and 52 on the radiographic set (successively on the overall population and on subjects with elevated hs-CRP and/or with at least one bone erosion at Baseline).

8.3.1.16 Dactylitis-free state based on the LDI and change in LDI from Baseline in the subgroup of subjects with dactylitis at Baseline

Dactylitis-free state is based on a Leeds Dactylitis Index (LDI) of 0, for subjects with dactylitis at Baseline (LDI>0) (Section 8.2.1.7).

The percentage of subjects with dactylitis free-state and the change from Baseline in LDI (on the subgroup of subjects with dactylitis at Baseline) will be analyzed at Week 4, 8, 12, 16 on the pooled Sample of PA0010 and PA0011 subjects (for subjects in the BKZ 160mg Q4W and the Placebo groups) and at Week 4, 8,12, 16, 20, 24, 36, 52 on PA0010 subjects.

8.3.1.17 Enthesitis-free state based on the LEI and change in LEI from Baseline in the subgroup of subjects with enthesitis at Baseline

Enthesitis-free state is based on a Leeds Enthesitis Index (LEI) of 0, for subjects with enthesitis at Baseline (LEI>0) (Section 8.2.1.6).

The percentage of subjects with enthesitis-free state and the change from Baseline in LEI (on the subgroup of subjects with enthesitis at Baseline) will be analyzed at Week 4, 8, 12, 16 on the pooled Sample of PA0010 and PA0011 subjects (for subjects in the BKZ160mg Q4W and the Placebo groups) and at Week 4, 8,12, 16, 20, 24, 36, 52 on PA0010 subjects.

8.3.1.18 Enthesitis-free state based on the SPARCC index and change in SPARCC from Baseline in the subgroup of subjects with enthesitis at Baseline

Enthesitis-free state is based on a SPARCC index of 0, for subjects with enthesitis at Baseline (SPARCC>0) (Section 8.2.1.12).

The percentage of subjects with enthesitis-free state and the change from Baseline in the SPARCC index (on the subgroup of subjects with enthesitis at Baseline) will be analyzed at Week 4, 8, 12, 16 on the pooled Sample of PA0010 and PA0011 subjects (for subjects in the BKZ160mg Q4W and the Placebo group) and at Week 4, 8,12, 16, 20, 24, 36, 52 on PA0010 subjects.

8.3.1.19 Proportion of subjects with a decrease of HAQ-DI from Baseline of at least 0.35 in the subgroup of subjects with Baseline HAQ-DI≥0.35

The proportion of HAQ-DI responders (subjects with a decrease of HAQ-DI from Baseline of at least 0.35 for subjects with HAQ-DI≥0.35 at Baseline) will be analyzed at Week 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 44, 52.

In addition, subgroup analyses will be performed on the HAQ-DI response at Week 16. Refer to Section 4.8 for the list of subgroups of interest. The same analysis method used for the subgroup analysis of the primary endpoint will be performed and the results will be presented similarly.

8.3.1.20 Change from Baseline in the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) in the subgroup of subjects with axial involvement defined by a score of ≥4 at Baseline

The BASDAI measure the disease activity of ankylosing spondylitis. The BASDAI is a validated self-reported instrument which consists of 6 horizontal Numeric Rating Scales, each with 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. To give each symptom equal weighting, the average of the 2 scores relating to morning stiffness is taken. The resulting 0 to 50 sum score is divided by 5 to give a final BASDAI score between 0 and 10, with lower scores indicating lower disease activity.

The BASDAI is calculated as follows:

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

where Q1 – Q6 are the 6 questions from the BASDAI questionnaire.

If 1 of the 2 morning stiffness measurements (ie, questions: “How would you describe the overall level of morning stiffness you have had from the time you wake up?” and “How long does your

morning stiffness last from the time you wake up?") is missing, the other one will be used for the morning stiffness calculation.

If 1 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 1 major symptom is missing, the total BASDAI score will be set to missing.

Change from Baseline in BASDAI will be analyzed for subjects with axial involvement at Baseline (BASDAI≥4) at Week 4, 8, 12, 16, 24, 36, 52.

8.3.1.21 Change from Baseline in the modified Nail Psoriasis Severity Index (mNAPSI) score in the subgroup of subjects with psoriatic nail disease at Baseline

Subjects with psoriatic nail disease will have a target nail selected at the Baseline visit for evaluation using the mNAPSI. The nail selected should be the most affected nail observed at Baseline and should be the only one assessed throughout the study. The target nail will be scored (0 to 3) for onycholysis/oil drop dyschromia, nail plate crumbling, and pitting and will be scored (0 for "no" or 1 for "yes") for leukonychia, nail bed hyperkeratosis, splinter hemorrhages, and red spots in the lunula. The total score (also called the mNAPSI score) will then range from 0 to 13 with higher scores indicative of more severe nail disease.

If any of the 7 response items that contribute to mNAPSI is present, while other items are missing (ie partial mNAPSI data), then the missing items are assumed to be 0 for the mNAPSI calculation. In some cases, the data may be captured in such a way that only non-zero component scores are present in the database. Again, those components that are not present are simply assumed to be 0 for the mNAPSI calculation.

Change from Baseline in mNAPSI and mNAPSI resolution (mNAPSI=0 at post-baseline assessments) will be analyzed for subjects with psoriatic nail disease at Baseline (mNAPSI score>0) at Week 4, 8, 12, 16, 24, 36, 52.

8.3.1.22 Change from Baseline in PsAID-12 total score

Method for derivation of PsAID-12 can be found in Section 8.2.1.13.

The proportion of subjects achieving a total score of PsAID-12 ≤ 4 - Patient acceptable symptom state (PASS), the proportion of subjects achieving a decrease of 3 or more points from Baseline on the total score (for subjects with minimal score of 3 at Baseline), the change from Baseline in the total score and in the individual scores for each PsAID-12 domain, will be analyzed at Week 4, 16, 24, 36, 52 (the proportions of subjects achieving PASS will also be analyzed at Baseline).

8.3.1.23 Change from Baseline in the Psoriatic Arthritis Quality of Life (PsAQoL) total score

PsAQoL is a disease specific Health-Related Quality of Life questionnaire comprised of 20 items. Each item is scored as 0 for "not true" or 1 for "true", so that the score ranges from 0 to 20 with higher scores indicating worse health-related quality of life.

If 6 or less item responses are missing, the missing responses will be imputed with the mean of available responses to calculate a total score. If more than 6 items responses are missing, the total score will be left missing.

Change from Baseline in PsAQoL total score will be analyzed at Week 4, 16, 24, 36, 52.

8.3.1.24 Change from Baseline in the SF-36 PCS and MCS scores, as well as the 8 domain scores (Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional, and Mental Health)

Scoring instructions of SF-36 norm-based T-scores for the 8 domains and the SF-36 PCS and MCS are described in Section 8.2.1.3.

Change from Baseline in SF-36 individual domain scores, MCS and PCS scores will be analyzed at Week 4, 16, 24, 36, 52.

8.3.1.25 Change from Baseline in Functional Assessment of Chronic Illness Therapy (FACIT)—Fatigue subscale score

The FACIT-Fatigue is a 40-item measure that assesses self-reported fatigue and its impact upon daily activities and function over 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 is the only one 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 must be recoded, so that responses range from worst (0) to the best (4) outcome.

Items 1 to 6 and 9 to 13 are negatively framed questions for which the coding needs to be reversed before deriving the score.

The FACIT-Fatigue subscale score is obtained by summing up the responses of all reversed / non reversed items 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:

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

The minimum clinically important difference for FACIT-Fatigue subscale score in patients with PsA was determined to be a 4-point positive change.

Change from Baseline in FACIT-Fatigue subscale score, as well as the proportion of subjects with minimum clinically important difference for FACIT-Fatigue subscale score in subjects with FACIT-Fatigue subscale score ≤48 at Baseline, will be analyzed at Week 4, 16, 24, 36, 52.

8.3.1.26 Change from Baseline in Work Productivity and Activity Impairment Questionnaire–Specific Health Problem (WPAI-SHP) v2.0 adapted to PsA scores

The WPAI-SHP is a subject-reported questionnaire that assesses subject's employment status, work absenteeism, work impairment while working (presenteeism), 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:

Questions:

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

Scores:

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

$$100 \times \frac{Q2\ hours}{(Q2\ hours + Q4\ hours)}$$

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

$$100 \times \frac{Q5\ Score}{10}$$

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

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

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

$$100 \times \frac{Q6\ Score}{10}$$

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 psoriatic arthritis:
  - 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.

The WPAI-SHP 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.

A negative number will indicate a reduction in the score/improvement for subjects.

The change from Baseline score is derived as post Baseline score minus Baseline score. A negative change score indicates a reduction in the score/improvement for the subject.

The change from Baseline in scores will be analyzed at Week 16, 24, 52.

The analysis of WPAI-SHP scores will be performed on observed cases only.

8.3.1.27 Responses to the EuroQol-5 Dimensions-3 Level (EQ-5D-3L), change from Baseline in VAS score

The Euro-Quality of Life 5-Dimensions 3 Level version (EQ-5D-3L) health questionnaire provides a descriptive profile and a single index value for health status (using a VAS).

The instrument is composed of a 5-item health status measure (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) and a 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. These scores are referred as "utility scores".

The EQ-5D VAS records the respondent's self-rated health status on a vertical 20cm scale, 0 to 100 graduated (0=worst imaginable health status, 100=best imaginable health status).

The response in each of the 5 items of the EQ-5D-3L and the values and the changes in VAS from Baseline will be presented at Week 4, 16, 24, 36, 52.

8.3.1.28 Change from Baseline in Patient's Global Assessment of Arthritis (PGA-arthritis) and Physician's Global Assessment of Arthritis (PhGA-Arthritis)

The physician's global assessment of arthritis (PhGA-arthritis) and the patient's global assessment of arthritis (PGA-arthritis) are clinical outcomes based on visual analog scales (VAS) ranged from 0 to100 (in mm).

- PGA-Arthritis: Subjects will assess the impact of their arthritis pain in answering the following question "Considering all the ways your arthritis affects you, please mark a vertical line on the scale below to show how you are feeling today." The subject should be asked to consider their arthritis symptoms and functional capacity in their response to this question. Subjects will score PGA-Arthritis in using a VAS where 0 is "very good, no symptoms" and 100 is "very poor, severe symptoms."
- PhGA-Arthritis: The investigator will assess how the subject's overall arthritis appears at the time of the visit using a numerical rating scale where 0 is "very good" and 100 is "very poor". This assessment will be based through the following question: "The Patient's arthritis at this time is: "Please mark a vertical line on the scale below to assess the overall status of the subject's arthritis signs and symptoms and the functional capacity of the subject."

Values and changes from Baseline in PhGA-Arthritis and PGA-Arthritis will be analyzed at Week 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 44, 52.

8.3.2 Analyses of other efficacy endpoints

Unless specified above, all other efficacy variables will be analyzed for all subjects:

- In by-visit tables, for the following treatment groups: Placebo/BKZ160 mg Q4W, BKZ 160mg Q4W, ADA 40mg Q2W.
- on the RS, for the overall study period (until Week 24 for the Week 24 analysis).

8.3.2.1 Time to ACR20/50/70 response

The time to ACR20/50/70 response will be successively analyzed:

- For the DBP, on the 3 treatment groups (for the Week 24 analysis).
- For the overall study period, on the 3 treatment groups (for the Week 52 analysis) They will be estimated and presented using the Kaplan Meier (KM) product limit method for each of treatment groups.

The KM plots of time to ACR20,50,70 response will be presented by treatment group. The line will start at 0 and will increase over time, representing time to achieving the response (reversed KM estimates).

The median time to response, including the 2 sided 95% CI, will be calculated for each treatment group.

For the analysis on the DBP, the hazard ratio of BKZ vs. placebo as well as the corresponding 95% CI and p-value will be displayed (using SAS® PROC PHREG stratified by region and bone erosion at Baseline)

Comparisons of BKZ vs. placebo will be analyzed based on a log-rank test stratified by region and bone erosion at Baseline (0, ≥1) (considering that the same stratification factors have been retained for the primary efficacy analysis). This comparison will be performed using SAS® PROC LIFETEST on subjects from both treatment groups.

The time to ACR20/50/70 endpoints will be analyzed using OC data.

8.3.2.2 Other endpoints (excluding time to ACR20/50/70 response)

For the binary variables, the following estimand structure (composite estimand) will be defined:

- Population = Subjects enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP.
- Subject-level outcome = The given variable and time point being summarized.
- IE handling = An IE is defined as discontinuation of study treatment prior to the time point being summarized. A composite strategy will be implemented in which a positive clinical outcome is defined as achieving the given variable at the specified time point and not discontinuing study treatment through that time point.
- Population-level summary measure = Unadjusted proportion of responder.

Subjects who have an IE, which will be considered associated to a treatment failure, will be considered as non-responders. Any missing data will also be considered as non-responders (NRI approach). For categorical variables, the worst-category will be imputed similarly instead of non-response.

For the continuous variables, the following estimand (hypothetical estimand) will be defined:

- Population = Subjects enrolled according to the protocol-specified inclusion/exclusion criteria and randomized to IMP.
- Subject-level outcome = The given variable and time point being summarized.
- IE handling = An IE is defined as discontinuation of study treatment prior to the time point being summarized. A hypothetical strategy will be implemented in which outcomes for subjects without an IE are as observed at the given time point, and outcomes for subjects with an IE are treated as though they had completed the randomized study treatment through the time point being summarized. A MI strategy will be used to impute data following an IE.
- Population-level summary measure = Unadjusted mean.

The analysis of other continuous efficacy variables will be performed according to the above estimand, using MI-MCMC/ monotone regression as imputation method for missing data for all visits up to Week 24 for the Week 24 analysis and all visits up to Week 52 for the Week 52 analysis (similarly as in Section 8.2.2).

The other efficacy variables will also be summarized based on OC and the modified composite estimand (for binary variables) and summarized by visit for all treatment groups as detailed below:

For binary endpoints using:

- The number and proportion of responders using the NRI approach.
- The mean proportion of responders in the multiple imputed datasets and the 95% CI (modified composite estimand). For ACR50 response at post-Week 16 visits for subjects responding at Week 16, the same imputed data as generated in the context of the analysis of the primary efficacy endpoint will be used.
  - For enthesitis and dactylitis free-state based on the pooled data of PA0010 and PA0011 subjects (for subjects in the BKZ 160mg Q4W and the Placebo group), the MI will be performed on the pooled dataset (using data from Baseline, Week 4, 8, 12 and 16) the same way as described in Section 4.2.2.2, but with the following covariates in that order for each SAS® PROC MI: region (North America, Western Europe and Eastern Europe in that order) and study-id ("1" for PA0011 and "0" for PA0010).
  - For enthesitis based on LEI, the MI will be run on all subjects and results reported for subjects with LEI>0 at Baseline ; for other binary variables derived from continuous value(s), the MI will be run on the subset of subjects (if applicable) for which the variable will be reported (ie: for LDI and Tender dactylitis count in the context of PASDAS calculation, on subjects with LDI>0 at Baseline)
  - For ACR50 based on Week 16 Baseline, the MI will be run on Week 16 and post-week 16 values, on the ATBS for the Placebo group. Week 16 value will be considered as the Baseline value regardless of whether the Placebo subjects switched to BKZ on a later visit.
  - For PASI90 based on Week 16 Baseline, the MI will be run on W16 and post-week 16 values, on the ATBS for the Placebo group on subjects with PSO involving at least 3% of BSA at Week 16. Week 16 value will be considered as the Baseline value regardless of whether the Placebo subjects switched to BKZ on a later visit.
- The observed number and proportion of responders on subjects with observed data

For categorical endpoints (ie, EQ-5D-3L dimensions score, DAPSA state, PASDAS categories) using:

- The number and percentage by endpoint category using the NRI approach (as performed for the primary analysis of the primary endpoint with the exception that the NRI is replaced by the worst category imputation)

- The mean percentage and corresponding 95% CI by endpoint category in the multiple imputed datasets (modified composite estimand). Regardless of imputed values, the value of the categorical endpoint after IE are by default set to the worst category (ie: High Disease Activity for DAPSA state)
- The observed number and percentage by category on subjects with observed data.

For continuous endpoints, on absolute values and changes from Baseline (or ratio from Baseline for hs-CRP) using:

- The mean (with geometric mean for hs-CRP and its 95% CI), SE, median, minimum, and maximum in the multiple imputed datasets (hypothetical estimand) (except for WPAI-SHP scores).
  - For the change from baseline in LDI, LEI and SPARCC based on the pooled data of PA0010 and PA0011 subjects (for subjects in the BKZ 160mg Q4W and the Placebo group), the MI will be performed on the pooled dataset (using data from Baseline, Week 4, 8, 12 and 16) the same way as described in Section 4.2.2.2, but with the following covariates in that order for each SAS® PROC MI: region (North America, Western Europe and Eastern Europe in that order) and study-id ("1" for PA0011 and "0" for PA0010).
  - For LEI, the MI will be run on all subjects and results reported for subjects with LEI>0 at Baseline ; for other continuous variables derived from continuous value(s), the MI will be run on the subset of subjects (if applicable) for which the variable will be reported (ie: for LDI, on subjects with LDI>0 at Baseline)
  - For Tender dactylitis count in the context of PASDAS calculation, if the MI does not converge, a conservative approach will be used by first imputing a value of 1 for all joints with missing tenderness score.
- The observed number of observations (n), the mean (with the geometric mean for hs-CRP), SD, median, CV (geometric CV for hs-CRP), minimum, and maximum.

For ACR20, ACR50, ACR70, PASI75, PASI90, and PASI100 responders over time, a line plot by treatment group will be produced.

8.3.2.3 Additional statistical analysis for other efficacy endpoints

Although statistical testing is not planned for other efficacy variables as per protocol, for selected other efficacy variables, it is of interest to perform statistical testing and calculate inferential statistics. The associated p-values will be considered nominal and not controlled for multiplicity.

For continuous variables, the MI-MCMC/monotone regression approach will be applied for the imputation model on the change from Baseline (hypothetical estimand). The analysis model will be based on ANCOVA with fixed effect of treatment, region, bone erosion at Baseline and Baseline value as covariates.

For responder variables, the analysis will follow the NRI approach (composite estimand). The analysis model will be based on a logistic model with fixed effect for treatment and bone erosion at Baseline and region as stratification variables.

Below is a list of other secondary efficacy variables for which these nominal p-values (with the timepoints in parentheses) will be calculated.

The results of these inferential tests (LSM differences for continuous variables, OR for binary variables, 95% CI and p-value) will be presented in a single table:

For the following endpoints, the p-value provided (based on the change from Baseline for continuous variables) will be based on the comparison of BKZ 160mg Q4W and placebo groups using the same statistical methods for binary variables as used for the primary analysis of ACR50 (Section 8.1.2) and for the continuous variables as used for the analysis of the ACR components (Section 8.1.4.4):

- ACR20/50/70 (Week 2, 4, 8, 12, 16)
- PASI75/90/100 (Week 2, 4, 8, 12, 16)
- ACR50/PASI90 composite (Week 2, 4, 8, 12, 16)
- ACR50/PASI100 composite (Week 2, 4, 8, 12, 16)
- TJC (Week 2, 4, 8, 12, 16)
- SJC (Week 2, 4, 8, 12, 16)
- MDA (Week 4, 8, 12, 16)
- PtAAP (Week 2, 4, 8, 12, 16)
- PsAQoL (Week 4,16)
- PsAID-12 (Week 4, 16)
- FACIT-Fatigue subscale (Week 4, 16)
- WPAI-SHP (Week 16) (on observed data)
- VdHmTSS (Week 16)
- DAPSA (Low disease activity or better) “response” (Week 2, 4, 8, 12, 16)
- PsARC (Week 2, 4, 8, 12, 16).

For the following endpoints, 2 p-values will be provided one for the comparison of BKZ 160mg Q4W and Placebo / BKZ 160mg Q4W groups and one for the comparison of BKZ 160mg Q4W and ADA 40mg Q2W groups:

- ACR20/50/70 (Week 16, 24, 52)
- PASI75/90/100 (Week 16, 24, 52)
- ACR50/PASI90 composite (Week 16, 24, 52)
- ACR50/PASI100 composite (Week 16, 24, 52)
- MDA (Week 16, 24, 52)
- PsARC (Week 16, 24, 52).

9 PHARMACOKINETICS AND IMMUNOLOGICAL (IMMUNOGENICITY) ANALYSES

9.1 Pharmacokinetics

Pharmacokinetic variables will be analyzed for all subjects in the PK-PPS.

Only subjects treated with BKZ will be included in the PK analyses. The Baseline observation for the Placebo group will be the Week 16 pre-dose observation. PK variables will be summarized at each scheduled visit by treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W).

- BKZ concentrations will be summarized by treatment group, globally and on subjects randomized in Japan.

No imputation will be used for missing samples. However, if plasma concentration measurements are below the limit of quantification (BLQ), then for the calculation of the derived statistics, the result will be set to ½ of LLOQ (ie, 1/2*0.250 = 0.125 µg/mL). Descriptive statistics including number of values, geometric mean, its 95% CI, geometric coefficient of variation, mean, SD, median, minimum, and maximum. Geometric mean and its 95% CI, geometric CV, mean and SD will be calculated if at least ⅔ of the values of interest is above the LLOQ and number of values ≥3 at the respective timepoint; otherwise, only number of values, median, minimum, and maximum will be presented.

- In addition, geometric mean of BKZ plasma concentration (with 95% CI) time curves will be plotted on linear and semi-logarithmic scales:
  - by treatment group (2 lines per graph, one line for each treatment),
- The table summary and figures will be primarily repeated by anti-BKZ antibody status (positive, negative, missing) by treatment group (3 lines per graph, one line for each anti-BKZ antibody status, one graph per treatment group). The missing group will not be displayed if ≥95% of subjects are categorized in the non-missing groups.

  The ADAb status (positive, negative, or missing) will be considered in a cumulative manner at each time point:
  - a subject will be counted as positive from the first visit at which the subject 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 subject has only negative ADAb samples or only one missing/inconclusive sample with negative ADAb samples up to that timepoint, the subject will be classified as negative. An exception remains for the Baseline Visit where only one sample would be available. If the sample is missing/inconclusive, then the sample will be classified as being negative for the cumulative ADAb status.
  - Otherwise, the subject will be classified in the missing ADAb category.
- The table summaries and figures will also be repeated by neutralizing anti-bimekizumab antibodies (NAb) status (ADAb negative, NAb positive, ADAb positive / NAb negative,

NAb missing) (Section 9.2.2) and treatment groups. Separate tables and figures will be generated for concentration data up to Week 16, 24 and 52.

For the concentration summary tables and figures by NAb status, 2 sets of summary tables and figures will be produced:

- For the DBP (up to Week 16), a table and figure summarizing the BKZ concentration for the BKZ 160 mg Q4W group. For these tables and figures, the NAb status will be determined on a subject-level basis (ie, not on a visit level) according to the categories in Section 9.2.2. The NAb status will be determined based on data from the Baseline visit (pre-dose sample on study Day 1) up to Week 16 NAb sample (eg, if a subject has no positive samples for IL17AA or IL17FF from Baseline up to Week 16, that subject will be assigned a NAb negative status for all visit of the DBP).
- For the DBP and the ATP (up to Week 24), tables and figures summarizing the BKZ concentration by treatment group and NAb status. For the Placebo / BKZ 160mg Q4W group, the NAb status will be determined based on data from sample at baseline visit (pre-dose sample on Week 16) up to Week 24. For the BKZ 160 mg Q4W group, the NAb status will be determined based on data from sample at Baseline visit (pre-dose sample on study Day 1) up to the NAb sample on Week 24 (eg, if a subject has no positive samples for IL17AA or IL17FF from Baseline up to Week 24, that subject will be assigned a NAb negative status for all visit of the DBP and ATP up to Week 24). The NAb status will be re-derived with updated data for the Week 52 analysis, followed by an update to the relevant tables and figures.

PK samples collected at scheduled visits and meeting the following requirements will be included in summaries and corresponding figures. Samples not meeting the following requirements will be excluded:

- For PK samples associated with Visit 3 (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 Visit 28 (Week 52) or the ET visit: include samples collected >14 days and < 42 days after the last/previous dose received.
- For PK samples associated with all other visits (Week 4, 8, 12, 16, 20, 24, 36): 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 1 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.

- The BKZ concentrations will also be listed for all subjects in the SS (for BKZ treated subjects). 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 where the reason for exclusion will be one or more of the following:
  - 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 1 sample obtained at the same visit: Includes all samples excluded due to multiple valid samples associated with a visit.

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

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 Immunological (Immunogenicity) analyses

9.2.1 Anti-bimekizumab antibody

The ADAb will be assessed using a 3-tiered assay approach: Screening, confirmatory, and titration assays. Samples confirmed as positive within the confirmatory assay will be further evaluated for the presence of NAb specific to binding of BKZ to IL-17AA, IL-17FF or both.

Screening, confirmatory and titer cut points of the respective assays will be determined by the bioanalytical laboratory. The relevant statistical reports will be provided as part of the bioanalytical reports.

The ADAb samples are analyzed on the SS but are not analyzed when subjects are on a treatment other than BKZ. For subjects who switch from placebo to BKZ, samples are analyzed at Baseline visit (prior to any dosing) and at the visit (Week 16) when the switch to BKZ occurs. The sample at the visit when the switch occurs will act as the Baseline for that treatment group.

The Screening cut point will be used to determine the ADAb status in the test sample 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).

The ADAb status for each visit will be derived as following:

- Sample values that are either NS or PS and NI 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 (200 µg/mL) 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 if it does not go into one of the above categories.

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 the target binding of BKZ (IL17A or IL17F or both) in-vitro.

The following rule will be implemented for by-visit ADAb summaries where applicable:

- From Baseline through Week 24 visit, if the ADAb sample is collected within ±14 days relative to the visit date at which the drug was administered (or ±14 days from a scheduled visit at which dosing was not performed), the ADAb result for that sample will be associated with the scheduled visit and summarized accordingly.
- From Week 24 visit, if the ADAb sample is collected within ±21 days relative to the visit date at which the drug was administered (or ±21 days from a scheduled visit at which dosing was not performed), the ADAb result for that sample will be associated with the scheduled visit and summarized accordingly.
- In all cases, this will include unscheduled assessments as described in Section 3.1 (if a dose was administered at an unscheduled visit). Samples collected outside this window will be excluded from the ADAb summaries and will be listed only.

The rules above will apply to by-visit summaries only. This rule will not apply for the Week 52, ET or SFU visits. Thus, all ADAb data obtained at these visits will be included in the by-visit summaries.

All other summaries of ADAb status will use all available data (scheduled and unscheduled).

In addition, the anti-BKZ antibody status will be further classified according to the ADAb categories outlined below:

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

- **Category 3: Pre ADAb positive – treatment reduced ADAb**: Includes subjects who are positive at Baseline, and antibody negative at all sampling points post treatment (including SFU).
- **Category 4: Pre ADAb positive – treatment unaffected ADAb positive**: Includes subjects who are positive at Baseline and are positive at any sampling point post treatment (including SFU) with titer values of the same magnitude as Baseline (ie less than a predefined fold difference from the Baseline value). For the purposes of this study, this is set at an increase of less than the validated Minimum Significant Ratio (MSR) of 2.07-fold from baseline.
- **Category 5: Pre ADAb positive – treatment boosted ADAb positive**: Includes subjects who are positive at Baseline and are positive at any sampling point post treatment (including SFU) with increased titer values compared to Baseline (greater than or equal to a predefined fold difference increase from Baseline value which is defined within the validation of the assay). For the purposes of this study, this is set at an increase greater than or equal to the validated Minimum Significant Ratio (MSR) of 2.07-fold from Baseline.

  Note: For any subject 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 subject will be considered as treatment boosted (ie, Category 5), assuming no other samples are available.
- **Category 6: Inconclusive**: Includes subjects who have a positive pre-treatment sample and some post-treatment samples are missing, while other post-treatment samples are ADAb negative or missing.
- **Category 7: Total treatment-emergent**: Category 2 and 5 combined: Includes subjects who are pre ADAb negative – treatment emergent ADAb positive (Category 2) and pre ADAb positive – treatment boosted ADAb positive (Category 5).
- **Category 8: Total prevalence of pre-ADAb positivity**: (Categories 3, 4, 5 and 6 combined): Subjects that are tested ADAb positive at Baseline.
- **Category 9: Missing:** Includes subjects who have a negative or a missing/inconclusive pre-treatment sample and some post-treatment samples are missing/inconclusive, while other post-treatment samples are ADAb negative or missing.

For the subjects randomized to placebo group, the Baseline will be the Week 16 pre-dose observation.

All the following analyses will be prepared on the SS on BKZ treated subjects by treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W):

- A summary table displaying the number and percentage of subjects with ADAb status - positive, negative, total of positive and negative , missing - at each visit (including SFU) and overall will be provided. For the overall summary at any visit, 2 summaries will be presented:

- Including any visit during the treatment period (as defined in Section 3.1). 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 Week 16 (for the BKZ 160 mg Q4W group), and to Week 24 and 52 by treatment group (BKZ 160 mg Q4W, Placebo / BKZ 160 mg Q4W).
- Including any visit during the study. Thus, this summary will include data obtained at both the SFU visit and at Baseline. Subjects who rolled over to the OLE study will not have a SFU visit per study protocol. For these subjects, the 'overall ADAb status including SFU' will then be considered as being identical to the 'overall ADAb status up to Week 52'.

For the overall ADAb status, a subject will be classified as:

- Positive if the subject has at least one positive sample at any time in the treatment period (regardless of having missing /inconclusive data)
- Negative if the subject has all the samples negative or only one missing/inconclusive sample with negative ADAb samples up to the timepoint of interest
- Missing if the subject has missed more than one sample result for ADAb assessment (or have more than one inconclusive sample) and all other available ADAb samples are negative during the period of interest (If there are ≥95% of subjects 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 subjects in each of the ADAb status (positive, negative, total of positive and negative, missing) will be repeated for the subjects randomized in Japan for Placebo / BKZ 160 mg Q4W and BKZ 160 mg Q4W groups (overall summaries only). The overall summary excluding SFU will be generated for up to Week 16, 24 and 52.
- A summary table displaying the number and percentage of subjects in each of the ADAb status (positive, negative, total of positive and negative, missing) by concomitant medications (use of cDMARDs at entry, use of MTX at entry, use of oral/systemic corticosteroids at entry) will be provided (overall summaries only, including summaries up to Week 16, 24, 52 and SFU).
- A table displaying the number (%) of subjects with the first occurrence of any ADAb positivity (ie: including Baseline visit) or ADAb treatment-emergent positivity during the study will be summarized by treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W). This summary will include the following categories:
  - Any ADAb+: ADA positive sample regardless of category during the treatment period,
  - ADAb Category 2: Pre ADAb negative – treatment-emergent ADAb positive,
  - ADAb Category 5: Pre ADAb positive – treatment-boosted ADAb positive (if this category represents at least 10% of the subjects, otherwise results are to be merged with category 2). For this category, the first occurrence of a boosted result is considered.

The table will include the number and percentage of subjects with first occurrence of any ADAb+ sample, and also subjects 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 subjects with any ADAb+ sample and treatment-emergent ADAb positive results at each time point.

- A boxplot of the ADAb titer by time of occurrence of ADAb positivity will be created for each treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W). The ADAb titer results will be presented on the log-scale. The time to occurrence is defined as the time in weeks from Baseline until the visit of interest when a sample is ADAb positive. Subjects who do not have any ADAb positivity will be excluded from the plot.
- A summary table of the number and percentage of subjects in each of the 9 ADAb categories will be tabulated.
- Some figures summarizing the time to achieve any ADAb positivity and treatment emergent ADAb positivity on a cumulative basis will be presented by treatment group (4 lines per plot: i.e. 2 lines for each treatment group, one line for time to achieve any ADAb positivity and one line for time to achieve treatment emergent ADAb positivity). Subjects will be considered to have an event at the timepoint 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).
  If ≥10% subjects are classified as ADAb category 5, the treatment emergent ADAb+ line will be further split by the following categories (thus the plot by treatment group will include 3 lines):
  - ADAb category 2: Pre ADAb negative – treatment-emergent ADAb positive,
  - ADAb category 5: Pre ADAb positive – treatment-boosted ADAb positive.
- Individual plots (one plot by subject) of BKZ concentrations/ADAb titer and ACR50 response (based on NRI) will be created. All 3 endpoints will be plotted on the Y-axis by visit (x-axis) for the full treatment period, including SFU when a subject has not progressed into the OLE. Plots should be labeled and grouped into the 9 ADAb categories and display vertical lines on the x-axis corresponding to the dosing incidences/visits.
  For the Week 24 analysis, the ADAb categories will be based on data up to Week 24.
- Individual plots (1 plot by subject) of BKZ concentrations/ADAb titer and PASI90 (on subjects with PSO involving at least 3% of BSA at Baseline) will also be created similarly as described above.
- Spaghetti plots of ADAb titer (y-axis) will be reported on the log-scale by visit (x-axis) by treatment group for all ADAb positive subjects. This plot will include the following ADAb categories:
  - Category 2: Pre ADAb negative – treatment-emergent ADAb positive,
  - Category 5: Pre ADAb positive – treatment-boosted ADAb positive.

The 3 plots described above will present ADAb titers using a semi-logarithmic (ADAb negative samples will therefore be excluded from the plots). Furthermore, all ADAb titer values<100 will be represented as 1 in these plots.

- Figures will summarize efficacy response (ACR50, ACR20 and PASI90 responders based on NRI, for PASI90 in the subgroup of subjects with PSO involving at least 3% BSA at Baseline) 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 ACR50 (ACR20 or PASI90) responders at Week 16. The plot will therefore display the percentage of ACR50 (ACR20 or PASI90) responders as a function of the number of subjects within each ADAb and titer categories. Subjects with negative ADAb results at Week 16 will be included in the 'negative' category on the x-axis. This plot will include data only from subjects randomized to BKZ 160 mg Q4W.
  The figure will be repeated using Week 24 and 52 ACR50, ACR20 and PASI90 responder rates and Week 24 and 52 ADAb titer quartiles, respectively, and will include all subjects receiving BKZ. The figure for Week 24 and 52 data will be split by treatment group. Thus, 2 figures per efficacy variable and time point will be presented.
- Two figures (side by side) summarizing time course of efficacy response (ACR50 responders based on NRI) for each treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W) by the following ADAb categories considered for the efficacy sub-group analysis (3 lines per plot) will be created:
  - ADAb positive - Defined as subjects having at least 2 ADAb positive samples during the study (including baseline and SFU up to the timepoint of interest) regardless of other ADAb negative samples and/or missing or inconclusive samples
  - ADAb negative - Defined as subjects for whom either (1) all samples (including baseline and SFU up to the timepoint of interest) are ADAb negative and there are no missing or inconclusive samples, (2) only 1 sample is ADAb positive and all other samples (including baseline and SFU up to the timepoint of interest) are ADAb negative or missing/inconclusive or (3) only one sample is missing/inconclusive and the remaining ADAb samples are negative.
  - ADAb Missing - Defined as subjects who do not fulfil the criteria for one of the 2 groups listed above.

  Separate figures will be presented for data up to Week 16 (for the BKZ 160mg Q4W group) and side by side by treatment group for data up to Week 24 and Week 52. The data will also be presented in a tabular format. The analyses will be repeated for ACR20 and PASI90 (for PASI90, in the subgroup of subjects with PSO involving at least 3% BSA at Baseline).

  If ≥95% of subjects are included in the non-missing groups, the missing group will not be tabulated and thus not displayed on the figure.

  Note: for efficacy by ADAb status summary up to a particular timepoint (Week 16, 24 and 52), only ADAb data up to the timepoint of interest will be utilized in deriving the overall ADAb status during that period. The percent of subjects in each of the ADAb categories considered for efficacy sub-group analyses will be summarized.

- All individual subject-level ADAb results will be listed including the Screening assay, confirmatory assay, ADAb 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 for ADAb measurements that were excluded from the by-visit summaries. The reason for exclusion will be one of the following:
  - Sample collected out of window relative to current dose (or visit),
  - More than one sample obtained at the same visit.
- 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 for excluded PK results.

9.2.2 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 > the drug tolerance limits of the NAb assays (100µg/mL) will be labeled 'inconclusive'. All inconclusive results will be regarded as missing.

Subjects will be assigned an overall Neutralizing anti-BKZ antibodies (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 subjects 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 classified as follows:
  - ADAb positive / NAb negative: ADAb positive subjects 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 subject 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 subject 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
- NAb Missing:

– >1 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.

Note: For PK and efficacy analyses up to a certain timepoint (Week 16, 24 and 52) by NAb status, only data up to the timepoint of interest will be utilized in deriving the subject's overall NAb status during that period (eg for the ACR50 by NAb up to Week 16, only NAb data from baseline up to Week 16 will be used to derive subject's overall NAb status during the DBP).

- A listing will be produced to summarize the NAb status overall in the study. The listing will be sorted by treatment group (Placebo/ BKZ 160mg Q4W vs. BKZ 160mg Q4W), subject identifier and visit and will summarize the following information for each subject assessed for NAb:
  - Visit
  - Study week
  - Laboratory sampling date and time
  - Time since previous dose (weeks)
  - The corresponding BKZ plasma concentration level at each visit (ug/mL)
  - Anti-BKZ 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
- A summary table will provide the following overall summary statistics by treatment group (Placebo/ BKZ 160mg Q4W vs. BKZ 160mg Q4W) (based on the total number of randomized subjects):
  - The number and percentage of subjects confirmed as anti-BKZ antibody positive and anti-BKZ antibody negative up to Week 16, 24, 52 (excluding SFU), 52 (including SFU)
  - The number and percentage of subjects who are NAb positive, NAb negative and missing up to Week 16, 24, 52 (excluding SFU) and 52 (including SFU).

  The NAb summary tables will be repeated with percentages calculated on a denominator based on the number of ADAb positive subjects during the above periods.

In addition, the following analyses will be performed to assess the impact of NAb status on efficacy variables:

- Two figures (side by side) summarizing efficacy response (ACR50 based on NRI) versus time for each treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W) by NAb status (ADAb negative, NAb positive, ADAb positive / NAb negative, missing) will be created. If ≥95% of subjects are included in the non-missing groups, the missing group will not be displayed in the figure. Separate figures will be presented for data up to Week 16 (for the BKZ 160mg Q4Wgroup), and to Week 24 and 52 by treatment group (BKZ 160mg Q4W vs. Placebo / BKZ 160mg Q4W). The analyses will be repeated for ACR20 and PASI90 (for

PASI90, in the subgroup of subjects with PSO involving at least 3% BSA at Baseline). The data will be also presented in a tabular format.

10 SAFETY ANALYSES

All safety summaries will be done on all subjects in the SS.

10.1 Extent of exposure

The extent of exposure will be evaluated based on the duration of exposure to study drug and the time at risk.

Summaries for exposure will be presented:

- For the DBP on the SS by treatment group,
- For the ATP on the ATS by treatment group and for all BKZ treated subjects
- For the AMP on the AMS by treatment group and for all BKZ treated subjects.

The following results will be presented:

- A descriptive summary of duration of study treatment exposure in days,
- A total study medication duration in subject-years,
- The number of subjects exposed to the following duration of study treatment (for the DBP and ATP together):
  - > 0 week
  - ≥ 16 weeks
  - ≥ 24 weeks
  - ≥ 36 weeks
  - ≥ 48 weeks
  - ≥ 52 weeks
- A descriptive summary of time at risk in days,
- A total time at risk in subject-years.

A by-subject listing of date of first and last dose, the duration of exposure and the time at risk will be performed.

Throughout the below sections, date of last clinical contact for each subject is defined as the maximum of the following dates:

- Date of last visit including SFU visit,
- Last imputed AE start date,
- Date of study termination/completion,
- Last date of study drug administration.

- Or any other collected date.

In calculation of exposure detailed below, 14 days refers to the time between 2 consecutive study treatment administrations.

10.1.1 Double-Blind Treatment Period

10.1.1.1 Study medication duration (days)

The study medication duration will be calculated as:

- For subjects with a date of last dose in the DBP +14 days beyond the date of first dose in the ATP (excluding SFU):

  $Date\ of\ first\ dose\ (ATP) - Date\ of\ first\ dose\ (DBP) + 1$

- For subjects with a date of last dose in the DBP +14 days before or on the same day as the date of first dose in the ATP (excluding SFU):

  $$Date\ of\ last\ dose\ (DBP) - Date\ of\ first\ dose\ (DBP) + 14$$

- For subjects who have died during the DBP:

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

- For other subjects:

  $Date\ of\ last\ dose\ (DBP) - Date\ of\ first\ dose\ (DBP) + 14$

- For other subjects who withdrew from study or ongoing in the study during DBP, the minimum between:

  $Date\ of\ last\ dose\ (DBP) - Date\ of\ first\ dose\ (DBP) + 14$

  $Date\ of\ the\ final\ visit\ (excluding\ SFU) - Date\ of\ first\ dose\ (DBP) + 1$

10.1.1.2 Time at risk (days)

The time at risk will be calculated as:

- For subjects who completed the DBP and continue in the ATP:

  $Date\ of\ first\ dose\ (ATP) - Date\ of\ first\ dose\ (DBP) + 1$

- For subjects who have died prior to the final visit of the DBP:

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

- For all other subjects (who discontinue on or prior to the final visit of the DBP), the minimum between:

  $Date\ of\ last\ clinical\ contact - Date\ of\ first\ dose\ (DBP) + 1$

  $$Date\ of\ last\ dose\ (DBP) - Date\ of\ first\ dose\ (DBP) + 141$$

  $Total\ number\ of\ days\ in\ the\ DBP\ (112\ days)$

Note: 140 days refer to 5 times the half-life of BKZ.

10.1.2 Active Treatment-blind Period

The exposure will be calculated for subjects in the ATS only.

10.1.2.1 Study medication duration (days)

The study medication duration will be calculated as:

- For subjects who completed the study, with a date of last dose +14 days beyond the date of final visit of the ATP (excluding SFU):

  $Date\ of\ final\ visit\ (ATP) - Date\ of\ first\ dose(ATP) + 1$

- For subjects who have died during the ATP:

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

For other subjects$Date\ of\ last\ dose\ (ATP) - Date\ of\ first\ dose(ATP) + 14$

10.1.2.2 Time at risk (days)

The time at risk will be calculated as:

- For subjects who completed the study (and continue into the OLE study):

  $Date\ of\ final\ visit\ (ATP) -\ Date\ of\ first\ dose\ (ATP) + 1$

- For subjects who have died prior to the final visit of the ATP:

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

- For all other subjects (subjects who discontinue early from the study, subjects who complete study as planned but chose not to continue into the OLE study, or subjects who are ongoing in the SFU period at the time of the data snapshot), the minimum between:

  $Date\ of\ last\ clinical\ contact - Date\ of\ first\ dose\ (ATP)\ + 1$

  $Date\ of\ last\ dose\ (ATP) - Date\ of\ first\ dose\ (ATP) + 141$

Note: 140 days refer to 5 times the half-life of BKZ.

10.1.3 Double-blind and the Active Treatment-blind Periods (combined)

10.1.3.1 Study medication duration (days)

The study medication duration (in days) for subjects who do not switch study treatment at Week 16 will be calculated as:

- For subjects who completed the study with a date of last dose +14 days beyond the date of final visit date (excluding SFU):

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

- For subjects who have died:

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

- For other subjects:

  $Date\ of\ last\ dose\ - Date\ of\ first\ dose\ + 14$

The study medication duration (in days) for Placebo subjects who switch study treatment at Week 16 will be calculated as:

For the DBP (attributed to the Placebo treatment):

- For subjects with a date of last dose +14 days beyond the date of final visit date (excluding SFU):

  $Date\ of\ first\ dose\ (ATP) - Date\ of\ first\ dose\ (DBP) + 1$

- For other subjects:

  $Date\ of\ last\ dose\ (DBP) - Date\ of\ first\ dose\ (DBP)\ + 14$

For the ATP (attributed to the BKZ treatment):

- Use the algorithm specified for the ATP.

10.1.3.2 Time at risk (days)

The time at risk (days) for subjects who do not switch study treatment at Week 16 will be calculated as:

- For subjects who completed the study (and continue into the OLE study):

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

- For subjects who have died prior to scheduled final visit date of the ATP:

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

- For all other subjects (subjects who discontinue early from the study, subjects who complete study as planned but chose not to continue into the OLE study, or subjects who are ongoing in the SFU period at the time of the data snapshot), the minimum between:

  $Date\ of\ last\ clinical\ contact - Date\ of\ first\ dose\ + 1$

  $Date\ of\ last\ dose\ - Date\ of\ first\ dose\ + 141$

The time at risk (days) for Placebo subjects who switch study treatment at Week 16 will be calculated as:

For the DBP (attributed to the Placebo treatment):

- For all subjects:

  $Date\ of\ first\ dose\ (ATP)\ - date\ of\ first\ dose\ (DBP) + 1$

For the ATP (attributed to the BKZ treatment):

- Use the algorithm specified for the ATP.

10.1.4 Exposure and time at risk in subject-years

The sum of all durations of exposure (in subject-years) will be calculated as:

$$\frac{\sum_{Subjects} Duration\ of\ exposure\ (days)}{365.25}$$

The sum of all times at risk (in subject-years) will be calculated as:

$$\frac{\sum_{Subjects} Times\ at\ risk\ (days)}{365.25}$$

10.1.5 Time at risk in subject-years by COVID-19 pandemic period

The time at risk will be determined by COVID-19 pandemic period as defined in Section 3.11.

10.1.5.1 Time at Risk in the pre-COVID-19 pandemic period

Only subjects starting study drug during the pre-COVID-19 pandemic period will have a time at risk calculated during that period.

- For subjects with date of last dose of study drug on or after 11mar2020:

  10mar2020 – date of first dose + 1

- For subjects who have died on or before 10mar2020:

  Date of death – date of first dose + 1 (If the date of death is later than 140 days following the last dose of BKZ, this rule does not apply).

- For subjects with date of last dose of study drug on or before 10mar2020 but with the earliest date between (date of last dose + 141) and (date of last clinical contact) on or after 11mar2020:

  10mar2020 – date of first dose + 1

- For subjects with date of last dose of study drug on or before 10mar2020 but with the earliest date between (date of last dose + 141) and (date of last clinical contact) on or before 10mar2020, minimum between:

  date of last dose + 141 – date of first dose

  date of last clinical contact – data of first dose +1

10.1.5.2 Time at Risk in the COVID-19 pandemic period

Subjects with study drug period overlapping the COVID-19 pandemic periods will have a time at risk calculated during that period.

- For subjects with date of last dose of study drug on or after 11mar2020, minimum between:

  Minimum(date of last dose + 141, COVID-19 pandemic end date) – maximum of (11mar2020, date of first dose)

  Minimum(date of last clinical contact, COVID-19 pandemic end date) – maximum of (11mar2020, date of first dose) +1

- For subjects who have died on or after 11mar2020:

  Minimum(Date of death, COVID-19 pandemic end date) – maximum of (11mar2020, date of first dose) + 1 (If the date of death is later than 140 days following the last dose of BKZ, this rule does not apply).

- For subjects with date of last dose of study drug on or before 10mar2020 but with the earliest date between (date of last dose + 141) and (date of last clinical contact) on or after 11mar2020, minimum between:

  Minimum(date of last dose+ 141, COVID-19 pandemic end date) – 11mar2020

  Minimum(date of last clinical contact, COVID-19 pandemic end date) – 11mar2020+1

10.1.5.3 Time at Risk in the post-COVID-19 pandemic period

Subjects with study drug period continuing after the COVID-19 pandemic periods will have a time at risk calculated during that period:

- For subjects with date of last dose of study drug on or after the end date of the COVID-19 pandemic period, minimum between:

  date of last dose + 141 – maximum of (COVID-19 pandemic end date, date of first dose)

  date of last clinical contact – maximum of (COVID-19 pandemic end date, date of first dose) +1

- For subjects who have died on or after *COVID-19 pandemic end date*:

  Date of death – maximum of (COVID-19 pandemic end date, date of first dose) + 1 (If the date of death is later than 140 days following the last dose of BKZ, this rule does not apply).

- For subjects with date of last dose of study drug on or before *COVID-19 pandemic end date* but with the earliest date between (date of last dose + 141) and (date of last clinical contact) on or after *COVID-19 pandemic end date*, minimum between :

  date of last dose + 141 – COVID-19 pandemic end date

  date of last clinical contact – COVID-19 pandemic end date+1

10.2 Adverse events

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

AEs (including SAEs) are characterized as either non-treatment or treatment-emergent according to the following criteria:

- Non treatment-emergent are the events with onset date and time prior to the very first administration of study medication (BKZ or placebo) or after a 140-day period after the final drug administration.
- Treatment-emergent AEs (TEAE) are those with onset date at or after the very first administration of study medication. The events that emerge within 140 days after the final drug administration, will also be considered as treatment-emergent (eg, in the case of premature discontinuation or during the SFU Period).

If an AE occurs on the same day as the treatment switch (Week 16, to BKZ for subjects randomized to Placebo) then the AE will be allocated to the Double-Blind Treatment. An exception from this general rule is made for the following types of AEs that will be allocated to the BKZ:

- Events that fulfill the anaphylaxis criteria for acute events (refer to Appendix 1)
- Events that fulfill the hypersensitivity reaction criteria (refer to Appendix 1)
- Events with a high-level term (HLT) of "Administration site reactions NEC"
- Events with an HLT of "Injection site reactions"

The incidence of TEAEs will be summarized descriptively by treatment group, MedDRA primary system organ class (SOC), HLT, and preferred term (PT).

AEs will be presented as "number of subjects (percentage of subjects) [number of events]". In this style of output, "[number of events]" will include all cases of an AE including repeat occurrences in individual subjects, while "number of subjects" will count each subject 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 BKZ 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 BKZ treatment group, and in the event of ties, PT will be sorted alphabetically.

TEAEs will be classified by period (DBP, ATP or AMP) according to their date of emergence.

Summary tables of TEAEs will be provided on:

- The DBP by treatment on the SS,
- The ATP by treatment and for all BKZ-treated subjects in the ATS,
- The AMP for all BKZ-treated subjects and the ADA group in the AMS.

10.2.1 Standard AE summaries

Tables for incidence of the following categories of TEAES will be provided:

- TEAEs (overview),
- TEAEs by SOC, HLT, PT,
- TEAEs by decreasing frequency of PT,
- TEAEs above the reporting threshold of 5% by PT, by SOC and PT
- TEAEs by maximum intensity,
- TEAEs by relationship to IMP,
- Related TEAEs by SOC, HLT, PT,
- Related TEAEs above the reporting threshold of 5% by PT, by SOC and PT
- TEAEs leading to study discontinuation, by SOC, HLT, PT,

- TEAEs leading to study drug discontinuation, by SOC, HLT, PT,
- Treatment-emergent SAEs (TESAEs) by SOC, HLT, PT,
- TESAEs by relationship by SOC, HLT, PT,
- TEAEs leading to Death by SOC, HLT, PT - Deaths will also be tabulated and listed,
- Non-serious TEAEs above the reporting threshold of 5% of subjects by relationship by PT.

Some selected tables of incidence will include the exposure adjusted incidence rate (EAIR) with associated 95% CI and the exposure adjusted event rate (EAER) (Section 10.2.3). The categories of TEAES concerned are listed below:

- TEAEs by SOC, HLT, PT,
- TEAEs leading to study discontinuation by SOC, HLT, PT,
- TEAEs leading to study drug discontinuation,
- TEAEs by timing of onset relative to ADAb status by SOC, HLT, PT (on subjects treated with BKZ). This 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 subjects in category 2 and b) the first post-baseline boosted ADAb titer result for subjects with titer results and the first post-Baseline positive ADAb result for subjects with positive ADAb at Baseline with no other samples with titer available for subjects 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 subjects in categories 2, 3, 4 and 6, and b) the first post-baseline boosted ADAb titer result for subjects with titer results and the first post-Baseline positive ADAb result for subjects with positive ADAb at Baseline with no other samples with titer available for subjects in category 5.
  - TEAEs for subjects who are ADAb negative at all timepoints (includes ADAb Category 1).
- TEAEs by NAb status (ADAb negative, NAb positive, ADAb positive / NAb, negative) on subjects treated with BKZ. TEAEs will be sorted by system organ class, high level term and preferred term.
  Note: for TEAE by onset relative to ADAb positivity and by NAb status, all available ADAb and NAb data at the time of IA cut-off, respectively, will be utilized to derive the subject-level ADAb/NAb status categories.
- TESAEs by SOC, HLT, PT.

By-subject listings on all AEs and all deaths will be provided on the Enrolled Set. AEs leading to study discontinuation will also be listed on the SS. AE occurring at first study drug administration (including for subjects randomized to Placebo, at first administration of BKZ at Week 16) will also be listed. A glossary listing will also be created for all TEAEs.

10.2.2 Exposure duration of AEs

The duration of each AE will be calculated as:

$$Duration\ of\ AE\ (days) = End\ date\ of\ AE - Start\ date\ of\ AE + 1$$

For the calculation of duration for AEs with missing start dates, the dates will be imputed as described in Section 4.2.3.

The time since first dose of IMP for each AE will be calculated for all TEAEs as:

$$Time\ since\ first\ dose\ (days) = Start\ date\ of\ AE - Date\ of\ first\ dose + 1$$

The time since the most recent dose of IMP for each AE will be calculated for all TEAEs as:

$$Time\ since\ the\ most\ recent\ dose\ (days) = Start\ date\ of\ AE - Date\ of\ most\ recent\ dose + 1$$

Time to first dose and time to most recent dose will not be calculated for pretreatment AEs.

For the calculation of time to first BKZ and time to most recent BKZ dose the same formulas as above will be used but the date of first BKZ dose will be used instead of the date of first dose of IMP. The time to first dose and time to first BKZ dose differs only for those subjects who receive placebo at Baseline after the time they switch to BKZ.

10.2.3 Exposure adjusted incidence rate (EAIR) and exposed adjusted event rate (EAER)

The EAIR is defined as the number of subjects (k) with a specific AE adjusted for the exposure and will be scaled to 100 patient-years:

$$EAIR = 100 \times \frac{k}{\sum_{i=1}^{n}\left(T_{Exp(i)}\right)}$$

where n is the total number of subjects and $T_{Exp(i)}$ is the time of exposure for each subject defined as:

- For subjects with the AE: years since first dose of IMP to the first occurrence for the AE of interest at the level of coding evaluated
- For subjects without the AE: the time at risk in years.

Exact Poisson 95% CIs for incidence rates are calculated using the relationship between the Poisson and the Chi-square distribution (Ulm, 1990; Fay and Feuer, 1997):

$$LCL = \frac{\chi^2_{2k,\alpha/2}}{2}$$

$$UCL = \frac{\chi^2_{2(k+1),1-\alpha/2}}{2}$$

where k is the number of subjects 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 $\chi 2$.

The EAER will be the number of AEs (N_{AE})including repeat occurrences in individual subjects divided by the total time at risk scaled to 100 patient-years and calculated using:

$$EAER = 100 \times \frac{N_{AE}}{\sum_{i=1}^{n}(T_{Risk(i)})}$$

where N_{AE} is the number of AE occurrences, n is the total number of subjects and $T_{Risk(i)}$ is the time at risk in years for each subject.

No CI will be computed for EAER.

10.2.4 Other safety topics of interest

Other safety topics of interest for this study include:

- Infections (serious, opportunistic, fungal and tuberculosis)
- Malignancies
- Major adverse cardiac events (MACE)
- Neutropenia
- Suicidal Ideation and Behavior (SIB)
- Inflammatory bowel disease
- Hypersensitivity (including Anaphylaxis)
- Hepatic events and drug-induced liver injury (DILI)

The analyses produced for the other safety topics of interest are based on the specifications described in the BKZ safety topic of interest document version from Apr-2021.

The incidence of other safety topics of interest will be summarized by MedDRA SOC, HLT, and PT. The EAIR with associated 95% CI and the EAER will be included in the summary tables.

Output tables will be produced for categories of other safety topics of interest as described below separately for the DBP in the SS, the ATP on the ATS and on the combined DBP and ATP in the AMS.

10.2.4.1 Infections (serious, opportunistic, fungal and TB)

Serious infections will be identified based on MedDRA classification (SOC "Infections and infestations") using the "Any SAE" table. A separate table does not need to be produced to summarize these events:

- Fungal infections will be summarized in a stand-alone table. The table will include all TEAEs (serious and non-serious) which code into the High-Level Group Term (HLGT) "Fungal infectious disorders".
- Opportunistic infections (including tuberculosis) will be summarized in a stand-alone table. The table will include all opportunistic infection TEAEs identified using UCB-defined search criteria. The process for identifying opportunistic infections is outlined in a separate document located in Appendix 2.

10.2.4.2 Malignancies

Two tables will be produced for malignancies:

- A table will be based on the criteria SMQ= "Malignant or unspecified tumours (SMQ)".
- A table will be based on the criteria SMQ=" Malignant tumours (SMQ)".

SMQ search should include all TEAEs which code to a PT included in the Scope=Narrow group within each SMQ.

Note that the events included in the "Malignancies" table will be a subset of the events included in the "Malignancies (including unspecified)" table. While the "Malignant tumours (SMQ)" is most relevant, "Malignant or unspecified tumours (SMQ)" must be reviewed for potential malignancies.

The output tables will include 2 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 High Level Term (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)".

10.2.4.3 Major adverse cardiac events (MACE)

The classification of an event as MACE is determined by an external cardiovascular event adjudication committee.

- Major adverse cardiac events will be presented in a table.
- 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 event types will be considered an extended MACE event.

Table 10–1: Type of Extended MACE events

Event Type Code	Event Type
1	Non-Fatal Myocardial Infarction (MI)
2	Non-Fatal Stroke: hemorrhagic
3	Non-Fatal Stroke: ischemic
4	Non-Fatal Stroke: embolic
5	Non-Fatal Stroke: undeterminable
6	Hospitalization or ER for Unstable Angina with urgent revascularization
8	Hospitalization for Heart Failure

Event Type Code	Event Type
10	Coronary Revascularization Procedures (eg percutaneous coronary intervention, coronary artery bypass grafting)
11	Urgent Revascularization Procedures (ie due to symptoms of brain ischemia or pending infarction)
18	Death due to Myocardial Infarction (MI)
19	Death due to Stroke
20	Sudden Cardiac Death
21	Other CV Death (eg heart failure, pulmonary embolism, cardiovascular procedure-related)
22	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 in total, for the full list of adjudicated cardiovascular events refer to the Bimekizumab-Event-Adjudication-Committee-Analysis-Plan), 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.

10.2.4.4 Neutropenia

A table will be created 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

10.2.4.5 Suicidal Ideation and Behavior Neuropsychiatric events

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 listing for SIB events as determined by the adjudication committee will be produced.

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. A separate listing will also be produced to summarize the adjudicated results of all events escalated to the full committee.

10.2.4.6 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 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 subjects 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 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.

10.2.4.7 Hypersensitivity (including Anaphylaxis)

A table will be prepared based on the MedDRA anaphylaxis Algorithm (see Appendix 1) for 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 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. In addition, a separate table will be prepared to summarize serious hypersensitivity events, identified using the SMQ "Hypersensitivity (SMQ)". All serious TEAEs which code to a PT included in the Scope=Narrow search will be included in this table. 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.

Furthermore, a table will be prepared to summarize injection site reactions, identified using the HLTs: "Administration site reactions NEC" and "Injection site reactions".

10.2.4.8 Hepatic events and DILI

A table for hepatic events will be created based on the SMQ of "Drug related hepatic disorders - comprehensive search (SMQ)". However, these 2 sub-SMQs are to be excluded: "Liver

neoplasms, benign (incl. cysts and polyps) (SMQ)" and "Liver neoplasms, malignant and unspecified (SMQ) ". For each of the above SMQs, include all TEAEs which code to a PT included in the Scope=Broad and/or Scope=Narrow.

Note that all AEs meeting the above criteria are to be included. It should not be limited to events that the investigator determined to be related to study drug.

Potential cases of Hy's Law will be reported separately in a liver function test table (Section 10.3).

10.2.4.9 Summary table and figure

The incidence of the below safety topics of interest occurring during the Double-Blind Treatment Period will be reported in one summary table and displayed graphically.

- Serious infection (Terms included in SOC "Infections and Infestations")
- Fungal infection
- Opportunistic infection
- Malignancy (Terms included in the SMQ "Malignant tumours (SMQ)")
- Adjudicated major adverse cardiac events (MACE)
- Neutropenia
- Adjudicated suicidal ideation and behavior (SIB)
- Definite adjudicated IBD
- Definite or probable adjudicated IBD
- Anaphylactic reaction
- Hypersensitivity reaction
- Hepatic events and drug-induced liver injury

For each of the above categories, the summary table and the graph will include the incidence, the EAIR by treatment group and the risk difference (BKZ minus placebo) with its 95% CI.

The risk difference (RD) will be derived as follows:

$$RD = \frac{\mathrm{Ab}}{\mathrm{Nb}} - \frac{\mathrm{Ap}}{\mathrm{Np}}$$

The lower and upper CI bounds for RD will be calculated as follows:

$$RD - Z_{\alpha/2} * SE\ (RD), RD + Z_{\alpha/2} * SE\ (RD)$$

Where:

Ab / Ap = Number of subjects in the BKZ / Placebo group (Safety Set) affected by the event

Nb / Np = Number of subjects in the BKZ / Placebo group (Safety Set)

$Z_{\alpha/2}$ = Critical value from the standard normal distribution (1.96 for a 95% CI).

10.2.5 Special considerations for interim data snapshots

As two interim analyses are planned, there will be 2 interim data snapshots before the final analysis:

- One after the last subject completed Week 24 (visit 14),
- One after the last subject completed Week 52 (visit 28).

Some subjects will likely have pending safety outcome data at the time of the interim data snapshots. Despite this, safety tables for the interim data analyses will be based on all available data at the time of the snapshot.

For the Week 24 analysis, AE summary tables and listings will only include AEs with an onset date prior to Week 24.

For the Week 52 analysis, some AEs including the ones with onset date prior to Week 24 may have been updated. For this analysis, all TEAEs tables created for the Week 24 analysis will need to be rerun.

Because of the long SFU Period (140 days from last dose), subjects who have completed the SFU visit (and do not, therefore, enroll into the OLE study) may have a longer period of exposure than subjects who complete the study and roll in to the OLE study. For those subjects, exposure ends at 20 weeks after the last dose of IMP.

For a subject completing Week 52 and enrolled into the OLE study, exposure ends at Week 52 (visit 28).

This potential difference is acknowledged and is not considered to have a meaningful impact on the AE summaries. Furthermore, the exposure-adjusted AE summaries (which will be incorporated into many of the AE tables) would account for these differences with EAIRs and EAERs.

After the last subject has completed the final SFU visit, the study database will be locked, and TFLs will be re-run including all subjects.

10.2.6 COVID-19 impact

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 dates (AE allocated to a period if is starts during the period).

The following categories of TEAEs 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 (pre – during – post) by region and overall

- All TEAEs leading to study discontinuation and/or permanent withdrawal of study medication by time of onset relative to COVID-19 pandemic period (pre – during – post) by region and overall
- All Serious TEAEs by time of onset relative to COVID-19 pandemic period (pre – during – post) by region and overall
- All COVID-19 related TEAEs by treatment group, separately for the DBP, ATP, and the AMP, by region and overall - COVID-19 related TEAEs will be identified by the 2 preferred terms “Corona virus infection” and “coronavirus test positive” and will include confirmed or suspected COVID-19 infections.

A listing of COVID-19 related AEs will be presented where these AEs are identified as described above. Additionally, the listing of AEs will include a column for COVID-19 relatedness and the time of onset of each AE relative to the COVID-19 pandemic will be included in all AE listings.

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 subject may therefore be counted in the denominator for several COVID-19 pandemic periods dependent on whether the subject 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 routine clinical laboratory evaluations specified below will be evaluated:

Table 10–2: 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 [b]
MCH	Creatinine	Urine drug screen[c]
MCHC	hs-CRP	
MCV	AST	
Platelet count	ALT	
RBC count	GGT	
WBC count	ALP	

Hematology	Biochemistry	Urinalysis
	Total bilirubin	
	Triglycerides[a]	
	Cholesterol[a]	
	HDL cholesterol[a]	
	LDL cholesterol[a]	
	LDH	
	Serum pregnancy testing[b]	

ALP=alkaline phosphatase; ALT=alanine aminotransferase; AST=aspartate aminotransferase; BUN=blood urea nitrogen; ET=early termination; GGT=gamma-glutamyl transferase; HDL=high density lipoprotein; hs-CRP=high sensitivity C-reactive protein; IMP=investigational medicinal product; IMS=International Menopause Society; LDH=lactate dehydrogenase; LDL=low density lipoprotein; 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

[a] Biochemistry will include triglycerides, cholesterol, HDL cholesterol, and LDL cholesterol at Baseline, Weeks 12 and 24, and at the ET visit.

[b] Pregnancy testing will consist of serum testing at the Screening visit for all women of childbearing potential. The pregnancy test will be urine at all other visits.

[c] Urine drug screen will be performed at Screening visit.

Biochemistry and hematology values will be flagged based on their reference range:

- Values that are below the lower limit of the reference range will be flagged as ‘L’ (low)
- Values that are above the upper limit of the reference range will be flagged as ‘H’ (high)
- All other values will be referenced as normal.

Markedly abnormal values for blood chemistry and hematology will be defined as laboratory values graded 3 or 4 according to the Rheumatology Common Toxicity Criteria (RCTC). Definitions of the markedly abnormal values are given in Table 10–3 and Table 10–4 and are based on the RCTC units. All units in the tables below will be converted to the standard units based on Clinical Data Interchange Standards Consortium (CDISC) standards.

Table 10–3: Definitions of markedly abnormal biochemistry values

Parameter name	Unit	Criteria	Abnormal designation
Creatinine	mmol/L	>3.0 x ULN or >3.0 x Baseline	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

Parameter name	Unit	Criteria	Abnormal designation
Potassium	mmol/L	>6.0 <3.0	AH AL
Sodium	mmol/L	>155 <130	AH AL
Triglyceride	mmol/L	>5.7	AH
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 10–4: Definitions of markedly abnormal hematology values

Parameter name	Unit	Criteria	Abnormal designation
Hemoglobin	g/L	<80 >40 above ULN or >40 above Baseline if Baseline is 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

Table 10–5: Definitions of markedly abnormal liver enzyme elevation values

Parameter name	Unit	Criteria	Abnormal designation
Alkaline Phosphatase		>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

Definitions of the markedly abnormal values for liver enzyme elevation are given in Table 10–5.

All laboratory summaries will be presented in standard international (SI) units and will be based on observed case values. In the case where laboratory values are below the LLOQ, then these will be set to the midpoint between 0 and the LLOQ for the purpose of summarizing the data.

In the case where laboratory data are expressed as above a fixed value (ie, ">xx.xx"), then the numeric portion of the value will be used (ie, xx.xx) to present and calculate statistics.

To define the minimum/maximum post-Baseline values, all post-Baseline assessments including those at unscheduled visits (but excluding any that occur more than 140 days after the last administration of study medication) will be used.

The following additional rule will be also applied :

Subjects who have completed PA0010 and have a PA0012 start date, only PA0010 results up to and including the PA0012 start date will be used to find the subjects minimum or maximum results. Additionally, subjects who are in PA0012 without a start date (date of first dose), only PA0010 results up to the Week 52 visit + 14 days will be used to find the subjects minimum or maximum results.

Different summary tables for hematology, biochemistry variables (except hs-CRP which will be analyzed in the efficacy section) and urinalysis (pH, Albumin (protein), glucose, blood, leukocyte esterase and nitrite) will be provided in the SS. For shift tables, results for the DBP and the ATP will be presented together for the overall study period.

As indicated in Section 3.1, for tables where data are summarized by visit only values occurring at scheduled visits will be included.

For each laboratory parameter, the below results will be presented for the overall study period, by treatment (Placebo/BKZ 160mg Q4W, BKZ 160mg Q4W and ADA 40mg Q2W):

- Observed values at Baseline and observed values and changes from Baseline at each scheduled post-Baseline visit
- Number and percentage of subjects by CTCAE grade (version 4.03) (when applicable) based on minimum/maximum post Baseline value (for blood chemistry and hematology). Subjects who meet the decreased potassium criterion of 3.0-<LLN, which is specified as the decreased potassium lab criterion for both CTCAE Grade 1 and Grade 2, will be counted as Grade 2.
- Shift in CTCAE grade from Baseline to minimum/maximum post-Baseline value (for blood chemistry and hematology)
- Number and percentage of subjects meeting the below criteria at any time during the study:
  - 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
- Number and percentage of subjects with potential Hy's Law cases based on the 2 following definitions:
  - [AST ≥3xULN or ALT ≥3xULN] and Total Bilirubin ≥2xULN
  - [AST ≥3xULN or ALT ≥3xULN] and Total Bilirubin ≥2xULN in the absence of ALP ≥2xULN

To meet either of the above criteria, a subject must experience the elevation in bilirubin and ALT or AST (and the absence of ALP elevation, if applicable) at the same visit.

The below results will be presented on the SS (for the DBP), on the ATS (for the ATP by treatment and for all BKZ-treated subjects), and on the AMS (for all the BKZ-treated subjects and on the ADA group). All values observed at any time while on treatment (ie, occurring at scheduled and unscheduled visits) will be included in the below tables. Baseline values and values observed more than 140 days after the last administration of study medication are not considered:

- Number and percentage of subjects with treatment-emergent markedly abnormal laboratory data overall (for blood chemistry and hematology), 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.
- Number and percentage of subjects with markedly abnormal liver enzyme elevation tests (data beyond the CTCAE Grade 3 thresholds as outlined in Table 10–5)

All laboratory data will be listed by treatment, subject and visit including changes from Baseline for numeric variables, flags for measurements outside the normal ranges, the relative study day, a flag for whether the test was not done and a flag for whether the subject was fasting.

The markedly abnormal laboratory data will also be listed separately.

Listings will also include Hepatitis B and C, human immunodeficiency virus (HIV), genomic proteomic/metabolomics, and genetic/epigenetic tests.

In addition, the laboratory results classified as Grade 3 or Grade 4 will be listed separately.

Urinalysis laboratory results and hepatic events will be listed separately.

10.4 Vital signs, physical examination findings, and other observations related to safety

10.4.1 Vital signs and physical examination findings

The vital signs (systolic BP (mmHg), diastolic BP (mmHg), and pulse rate (beats/min)) will be summarized as:

- Absolute and change from Baseline at each scheduled visit on the SS for the full study period by randomized treatment
- For systolic and diastolic blood pressure, number and percentage of subjects experiencing at least one markedly abnormal value as defined in Table 10–6.

Table 10–6: 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

Parameter (unit)	Markedly abnormal low	Markedly abnormal high
Diastolic blood pressure (mmHg)	<50 and a decrease from Baseline of ≥15	>105 and an increase from Baseline of ≥15

For the Baseline visit where several BP assessments are taken, the individual observed data assessments as well as the mean of these assessments will be summarized and listed (pre-dose, 30 min post-dose and 1-hour post-dose).

Treatment-emergent markedly abnormal values will be analyzed on the SS (for the DBP, by treatment), on the ATS (for the ATP by treatment and for all BKZ-treated subjects), and on the AMS (for all BKZ-treated subjects and the ADA group).

Vital signs measurements will also be listed by visit and timing relative to dosing including changes from Baseline. The listing will also include details of abnormal value based on flags as defined in Table 10–6 ("L" for markedly abnormal low and "H" for markedly abnormal high).

Similarly, physical examination findings together with the details of abnormalities (when applicable) will be listed by treatment group, subject and visit.

Body weight will be only listed.

10.4.2 Electrocardiograms

The following ECG variables will be assessed at Baseline (Screening visit), Week 16, Week 52 and at the end of SFU:

- Heart Rate (bpm)
- PR interval (ms)
- RR interval (ms)
- QRS duration (ms)
- QT interval (ms)
- QTcF interval (ms)
- QTcB interval (ms)

The date and time of the ECG will be recorded in the eCRF together with the Investigator interpretation and details of any abnormalities. A summary of the number and percentage of subjects with normal, abnormal not clinically significant, and abnormal clinically significant ECG results will be presented for Baseline, Week 16, Week 52 timepoints and at end of SFU.

Absolute values at Baseline, Week 16, Week 52 and end of SFU and change from Baseline at Week 16, Week 52 and at end of SFU of all ECG variable will also be summarized.

QTc outliers are defined as QTcF/QTcB values following dosing that are > 450 ms or that have increased from Baseline > 30 ms. These outliers will be summarized in a dedicated table using the following categories:

- Value >450 ms, >480 ms, >500 ms
- Increase from Baseline >30 ms, >60 ms, >90 ms

- Value >450 ms and increase from Baseline >30 ms, value >500 ms and increase from Baseline >60 ms.

Number and percentage of subjects who meet the ECG outlier criteria defined above at Week 16, Week 52 and end of treatment will be summarized.

All ECG tables will be provided on the SS by randomized treatment.

All 12-lead ECG data where QTcF and QTcB outliers are highlighted will also be listed.

10.5 Other safety variables

10.5.1 Electronic Columbia Suicide Severity Rating Scale (eC-SSRS)

The eC-SSRS is an assessment tool that evaluates suicidal ideation and behavior. The eC-SSRS contains 9 categories with binary responses (yes/no):

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

Following composite endpoints based on the above categories are defined as:

- Suicidal ideation: A "yes" answer at any time during treatment to any 1 of the 5 suicidal ideation questions (Categories 1-5).
- Suicidal behavior: A "yes" answer at any time during treatment to any 1 of the 4 suicidal behavior questions (Categories 6-9).
- Suicidal ideation or behavior: A "yes" answer at any time during treatment to any 1 of the 9 suicidal ideation and behavior questions (Categories 1-9).

Self-injurious behavior without suicidal intent is defined as event in the category non-suicidal self-injurious behavior (in category 9).

The visit identification for each assessment of eC-SSRS will not be part of the data collected and the visit date in eCRF will then be used to assign a visit identification to the eC-SSRS data.

For each subject, eC-SSRS assessment dates will be compared to eCRF visit dates.

In the Study Data Tabulation Model:

- If an 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. Otherwise, the eC-SSRS visit identification will be considered as “unscheduled”.

In the analysis datasets:

- If an eC-SSRS assessment date is matching an eCRF scheduled visit date within a window of +/- 2 days, then the eC-SSRS visit identification will be the one corresponding to the eCRF visit date. Otherwise, the eC-SSRS visit identification will stay as “unscheduled”.

The incidence of subjects with suicidal ideation, suicidal behavior, suicidal behavior or ideation, and self-injurious behavior without suicidal intent will be summarized by treatment and by visit for the SS during the DBP and the ATP. A by-subject listing of the eC-SSRS questionnaire data will be provided.

10.5.2 Assessment of tuberculosis

The laboratory test results from the TB assessment performed by interferon gamma release assay (for non-Japanese subjects) and by T-Spot test (for Japanese subjects) will be summarized by treatment group for Baseline (Screening) and Week 48. They will also be listed.

The results from the ‘Evaluation of signs and symptoms of tuberculosis’ questionnaire data will be also listed.

10.5.3 Change from Baseline in the Patient Health Questionnaire-9 (PHQ-9)

The PHQ-9 is a multipurpose instrument for Screening, diagnosing, monitoring, and measuring the severity of depression. The PHQ-9 score is based on 9 questions assessing the depression over the last 2 weeks through 9 criteria. Each criterion is scored from 0 (not at all) to 3 (nearly every day). The total score for the 9 questions is added up and provides the PHQ-9 score and can range from 0 to 27 with higher scores indicating worse state. If 1 of the 9 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 below 5 is considered to be none to minimal symptom of depression,
- A score of 5 to 9 is considered to be mild symptoms of depression,
- A score of 10 to 14 is considered to be moderate symptoms of depression,
- A score of 15 to 19 is considered to indicate moderately severe depression,
- A score ≥20 is considered to be severe depression.

The change from Baseline in PHQ-9 will be analyzed as Week 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52.

The percentage of subjects with scores below 5, between 5 and 9, between 10 and 14, between 15 and 19, and ≥20 in PHQ-9 will also be analyzed by visit.

Only observed will be presented for PHQ-9. results

11 REFERENCES

Fleiss JL, Tytun A, Ury SHK. A simple approximation for calculating sample sizes for comparing independent proportions. Biometrics. (1980): 36:343-6.

Moser, BK; Stevens, GR; Watts, CL. The Two-Sample T Test Versus Satterthwaite's Approximate F Test. Communication in Statistics- Theory and Methods 18(11):3963-3975

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.

Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio. American Journal of Epidemiology (1990);131(2):373-375.

Hosmer, D. W., Jr., and Lemeshow, S. (2000). Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons.

Maruish M, Kosinki M, Bjorner J, Gandek B, Turner-Bowker D, Ware J. User's manual for the SF36v2 Health Survey

12 APPENDICES

12.1 Appendix 1: MedDRA algorithmic approach to anaphylaxis (MedDRA V19.0)

The SMQ Anaphylactic reaction consists of 3 parts:

- A **narrow search** containing PTs that represent core anaphylactic reaction terms (Category A – core anaphylactic terms):
  - Anaphylactic reaction
  - Anaphylactic shock
  - Anaphylactic transfusion reaction
  - Anaphylactoid reaction
  - Anaphylactoid shock
  - Circulatory collapse
  - Dialysis membrane reaction
  - Kounis syndrome
  - Shock
  - Shock symptom
  - Type I hypersensitivity
- 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

Category B (Upper Airway / Respiratory Terms)

  - 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
- Mouth swelling
- Nasal obstruction
- Oedema mouth
- Oropharyngeal spasm
- Oropharyngeal swelling
- Respiratory arrest
- Respiratory distress
- Respiratory failure
- Reversible airways obstruction
- Sensation of foreign body
- Sneezing
- Stridor
- Swollen tongue
- Tachypnoea
- Throat tightness
- Tongue oedema
- Tracheal obstruction
- Tracheal oedema
- Upper airway obstruction
- Wheezing

Category C (Angioedema / Urticaria / Pruritus / Flush tems)

- 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
-

Category D (Cardiovascular / Hypotension Terms)

- 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 one of the following where both occur on either the same day as when an injection was administered or one day after, and for scenarios where two events must have been reported, both events must have occurred within one day of each other**:**
  - 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)]

12.2 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 UCB Opportunistic infections document (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) (IF APPLICABLE)

13.1 AMENDMENT 1

Rationale for the amendment

The primary purpose of this amendment was to implement changes in response to protocol amendments, discussions and feedback provided at meetings between UCB and Parexel technical teams for procedural clarifications, COVID-19 global pandemic impact and updates in guidelines for BKZ AE of special monitoring conventions. The main changes are described below:

13.1.1 Efficacy variables (Section 2.2.1)

Secondary efficacy variables : change of analysis population for the following endpoints:

- "Enthesitis-free state based on the Leeds Enthesitis Index (LEI) at Week 16 in the subgroup of subjects with enthesitis at Baseline" to be analyzed in the pooled population of PA0010 and PA0011.
- "Dactylitis-free state based on the Leeds Dactylitis Index (LDI) at Week 16 in the subgroup of subjects with dactylitis at Baseline" to be analyzed in the pooled population of PA0010 and PA0011.

Other efficacy variables : New variables

- Composite endpoint composed of ACR50 and PASI100 in subjects with PSO involving at least 3% BSA at Baseline
- Proportion of Psoriatic Arthritis Response Criteria (PsARC) responders
- Psoriatic Arthritis Disease Activity Score (PASDAS) categories
- Change from Baseline in the PASDAS
- Proportion of subjects with a decrease of HAQ-DI from Baseline of at least 0.35 in the subgroup of subjects with Baseline HAQ-DI≥0.35
- Proportion of FACIT-Fatigue subscale responders (subjects with a minimum clinically important difference for FACIT-Fatigue subscale score, defined as an increase of ≥4) in subjects with FACIT-Fatigue subscale score ≤48 at Baseline

Other efficacy variables : change of analysis population for the following endpoints:

- Dactylitis-free state based on the LDI in the subgroup of subjects with dactylitis at Baseline
- Enthesitis-free state based on the LEI in the subgroup of subjects with enthesitis at Baseline
- Enthesitis-free state based on the SPARCC index in the subgroup of subjects with enthesitis at Baseline
- Change from Baseline in the LEI in the subgroup of subjects with enthesitis at Baseline
- Change from Baseline in the SPARCC index in the subgroup of subjects with enthesitis at Baseline

- Change from Baseline in the LDI in the subgroup of subjects with dactylitis at Baseline

Other efficacy variables : Additional detail provided as below:

- Disease Activity Index for Psoriatic Arthritis (DAPSA) score **categories**
- Change from Baseline in PsAID-12 **total score, as well as** the individual domain scores
- PsAID-12 responders (**decrease** from Baseline in PsAID-12 total score ≥3) **in subjects with PsAID-12 total score>3 at Baseline**
- Change from Baseline in Functional Assessment of Chronic Illness Therapy (**FACIT)**-Fatigue **subscale score**

Other efficacy variable moved to the safety section:

- Change from baseline in PHQ-9

Other efficacy variable removed:

- Change from Baseline in the EQ-5D-3L utility score

Change of analysis method for the supportive analysis of primary endpoint, based on the analysis of the individual components of the ACR:

- Individual components of the ACR to be analyzed using the referenced based multiple imputation method.

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 variable:

- Change from Baseline in the Patient Health Questionnaire-9 (PHQ-9)

13.1.4 Protocol deviation (Section 3.5)

New category of protocol deviation:

- COVID-19 related IPD

13.1.5 Analysis sets (Section 3.6)

Additional subjects to be excluded from the PPS and the PK-PPS:

- Subject excluded from the FAS

New analysis set:

- COVID-19-free Set (subjects in the RS who had no COVID-19 impact up to the primary efficacy endpoint)

13.1.6 Strategy for handling missing data for efficacy analyses (Section 4.2.1)

Definition of an addition type of endpoint:

- Composite continuous endpoint

New table of continuous endpoints :

- New table listing for each non composite continuous endpoint, the minimum and maximum value and whether the endpoint should be integer.

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).

MI process

- All MI now performed up to Week 24 for the interim analysis 1 and Week 52 for the interim analysis 2

New statistics

- New method added describing how to obtain the estimated proportion of responders (ie, estimated responder rate) and the difference in the proportion of responders between BKZ and along with 2-sided 95% CIs.

Worst category imputation for categorical variables

- Worst category imputation introduced for categorical variable as the equivalence of the NRI concept for binary variables.

No order of priority in the analysis of other endpoints:

- Other efficacy endpoints will be analyzed using the same method as for the secondary efficacy variables

13.1.7 Data reported in interim analysis 1 (Section 4.3)

- For the efficacy analysis, only data collected up to W24 will be reported in tables.
- For the safety analysis, all data collected up to the cut-off point fixed for this analysis will be reported.

13.1.8 Testing hierarchy (Section 4.5)

- H7 is now: Enthesitis-free state superior to Placebo (based on pooled PA0010 and PA0011 W16 data)
- H8 is now: Dactylitis-free state superior to Placebo (based on pooled PA0010 and PA0011 W16 data)

13.1.9 Changes related to COVID-19 (Section 5.2)

- Additional disposition, demographic and baseline characteristics, efficacy, and safety analyses to be performed by COVID-19 pandemic periods (pre COVID-19 pandemic period / COVID-19 pandemic period / Post COVID-19 pandemic period).
- Now analyses on visits, endpoint collection impacted by the COVID-19 pandemic

13.1.10 Statistical analysis of the primary efficacy variables (Section 8.1)

Additional details for subgroup analyses

- NRI used instead of MI for subgroup analyses
- Addition details provided for subgroup analyses based on Bone Erosion at Baseline, region and ADAb categories
- Observed results to be provided in the context of subgroup analysis.

13.1.11 Statistical analysis of the secondary efficacy variables (Section 8.2)

Two extra rules added for the LDI calculation:

- Circumferences of 0 will be considered as missing. Circumferences ≤15 will be assumed to be in cm instead of mm and will be multiplied by 10 before being used in summaries/analyses.
- If a digit is recorded as 'not affected' but has circumference and contralateral circumferences collected, the tenderness score will be considered as 1 if the percent of difference between circumference and contralateral circumference is ≥to 10%.
- Introduction of the tenderness dactylitis count to be used for the PAS score calculation.

IGA

- New supportive analysis for subjects with missing IGA at some post-baseline visits when BSA<3% (to be considered as responders)
- New Analysis of BSA by visit

13.1.12 Statistical analysis of the other efficacy variables (Section 8.3)

ACR50 responders at Week 16 and maintaining response at post-Week 16 visits up to Week 52

- Extension of this analysis to all post-W16 visits and not only W52

WPAI-SHP

- Analysis of WPAI-SHP to be performed on OC only.

Additional analyses for other efficacy endpoints

- Addition of some extra analyses of efficacy endpoint with statistical testing

13.1.13 Extent of exposure (Section 10)

- Now also calculated by COVID-19 pandemic period

13.1.14 Safety topic of interest (Section 10.2.4)

- Hypersensitivity now includes injection site reactions in addition to anaphylaxis

13.2 AMENDMENT 2

Rationale for the amendment

The primary purpose of this amendment was to address discrepancies of the SAP amendment 1 vs. the table shells that are going to be produced for the Week 24 analysis.

Modifications and changes

Global change

Interim analysis 1 is now referenced as Week 24 analysis throughout the document. Interim analysis 2 is now referenced as Week 52 analysis throughout the document.

Change #1

Section 2.2.4: Immunological variables (last sentence)

These 3 variables will be assessed at Baseline, Week 4, 6, 8, 12, 16, 20, 24, 36, 52, end of treatment and at SFU visit.

Has been changed to:

These 3 variables will be assessed at Baseline, Week 4, 8, 12, 16, 20, 24, 36, 52, end of treatment and at SFU visit.

Change #2

Section 3.3: Definition of Baseline value (second paragraph from the end)

For subjects who switch treatment from placebo to BKZ at Week 16 (Visit 10), an additional Baseline value (called Week 16 Baseline) will be defined for efficacy variables. This Baseline value will be the last value collected before and up to the date of treatment switch (planned at Week 16), which corresponds to the beginning of the ATP.

Has been changed to:

For subjects who switch treatment from placebo to BKZ at Week 16 (Visit 10), an additional Baseline value (called Week 16 Baseline) will be defined for efficacy variables. **~~This Baseline value will be the last value collected before and up to the date of treatment switch (planned at Week 16), which corresponds to the beginning of the ATP.~~** **The Week 16 value will be**

considered as the Baseline value regardless of whether the Placebo subjects switched to BKZ on a later visit.

Change #3

Section 3.5: Protocol deviations

An important protocol deviation (IPD) is defined as a protocol deviation (PD) likely to have a meaningful impact on the study conduct, the primary efficacy outcome, the key safety or the PK outcomes for an individual subject. The IPDs will be identified and classified by the deviation types defined in the appropriate protocol-specific document. All PDs will be reviewed as part of the ongoing data cleaning and data evaluation process and IPD will be identified and documented prior to unblinding to confirm exclusion from Per Protocol Set (PPS) and the Pharmacokinetics Per Protocol Set (PK-PPS).

An IPD will not always necessitate the removal of a subject from the PPS. Depending on the type of PD, the subject may remain in the PPS. The IPD document developed by the clinical study team should clearly state which PD will result in a removal from the 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 DBP, ie only subjects with IPD observed prior to or at Week 16 (Visit 10) will be excluded from the PPS. Subjects with IPD after Week 16 will not be excluded from the PPS.

A specific category of PD as a consequence of the Coronavirus Disease 2019 (COVID-19) pandemic (called COVID-19 related PD) will be assessed based especially on the information collected on a dedicated electronic case report form (eCRF) page (and other sources).

Has been changed to:

An important protocol deviation (IPD) is defined as a protocol deviation (PD) likely to have a meaningful impact on the study conduct, the primary efficacy outcome, the key safety or the PK outcomes for an individual subject. The IPDs will be identified and classified by the deviation types defined in the appropriate protocol-specific document. All PDs will be reviewed as part of the ongoing data cleaning and data evaluation process and IPD will be identified and documented prior to unblinding to confirm exclusion from Per Protocol Set (PPS) and the Pharmacokinetics Per Protocol Set (PK-PPS).

An IPD will not always necessitate the removal of a subject from the PPS. Depending on the type of PD, the subject may remain in the PPS. The IPD document developed by the clinical study team should clearly state which PD will result in a removal from the PPS.

In addition to IPDs resulting in the exclusion from the PPS, subjects who reduce the dose or dosing frequency of certain medications due to intolerance, AE, side-effects or receive new prohibited medication for AE will be removed from the PPS. While this is not an IPD, as it is allowed per protocol for safety reasons, these subjects will be removed from the PPS as this non-PD could have an affect the primary efficacy outcome in the same way as flagged IPDs resulting in exclusion from the 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 DBP, ie only subjects with IPD **(or the non-PD above)** observed prior to or at Week 16 (Visit 10) will be excluded from the PPS. Subjects with IPD after Week 16 will not be excluded from the PPS.

A specific category of PD as a consequence of the Coronavirus Disease 2019 (COVID-19) pandemic (called COVID-19 related PD) will be assessed based especially on the information collected on a dedicated electronic case report form (eCRF) page (and other sources).

Change #4

Section 3.6.7: Per Protocol Set

The Per-Protocol Set (PPS) will consist of all subjects in the RS who had no IPD affecting the primary efficacy variable (only IPD observed prior to Week 16 are considered for exclusion from the PPS). The IPDs will be predefined and subjects with IPDs will be evaluated during ongoing data cleaning and data evaluation meetings prior to unblinding of the data (Section 3.5). Exclusion from the FAS will be considered as IPDs that also result in exclusion from the PPS.

Supportive analysis of the primary efficacy variable will be performed on the PPS.

Has been changed to:

The Per-Protocol Set (PPS) will consist of all subjects in the RS who had no IPD **or non-PD related to prohibited medications** affecting the primary efficacy variable (only IPD**/non-PD related to prohibited medications** observed prior to Week 16 are considered for exclusion from the PPS). The ~~**IPDs**~~ **deviations** will be predefined and subjects with ~~**IPDs**~~ **deviations** will be evaluated during ongoing data cleaning and data evaluation meetings prior to unblinding of the data (Section 3.5). Exclusion from the FAS will be considered as **an** IPD~~**s**~~ that also result**s** in exclusion from the PPS.

Supportive analysis of the primary efficacy variable will be performed on the PPS.

Change #5

Section 4.2.1: Strategy for handling missing data for efficacy analyses

Added line to Table 4.1

Efficacy endpoint	Minimum	Maximum	Integer value
Tender dactylitis count	**0**	**20**	**Yes**

Change #6

Section 4.2.2.3: MI-MCMC / Reference-based imputation (steps of the procedure)

In the case of continuous endpoints, the procedure will be implemented as follows on the raw values:

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 the placebo-treated subjects only.

a. Initialization. Set t=1 (Baseline visit).

b. Iteration. Set t=t+1. Create a dataset combining records from BKZ and placebo subjects with columns for covariates (bone erosion and region) and outcomes at visits 1 to t. Outcomes for all BKZ randomized subjects are set to missing at visit t and set to observed or previously imputed values at visits 1 to t-1. Outcomes for placebo-treated subjects 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.

c. Imputation. Impute missing values for visit t using previous outcomes for visits 1 to t-1, bone erosion at Baseline, and region. Note that only placebo data will be used to estimate the imputation model since no outcome is available for BKZ-treated subjects at visit t. As a consequence, the input dataset should include all subjects from Placebo but only subjects from the BKZ arm that have values at timepoint t missing.

d. Repeat steps 2a-2d for all timepoints, (…)

Has been changed to:

In the case of continuous endpoints, the procedure will be implemented as follows on the raw values:

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 **(Baseline)**,…, Week 16 (Week 16 being the time point of interest) using data from the placebo-treated subjects only.

a. Initialization. Set t=1 (Baseline visit).

Create a dataset combining all records from BKZ and placebo subjects with columns for covariates (bone erosion and region) and outcome at Baseline.

Impute missing values at Baseline using bone erosion at Baseline, and region. Note that both placebo and BKZ data will be used to estimate the imputation model.

b. Iteration. Set t=t+1. Create a dataset combining records from BKZ **subjects with missing data at visit t** and **all** placebo subjects with columns for covariates (bone erosion and region) and outcomes at visits 1 to t. **In this dataset**, outcomes for ~~**all**~~ BKZ randomized subjects are ~~**set to**~~ missing at visit t and ~~**set to**~~ observed or previously imputed ~~**values**~~ at visits 1 to t-1. Outcomes for placebo-treated subjects are ~~**set to**~~ observed **or missing** at visit t ~~**or**~~ **and** observed or previously imputed ~~**values**~~ at visits 1 to t-1. The outcomes should be sorted in chronological order in the model.

c. Imputation. Impute missing values for visit t using previous outcomes for visits 1 to t-1, bone erosion at Baseline, and region. Note that only placebo data will be used to estimate the imputation model since no outcome is available for BKZ-treated subjects at visit t. As a

consequence, the input dataset should include all subjects from Placebo but only subjects from the BKZ arm that have values at timepoint t missing.

d. Repeat steps **1b-1c** for all timepoints **after Baseline up to Week 16**, (…)

Change #7

Section 4.2.2.4: Tipping point analysis (Updated line to Table 4.3)

ACR component	Range	Delta
hs-CRP	Observed range of the $\log_e$ transformed values for all **BKZ and Placebo** subjects (at Week 16)	5% of the observed range

Change #8

Section 4.3.1 Interim analyses and final analysis (first bullet)

The Week 24 analysis is the first interim analysis after all randomized subjects have completed the DBP and the ATP up to Week 24 or have withdrawn from the IMP or the study before Week 24. The purpose of this analysis is to perform a comprehensive evaluation of all available data for the 3 treatment arms to prepare a regulatory submission for a marketing authorization application based on this analysis. This interim analysis will cover the analysis of all efficacy in tables on data collected up to Week 24 and the analysis of safety in tables on data collected up to the cut-off date (defined as Week 24 visit) for the last subject (all available data will be reported in listings). The type of efficacy and safety tables provided for the first interim analysis is detailed in Table 4-4.

Has been changed to:

The Week 24 analysis is the first interim analysis after all randomized subjects have completed the DBP and the ATP up to Week 24 or have withdrawn from the IMP or the study before Week 24. The purpose of this analysis is to perform a comprehensive evaluation of all available data for the 3 treatment arms to prepare a regulatory submission for a marketing authorization application based on this analysis. This interim analysis will cover the analysis of all efficacy ~~in~~ **and non-safety by visit tables** on data collected up to Week 24 and the analysis of safety in tables on data collected up to the cut-off date (defined as Week 24 visit) for the last subject (all available data will be reported in listings). The type of efficacy and safety tables provided for the first interim analysis is detailed in Table 4-4.

Change #9

Section 4.8 Examination of subgroups (5th bullet from the end)

- Concomitantly receiving MTX vs. cDMARDs at Baseline (concomitant MTX, no concomitant MTX and cDMARDs at Baseline, no concomitant MTX and no cDMARDs at Baseline)

Has been changed to:

- Concomitantly receiving MTX **at Baseline** vs. **other** cDMARDs at Baseline (~~**concomitant**~~ MTX **at Baseline**, no ~~**concomitant**~~ MTX **at Baseline** and cDMARDs at Baseline, no ~~**concomitant**~~ MTX **at Baseline** and no cDMARDs at Baseline)

Change #10

Section 5.1 Subject Disposition (New text after paragraph 4).

The numbers and percentages of these subjects who either complete the SFU visit or not will additionally be presented.

The numbers and percentages of randomized subjects entering the Open Label Extension (OLE) study and those not entering the OLE study will be presented. Of the randomized subjects not entering the OLE study,

Change #11

Section 5.3 Protocol deviations.

The definition of an IPD is given in Section 3.5

A summary displaying the number and percentage of subjects with an IPD will be provided by treatment group separately for the DBP on the RS and on the ATP on the ATS. This will include a summary of subjects excluded from the PPS (DBP only) and the PK-PPS due to IPD. The summary will be provided 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, COVID-19 related IPD).

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

A by-subject listing of IPDs will be provided for all subjects in the RS.

A by-subject listing of COVID-19 related PDs will be provided for all subjects in the RS.

Has been changed to:

The definition of an IPD is given in Section 3.5

A summary displaying the number and percentage of subjects with an IPD will be provided by treatment group separately for the DBP on the RS and on the ATP on the ATS. This will include a summary of subjects excluded from the PPS (DBP only) and the PK-PPS ~~**due to IPD**~~. The summary will be provided 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, COVID-19 related IPD) **as well as the number and percentage of subjects excluded from the PPS due to reason other than PD**.

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

A by-subject listing of IPDs will be provided for all subjects in the RS.

A by-subject listing of COVID-19 related PDs will be provided for all subjects in the RS.

A by-subject listing of subjects excluded from the PPS for reasons other than IPD will be provided.

Change #12

Section 6.2.2 Other baseline characteristics (14th bullet)

- NSAID at Baseline (Yes, No)
- Prior cDMARDs (0,1, >=2)

Has been changed to:

- NSAID at Baseline (Yes, No)
- **Past cDMARDs Therapy (Yes, No)**
- Prior cDMARDs (0,1, >=2)

Change #13

Section 7 Measure of treatment compliance (last sentence)

A summary of percent treatment compliance categorized as ≤75% and >75% will be provided by treatment group, as well as a by-subject listing of treatment compliance.

Has been changed to:

A summary of percent treatment compliance categorized as **<75%** and **>=75%** will be provided by treatment group, as well as a by-subject listing of treatment compliance.

Change #14

Section 8.1.2 Primary analysis of the primary efficacy variable (paragraph 8) and Section 8.2.2 Primary analysis of secondary efficacy variables (last sentence)

For the country specific analyses performed on subjects randomized in Japan, the region factor will not be considered as a covariate for the modelling.

Has been changed to:

For the country specific analyses performed on subjects randomized in Japan, the region factor will not be considered as a covariate for the modelling.

Considering that the number of subjects randomized in Japan is low (less than 10% of the Randomized Set), statistical models might not converge. If a model (Logistic Model or Mixed model) is not converging, all related adjusted statistics and p-value will not be presented : "NE" for "Not Evaluable" will be displayed instead.

Change #15

Section 8.2.1.4 MDA at Week 16 (last but one paragraph)

The following rule will be applied for subjects with BSA <3 at Baseline: Subjects with BSA <3 at Baseline will always meet the criteria PASI ≤1 or BSA ≤3 except in the cases where a BSA score ≥3 is observed.

Has been changed to:

The following rule will be applied for subjects with BSA <3 at Baseline: Subjects with BSA <3 at Baseline will always meet the criteria PASI ≤1 or BSA ≤3 except in the cases where a BSA score **>3** is observed.

Change #16

Section 8.2.1.5 Van den Hedje modified Total Sharp Score (vdHmTSS) : Change from Baseline at Week 16 (after the sixth bullet from the end).

Radiographic assessments should be conducted using a standardized imaging methodology as defined in an imaging charter. For the analysis, the average of the scores (ES, JSN score and total score) from the 2 independent reviewers will be utilized.

Has been changed to:

Radiographic assessments should be conducted using a standardized imaging methodology as defined in an imaging charter. **Subjects reaching Week 52 will have 2 sets of reads for Baseline and Week 16 as x-rays for these subjects are read more than once in different sessions to reduce within subject variability by each of the independent reviewers. For the interim analysis** first set of reads (initial Baseline and Week 16 read)**will be used for all subjects.** For the analysis, the average of the scores (ES, JSN score and total score) from the 2 independent reviewers will be utilized. They will be calculated as:

Change #17

Section 8.2.3 Supportive analyses of the secondary efficacy variables (second paragraph)

For PASI90, the MI will be run on the PASI score on subjects involving at least 3% of BSA at Baseline.

Has been changed to:

For PASI90, the MI will be run on the PASI score on subjects involving at least 3% of BSA at Baseline.

For the MDA (and VLDA) MI analysis, subjects with BSA <3 at Baseline will always meet the criteria PASI ≤1 or BSA ≤3 except in the cases where a BSA score >3 is observed. Subjects involving at least 3% of BSA at Baseline will have their BSA values imputed using MI when deriving MDA (VLDA).

Change #18

Section 8.2.3 Supportive analyses of the secondary efficacy variables (last sentence)

To assess the impact of the COVID-19 pandemic, the analysis of secondary efficacy endpoints included in the hierarchy with be repeated on the COVID-19-Free Set using the primary analysis method.

Has been changed to:

To assess the impact of the COVID-19 pandemic, the analysis of secondary efficacy endpoints included in the hierarchy with be repeated on the COVID-19-Free Set **(On subjects from the Radiographic Set included in the COVID-19-Free-Set for the analysis of VdHmTSS)** using the primary analysis method.

Change #19

Section 8.3.1.6 : Psoriatic Arthritis Disease Activity Score (PASDAS) categories and change from Baseline in PASDAS

PASDAS= ((0.18 × √Physician global VAS) + (0.159 × √Patient global VAS)
− (0.253 × √SF36 PCS) + (0.101 × LN(SJC + 1)) + (0.048 × LN(TJC + 1))
+ (0.23 × LN(LEI + 1)) + (0.377 × LN (Tender dactylitis count + 1))
+ (0.102 x LN(hs − CRP + 1) + 2))

Note: LN=natural logarithm.

If any of the PASDAS component is missing, PASDAS will be set to missing.

(…)

The following categories are used to define the level of disease activity:

• Remission: PASDAS <1.9

• Low disease activity: PASDAS 1.9 to <3.2

• Moderate disease activity: PASDAS 3.2 to <5.4

• High disease activity: PASDAS ≥5.4.

The PASDAS categories and the change from Baseline in PASDAS score will be analyzed at Week 4, 8, 12, 16, 20, 24, 28, 32, 36, 52.

Has been changed to:

PASDAS=((0.18 × √Physician global VAS) + (0.159 × √Patient global VAS)
− (0.253 × √SF36 PCS)
+ (0.101 × LN(SJC + 1))
+ (0.048 × LN(TJC + 1))
+ (0.23 × LN(LEI + 1))
+ (0.377 × LN (Tender dactylitis count + 1))
+ (0.102 x LN(hs−CRP + 1))
+ 2)***1.5**

Note: LN=natural logarithm.

If Tender Dactylitis count is missing because LDI=0 at Baseline, Tender Dactylitis count will be replaced by 0 in the formula.

If **for any other reason a** ~~**any of the**~~ PASDAS component is missing, PASDAS will be set to missing.

(…)

The following categories are used to define the level of disease activity:

• Remission: PASDAS ≤1.9

• Low disease activity: PASDAS >1.9 to <3.2

• Moderate disease activity: PASDAS 3.2 to <5.4

• High disease activity: PASDAS ≥5.4.

The PASDAS categories and the change from Baseline in PASDAS score will be analyzed at **Baseline, Week 4, Week 16, 24, 36, 52.**

Change #20

Section 8.3.2.2 : Other endpoints (excluding time to ACR20/50/70 response)

(….)

For binary endpoints using:

- The number and proportion of responders using the NRI approach.
- The mean proportion of responders in the multiple imputed datasets and the 95% CI (modified composite estimand). For ACR50 response at post-Week 16 visits for subjects responding at Week 16, the same imputed data as generated in the context of the analysis of the primary efficacy endpoint will be used.
  - For enthesitis and dactylitis free-state based on the pooled data of PA0010 and PA0011 subjects (for subjects in the BKZ 160mg Q4W and the Placebo group), the MI will be performed on the pooled dataset (using data from Baseline, Week 4, 8, 12 and 16) the same way as described in 4.2.2.2, but with the following covariates in that order for each SAS® PROC MI: region (North America, Western Europe and Eastern Europe in that order) and study-id (“1” for PA0011 and “0” for PA0010).

- For enthesitis based on LEI, the MI will be run on all subjects and results reported for subjects with LEI>0 at Baseline ; for other binary variables derived from continuous value(s), the MI will be run on the subset of subjects (if applicable) for which the variable will be reported (ie: for LDI, on subjects with LDI>0 at Baseline)

- The observed number and proportion of responders on subjects with observed data.

Has been changed to:

For binary endpoints using:

- The number and proportion of responders using the NRI approach.
- The mean proportion of responders in the multiple imputed datasets and the 95% CI (modified composite estimand). For ACR50 response at post-Week 16 visits for subjects responding at Week 16, the same imputed data as generated in the context of the analysis of the primary efficacy endpoint will be used.
  - For enthesitis and dactylitis free-state based on the pooled data of PA0010 and PA0011 subjects (for subjects in the BKZ 160mg Q4W and the Placebo group), the MI will be performed on the pooled dataset (using data from Baseline, Week 4, 8, 12 and 16) the same way as described in 4.2.2.2, but with the following covariates in that order for each SAS® PROC MI: region (North America, Western Europe and Eastern Europe in that order) and study-id ("1" for PA0011 and "0" for PA0010).
  - For enthesitis based on LEI, the MI will be run on all subjects and results reported for subjects with LEI>0 at Baseline ; for other binary variables derived from continuous value(s), the MI will be run on the subset of subjects (if applicable) for which the variable will be reported (ie: for LDI, on subjects with LDI>0 at Baseline)
  - **For ACR50 based on Week 16 Baseline, the MI will be run on Week 16 and post-week 16 values, on the ATBS for the Placebo group. Week 16 value will be considered as the Baseline value regardless of whether the Placebo subjects switched to BKZ on a later visit.**
  - **For PASI90 based on Week 16 Baseline, the MI will be run on W16 and post-week 16 values, on the ATBS for the Placebo group on subjects with PSO involving at least 3% of BSA at Week 16. Week 16 value will be considered as the Baseline value regardless of whether the Placebo subjects switched to BKZ on a later visit.**
- The observed number and proportion of responders on subjects with observed data.

Change #21

Section 8.3.2.2 Other endpoints (excluding time to ACR20/50/70 response) (paragraph about analysis of categorical endpoints)

For categorical endpoints (ie, EQ-5D-3L dimensions score, DAPSA state) using:

- The number and percentage by endpoint category using the NRI approach (as performed for the primary analysis of the primary endpoint with the exception that the NRI is replaced by the worst category imputation)
- The mean percentage by endpoint category in the multiple imputed datasets (modified composite estimand). Regardless of imputed values, the value of the categorical endpoint after IE are by default set to the worst category (ie: High Disease Activity for DAPSA state)

Has been changed to:

For categorical endpoints (ie, EQ-5D-3L dimensions score, DAPSA state, **PASDAS categories**) using:

- The number and percentage by endpoint category using the NRI approach (as performed for the primary analysis of the primary endpoint with the exception that the NRI is replaced by the worst category imputation)
- The mean percentage **and corresponding 95% CI** by endpoint category in the multiple imputed datasets (modified composite estimand). Regardless of imputed values, the value of the categorical endpoint after IE are by default set to the worst category (ie: High Disease Activity for DAPSA state)

Change #22

Section 8.3.2.3 : Additional statistical analysis for other efficacy endpoints (3rd and 4th bullet from the end)

- ACR50/PASI90 composite (Week16, 20, 52)
- ACR50/PASI100 composite (Week16, 20, 52)

Has been changed to:

- ACR50/PASI90 composite (Week16, ~~20~~ **24**, 52)
- ACR50/PASI100 composite (Week16, ~~20~~ **24**, 52)

Change #23

Section 9.1 : Pharmacokinetics (paragraph 3)

No imputation will be used for missing samples. However, if plasma concentration measurements are below the limit of quantification (BLQ), then for the calculation of the derived statistics, the result will be set to ½ of LLOQ (ie, 1/2*0.250 = 0.125 ug/mL). Descriptive statistics including number of values, geometric mean, its 95% CI, geometric coefficient of variation, mean, SD, median, minimum, and maximum. Geometric mean and its 95% CI, geometric CV, mean and SD will be calculated if at least ⅔ of the values of interest is above the LLOQ ; otherwise, only number of values, median, minimum, and maximum will be presented.

Has been changed to:

No imputation will be used for missing samples. However, if plasma concentration measurements are below the limit of quantification (BLQ), then for the calculation of the derived statistics, the result will be set to ½ of LLOQ (ie, 1/2*0.250 = 0.125 ug/mL). Descriptive statistics including number of values, geometric mean, its 95% CI, geometric coefficient of variation, mean, SD, median, minimum, and maximum. Geometric mean and its 95% CI, geometric CV, mean and SD will be calculated if at least ⅔ of the values of interest is above the LLOQ **and number of values>=3**; otherwise, only number of values, median, minimum, and maximum will be presented.

Change #24

Section 9.1 : Pharmacokinetics (Third bullet)

- The table summary and figures will be primarily repeated by anti-BKZ antibody status (positive, negative, missing) by treatment group (3 lines per graph, one graph per treatment group). The missing group will not be displayed if ≥95% of subjects are categorized in the non-missing groups.

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

  – a subject will be counted as positive from the first visit at which the subject 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 subject has only negative ADAb samples or only one missing/inconclusive sample with negative ADAb samples up to that timepoint, the subject will be classified as negative.

Has been changed to:

- The table summary and figures will be primarily repeated by anti-BKZ antibody status (positive, negative, missing) by treatment group (3 lines per graph, one graph per treatment group). The missing group will not be displayed if ≥95% of subjects are categorized in the non-missing groups.

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

  – a subject will be counted as positive from the first visit at which the subject 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 subject has only negative ADAb samples or only one missing/inconclusive sample with negative ADAb samples up to that timepoint, the subject will be classified as negative. **An exception remains for the Baseline Visit where only one sample would be available. If the sample is missing/inconclusive, then the sample will be classified as being negative for the cumulative ADAb status.**

Change #25

Section 9.1 : Pharmacokinetics (last sentence)

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

Has been changed to:

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

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.

Change #26

Section 9.2.1 : Anti-bimekizumab antibody (line before the description of the ADAb categories)

Summaries of cumulative ADAb status and time to treatment-emergent positivity will use all available data.

Has been changed to:

All other Summaries of ~~**cumulative**~~ ADAb status ~~**and time to treatment-emergent positivity**~~ will use all available data **(scheduled and unscheduled)**.

Change #27

Section 9.2.1 : Anti-bimekizumab antibody (ADAb categories)

- **Category 1: Pre ADAb negative – treatment-emergent ADAb negative**: Includes subjects who are negative at Baseline and antibody negative at all sampling points post treatment (including SFU). This group also includes subjects who have missing/inconclusive pre-treatment sample (eg either missing/inconclusive or insufficient volume) at baseline with all post-baseline samples as ADAb negative.

(…)

- **Category 6: Inconclusive**: Includes subjects who have a positive pre-treatment sample and some post-treatment samples are missing, while other post-treatment samples are ADAb negative or missing.

(…)

- **Category 9: Missing:** Includes subjects who have a negative or a missing/inconclusive pre-treatment sample and some post-treatment samples are missing/inconclusive, while other post-treatment samples are ADAb negative.

Has been changed to:

- **Category 1: Pre ADAb negative – treatment-emergent ADAb negative**: Includes subjects who are negative at Baseline and antibody negative at all sampling points post treatment (including SFU), **one post-baseline missing/inconclusive sample is allowed for subjects with pre-ADAb negative sample**. This group also includes subjects who have missing/inconclusive pre-treatment sample (eg either missing/inconclusive or insufficient volume) at baseline with all post-baseline samples as ADAb negative.

(…)

- **Category 6: Inconclusive**: Includes subjects who have a positive pre-treatment sample and some post-treatment samples are missing, while other post-treatment samples are ADAb negative **or missing**.

(…)

- **Category 9: Missing:** Includes subjects who have a negative or a missing/inconclusive pre-treatment sample and some post-treatment samples are missing/inconclusive, while other post-treatment samples are ADAb negative **or missing**.

Change #28

Section 9.2.1 : Anti-bimekizumab antibody (First Bullet)

- A summary table displaying the number and percentage of subjects with ADAb status (positive, negative or missing) at each visit and overall will be provided. For the overall summary at any visit, 2 summaries will be presented:
  - Including any visit during the treatment period (as defined in Section 3.1). 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 Week 16, 24 and 52 by treatment group (BKZ 160 mg Q4W, Placebo / BKZ 160 mg Q4W).
  - Including any visit during the study. Thus, this summary will include data obtained at both the SFU visit and at Baseline.

Has been changed to:

- A summary table displaying the number and percentage of subjects with ADAb status (positive, negative or missing) at each visit and overall will be provided. For the overall summary at any visit, 2 summaries will be presented:
  - Including any visit during the treatment period (as defined in Section 3.1). 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 Week 16, 24 and 52 by treatment group (BKZ 160 mg Q4W, Placebo / BKZ 160 mg Q4W).
  - Including any visit during the study. Thus, this summary will include data obtained at both the SFU visit and at Baseline**. Subjects who rolled over to the OLE study will not have a SFU visit per study protocol. For these subjects, the 'overall ADAb status**

including SFU' will then be considered as being identical to the 'overall ADAb status up to Week 52'.

Change #29

Section 9.2.1 : Anti-bimekizumab antibody (Second, third and fourth bullet)

- The above summary table displaying the number and percentage of subjects in each of the ADAb status (positive, negative, missing) will be repeated by region for Placebo / BKZ 160 mg Q4W and BKZ 160 mg Q4W groups. The overall summary excluding SFU will be generated for up to Week 16, 24 and 52.
- A summary table displaying the number and percentage of subjects in each of the ADAb status (positive, negative, missing) by concomitant medications (use of cDMARDs at entry, use of MTX at entry, use of oral/systemic corticosteroids at entry) will be provided (overall summaries only, including summaries up to Week 16, 24 and 52).
- A table displaying the number (%) of subjects with the first occurrence of ADAb treatment-emergent positivity during the study (ie: including Baseline visit) will be summarized. This summary will include the following categories:
  - Any ADAb+: ADA positive sample regardless of category during the treatment period,
  - ADAb Category 2: Pre ADAb negative – treatment-emergent ADAb positive,
  - ADAb Category 5: Pre ADAb positive – treatment-boosted ADAb positive (if this category represents at least 10% of the subjects, otherwise results are to be merged with category 2). For this category, the first occurrence of a boosted result is considered.

Has been changed to:

- The above summary table displaying the number and percentage of subjects in each of the ADAb status (positive, negative, **total of positive and negative**, missing) will be repeated by region for Placebo / BKZ 160 mg Q4W and BKZ 160 mg Q4W groups. The overall summary excluding SFU will be generated for up to Week 16, 24 and 52.
- A summary table displaying the number and percentage of subjects in each of the ADAb status (positive, negative, **total of positive and negative**, missing) by concomitant medications (use of cDMARDs at entry, use of MTX at entry, use of oral/systemic corticosteroids at entry) will be provided (overall summaries only, including summaries up to Week 16, 24 ~~and~~ 52 **and SFU**).
- A table displaying the number (%) of subjects with the first occurrence of **any ADAb positivity (ie including Baseline visit) or** ADAb treatment-emergent positivity during the study ~~**(ie: including Baseline visit)**~~ will be summarized. This summary will include the following categories:
  - Any ADAb+: ADA positive sample regardless of category during the treatment period,
  - ADAb Category 2: Pre ADAb negative – treatment-emergent ADAb positive,

– ADAb Category 5: Pre ADAb positive – treatment-boosted ADAb positive (if this category represents at least 10% of the subjects, otherwise results are to be merged with category 2). For this category, the first occurrence of a boosted result is considered.

Change #30

Section 9.2.1 : Anti-bimekizumab antibody (eighth, nineth, and tenth bullet)

- Individual plots (one plot by subject) of BKZ concentrations/ADAb titer and ACR50 response (based on NRI) will be created. All 3 endpoints will be plotted on the Y-axis by visit (x-axis) for the full treatment period, including SFU where a patient has not progressed into the OLE. Plots should be labeled and grouped into the 9 ADAb categories and display vertical lines on the x-axis corresponding to the dosing incidences/visits.
  For the Week 24 interim analysis, the ADAb categories will be based on data up to Week 24.
- Individual plots (1 plot by subject) of BKZ concentrations/ADAb titer and PASI90 will also be created similarly as described above.
- Spaghetti plots of ADAb titer (y-axis) will be reported on the log-scale by visit (x-axis) by treatment group for all ADAb positive subjects. This plot will include 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).

Has been changed to:

- Individual plots (one plot by subject) of BKZ concentrations/ADAb titer and ACR50 response (based on NRI) will be created. All 3 endpoints will be plotted on the Y-axis by visit (x-axis) for the full treatment period, including SFU where a patient has not progressed into the OLE. Plots should be labeled and grouped into the 9 ADAb categories and display vertical lines on the x-axis corresponding to the dosing incidences/visits.
  For the Week 24 interim analysis, the ADAb categories will be based on data up to Week 24.
- Individual plots (1 plot by subject) of BKZ concentrations/ADAb titer and PASI90 will also be created similarly as described above.
- Spaghetti plots of ADAb titer (y-axis) will be reported on the log-scale by visit (x-axis) by treatment group for all ADAb positive subjects. This plot will include the following ADAb categories:
  – Category 2: Pre ADAb negative – treatment-emergent ADAb positive,
  – Category 5: Pre ADAb positive – treatment-boosted ADAb positive.

  The 3 plots **described above** will ~~**be**~~ present~~**ed**~~ **ADAb titers** using a semi-logarithmic scale ~~**for the ADAb titers**~~ (ADAb negative samples will therefore be excluded from the plot**s**).
  Furthermore, all ADAb titer values<100 will be represented as 1 in these plots.

Change #31

Section 9.2.1 : Anti-bimekizumab antibody (last paragraph)

- Finally, all individual subject-level ADAb results will be listed including the Screening assay, confirmatory assay, ADAb 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 for ADAb measurements that were excluded from the by-visit summaries. The reason for exclusion will be one of the following:
  - Sample collected out of window relative to current dose (or visit),
  - More than one sample obtained at the same visit.

Has been changed to:

- ~~**Finally,**~~ All individual subject-level ADAb results will be listed including the Screening assay, confirmatory assay, ADAb 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 for ADAb measurements that were excluded from the by-visit summaries. The reason for exclusion will be one of the following:
  - Sample collected out of window relative to current dose (or visit),
  - More than one sample obtained at the same visit.
- **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 for excluded PK results.**

Change #32

Section 9.2.2 : Neutralizing anti-bimekizumab antibodies (first bullet)

- NAb negative: No NAb positive samples for IL-17AA and IL-17FF at Baseline or post-Baseline. This group will also include subjects 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 classified as follows:
  - ADAb positive / NAb negative: ADAb positive subjects 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 subject has all the samples as ADAb negative or only one missing/inconclusive sample with all other available samples as negative ADAb.

Has been changed to:

- NAb negative: No NAb positive samples for IL-17AA and IL-17FF at Baseline or post-Baseline. This group will also include subjects 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 classified as follows:

- ADAb positive / NAb negative: ADAb positive subjects 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 subject 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 subject to the neutralizing assay.**

Change #33

Section 9.2.2 : Neutralizing anti-bimekizumab antibodies (Third bullet)

- NAb Missing:
  - >1 relevant NAb samples are missing and other available NAb samples during the period of interest are negative, eg, missing or insufficient sample left for NAb testing.

Has been changed to:

- NAb Missing:
  - >1 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.

Change #34

Section 10.2.1 : Standard AE summaries (18th bullet)

- TEAEs by timing of onset relative to ADAb status by SOC, HLT, PT (on subjects treated with BKZ). This will include columns for the following:
  - TEAEs starting before the first ADAb positive result (includes ADAb categories 2, 4 and 5) where TEAEs have occurred before the following events: a) the first positive ADAb result for subjects in category 2 and b) the first post-baseline boosted ADAb titer result for subjects with titer results and the first post-Baseline positive ADAb result for subjects with positive ADAb at Baseline with no other samples with titer available for subjects in category 5.

Has been changed to:

- TEAEs by timing of onset relative to ADAb status by SOC, HLT, PT (on subjects treated with BKZ). This will include columns for the following:
  - TEAEs starting before the first ADAb positive result (includes ADAb categories 2, ~~4~~ and 5) where TEAEs have occurred before the following events: a) the first positive ADAb result for subjects in category 2 and b) the first post-baseline boosted ADAb titer result for subjects with titer results and the first post-Baseline positive ADAb result for subjects

with positive ADAb at Baseline with no other samples with titer available for subjects in category 5.

Change #35

Section 10.3 : Clinical laboratory evaluations (2nd bullet after Table 10-5)

- Number and percentage of subjects by CTCAE grade (version 4.03) (when applicable) based on minimum/maximum post Baseline value (for blood chemistry and hematology).

Has been changed to:

- Number and percentage of subjects by CTCAE grade (version 4.03) (when applicable) based on minimum/maximum post Baseline value (for blood chemistry and hematology). **Subjects who meet the decreased potassium criterion of 3.0-<LLN, which is specified as the decreased potassium lab criterion for both CTCAE Grade 1 and Grade 2, will be counted as Grade 2.**

13.3 AMENDMENT 3

Rationale for the amendment

The primary purpose of this amendment was to address discrepancies of the SAP amendment 2 vs. the table shells that are going to be produced for the Week 52 analysis.

Modifications and changes

Change #1

Section 4.2.2.2: MI – MCMC/Monotone Regression

(…) For the MCMC method (when imputing intermittent missing values), randomization strata will be re-coded as indicator variables and will always be specified in the following order: North America region (1 for North America, 0 otherwise), Western Europe region (1 for Western Europe,0 otherwise), Eastern Europe region (1 for Eastern Europe, 0 otherwise), For each imputation model, Bone erosion at Baseline (0 if bone erosion at Baseline =0 ; 1 if bone erosion at Baseline≥1).

For each imputation model, the randomization strata (region and bone erosion at Baseline) may be dropped to facilitate convergence if required as described below: If the MI fails to converge when adjusting on the stratification variables specified above, or if the percentage of randomized subjects is < 10% in either of the Asia or Western Europe regions, then the Western Europe and Asia region will be combined and the binary variables for region that will be left in the MI will then be for North America and Eastern Europe regions.

Has been changed to:

(…) For the MCMC method (when imputing intermittent missing values), randomization strata will be re-coded as indicator variables and will always be specified in the following order: North America region (1 for North America, 0 otherwise), Western Europe region (1 for Western Europe,0 otherwise), Eastern Europe region (1 for Eastern Europe, 0 otherwise),

For each imputation model, Bone erosion at Baseline (0 if bone erosion at Baseline =0 ; 1 if bone erosion at Baseline≥1).

Bone erosions will be read for all visits both at the Week 24 analysis and the Week 52 / final analyses to ensure that unnecessary bias is not introduced. Given there may be differences at Baseline for the two sets of reads, the following rules will be applied for the MI process:

- **For the Week 24 analysis, the first set of reads will be used.**
- **For the Week 52 and final analyses, Week 16 specific outputs will continue to use the first set of reads. Outputs involving over time data up to Week 52 will use the second set of reads.**

For each imputation model, the randomization strata (region and bone erosion at Baseline) may be dropped to facilitate convergence if required as described below: If the MI fails to converge when adjusting on the stratification variables specified above, or if the percentage of randomized subjects is < 10% in either of the Asia or Western Europe regions, then the Western Europe and Asia region will be combined and the binary variables for region that will be left in the MI will then be for North America and Eastern Europe regions.

Change #2

Section 4.2.2.2: MI – MCMC/Monotone Regression

If a variable is dropped in order to allow convergence for one model in the study, that variable does not have to be dropped from other models in the study of the model converges without dropping the variable. That is, model convergence should be evaluated for each efficacy table independently. Furthermore, this means that a table based on a single timepoint may have different variables in the imputation model than a separate by-visit table for the same variable.

Has been changed to:

If a variable is dropped in order to allow convergence for one model in the study, that variable does not have to be dropped from other models in the study of the model converges without dropping the variable. That is, model convergence should be evaluated for each efficacy table independently. ~~**Furthermore, this means that a table based on a single timepoint may have different variables in the imputation model than a separate by-visit table for the same variable.**~~

Change #3

Section 4.2.2.2: MI – MCMC/Monotone Regression

Step 2: Standard rounding rules will be applied to the imputed values after each SAS® PROC MI. If an imputed value falls outside of the range for the given variable (as listed in Table 4-1) the value will be updated to be within the predefined range.

Has been changed to:

Step 2: ~~**Standard rounding rules will be applied to the imputed values after each SAS® PROC MI.**~~ If an imputed value falls outside of the range for the given variable (as listed in Table 4-1) the value will be updated to be within the predefined range.

Change #4

Section 4.2.2.5: Treatment Policy Strategy

Subjects 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-MCMC / monotone regression (Section 4.2.2.2).

Has been changed to:

Subjects 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-MCMC / monotone regression (Section 4.2.2.2) **based on non missing data collected up to Week 16.**

Change #5

Section 4.8 Examination of subgroups

In the context of submission to Japanese health authorities, a subset of tables and figures will be provided in the subgroup of subjects randomized in Japan.

Has been changed to:

In the context of submission to Japanese health authorities, a subset of tables and figures will be provided in the subgroup of subjects randomized in Japan **(when MI is used, data for subjects randomized in Japan will be extracted from the MI datasets generated for the full population)**.

Change #6

Section 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, dates of visit, relationship to COVID-19, impact category and a narrative (short description) of the event. These data will be summarized for non-randomized subjects and by treatment group and overall, for enrolled subjects.

Has been changed to:

A listing of visits affected by COVID-19 will be presented based on the Enrolled Set including the visit, dates of visit, relationship to COVID-19, impact category and a narrative (short description) of the event. These data will be summarized ~~**for non-randomized subjects and**~~ by treatment group and overall, for ~~**enrolled**~~ **randomized** subjects.

Change #7

Section 8.1.2 Primary analysis of the primary efficacy variable

A logistic regression model will be used to assess the treatment effect on ACR50 response at Week 16. The model will include a fixed effect for treatment. The suitability of including randomization stratification variables (bone erosion at Baseline and region) will be assessed using Pearson and Deviance and The Hosmer-Lemeshow Goodness-of-Fit Tests (Hosmer and Lemeshow, 2000).

Has been changed to:

A logistic regression model will be used to assess the treatment effect on ACR50 response at Week 16. The model will include a fixed effect for treatment. The suitability of including randomization stratification variables (bone erosion at Baseline and region) will be assessed using Pearson and Deviance and The Hosmer-Lemeshow Goodness-of-Fit Tests (Hosmer and Lemeshow, 2000).

In the context of the Week 24 analysis, actual bone erosion at Baseline as a covariate will be based on data from the first set of reads. For Week 52 analysis, data from the second set of reads will be used.

Change #8

Section 8.1.3 Subgroup 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.

Has been changed to:

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

The variables for subgroup analyses are defined in Section 4.8.

In the context of the Week 24 analysis, actual bone erosion at Baseline used as a covariate and/or a subgroup will be based on data from the first set of reads. For the Week 52/final analyses, data from the second set of reads will be used.

Change #9

Section 8.2.1.5 Van der Heijde modified Total Sharp Score (vdHmTSS): Change from Baseline at Week 16.

(…)

Radiographic assessments should be conducted using a standardized imaging methodology as defined in an imaging charter. Subjects reaching Week 52 will have 2 sets of reads for Baseline and Week 16 as x-rays for these subjects are read more than once in different sessions to reduce within subject variability by each of the independent reviewers. For the Week 24 analysis first set of reads (initial Baseline and Week 16 read) will be used for all subjects.

(…)

In the primary analysis of these 2 endpoints, missing data at Week 16 will be imputed based on value at Screening (as there is no value collected in between) using MI-MCMC / Monotone regression. Since VdHmTSS is only collected at Baseline prior to Week 16, only the second MI step to address monotone missing pattern and generating 100 imputations will be processed (Section 4.2.2.2).

Has been changed to:

(…)

Radiographic assessments should be conducted using a standardized imaging methodology as defined in an imaging charter. Subjects reaching Week 52 will have 2 sets of reads for Baseline and Week 16 as x-rays for these subjects are read more than once in different sessions to reduce within subject variability by each of the independent reviewers.

Bone erosions will be read for all visits both at the Week 24 analysis and the Week 52/final analyses to ensure that unnecessary bias is not introduced. Given there may be differences for the two sets of reads, the following rules will be applied for the derivation:

- For the ~~**Week 24**~~ analysis **involving data up to Week 16 only (Week 24 analysis), the** first set of reads ~~**(initial Baseline and Week 16 read)**~~ will be used for all subjects.
- **For the analysis involving data up to Week 52, the second set of reads will be used for all subjects.**

(…)

For the Week 24 interim analysis only, in the primary analysis of these 2 endpoints, missing data at Week 16 will be imputed based on value at Screening (as there is no value collected in between) using MI-MCMC / Monotone regression. Since VdHmTSS is only collected at Baseline prior to Week 16, only the second MI step to address monotone missing pattern and generating 100 imputations will be processed (Section 4.2.2.2).

Change #10

Section 8.3.2.2 Other endpoints (excluding time to ACR20/50/70 response)

For binary endpoint: (…)

- For enthesitis based on LEI, the MI will be run on all subjects and results reported for subjects with LEI>0 at Baseline ; for other binary variables derived from continuous value(s), the MI will be run on the subset of subjects (if applicable) for which the variable will be reported (ie: for LDI, on subjects with LDI>0 at Baseline)

(…)

For continuous endpoints, on absolute values and changes from Baseline (or ratio from Baseline for hs-CRP) using: (…)

– For LEI, the MI will be run on all subjects and results reported for subjects with LEI>0 at Baseline ; for other continuous variables derived from continuous value(s), the MI will be run on the subset of subjects (if applicable) for which the variable will be reported (ie: for LDI, on subjects with LDI>0 at Baseline)

Has been changed to:

For binary endpoint: (…)

– For enthesitis based on LEI, the MI will be run on all subjects and results reported for subjects with LEI>0 at Baseline ; for other binary variables derived from continuous value(s), the MI will be run on the subset of subjects (if applicable) for which the variable will be reported (ie: for LDI **and Tender dactylitis count in the context of PASDAS calculation,** on subjects with LDI>0 at Baseline)

(…)

For continuous endpoints, on absolute values and changes from Baseline (or ratio from Baseline for hs-CRP) using: (…)

– For LEI, the MI will be run on all subjects and results reported for subjects with LEI>0 at Baseline ; for other continuous variables derived from continuous value(s), the MI will be run on the subset of subjects (if applicable) for which the variable will be reported (ie: for LDI, on subjects with LDI>0 at Baseline)

– **For Tender dactylitis count in the context of PASDAS calculation, if the MI does not converge, a conservative approach will be used by first imputing a value of 1 for all joints with missing tenderness score.**

Change #11

Section 8.3.2.3 Additional statistical analysis for other efficacy endpoints

- MDA (Week 2, 4, 8, 12, 16)
- PtAAP (Week 2, 4, 8, 12, 16)
- PsAQoL (Week 2, 4,8,12,16)
- PsAID-12 (Week 2, 4, 8, 12, 16)
- FACIT-Fatigue subscale (Week 2, 4, 8, 12, 16)

Has been changed to:

- MDA (Week ~~**2,**~~ 4, 8, 12, 16)
- PtAAP (Week 2, 4, 8, 12, 16)
- PsAQoL (Week ~~**2,**~~ 4, ~~**8, 12**~~,16)
- PsAID-12 (Week ~~**2,**~~ 4, ~~**8, 12**~~,16)
- FACIT-Fatigue subscale (Week ~~**2,**~~ 4, ~~**8, 12**~~,16)

Change #12

Section 9.1 Pharmacokinetics

(…)

- BKZ concentrations will be summarized by treatment group, globally and by region (Japan vs. rest of the world excluding Japan).

No imputation will be used for missing samples. However, if plasma concentration measurements are below the limit of quantification (BLQ), then for the calculation of the derived statistics, the result will be set to ½ of LLOQ (ie, 1/2*0.250 = 0.125 ug/mL). Descriptive statistics including number of values, geometric mean, its 95% CI, geometric coefficient of variation, mean, SD, median, minimum, and maximum. Geometric mean and its 95% CI, geometric CV, mean and SD will be calculated if at least ⅔ of the values of interest is above the LLOQ and number of values ≥3 ; otherwise, only number of values, median, minimum, and maximum will be presented.

- In addition, geometric mean of BKZ plasma concentration (with 95% CI) time curves will be plotted on linear and semi-logarithmic scales:
  - by treatment group (2 lines per graph, one for each treatment),
  - by treatment group and region (Japan vs. rest of the world excluding Japan) (for each treatment: 2 lines per graph, one for each region).
- The table summary and figures will be primarily repeated by anti-BKZ antibody status (positive, negative, missing) by treatment group (3 lines per graph, one graph per treatment group). The missing group will not be displayed if ≥95% of subjects are categorized in the non-missing groups.

(…)

PK samples collected at scheduled 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:

(…)

- The BKZ concentrations will also be listed for all subjects in 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 where the reason for exclusion will be one or more of the following:

(…)

Has been changed to:

Section 9.1 Pharmacokinetics

(…)

- BKZ concentrations will be summarized by treatment group, globally and **on subjects randomized in Japan** ~~**by region (Japan vs. rest of the world excluding Japan).**~~

No imputation will be used for missing samples. However, if plasma concentration measurements are below the limit of quantification (BLQ), then for the calculation of the derived statistics, the result will be set to ½ of LLOQ (ie, 1/2*0.250 = 0.125 **µg**/mL). Descriptive statistics including number of values, geometric mean, its 95% CI, geometric coefficient of variation, mean, SD, median, minimum, and maximum. Geometric mean and its 95% CI, geometric CV, mean and SD will be calculated if at least ⅔ of the values of interest is above the LLOQ and number of values ≥3 **at the respective timepoint**; otherwise, only number of values, median, minimum, and maximum will be presented.

- In addition, geometric mean of BKZ plasma concentration (with 95% CI) time curves will be plotted on linear and semi-logarithmic scales:
  - by treatment group (2 lines per graph, one **line** for each treatment),
  - ~~**by treatment group and region (Japan vs. rest of the world excluding Japan) (for each treatment: 2 lines per graph, one for each region).**~~
- The table summary and figures will be primarily repeated by anti-BKZ antibody status (positive, negative, missing) by treatment group (3 lines per graph, **one line for each anti-BKZ antibody status,** one graph per treatment group). The missing group will not be displayed if ≥95% of subjects are categorized in the non-missing groups.

(…)

PK samples collected at scheduled 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:

(…)

- The BKZ concentrations will also be listed for all subjects in the SS **(for BKZ treated subjects).** 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 where the reason for exclusion will be one or more of the following:

(…)

Change #13

Section 9.2.1 Anti-bimekizumab antibody

- A summary table displaying the number and percentage of subjects with ADAb status (positive, negative or missing) at each visit and overall will be provided. For the overall summary at any visit, 2 summaries will be presented:
  - Including any visit during the treatment period (as defined in Section 3.1). 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 Week 16, 24 and 52 by treatment group (BKZ 160 mg Q4W, Placebo / BKZ 160 mg Q4W).

(…)

- The above summary table displaying the number and percentage of subjects in each of the ADAb status (positive, negative, total of positive and negative, missing) will be repeated by region for Placebo / BKZ 160 mg Q4W and BKZ 160 mg Q4W groups. The overall summary excluding SFU will be generated for up to Week 16, 24 and 52.

(…)

- A table displaying the number (%) of subjects with the first occurrence of any ADAb positivity (ie: including Baseline visit) or ADAb treatment-emergent positivity during the study will be summarized.

(…)

- A boxplot of the ADAb titer by time of occurrence of ADAb positivity will be created for each dose group. The ADAb titer results will be presented on the log-scale. The time to occurrence is defined as the time in weeks from Baseline until the visit of interest when a sample is ADAb positive. Subjects who do not have any ADAb positivity will be excluded from the plot.

(…)

- Some figures summarizing the time to achieve any ADAb positivity and treatment emergent ADAb positivity on a cumulative basis will be presented by treatment group (2 lines per plot, one plot per treatment group). Subjects will be considered to have an event at the timepoint 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).
  If ≥10% subjects are classified as ADAb category 5, the lines will be further split by the following categories (thus the plot by treatment group will include 3 lines):
  - ADAb category 2: Pre ADAb negative – treatment-emergent ADAb positive,
  - ADAb category 5: Pre ADAb positive – treatment-boosted ADAb positive.
- Individual plots (one plot by subject) of BKZ concentrations/ADAb titer and ACR50 response (based on NRI) will be created. All 3 endpoints will be plotted on the Y-axis by visit (x-axis) for the full treatment period, including SFU where a patient has not progressed into the OLE. Plots should be labeled and grouped into the 9 ADAb categories and display vertical lines on the x-axis corresponding to the dosing incidences/visits.
  For the Week 24 analysis, the ADAb categories will be based on data up to Week 24.
- Individual plots (1 plot by subject) of BKZ concentrations/ADAb titer and PASI90 will also be created similarly as described above.

(…)

- Figures will summarize efficacy response (ACR50, ACR20 and PASI90 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 ACR50 (ACR20 or PASI90) responders at Week 16. The plot will therefore display the percentage of ACR50 (ACR20 or PASI90) responders as a function of the number of subjects within each ADAb and titer categories. Subjects with negative ADAb results at Week 16 will be included in the 'negative' category on the x-axis. This plot will include data only from subjects randomized to BKZ 160 mg Q4W.
  The figure will be repeated using Week 24 and 52 data (ie Week 24 and 52 ACR50 (and ACR20, PASI90) responder rates and Week 24 and 52 ADAb titer quartiles, respectively) and will include all subjects receiving BKZ. The figure for Week 24 and 52 data will be split by treatment group. Thus, 2 figures per efficacy variable and time point will be presented.
- Two figures (side by side) summarizing time course of efficacy response (ACR50 responders based on NRI) for each treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W) by the following ADAb groups (3 lines per plot) will be created:
  - ADAb positive - Defined as subjects having at least 2 ADAb positive samples during the study (including baseline and SFU) regardless of other ADAb negative samples and/or missing or inconclusive samples
  - ADAb negative - Defined as subjects for whom either (1) all samples (including baseline and SFU) are ADAb negative and there are no missing or inconclusive samples, (2) only 1 sample is ADAb positive and all other samples (including baseline and SFU) are ADAb negative or missing/inconclusive or (3) only one sample is missing/inconclusive and the remaining ADAb samples are negative.
  - Missing - Defined as subjects who do not fulfil the criteria for one of the 2 groups listed above.

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

  Separate figures will be presented for data up to Week 16 (for the BKZ 160mg Q4W group) and side by side by treatment group for data up to Week 24 and Week 52. The data will also be presented in a tabular format. The analyses will be repeated for ACR20 and PASI90.

Has been changed to:

- A summary table displaying the number and percentage of subjects with ADAb status ~~(~~ positive, negative, **total of positive and negative** ~~**or**~~, missing~~)~~ - at each visit **(including SFU)** and overall will be provided. For the overall summary at any visit, 2 summaries will be presented:
  - Including any visit during the treatment period (as defined in Section 3.1.1). 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 Week 16 **(for the BKZ 160 mg Q4W group), and to Week** 24 and 52 by treatment group (BKZ 160 mg Q4W, Placebo / BKZ 160 mg Q4W).

(…)

- The above summary table displaying the number and percentage of subjects in each of the ADAb status (positive, negative, total of positive and negative, missing) will be repeated ~~by region~~ **for the subjects randomized in Japan** for Placebo / BKZ 160 mg Q4W and BKZ 160 mg Q4W groups **(overall summaries only).** The overall summary excluding SFU will be generated for up to Week 16, 24 and 52.

(…)

- A table displaying the number (%) of subjects with the first occurrence of any ADAb positivity (ie: including Baseline visit) or ADAb treatment-emergent positivity during the study will be summarized **by treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W).**

(…)

- A boxplot of the ADAb titer by time of occurrence of ADAb positivity will be created for each **treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W)** ~~dose group~~. The ADAb titer results will be presented on the log-scale. The time to occurrence is defined as the time in weeks from Baseline until the visit of interest when a sample is ADAb positive. Subjects who do not have any ADAb positivity will be excluded from the plot.

(…)

- Some figures summarizing the time to achieve any ADAb positivity and treatment emergent ADAb positivity on a cumulative basis will be presented by treatment group (~~2~~ **4** lines per plot, ~~one plot per treatment group~~. **e.g 2 lines for each treatment group, one line for time to achieve any ADAb positivity and one line for time to achieve treatment emergent ADAb positivity).** Subjects will be considered to have an event at the timepoint 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).
  If ≥10% subjects are classified as ADAb category 5, **the treatment-emergent ADAb+** ~~the~~ line~~s~~ will be further split by the following categories (thus the plot by treatment group will include 3 lines):
  - ADAb category 2: Pre ADAb negative – treatment-emergent ADAb positive,
  - ADAb category 5: Pre ADAb positive – treatment-boosted ADAb positive.
- Individual plots (one plot by subject) of BKZ concentrations/ADAb titer and ACR50 response (based on NRI) will be created. All 3 endpoints will be plotted on the Y-axis by visit (x-axis) for the full treatment period, including SFU ~~where~~ **when** a ~~patient~~ **subject** has not progressed into the OLE. Plots should be labeled and grouped into the 9 ADAb categories and display vertical lines on the x-axis corresponding to the dosing incidences/visits.
  For the Week 24 analysis, the ADAb categories will be based on data up to Week 24.

- Individual plots (1 plot by subject) of BKZ concentrations/ADAb titer and PASI90 **(on subjects with PSO involving at least 3% of BSA at Baseline)** will also be created similarly as described above.

(…)

- Figures will summarize efficacy response (ACR50, ACR20 and PASI90 responders based on NRI**, for PASI90 in the subgroup of subjects with PSO involving at least 3% BSA at Baseline)** 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 ACR50 (ACR20 or PASI90) responders at Week 16. The plot will therefore display the percentage of ACR50 (ACR20 or PASI90) responders as a function of the number of subjects within each ADAb and titer categories. Subjects with negative ADAb results at Week 16 will be included in the 'negative' category on the x-axis. This plot will include data only from subjects randomized to BKZ 160 mg Q4W.
  The figure will be repeated using Week 24 and 52 ~~**data ie Week 24 and 52**~~ ACR50, ~~**(and**~~ ACR20 **and** PASI90 responder rates and Week 24 and 52 ADAb titer quartiles, respectively, and will include all subjects receiving BKZ. The figure for Week 24 and 52 data will be split by treatment group. Thus, 2 figures per efficacy variable and time point will be presented.
- Two figures (side by side) summarizing time course of efficacy response (ACR50 responders based on NRI) for each treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W) by the following ADAb ~~**groups**~~ **categories considered for the efficacy sub-group analysis** (3 lines per plot) will be created:
  - ADAb positive - Defined as subjects having at least 2 ADAb positive samples during the study (including baseline and SFU **up to the timepoint of interest**) regardless of other ADAb negative samples and/or missing or inconclusive samples
  - ADAb negative - Defined as subjects for whom either (1) all samples (including baseline and SFU **up to the timepoint of interest**) are ADAb negative and there are no missing or inconclusive samples, (2) only 1 sample is ADAb positive and all other samples (including baseline and SFU **up to the timepoint of interest**) are ADAb negative or missing/inconclusive or (3) only one sample is missing/inconclusive and the remaining ADAb samples are negative.
  - **ADAb** Missing - Defined as subjects who do not fulfil the criteria for one of the 2 groups listed above.

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

  Separate figures will be presented for data up to Week 16 (for the BKZ 160mg Q4W group) and side by side by treatment group for data up to Week 24 and Week 52. The data will also be presented in a tabular format. The analyses will be repeated for ACR20 and PASI90 **(for PASI90, in the subgroup of subjects with PSO involving at least 3% BSA at Baseline).**

  If ≥95% of subjects are included in the non-missing groups, the missing group will not be tabulated and thus not displayed on the figure.

Change #14

Section 9.2.9 Neutralizing anti-bimekizumab antibodies

- Two figures (side by side) summarizing efficacy response (ACR50 based on NRI) versus time for each treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W) by NAb status (ADAb negative, NAb positive, ADAb positive / NAb negative, missing) will be created. If ≥95% of subjects are included in the non-missing groups, the missing group will not be displayed in the figure. Separate figures will be presented for data up to Week 16 (for the BKZ 160mg Q4Wgroup), and to Week 24 and 52 by treatment group (BKZ 160mg Q4W vs. Placebo / BKZ 160mg Q4W). The data will also be presented in a tabular format. The analyses will be repeated for ACR20 and PASI90. The data will be also presented in a tabular format.

Has been changed to:

- Two figures (side by side) summarizing efficacy response (ACR50 based on NRI) versus time for each treatment group (Placebo / BKZ 160mg Q4W and BKZ 160mg Q4W) by NAb status (ADAb negative, NAb positive, ADAb positive / NAb negative, missing) will be created. If ≥95% of subjects are included in the non-missing groups, the missing group will not be displayed in the figure. Separate figures will be presented for data up to Week 16 (for the BKZ 160mg Q4Wgroup), and to Week 24 and 52 by treatment group (BKZ 160mg Q4W vs. Placebo / BKZ 160mg Q4W). **~~The data will also be presented in a tabular format.~~** The analyses will be repeated for ACR20 and PASI90 **(for PASI90, in the subgroup of subjects with PSO involving at least 3% BSA at Baseline)**. The data will be also presented in a tabular format.

Change #15

Section 10.2.4 Other safety topics of interest

New section : Section 10.2.4.9 Summary table and figure

Change #16

Section 10.2.5 Special considerations for interim data snapshots

Because of the long SFU Period (140 days from last dose), subjects who have completed the SFU visit (and do not, therefore, enroll into the OLE study) may have a longer period of exposure than subjects who complete the study and roll in to the OLE study. For those subjects, exposure ends at 20 weeks after Week 48 (last dose of BKZ) making 68 weeks of exposure.

Has been changed to:

Because of the long SFU Period (140 days from last dose), subjects who have completed the SFU visit (and do not, therefore, enroll into the OLE study) may have a longer period of exposure than subjects who complete the study and roll in to the OLE study. For those subjects,

exposure ends at 20 weeks after ~~**Week 48 (last dose of BKZ) making 68 weeks of exposure**~~ **the last dose of IMP.**

Change #17

Section 10.3 Clinical laboratory evaluations

(…)

To define the minimum/maximum post-Baseline values, all post-Baseline assessments including those at unscheduled visits (but excluding any that occur more than 140 days after the last administration of study medication) will be used.

Has been changed to:

(…)

To define the minimum/maximum post-Baseline values, all post-Baseline assessments including those at unscheduled visits (but excluding any that occur more than 140 days after the last administration of study medication) will be used.

The following additional rule will be also applied :

Subjects who have completed PA0010 and have a PA0012 start date, only PA0010 results up to and including the PA0012 start date will be used to find the subjects minimum or maximum results. Additionally, subjects who are in PA0012 without a start date (date of first dose), only PA0010 results up to the Week 52 visit + 14 days will be used to find the subjects minimum or maximum results.

Change #18

Section 10.5.2 Assessment of tuberculosis

The laboratory test results from the TB assessment performed by interferon gamma release assay will be summarized by treatment group for Baseline (Screening) and Week 48. They will also be listed.

The results from the 'Evaluation of signs and symptoms of tuberculosis' questionnaire data will be also listed.

Has been changed to:

The laboratory test results from the TB assessment performed by interferon gamma release assay **(for non-Japanese subjects) and by T-Spot test (for Japanese subjects)** will be summarized by treatment group for Baseline (Screening) and Week 48. They will also be listed.

The results from the 'Evaluation of signs and symptoms of tuberculosis' questionnaire data will be also listed.

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: PA0010 SAP AMEND3.0 FINAL

Version: 1.0

Document Number: CLIN-000190125

Title: pa0010-sap-amend-3

Approved Date: 06 May 2022

Document Approvals	
Approval Verdict: Approved	Name: [redacted] Capacity: Document Author Date of Signature: 06-May-2022 14:58:17 GMT+0000
Approval Verdict: Approved	Name: [redacted] Capacity: Subject Matter Expert Date of Signature: 06-May-2022 15:12:36 GMT+0000
Approval Verdict: Approved	Name: [redacted] Capacity: Clinical Date of Signature: 06-May-2022 15:44:31 GMT+0000