STATISTICAL ANALYSIS PLAN

Study: AS0013

Product: Bimekizumab

A MULTICENTER, PHASE 2A, RANDOMIZED, INVESTIGATOR-BLIND, SUBJECT-BLIND, PARALLEL-GROUP STUDY TO EVALUATE THE EFFICACY AND SAFETY OF BIMEKIZUMAB AND CERTOLIZUMAB PEGOL IN SUBJECTS WITH ACTIVE ANKYLOSING SPONDYLITIS

SAP/Amendment Number	Date
Final SAP	17 January 2018
Amendment 1	15 August 2018
Amendment 2	23 May 2019
Amendment 3	19 June 2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

ACP	above cut point
ADA	anti-drug antibodies
AE	adverse event
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ALQ	above the limit of quantification
AS	ankylosing spondylitis
ASAS	Assessment in SpondyloArthritis International Society
ASAS-PR	Assessment in SpondyloArthritis International Society partial remission
ASAS20	Assessment in SpondyloArthritis International Society 20%
ASAS40	Assessment in SpondyloArthritis International Society 40%
ASDAS	Ankylosing Spondylitis Disease Activity Score
ASDAS-CII	Ankylosing Spondylitis Disease Activity Score – clinically important improvement
ASDAS-HD	Ankylosing Spondylitis Disease Activity Score – high disease activity
ASDAS-ID	Ankylosing Spondylitis Disease Activity Score – inactive disease
ASDAS-MI	Ankylosing Spondylitis Disease Activity Score – major improvement
ASDAS-vHD	Ankylosing Spondylitis Disease Activity Score – very high disease activity
AST	aspartate aminotransferase
ATC	anatomical therapeutic chemical
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BASFI	Bath Ankylosing Spondylitis Functional Index
BCP	below cut point
BKZ	bimekizumab
BLQ	below the limit of quantification
BMI	body mass index
BP	blood pressure
CI	confidence intervals
CP	confirmed positive
C-SSRS	Columbia-Suicide Severity Rating Scale
CSR	clinical study report
CT	controlled terminology
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation

CZP	certolizumab pegol
DEM	data evaluation meeting
DMC	data monitoring committee
ECG	electrocardiogram
eCRF	electronic case report form
ES	enrolled set
EW	early withdrawal
FAS	full analysis set
FDA	Food and Drug Administration
geoCV	geometric coefficient of variation
HADS	Hospital Anxiety and Depression Scale
HADS-A	Hospital Anxiety and Depression Scale – Anxiety
HADS-D	Hospital Anxiety and Depression Scale – Depression
HIV	human immunodeficiency virus
HLA-B27	human leukocyte antigen B27
HLGT	high level group term
HLT	high level term
hs-CRP	high sensitivity C-reactive protein
IA1	Interim Analysis 1
IA2	Interim Analysis 2
IA3	Interim Analysis 3
ICH	International Council for Harmonization
IL-17A	interleukin 17-A
IL-17F	interleukin 17-F
IMP	investigational medicinal product
IPD	important protocol deviation
LLN	lower limit of normal
LLT	low level term
LLOQ	lower limit of quantification
LSMeans	least squares means
MACE	major adverse cardiovascular event
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model Repeated Measures
mNY	modified New York

n	number of available observations
NCP	not confirmed positive
NRS	numeric rating scale
PCS	potentially clinically significant
PD	pharmacodynamics
PDILI	potential drug-induced liver injury
PD-PPS	pharmacodynamics per-protocol set
PET-CT	positron emission tomography-computed tomography
PET-MRI	positron emission tomography-magnetic resonance imaging
PhGADA	physician's global assessment of disease activity
PK	pharmacokinetics
PK-PPS	pharmacokinetic per-protocol set
PoC	proof-of-concept
PPS	per-protocol set
PT	preferred term
Q2W	every 2 weeks
Q4W	every 4 weeks
QTcF	Fridericia's QTc interval
RCTC	Rheumatology Common Toxicity Criteria
RS	randomized set
SAE	serious adverse event
SAP	statistical analysis plan
Sc	subcutaneous
SD	standard deviation
SDTM	Study Data Tabulation Model
SE	standard error
SFU	safety follow-up
SIJ	sacroiliac joint
SMQ	standardized MedDRA query
SOC	system organ class
SS	safety set
SUV_{auc}	standard uptake value corrected for individual integrated whole blood activity concentration
TB	tuberculosis
TEAE	treatment-emergent adverse event

TEMA	treatment-emergent markedly abnormal
TFLs	tables, figures and listings
TNF	tumor necrosis factor
ULN	upper limit of normal
VOI	volume of interest
VUMC	VU University Medical Center
WHODD	World Health Organization Drug Dictionary

1 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to provide all information that is necessary to perform the required statistical analysis of study AS0013. It also defines the summary tables, figures and listings (TFLs) to be included in the final clinical study report (CSR) according to the protocol.

This SAP is based on, and assumes familiarity with, the following documents:

- Protocol amendment 1, dated 30 May 2017
- Protocol amendment 2, dated 5 March 2018
- Protocol amendment 3, dated 27 February 2019

Unless specified in the sections below, the study will be analyzed as described in the most recent version of the protocol. If a future protocol amendment necessitates a substantial change to the statistical analysis of the study data, this SAP will be amended accordingly. In addition, if analysis definitions need to be modified or updated prior to database lock, a SAP amendment will be required. If, after database lock, additional analyses are required to supplement the planned analyses described in this SAP, these changes will be described in the CSR together with the associated rationale.

The content of this SAP is compatible with the International Council for Harmonization (ICH)/Food and Drug Administration (FDA) E9 Guidance documents (ICH, 1998).

UCB is the Sponsor and PAREXEL is the Contract Research Organization for this study.

2 PROTOCOL SUMMARY

2.1 Study objectives

2.1.1 Primary objective

The primary objective of the study is to evaluate the efficacy of bimekizumab (BKZ) administered subcutaneously (sc) every 2 weeks (Q2W) for 12 weeks compared to certolizumab pegol (CZP; Cimzia®) in the treatment of adult subjects with active ankylosing spondylitis (AS).

2.1.2 Secondary objectives

The secondary objectives of the study are as follows:

- To assess the safety and tolerability of BKZ

2.1.3 Other objectives

The other objectives of the study are as follows:

- To evaluate the effect of BKZ or CZP on changes in bone formation
- To assess the pharmacokinetics (PK) and immunogenicity of BKZ
- To assess additional biomarker, clinical, and imaging data as available
- To assess the efficacy and safety of BKZ or CZP during the Treatment Extension Period

2.2 Study variables

2.2.1 Efficacy variables

2.2.1.1 Primary efficacy variable

The primary efficacy variable for this study is the change from Baseline in Ankylosing Spondylitis Disease Activity Score (ASDAS) at Week 12.

2.2.1.2 Secondary efficacy variables

The secondary efficacy variables for this study are as follows:

- Ankylosing Spondylitis Disease Activity Score-inactive disease (ASDAS-ID) at Week 12
- Ankylosing Spondylitis Disease Activity Score-major improvement (ASDAS-MI) at Week 12

2.2.1.3 Other efficacy variables

Other efficacy variables, assessed over the Treatment Period and Treatment Extension Period, are as follows:

- Change from Baseline in ASDAS
- Assessment in SpondyloArthritis International Society (ASAS) 20% (ASAS20) response
- ASAS 40% (ASAS40) response
- Time to ASAS20 response
- Time to ASAS40 response
- ASAS partial remission
- Change from Baseline in Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)
- Change from Baseline in total Hospital Anxiety and Depression Scale – Anxiety (HADS-A) and total Hospital Anxiety and Depression Scale – Depression (HADS-D)
- Change from Baseline in Physician's Global Assessment of Disease Activity (PhGADA)
- Change from Baseline in Patient's Global Assessment of Disease Activity (PGADA)
- Changes in bone formation as measured by standardized uptake value corrected for individual integrated whole blood activity concentration (SUV_{auc}) for each PET-positive lesion identified together with the maximum of the SUV_{auc} values across all lesions within the spine and sacroiliac joint (SIJ). SUV_{auc} will be derived from positron-emission tomography magnetic resonance imaging (PET-MRI) or positron-emission tomography computed tomography (PET-CT) scans at Baseline, Week 12, and Week 48

2.2.2 Safety variables

2.2.2.1 Primary safety variables

The primary safety variables for this study are as follows:

- Incidence of adverse events (AEs) and serious adverse events (SAEs)

- Withdrawal due to AEs

2.2.2.2 Other safety variables

The other safety variables for this study are as follows:

- Change from Baseline in vital signs (blood pressure, temperature, pulse rate) and body weight
- Change from Baseline in physical examination
- Change from Baseline in standard 12-lead electrocardiogram (ECG) intervals (RR, PR, QRS, QT, and QT intervals corrected for heart rate using Fridericia's formula [QTcF])
- Change from Baseline in clinical laboratory values (hematology, biochemistry, and urinalysis)

2.2.3 Pharmacokinetic variables

The PK variables are plasma concentrations of BKZ and CZP.

2.2.4 Pharmacodynamic variables

The pharmacodynamic (PD) variables are the blood or blood derivative (eg, plasma) concentrations of cytokines and chemokines of relevance to interleukin-17A (IL-17A) or interleukin-17F (IL-17F) signaling pathway, tumor necrosis factor (TNF) signaling pathway, AS biology, and bone metabolism. Additional variables may include, but will not be limited to, serum complement concentrations and mononuclear cell subtypes.

2.2.5 Immunological variables

Anti-BKZ antibody and anti-CZP antibody (collectively anti-drug antibody [ADA]) detection prior to and following study treatment will be evaluated.

2.2.6 Nonhereditary pharmacogenomic variables

Where local regulations permit, blood and urine will be collected and stored for up to 20 years to allow for potential exploratory analyses of ribonucleic acid, proteins, and metabolite biomarkers relevant to AS, bone metabolism, and the inflammatory and immune response processes. The nature and format of these tentative analyses will be determined at a later stage and will not form part of the CSR and are not further described in this SAP.

2.2.7 Pharmacogenetic variables

Additional blood samples will be collected from subjects who consent to participate in the pharmacogenetic substudy and stored at -80°C for up to 20 years. Pharmacogenetic biomarkers may be measured to evaluate the relationship to response to treatment with BKZ, AS disease biology, bone metabolism, 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 results of these analyses will not form part of the CSR and are not further described in this SAP.

2.3 Study design and conduct

AS0013 is a multicenter, randomized, subject-blind and investigator-blind, parallel-group study to evaluate the efficacy and safety of BKZ compared to CZP in adult subjects with active adult-onset AS.

It is planned that at least 60 subjects will be randomized to 1 of 2 treatment arms in a 2:1 ratio and will receive either BKZ or CZP up to Week 44 (final dose of investigational medicinal product [IMP]).

The study will include a Screening Period of 2 to 4 weeks, a 12-week Treatment Period, a 36-week Treatment Extension Period, and a 20-week Safety Follow-up (SFU) Period (starting after the final dose of IMP). Therefore, the maximum duration of the study will be 68 weeks.

At the start of the Treatment Period, eligible subjects will be randomized in a 2:1 ratio to receive the following blinded study treatments:

- BKZ 160mg sc Q2W from Week 0 through Week 10. In addition, subjects will receive 1 placebo injection at Baseline (Visit 2), Week 2 (Visit 3), and Week 4 (Visit 3) in order to maintain the blind versus the CZP loading dose at these visits.
- CZP 400mg sc Q2W at Weeks 0, 2, and 4 (loading dose) followed by CZP 200mg sc Q2W at Weeks 6 to 10.

After completing the 12-week Treatment Period, subjects will enter a 36-week Treatment Extension Period and will receive the following treatments:

- Subjects randomized to BKZ during the Treatment Period will receive BKZ 320mg sc every 4 weeks (Q4W) from Week 12 to Week 44.
- Subjects randomized to CZP during the Treatment Period will receive CZP 400mg Q4W from Week 12 to Week 44.

Subjects not responding to treatment during the Treatment Period will be withdrawn from the study as per Investigator's discretion and will not enter the Treatment Extension Period.

In a substudy of AS0013, a PET-MRI or PET-CT scan of the entire spine will be performed at selected sites in approximately 25 subjects at Screening and during the study at Week 12 and Week 48 or Early Withdrawal Visit if PET positive lesions were observed in the previous scan.

All subjects who complete the study or who discontinue from the study early, including those withdrawn from the IMP treatment, will have a SFU Visit at 20 weeks after their final dose of IMP.

Three unblinded interim analyses will be conducted:

- Interim Analysis 1 (IA1): when approximately 45 subjects have completed the Week 4 visit during the Treatment Period.
- Interim Analysis 2 (IA2): when all subjects not participating in the PET-MRI/PET-CT sub-study have completed the Week 12 visit at the end of the Treatment Period.
- Interim Analysis 3 (IA3): when the last randomized subject has completed the Week 12 visit at the end of the Treatment Period, or the subject has discontinued prematurely from the study.

Further details on the interim analyses planned for this study are included in the interim SAP.

A data monitoring committee (DMC) will review the data on an ongoing basis. The composition and roles of the DMC are described in a separate DMC Charter. The analyses required and data to be presented are described in a separate DMC SAP.

2.4 Determination of sample size

A sufficient number of subjects will be enrolled to ensure that at least 60 subjects are available in the main study at Week 12 to compare the change from Baseline in ASDAS between BKZ and CZP. This will also provide at least 25 subjects enrolled in the PET-CT or PET-MRI substudy from multiple sites. Subjects will be randomized in a 2:1 ratio to receive BKZ or CZP respectively.

The sample size was calculated to provide at least 80% power to detect a difference between treatment groups in the mean change from Baseline in ASDAS at Week 12 of 0.89 (group mean change from Baseline for CZP and BKZ being -1.78 and -2.67 respectively) with a common standard deviation (SD) of 1.17 using frequentist methods for the comparison of the treatment group mean change from Baseline in ASDAS at Week 12 with Baseline ASDAS as a covariate (Overall and Starbuck, 1979).

The correlation between Baseline and Week 12 raw ASDAS was assumed to be 0.35. Note that the informative prior for the model intercept to be used in the Bayesian modelling of the primary efficacy variable, given by β_{CZP} ~ Normal (-1.78, var=0.0605), contributes an effective sample size of approximately 20 CZP subjects. Prior data conflict will be investigated, and a vague prior will be used for the model intercept coefficient if prior data conflict is evident.

Note that a clarification regarding the sample size for this study was provided in protocol amendment 3. As indicated above, the planned sample size of at least 60 subjects provides sufficient power to detect meaningful treatment differences in the primary efficacy variable whilst also providing at least 25 subjects for the PET-MRI or PET-CT substudy. However, due to complexities in the initiation of sites capable of performing PET-MRI or PET-CT scans, more than the planned 35 non-PET-MRI/CT substudy subjects were screened and randomized before it was possible to close enrolment at the sites not participating in the substudy. Owing to the novelty of the PET-MRI/CT outcomes within this disease area, it is considered important to ensure at least 25 evaluable subjects are included in the assessment of this exploratory objective to maximize the possibility of identifying the salient factors or variables that might differ between the treatment groups and, therefore, the total sample size is increased to at least 72 subjects. This sample size increase will not negatively impact the power to detect a difference in the primary efficacy variable at Week 12 (ie, it will remain >80%), but will ensure that the target number of subjects are included in the exploratory analysis of changes in bone formation at Week 12 and Week 48. Since the analyses of PET-MRI/CT data are exploratory and treatment effect sizes are unknown, formal power calculations were not performed. Clinical expertise does, however, indicate that a sample size of at least 25 subjects in the substudy will characterize broad differences between treatment groups.

3 DATA ANALYSIS CONSIDERATIONS

3.1 General presentation of summaries and analyses

Statistical evaluation will be performed by PAREXEL and supervised by the Exploratory Statistics and Global Statistical Sciences Departments of UCB.

The datasets will follow the UCB analysis data model data specifications.

All analyses will be performed using SAS version 9.2 or higher (SAS Institute, Cary, North Carolina, USA). R Version 2.10.1 (R Development Core Team) or later, and/or OpenBUGS Version 3.0.6 or later may be used for complementary analysis.

Descriptive statistics will be used to provide an overview of the Baseline, efficacy, PD, PK and safety results. Formal statistical testing will be conducted for the primary and secondary efficacy variables only in this study. Other efficacy variables will be summarized descriptively by treatment group.

Continuous variables will be summarized by treatment, visit and time point (where applicable) including number of subjects (n), mean, SD, median, minimum and maximum. First and third quartiles (Q1 and Q3 respectively) and 95% confidence intervals (CI) for the mean will also be included where stated in the SAP.

Categorical variables will be summarized by treatment, visit and time point (where applicable) with frequency counts and percentages. Geometric coefficient of variation (geoCV), geometric mean and 95% CI for the geometric mean will also be presented in the descriptive statistics for the PK concentration and hs-CRP data. In all outputs the 95% confidence limits will be restricted to the possible values that the variable can take.

When reporting relative frequencies or other percentage values, the following rules apply:

- For values where all subjects fulfill certain criteria, the percentage value will be displayed as 100
- For values where the absolute frequency is zero, there will be no percentage displayed at all
- All other percentage displays will use 1 decimal place
- Unless otherwise stated, the denominator for the percentages will be based on the number of subjects in the respective analysis set and treatment group.

Percentages displayed based on continuous data (eg, percentage changes from Baseline) will be displayed to 1 decimal place.

When reporting descriptive statistics, the following rules will apply in general (with the exception of PK and hs-CRP data, for which additional rules are stated below):

- n will be an integer
- Mean, SD, median and CIs will use 1 decimal place more (or 1 more significant figure, depending on the original data reporting format) than the original data
- Coefficient of variation will be reported as a percentage to 1 decimal place

- Minimum and maximum will be reported using the same number of decimal places as the original data
- If no subjects have data at a given time point, then only n=0 will be presented. If n<3, then only n, minimum and maximum will be presented. If n=3, then only n, median, minimum and maximum will be presented. The other descriptive statistics will be left blank.

When reporting individual values and descriptive statistics for PK concentration and hs-CRP data, the following rules will apply with regard to rounding and precision:

- Individual values will be reported to the same level of precision as received from the bioanalytical laboratory
- Descriptive statistics will be reported to the same level of precision as the individual data for the minimum and maximum, to 1 additional significant figure for the mean (arithmetic and geometric), median, Q1, Q3, SD, and to 2 additional significant figures for the 95% CI for the geometric mean
- The geoCV will be reported as a percentage to 1 decimal place

All statistical tests will be carried out 2-tailed at the 5% level of significance unless otherwise stated. The following rules will apply for the presentation of any results relating to inferential statistical analysis:

- P-values will be presented to 4 decimal places; p-values less than 0.0001 will be presented as "<0.0001" and p-values greater than 0.9999 will be presented as ">0.9999"
- Posterior probabilities associated with the Bayesian analyses will be presented as percentages to 1 decimal place
- Means, SDs, medians, 95% credible intervals and highest posterior density intervals obtained from the Bayesian analyses will be presented to 1 decimal place

All statistical output will be presented in statistical appendices where appropriate.

Time to response data (Section 8.3.2.2) will be analyzed for the Treatment Period and across the study (Treatment Period + Treatment Extension Period). All other data tabulations will be performed by treatment group, visit and time point (where applicable) and no separate analyses based on study period will be presented.

The treatment groups will be displayed as follows in the TFLs, where BKZ refers to bimekizumab and CZP refers to certolizumab pegol:

- BKZ
- CZP

Data listings containing all documented data and all derived data will be generated. Subjects enrolled in the PET-MRI/PET-CT substudy will be flagged in the data listings.

If Study Data Tabulation Model (SDTM) controlled terminology (CT) is defined for selected variables, it will be displayed similarly in the TFLs.

Yes/No responses will be presented as Y/N consistent with the YN CT and Male and Female will be replaced with M/F respectively in the TFLs.

3.2 General study level definitions

3.2.1 Relative day

Relative day for an event will be derived with the date of the first administration of IMP as reference date.

Relative day for an event or measurement occurring before the reference date will be calculated as follows:

$$Relative\ Day\ =\ Event\ Date - Reference\ Date$$

The relative day for an event or measurement occurring on or after the reference date, up to the date of the last administration of IMP, will be calculated as follows:

$$Relative\ Day\ =\ (Event\ Date - Reference\ Date)\ +\ 1$$

For events or measurements occurring after the date of the last administration of IMP, relative day will be prefixed with '+' in the data listings and will be calculated as follows:

$$Relative\ Day\ =\ Event\ Date - Date\ of\ Last\ Administration$$

There is no relative Day 0. Relative day will not be calculated for partial dates in cases where relative day is shown in a subject data listing. In such cases, relative day should be presented as '-' in the subject data listings.

3.2.2 Study periods

For each subject completing the study, the maximum study duration will be 68 weeks and will comprise the following:

- Screening Period: ≥2 weeks with a maximum of 4 weeks prior to randomization
- Treatment Period: 12 weeks
- Treatment Extension Period: 36 weeks
- Safety Follow-Up Period: 20 weeks after the last dose of IMP (scheduled at Week 44, 4 weeks prior to the end of the Treatment Extension Period)

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

3.3 Definition of Baseline values

Baseline will be the last available value prior to the first administration of IMP. Scheduled or unscheduled measurements can be used as the Baseline value. Measurement-specific Baseline definitions (based on the schedule of assessments in the protocol) are presented in Table 3–1.

Table 3–1: Definition of Baseline

Category	Measurement	Definition of Baseline
Efficacy	BASDAI	Visit 2 (Week 0); Screening result if missing at Week 0
	PGADA	Visit 2 (Week 0)

Table 3–1: Definition of Baseline

Category	Measurement	Definition of Baseline
	hs-CRP	Visit 2 (Week 0); Screening result if missing at Week 0
	PET-MRI/PET-CT	Visit 1 (Screening)
	BASFI	Visit 2 (Week 0); Screening if missing at Week 0
	Total and nocturnal spinal pain numeric rating scale (NRS)	Visit 2 (Week 0); Screening if missing at Week 0
	PhGADA	Visit 2 (Week 0)
	HADS	Visit 2 (Week 0); Screening if missing at Week 0
Safety	Vital signs	Visit 2 (Week 0); Screening if missing at Week 0
	12-lead ECG	Visit 2 (Week 0); Screening if missing at Week 0
	Clinical laboratory tests	Visit 2 (Week 0); Screening if missing at Week 0
	Body weight	Visit 1 (Screening)
	C-SSRS	Visit 2 (Week 0); Screening if missing at Week 0
Immunological	ADA	Visit 2 (Week 0)
Pharmacodynamic	Cytokines, chemokines, serum complement levels, mononuclear cell subtypes	Visit 2 (Week 0)

ADA=anti-drug antibody; BASDAI=Bath Ankylosing Spondylitis Disease Activity Index; BASFI=Bath Ankylosing Spondylitis Functional Index; C-SSRS=Columbia-Suicide Severity Rating Scale; ECG=electrocardiogram; HADS=Hospital Anxiety and Depression Scale; hs-CRP=high sensitivity C-reactive protein; PET-CT=positron-emission tomography-computed tomography; PET-MRI=positron-emission tomography-magnetic resonance imaging; PGADA=Patient's Global Assessment of Disease Activity; PhGADA=Physician's global assessment of disease activity.

The change from Baseline to any subsequent post-Baseline visit will be calculated as the simple difference between that post-Baseline visit's value and the Baseline value, as below:

$$Post\ Baseline\ Visit\ Value - Baseline\ Visit\ Value$$

The percentage change from Baseline to any subsequent post-Baseline visit will be calculated as follows:

$$100 \times (Post\ Baseline\ Visit\ Value - Baseline\ Visit\ Value) \ / \ (Baseline\ Visit\ Value)$$

The same Baseline value will be used in the calculations of changes from Baseline and percentage changes from Baseline at all visits during the Treatment Period and Treatment Extension Period.

3.4 Protocol deviations

Important protocol deviations (IPDs) are deviations from the protocol which potentially could have a meaningful impact on the primary objective of the study. The criteria for identifying such protocol deviations will be defined within the IPD document which is part of the data cleaning plan. Important protocol deviations will be classified according to the following categories:

- Inclusion/exclusion criteria deviations
- Administration of prohibited concomitant medications
- Deviations relating to withdrawal criteria
- Visit schedule deviations
- Study IMP administration deviations (including incorrect treatment received, handling and storage deviations and incorrect dosage received)
- Treatment non-compliance
- Procedural non-compliance
- Missing data
- Other

Important protocol deviations will be reviewed as part of the ongoing data cleaning process and data evaluation.

At least 4 data evaluation meetings (DEMs) will be performed at the following times, prior to each of the following analysis milestones:

- IA1
- IA2
- IA3
- The final analysis after all data have been verified/coded/entered into the database

Additional DEMs may be conducted as deemed necessary.

The purpose of these DEM reviews will be to review all protocol deviations, define the analysis sets as defined in Section 3.5, and check the quality of the data. The reviews will also help decide how to manage problems in the subjects' data (eg, missing values and withdrawals).

Accepted deviations from theoretical (scheduled) time points will be described in the appropriate documents and included in the Study Master File. After the pre-analysis review, resolution of all issues, and documentation of all decisions (including inclusion into each of the analysis sets) at the final DEM, the database will be locked.

3.5 Analysis sets

3.5.1 Enrolled Set

The Enrolled Set (ES) will consist of all subjects who have given informed consent (ie, all subjects screened). The ES will therefore include screening failures.

3.5.2 Randomized Set

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

3.5.3 Safety Set

The Safety Set (SS) will consist of all randomized subjects who receive at least 1 dose (full or partial) of the IMP. The safety analysis will be conducted on the SS.

3.5.4 Full Analysis Set

The Full Analysis Set (FAS) will consist of all randomized subjects who received at least 1 dose (full or partial) of the IMP and had a valid Baseline and a post-Baseline measurement for at least one efficacy variable (as defined in Section 2.2.1). The FAS will be used for the supportive and sensitivity analyses of the primary efficacy variable, the analyses of the secondary efficacy variables, other efficacy variables and the primary efficacy component data.

3.5.5 Per-Protocol Set

The Per-Protocol Set (PPS) will consist of all subjects in the FAS who have no IPDs affecting the primary efficacy variable. The subjects with IPDs will be predefined and evaluated during a DEM prior to unblinding of the data. The PPS will be used for the primary analyses and supportive and sensitivity analyses of the primary efficacy variable, for the analysis of the component variables of the primary efficacy variable and for the analysis of the secondary efficacy variables.

Protocol deviations will be reviewed at the DEMs and any exclusions from analysis sets will be documented as described in Section 3.4. It is anticipated that the PPS for the final analysis will differ from the PPS defined at the interim analyses due to additional information obtained on subjects during the SFU Period. Similarly, the PPS as defined at IA1, IA2 and IA3 may differ across the 3 IAs, due to potential additional data or information obtained during the time between each analysis.

Following the final DEM, the inclusion into each of the analysis sets for all subjects will be confirmed. The DEM minutes will document the final agreed PPS to be applied to the final analysis, as well as document any changes to the PPS relative to the previously agreed PPS at the interim(s), if applicable.

3.5.6 Pharmacokinetics Per-Protocol Set

The Pharmacokinetics Per-Protocol Set (PK-PPS) will consist of all randomized subjects who receive at least 1 dose (full or partial) of the IMP and have at least 1 quantifiable post-dose plasma concentration. The summaries of the PK concentration data and the immunological data will be based on the PK-PPS.

3.5.7 Pharmacodynamics Per-Protocol Set

The Pharmacodynamics Per-Protocol Set (PD-PPS) is a subset of the FAS, consisting of those subjects who had no IPDs potentially affecting the PD data, as confirmed during a pre-analysis review of the data prior to database lock.

3.5.8 PET Per-Protocol Set

The PET Per-Protocol Set (PET-PPS) will consist of all randomized subjects who receive at least 1 dose of the IMP and have evaluable PET-MRI or PET-CT scan data at Baseline and at least 1 of the post-Baseline assessments.

3.6 Treatment assignment and treatment groups

Listings and summaries will be presented by treatment group and/or overall, as applicable (Table 14–4).

The following order will be used in the TFLs:

- Not randomized
- CZP
- BKZ
- All subjects

The dosing schedule and number of injections given at each visit for each treatment group are described in Table 3–2.

The IMP will be administered in the clinic as 2 sc injections. Suitable areas for sc injection are the lateral abdominal wall, upper outer thigh, and upper arm. During each dosing visit, each of the 2 injections will be administered at a separate injection site.

In addition, subjects randomized to BKZ will receive 1 placebo injection at Baseline, Week 2, and Week 4 in order to maintain the blind vs the CZP loading dose at these visits.

The minimum time between doses should be no less than 10 days during the Treatment Period and no less than 22 days during the Treatment Extension Period.

Table 3–2: Administration of investigational medicinal product

Visit	Treatment group	
	Bimekizumab	**Certolizumab Pegol**
Treatment Period		
Baseline, Weeks 2, 4	160mg	400mg
Weeks 6, 8, 10	160mg	200mg
Treatment Extension Period		
Weeks 12 to 44, every 4 weeks	320mg	400mg

In the case of dosing administration errors during the Treatment Period, all statistical analyses of safety data will be conducted according to the actual treatment received, ie, treatment assignment for the SS will be based on actual treatment with the following rule:

- Subjects who received any doses of a treatment to which they were not randomized will be assessed at the DEM for the treatment group assignment, as it will depend on how many doses of each of the treatments were taken and at what time point during the study

Subjects receiving the incorrect treatment at a particular visit will be excluded from the PPS, PD-PPS and the PK-PPS as this would be considered as an IPD. Subjects may be excluded from the analysis at a particular visit (or visits) only or overall, following discussion at the DEMs, and will be included in the analyses according to their randomized treatment (which would be equivalent to an analysis according to their actual treatment).

All summaries of the PK and immunological data will be conducted according to the actual treatment received, ie, treatment assignment for the PK-PPS will be based on actual treatment.

Similarly, all statistical analyses of the PD data will be conducted according to the actual treatment received, ie, treatment assignment for the PD-PPS will be based on actual treatment.

3.7 Center pooling strategy

Due to the small number of subjects per site, it is planned to pool all subjects within the study. The data summaries and statistical analyses will not be performed by site or country.

3.8 Coding dictionaries

Adverse events and medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA®) version 19.0.

Medications will be coded according to the World Health Organization Drug Dictionary (WHODD) version SEP2015. Medical procedures will not be coded.

3.9 Changes to protocol-defined analyses

The following changes relative to the protocol-defined analyses have been included in the SAP:

- The PD-PPS was included as an additional analysis set in the SAP and was not part of the clinical study protocol. This was added as it was deemed appropriate to review protocol deviations separately in relation to the PD data and to define a separate analysis set accordingly.
- The HADS data are not included under the list of Other efficacy variables in the protocol. These data are, however, collected as supportive efficacy data and therefore the changes from Baseline in the total HADS-A and HADS-D scores are included in this SAP under Section 2.2.1.3 for Other efficacy variables.
- The PhGADA data are not included under the list of Other efficacy variables in the protocol. These data are, however, collected as supportive efficacy data and therefore the change from Baseline in PhGADA is included in this SAP under Section 2.2.1.3 for Other efficacy variables.
- The protocol defines the FAS as all randomized subjects who receive at least 1 dose of the IMP and have a valid measurement of the primary efficacy variable at Baseline and at least 1

post-Baseline efficacy assessment. The definition of the FAS in the SAP has been updated to include all randomized subjects who received at least 1 dose (full or partial) of the IMP and had a valid Baseline and a post-Baseline measurement for at least 1 efficacy variable. Since the primary efficacy variable is a composite endpoint based on BASDAI, PGADA and hs-CRP, the use of the protocol definition would omit any subjects with at least one missing component of the composite endpoint from the efficacy analyses.

4 STATISTICAL/ANALYTICAL ISSUES

4.1 Adjustments for covariates

The Bayesian analyses will employ a linear regression model adjusting for Baseline ASDAS as a covariate. Further exploratory analyses may be performed for the primary efficacy variable adjusting for other Baseline covariates.

4.2 Handling of dropouts or missing data

4.2.1 Efficacy data

4.2.1.1 ASDAS derivation

If 1 component for the ASDAS is missing at a given visit, and the visit prior to the visit with the missing component data is not the Baseline or Screening visit, the following will apply: the missing component will be imputed by carrying the last observation for the given component forward, and the ASDAS will be calculated accordingly.

If no value is available for that component before the missing visit, or if the visit prior to the visit with the missing component data is either the Baseline or Screening visit, then the next observation will be carried backwards, and the ASDAS will be calculated accordingly.

If an unscheduled or repeat visit with non-missing data for the given ASDAS component is available and this visit is closer to the scheduled visit where the component data is missing, then the component data from the unscheduled/repeat visit will be used in the derivation of the ASDAS score.

If more than one component for the ASDAS is missing, ASDAS will not be calculated and treated as missing.

Any measurements of hs-CRP that are below 2 mg/L will be imputed with a fixed value of 2 mg/L for the purpose of deriving the ASDAS.

The statistical analysis will be performed based on the resulting derived values of ASDAS.

4.2.1.2 hs-CRP

Any measurements of hs-CRP that are below the limit of quantification (BLQ) will be imputed with the midpoint between zero and the lower limit of quantification (LLOQ) for the purpose of the calculation of summary statistics and changes from Baseline. Measurements for hs-CRP that are above the upper limit of quantification (ALQ), if applicable, will be imputed to the upper quantification limit.

4.2.1.3 ASDAS component data

Except for the descriptive statistics of hs-CRP as described above, no imputation of missing data will be performed for the summaries of the ASDAS component efficacy variables.

4.2.1.4 ASAS20, ASAS40 and ASAS-PR

For all non-missed visits, in the calculation of ASAS20, ASAS40 and ASAS-PR responses, if any of the component data are missing, then the following rules will be applied:

- If all the component values are missing from Baseline to the post-Baseline visit being considered, the percentage change from Baseline to that visit for the given component will be imputed as 0%. The unit improvement will be imputed as 0.
- If the component value at a given post-Baseline visit is missing and the Baseline value is present, the missing component will be replaced by the last non-missing post-Baseline observation (LOCF) for that component.

4.2.1.5 BASFI data

In case of missing answers to 1 or 2 of the single items within the BASFI questionnaire, the BASFI score will be calculated by imputing missing items with the mean of the completed items. Then, the BASFI score will be calculated as described in Section 8.3.3.1. If more than 2 of the items are missing, the BASFI score will be left missing.

4.2.1.6 HADS data

Any missing data for the total score for HADS-A and HADS-D will be handled as follows:

- If a maximum of 1 item is missing in HADS-A or HADS-D, the missing item will be imputed with the mean score from the remaining completed items within the HADS-A or HADS-D respectively.
- If more than 1 item is missing in HADS-A or HADS-D, the total score will not be calculated.

4.2.1.7 Spinal pain and PhGADA data

Missing data for Total and nocturnal spinal pain NRS and PhGADA data will not be imputed and will be analyzed as far as data are available.

4.2.2 Pharmacodynamic data

For the PD data, measurements that are BLQ will be imputed with half of the LLOQ for the purpose of calculating summary statistics and changes from Baseline. Measurements that are ALQ, if applicable, will be imputed to the upper quantification limit.

4.2.3 Safety laboratory data

The rules for handling values that are BLQ or ALQ in the safety laboratory data will be the same as those described for PD data in Section 4.2.2.

4.2.4 Pharmacokinetic data

Concentrations that are BLQ will be imputed with half of the LLOQ for the respective analyte (BKZ or CZP) for the purpose of calculating the geometric mean and its 95% CI, the geoCV, the arithmetic mean and SD for summaries and figures.

For the individual figures, any concentrations that are BLQ will be regarded as missing, with the exception of pre-dose BLQ concentrations at Visit 2 (Week 0), which will be imputed with zero for linear scale plots.

4.2.5 Dates and times

Partial dates may be imputed for the following reasons:

- Classification of AEs as treatment-emergent
- Classification of medications as prior or concomitant

Imputed dates will not be shown in the listings; all dates will be displayed as reported in the database.

The following rules will be applied for partial start dates:

- If only the month and year are specified and the month and year of the first dose of IMP is not the same as the month and year of the start date then the 1st of the month will be used, or the date of Screening if this is later (if the latter imputation results in an end date that is earlier than the start date, then the 1st of the month will be used). If time is missing this will be imputed as 00:00 h
- If only the month and year are specified and the month and year of the first dose of IMP is the same as the month and year of the start date, then the date of the first dose of IMP will be used. If this results in an imputed start date that is after the specified end date, then the 1st of the month will be used, or the date of Screening if this is later (if the latter imputation results in an end date that is earlier than the start date, then the 1st of the month will be used). If the imputed date is the same as the date of the first dose of IMP then time will be imputed as the start time of the first injection (ie, an AE will be regarded as treatment-emergent and a medication will be classified as concomitant)
- If only the year is specified, and the year of the first dose of IMP is not the same as the year of the start date then January 01 of the year of the start date will be used. If time is missing this will be imputed as 00:00 h
- If only the year is specified, and the year of the first dose of IMP is the same as the year of the start date, then the date of the first dose of IMP will be used. If this results in an imputed start date that is after the specified end date, then January 01 of the year of the start date will be used, or the date of Screening if this is later (if the latter imputation results in an end date that is earlier than the start date, then January 01 will be used). If the imputed date is the same as the date of the first dose of IMP then time will be imputed as the start time of the injection (ie, an AE will be regarded as treatment-emergent and a medication will be classified as concomitant)

The following rules will be applied to partial stop dates:

- If only the month and year are specified, then the last day of the month will be used
- If only the year is specified, then December 31 of the known year will be used
- If the stop date is completely unknown, the stop date will not be imputed

Missing or partially missing dates and/or times will be imputed as described in Table 4–1 for the calculation of duration of each AE. AE duration will be computed in and reported in day and time format: xx d hh:mm.

Table 4–1: Calculation rules for duration of adverse events

Data availability	Onset date/time	Outcome date/time	Calculation rules
Complete data	D1/T1	D2/T2	Duration = [(D2 – D1)*24 + (T2 – T1)]/24 d
End time missing	D1/T1	D2/--	End time is substituted by time 23:59h (=23.98 in decimal format) Duration = <[(D2 – D1)*24 + (23.98 – T1)]/24 d
Start time missing	D1/--	D2/T2	Onset time is substituted by time 00:00h Duration = <[(D2 – D1)*24 + T2]/24 d
Start and end time missing	D1/--	D2/--	Duration = <D2 – D1 + 1
Start day and time missing	--/--	D2/T2	Duration = [(D2 – D0)*24 + (T2 – T0)]/24 d For a subject in the SS, D0 and T0 are the date and time of first administration of IMP and for screen failures, D0 is the date of the Screening visit and T0 = 00:00h
End day and time missing	D1/T1	--/--	If the stop date is missing, duration will not be calculated.
Start and end date missing	--/--	--/--	If the stop date is missing, duration will not be calculated.

4.3 Handling of repeated and unscheduled measurements

All repeated and unscheduled measurements will be presented in the data listings, where applicable. The following general rules will apply to all repeated and unscheduled measurements:

- For repeated measurements obtained prior to the first dose of IMP the latest value (which may be scheduled or unscheduled) will be used in the calculation of the descriptive statistics
- For repeated measurements obtained at the designated Baseline visit, the latest value (which may be scheduled or unscheduled) will be defined as the Baseline provided that this occurred prior to the first dose of IMP
- For repeated measurements obtained at any time point after the first dose of IMP, the values will not be included in the calculation of descriptive statistics or changes from Baseline.
- Unscheduled measurements performed for the Early Withdrawal (EW) visit will be assigned to the appropriate visit (Section 4.4) and analyzed accordingly

4.4 Handling of measurements obtained at the EW visit

Subjects withdrawing early from the study should undergo the same assessments scheduled for the Week 48 visit as soon as possible after withdrawal (as an EW visit) and then enter the SFU Period, completing study participation with the SFU Visit 20 weeks after their final dose of IMP.

The following rules will apply with regard to the inclusion of the results obtained at the EW visit in the descriptive summaries:

- Any measurements conducted at the EW visit should be included in the summaries for the respective scheduled visit, if the EW visit occurs at the time of the next scheduled visit. For example, if the EW visit occurs on Day 42, the results would be summarized together with the Visit 6 (Week 6, Day 42) results
- If the EW visit does not correspond to the day of a scheduled visit, the assessments of the EW visit should be mapped to the nearest scheduled visit, relative to the Baseline visit date, following the last scheduled visit where assessments are available
- If the date of the EW visit is equidistant between 2 scheduled visits at which no scheduled assessments were performed, the assessments from the EW visit will be mapped to the earliest of these visits
- If an EW visit mapping results in data being mapped to a visit where the specific assessment is not actually collected per the protocol schedule of assessments, these data will not be included in the summary statistics and will be listed only
- The only exception to the above rule is for ADA (for both BKZ and CZP) assessments, in which all early withdrawal visit assessments will be assigned to the next scheduled visit at which ADA are assessed. The rationale for this is that ADA positivity is summarized over a given study period. As part of that summary, a table indicating the first visit at which ADA positivity is observed will be presented. In order to match the number of subjects who were ADA positive at specific visits with the overall positivity for the period, it is necessary to ensure that ADA positivity is attributed to a visit where ADA antibody assessments were performed

Assessments from the EW visit will be displayed as the mapped visit and will be flagged in the by-visit data listings.

4.5 Interim analyses and data monitoring

4.5.1 Data monitoring committee

The conduct of the study and monitoring of the safety data will be supervised by an independent DMC to optimize subjects' safety and to identify any safety signals. Full details of the composition and roles of the DMC will be provided in a separate DMC Charter.

The data to be reviewed by the DMC will be as detailed in the DMC Charter and will include unblinded safety data. The details regarding the outputs to be produced and the analyses to be performed will be provided in a separate DMC SAP.

4.5.2 Interim analyses

Three unblinded informal IAs are planned for this study.

IA1 will be performed when approximately 45 subjects have completed the Week 4 visit.

IA2 will be performed when the last randomized subject not participating in the PET-MRI/PET-CT has completed the Week 12 visit at the end of the Treatment Period or has withdrawn prematurely from the study.

IA3 will be performed when the last randomized subject has completed the Week 12 visit at the end of the Treatment Period or has withdrawn prematurely from the study.

The purpose of the IAs is for key Sponsor personnel to review the results of the primary efficacy, selected secondary and other efficacy analyses and safety outcomes to facilitate additional Clinical Planning or Portfolio Management decisions.

All analyses and unblinding instructions will be pre-specified in the Interim Analysis SAP.

4.6 Multicenter studies

The data summaries and statistical analyses will not be performed by center.

4.7 Multiple comparisons/multiplicity

Not applicable.

4.8 Use of an efficacy subset of subjects

In order to investigate the impact on the efficacy analyses, the PPS will be used to evaluate the subset of subjects in the FAS with no IPDs.

4.9 Active-control studies intended to show equivalence

Not applicable.

4.10 Examination of subgroups

Not applicable.

5 STUDY POPULATION CHARACTERISTICS

5.1 Subject disposition

The number and percentage of subjects who were randomized into the study (and into the PET-MRI/PET-CT substudy), subjects who completed or prematurely discontinued the study, as well as the primary reason for discontinuation will be presented by treatment group and for all subjects, based on the RS. A subject who completed the study is defined as a subject who completed all visits up to the last scheduled study visit, ie, Visit 18 (Week 64). An additional summary of subject disposition in the PET-MRI/PET-CT substudy will also be presented separately by treatment group and for all subjects, based on the RS.

In addition, disposition of subjects during the Treatment Period and Treatment Extension Period will be summarized separately. The number and percentage of subjects who completed Week 4 and Week 12 of the Treatment Period, or prematurely discontinued the study during the Treatment Period, together with the primary reason for discontinuation will be presented by treatment group. The number and percentage of subjects who completed Week 48 (ie, the Treatment Extension Period) or prematurely discontinued during the Treatment Extension Period, together with the primary reason for discontinuation will be presented by treatment group. A subject is defined as completing Week 4, Week 12 and Week 48, respectively, if adequate efficacy data are available in the database to calculate the primary efficacy variable at the respective visit.

The number and percentage of subjects who discontinued due to AEs will be separately summarized by treatment group and for all subjects, based on the RS. This will be used for European Union Drug Regulating Authorities Clinical Trials (EudraCT) reporting.

The number and percentage of subjects included into each of the analysis sets will be summarized by treatment group and for all subjects based on all subjects screened. Percentages will be based on the RS for the purpose of this summary.

Screen failure reasons will be summarized, based on all subjects screened. A listing of subjects who did not meet study eligibility criteria (including glossary) will also be presented for all subjects screened.

In addition, the following listings will be presented by treatment group:

- Subject disposition (ES [all subjects screened])
- Study discontinuation (RS)
- Visit dates (SS)
- Subject analysis sets (ES [all subjects screened])

The listing of subject disposition will include:

- the date of informed consent
- for subjects participating in the respective pharmacogenomics, pharmacogenetic and PET MRI/CT substudies: date of respective informed consent
- date and time of first and last dose of IMP
- date of premature study termination and primary reason (if applicable)
- date of final contact
- if applicable, the date and reason for premature unblinding

The listing of study discontinuation will include the reason for discontinuation and the number of days on IMP.

The number of days on IMP will be calculated as follows:

$$Number\ of\ days\ on\ IMP = (Date\ of\ Last\ Dose\ Received - Date\ of\ First\ Dose\ Received) + 1$$

5.2 Protocol deviations

Important protocol deviations will be identified and classified by the deviations types listed in the IPD document.

A listing of all IPDs will be presented for all subjects in the RS and will include the deviation type and description.

The number and percentage of subjects with IPDs will be summarized by treatment group and for all subjects for each deviation type, based on the RS. This summary will be repeated for the FAS (if this analysis set is different to the RS).

6 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

6.1 Derivation of variables

6.1.1 Calculation of body mass index

The body mass index (BMI) in kg/m^2 is calculated based on the height (in m) and the weight (in kg) using the following formula:

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

The BMI will be automatically derived by and recorded on the electronic case report form (eCRF).

6.1.2 Classification of age categories

Age will be classified into categories based on requirements for European Union Drug Regulating Authorities Clinical Trials (EudraCT) and clinicaltrials.gov reporting.

For the EudraCT reporting, the categories will include:

- 18 to <65 years
- 65 to <85 years
- ≥85 years

For the clinicaltrials.gov reporting, the categories will include:

- ≤18 years
- 19 to <65 years
- ≥65 years

6.1.3 Duration of disease

The duration of disease will be calculated as follows and will be presented in years to 1 decimal place:

$$Duration = Date\ of\ Screening - Date\ of\ Diagnosis$$

In the event that the date of diagnosis is incomplete, it will be imputed to the most recent feasible date (ie, the imputed date must be before the date of Screening):

- If only the day is missing, it will be imputed to the last day of the known month
- If the day and month are missing, it will be imputed to December 31 in the known year
- If the date of diagnosis is completely missing or if the imputed date is not feasible, the duration of disease will not be calculated

6.2 Demographics

A by-subject listing of Baseline demographic characteristics will be presented by treatment group. This will include the year of birth (if available), age (in years), sex, race, ethnicity, height (in cm), weight at Screening (in kg) and body mass index (BMI) for all subjects screened. The

age and BMI will be obtained from the electronic case report form (eCRF) and will not be re-calculated for the reporting of demographic data.

All Baseline demographic characteristics (with the exception of year of birth) will be summarized by treatment group and for all subjects based on the SS. The summary of age will include descriptive statistics and categorical summaries, the latter based on requirements for EudraCT and clinicaltrials.gov reporting. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/CT substudy. Childbearing potential will be listed for the SS.

6.3 Other Baseline characteristics

The following Baseline disease characteristics will be presented in a by-subject listing for the SS:

- Duration of disease (as calculated in Section 6.1.3), including confirmation of sacroiliac joint X-ray and the date on which the sacroiliac joint X-ray was performed
- hs-CRP
- BASDAI score
- PGADA
- ASDAS
- ASDAS disease activity states (ASDAS-ID, ASDAS-Moderate Disease activity [ASDAS-MD], ASDAS-High Disease activity [ASDAS-HD] and ASDAS-Very High Disease activity [ASDAS-vHD])
- BASFI score
- Total and nocturnal spinal pain NRS
- PhGADA
- Total HADS-A and HADS-D scores

The calculation of Baseline ASDAS and the classification of the ASDAS disease activity states will be based on the Baseline BASDAI, PGADA and hs-CRP values. The Baseline for each of the variables listed above is defined in Table 3–1.

Baseline disease characteristics will be summarized using descriptive statistics by treatment group and for all subjects based on the SS. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy and will also include summary statistics for SUV_{auc} in the SIJ and spine at Baseline derived from the PET-MRI/PET-CT scans performed at Screening. SUV_{auc} values for each of the PET-positive lesions identified in the SIJ and spine will be pooled across subjects and across the SIJ and anatomical locations in the spine for the purpose of this summary. In addition, the maximum SUV_{auc} value across the SIJ and spine for each subject at Baseline will be summarized.

The above Baseline disease characteristics will be listed for the SS.

Subject lifestyle details (alcohol use, smoking history and caffeine use) will be listed, based on the SS.

6.4 Medical history and concomitant diseases

Medical history (except for AS) will be listed for the RS and summarized for the SS by treatment group and for all subjects, MedDRA® system organ class (SOC) and preferred term (PT). The reported term will be included in the listing. The summary will include the number and percentage of subjects and will be sorted alphabetically by SOC and by descending incidence of PT within each SOC, based on the 'All Subjects' column. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/CT substudy. A glossary of all medical history conditions will be presented for the RS including the SOC, PT and reported term.

Procedure history will be listed separately for the RS by treatment group for previous procedures relating to AS and previous procedures not relating to AS. Concomitant medical procedures performed during the study will be listed for the RS.

6.5 Ankylosing spondylitis history

Date of initial diagnosis of AS and the date that AS symptoms first started will be listed.

6.6 Prior and concomitant medications

Prior and concomitant medications (Section 6.6.1 and Section 6.6.2) will be listed for the RS by treatment group and subject, and summarized for the SS by WHODD Anatomical Main Group (Level 1), Pharmacological Subgroup (Level 3) and PT. The reported term will be included in the listing, together with flags to identify AS, rescue and prohibited medications, respectively. The flags for prohibited, AS and rescue medications will be confirmed and documented at the final DEM.

Prohibited medications will be listed separately.

Separate summaries will be presented for prior medications and concomitant medications and these will be presented by treatment group and for all subjects. Prior medications which continued into the study period are also classified as concomitant and are included in both summaries.

Separate summaries will be presented for the following, identified using WHODD Anatomical Main Group codes, the indication recorded on the eCRF and medical review of the recorded data:

- Prior and concomitant AS medications
- Rescue medication during study

All AS medications and rescue medication given during the study will be identified via medical review of the concomitant medications and confirmed at the DEMs prior to unblinding.

These summaries will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy.

All tabulations will be sorted alphabetically by Level 1 term, alphabetical Level 3 term within Level 1 and decreasing frequency of PT in the 'All Subjects' column.

A glossary of all prior and concomitant medications will be presented for the SS including the Anatomical Main Group (Level 1), Pharmacological Subgroup (Level 3), PT and reported Term.

Any medications with partially missing dates and/or times will be handled as described in Section 4.2.5 to classify them as prior or concomitant.

6.6.1 Prior medication definition

A medication with a start date that is prior to the date of first dose of IMP will be classified as a 'prior medication'.

6.6.2 Concomitant medication definition

A medication with a start date that is on or after the date of first dose of IMP will be classified as a 'concomitant medication'.

Any medication that started prior to the first dose of IMP and continued after will be classified as both prior and concomitant. Such medications will therefore be counted in the tabulations for both prior and concomitant medication.

7 MEASUREMENTS OF TREATMENT COMPLIANCE

During the Treatment Period of this study, IMP will be administered in the clinic and compliance will be determined at the visit by study personnel.

The number of injections administered at each visit will be included in the listing of IMP administration as specified in Section 11.1.

In addition, the percentage of planned injections administered will be calculated and presented on the listing.

This will be calculated as follows:

$$\%\ Planned\ Injections\ Administered = \frac{Total\ Number\ of\ Injections\ Received}{Total\ Number\ of\ Planned\ Injections} \times 100$$

where the planned number of injections for subjects who have completed the study is 14 for the Treatment Period and 16 for the Treatment Period (total of 30 injections). For subjects who have discontinued prematurely from the study, the planned number of injections will be the number of planned injections up to the date of discontinuation.

Any deviations from the planned dosing schedule will be addressed at the DEMs (including evaluation of any impact on the statistical analysis) and described in the CSR. No formal calculations of compliance will be presented as all IMP will be administered on site.

8 EFFICACY ANALYSES

8.1 Primary efficacy variable

8.1.1 Derivation of the primary efficacy variable

The ASDAS will be calculated based on individual component assessments of spinal pain, duration of morning stiffness and peripheral pain/swelling from the BASDAI together with the PGADA and hs-CRP values at the corresponding visit.

8.1.1.1 Bath Ankylosing Spondylitis Disease Activity Index

The BASDAI is the most commonly used instrument to measure the disease activity of AS from the subject's perspective. The BASDAI is a validated self-reported instrument which consists of

6 questions scored on a 0 to 10-unit horizontal numeric rating scale (NRS) to measure the severity of the 5 major symptoms: fatigue, spinal and peripheral joint pain and swelling, enthesitis, and morning stiffness (both severity and duration, respectively) over the last week.

Questions 2, 3 and 6 of the BASDAI are used in the calculation of ASDAS.

Question 2 asks [redacted]

Question 3 of the BASDAI [redacted]

Question 6 of the BASDAI [redacted] (Section 8.1.1.4).

The full BASDAI questionnaire is shown in Section 14.1.

8.1.1.2 Patient's Global Assessment of Disease Activity

Subjects will provide a global assessment of their disease activity in response to the question "How active was your spondylitis on average during the last week?" using a NRS where 0 is "not active" and 10 is "very active".

8.1.1.3 High-sensitivity C-reactive protein

Blood will be collected for the measurement of hs-CRP.

8.1.1.4 Calculation of Ankylosing Spondylitis Disease Activity Score

Each of the individual component assessments described above in Section 8.1.1.1, Section 8.1.1.2 and Section 8.1.1.3 are multiplied by a validated factor (van der Heijde et al, 2005) as listed below:

- 0.121 x Total spinal pain (BASDAI Question 2 result)
- 0.058 x Duration of morning stiffness (BASDAI Question 6 result)
- 0.110 x PGADA
- 0.073 x Peripheral pain/swelling (BASDAI Question 3 result)
- 0.579 x (natural logarithm of the hs-CRP [mg/L] + 1)

The results of these individual calculations or weighted components are then summed to obtain the ASDAS.

Handling of missing data and values of hs-CRP that are below 2 mg/L will be handled as described in Section 4.2.1.1.

ASDAS will be calculated at Baseline and at each post-Baseline visit per the schedule of assessments in the protocol.

The primary efficacy analysis will be based on the change from Baseline in ASDAS at Week 12.

8.1.2 Primary analysis of the primary efficacy variable

The primary efficacy analysis of the primary efficacy variable will be based on the PPS.

The primary efficacy analysis is the comparison of the change from Baseline in ASDAS at Week 12 in the BKZ treatment group versus the CZP treatment group in the PPS.

8.1.2.1 Bayesian modeling

The primary statistical analysis of the primary efficacy variable at Week 12 will be conducted following a Bayesian paradigm. The Bayesian analysis will be performed using a linear regression model including treatment group and Baseline ASDAS (mean centered using the overall ASDAS mean). The model will be parameterized such that the effect of CZP is represented by the intercept parameter whilst the effect of BKZ is additive to this. An informative prior will be used for the model for the intercept coefficient which, in the primary model, is equal to the mean change from Baseline in ASDAS in the CZP group when the centered Baseline mean is zero (βCZP ~ Normal [-1.78, var=0.0605]). Vague priors will be used for the other model coefficients, ie, centered Baseline coefficient and additive effect of BKZ.

Proof of Concept (PoC) will be declared if the posterior probability of a positive difference in the mean change from Baseline in ASDAS at Week 12 between the CZP and BKZ treatment groups (CZP-BKZ) is 0.975 or greater (ie, the lower bound of the 95% credible interval of the difference between treatment groups in the mean change from Baseline in ASDAS is zero or greater).

The model to be used is as follows:

$$y_i \sim \beta_{CZP} + \beta_{bl} x_{i,bl} + \beta_{BKZ} x_{i,BKZ}$$

Where i indexes the $i = 1, \ldots, N$ subjects, y_i is the change from Baseline in ASDAS at Week 12 for subject i, $x_{i,bl}$ is the centered Baseline ASDAS for subject i, $x_{i,\mathrm{BKZ}}$ is a variable equal to 1 if subject i was randomized to the BKZ group and 0 if not, β_{CZP} is the model intercept, β_{bl} is the coefficient for the centered Baseline ASDAS, and β_{BKZ} is the additive effect of BKZ in addition to the effect of CZP.

The prior distributions for the model parameters are:

$$\beta_{CZP} \sim N(-1.78, \sigma^2_{CZP} = 0.0605)$$
$$\beta_{bl} \sim N(0, \sigma^2_{\mathrm{bl}} = 10^6)$$
$$\beta_{BKZ} \sim N(0, \sigma^2_{BKZ} = 10^6)$$

where the prior for β_{CZP} is worth approximately 20 subjects (Section 2.4). The sample sizing for this study was based on the assumption that σ^2=1.1 which was based on the CZP data available from a single internal study (Section 8.1.2.2).

In addition to the primary Bayesian analysis where subjects with missing data will be excluded from the analysis, a supportive analysis based on imputed or predicted values for the change from Baseline in ASDAS at Week 12 will be performed.

Predicted Week 12 values for the subjects who do not have ASDAS values for the Week 12 visit will be obtained from a Mixed Model Repeated Measures (MMRM) analysis. The model will include treatment group, visit and the treatment group by visit interaction as fixed effects (visit as a categorical variable) and Baseline ASDAS as a continuous covariate. Subject-level random intercepts and slopes (visit as a continuous variable) will be included in the model and a 2 x 2

unstructured covariance matrix for these random effects will be used, along with a simple covariance structure for the within-subject errors.

If the model fails to converge then the simpler subject-level model including the intercept as only random effect will be fitted. If there are still problems with convergence and/or if the estimated response covariance matrices (random effects and response) are not positive definite, then alternative models will be considered, eg, allowing the random effects covariance parameters to be unbounded, or modelling the response covariance structure without using random effects.

The combination of observed and predicted values of change from Baseline in ASDAS at Week 12 obtained from the MMRM will then be modelled as the response variable in this supportive Bayesian analysis.

8.1.2.2 Building the prior

The CZP prior for this study has been built based on UCB study AS0001.

Study AS0001 was a Phase 3, multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of 2 dose regimens of CZP in subjects with active axial spondyloArthritis (axial SpA). Subjects were randomized in a 1:1:1 ratio to receive one of three treatment regimens:

- CZP administered subcutaneously (sc) at the dose of CZP 400mg every 2 weeks (Q2W) at Weeks 0, 2, and 4 followed by CZP 200mg Q2W sc;
- CZP administered sc at the dose of CZP 400mg Q2W at Weeks 0, 2, and 4 followed by CZP 400mg every 4 weeks (Q4W) sc;
- Placebo.

In this study AS0001, one of the secondary efficacy variables was the change from Baseline in ASDAS; the primary efficacy variable for the current study AS0013.

To build the prior using these data, subjects in the CZP 400mg treatment group were removed since their CZP responses were sufficiently different from those of the CZP 200mg subjects (leaving 105 CZP 200mg subjects with Baseline and Week 12 data available), and the linear model detailed in Section 8.1.2.1 was fitted to just the Placebo and CZP 200mg treatment groups with Placebo as the additive treatment effect (a treatment group by Baseline ASDAS interaction was also included in the model for this historical).

The estimated model was used to estimate the distribution $\beta_{CZP} \sim N(-1.78, \sigma^2_{CZP} = 0.0078)$ and might be used as a prior for the CZP mean response. However, in order to take into account expected study to study variation in the estimated CZP parameter β_{CZP}, this prior was discounted (the SD was inflated) to arrive at the prior given in Section 8.1.2.1 (by inflating the SD by 2.788 to give an effective sample size of approximately 20 subjects).

The historical data in study AS0001 gave the opportunity to build an informative prior for the overall effect of Baseline ASDAS in the proposed study (ignoring treatment group since a Baseline by treatment group interaction is not expected). However, since the AS0001 study only included Placebo and CZP treatment group data there is currently no data available with which to confirm whether or not the estimated Baseline ASDAS coefficient of -0.66 for CZP is also similar for BKZ. Accordingly, whilst a Baseline by treatment group interaction in the proposed

study is not expected, to account for this uncertainty a non-informative prior was used for the Baseline ASDAS coefficient.

8.1.2.3 Prior data conflict

Prior data conflict will be investigated for all informative priors through use of graphical displays. If there is evidence of prior data conflict for a given parameter, then a vague prior will be used for that parameter.

8.1.3 Presentation of the primary analysis of the primary efficacy variable

The Bayesian analysis will be implemented using multiple chains each utilizing different starting values for the model specified parameters. The posterior samples from each chain will be combined into one dataset from which the posterior summaries will be estimated. Missing data will be handled as described in Section 4.2.1

The posterior distributions of the mean changes from Baseline in ASDAS in the CZP and BKZ groups and of the difference in mean changes from Baseline between the treatment groups (CZP – BKZ) will be summarized using means, medians, SDs, 95% credible intervals and 95% highest posterior density intervals. The posterior probability that the difference between treatment groups in the mean change from Baseline in ASDAS at Week 12 is greater than zero will also be presented.

The posterior probability that BKZ achieves a lower ASDAS score at Week 12 compared to CZP will be derived from this distribution.

8.1.4 Presentation of Ankylosing Spondylitis Disease Activity Score

All component data, the derived ASDAS values and changes from Baseline at each visit will be listed by treatment group and subject, including each of the ASDAS disease activity states and disease improvement categories defined in Section 8.2.1.1, based on the FAS. Observed ASDAS values and changes from Baseline will be summarized by treatment group and visit, using descriptive statistics, based on the PPS and repeated for the FAS.

Observed mean values and changes from Baseline, including 95% CI for the means, will be plotted versus time, with both treatment groups overlaid on the same axes, based on the PPS and repeated for the FAS.

8.1.5 Presentation of component variables of Ankylosing Spondylitis Disease Activity Score

All data presentations in this section will be based on the PPS and repeated for the FAS.

8.1.5.1 Bath Ankylosing Spondylitis Disease Activity Index

The BASDAI score ranges from 0 to 10, with lower scores indicating lower disease activity.

The BASDAI questionnaire contains 6 questions, each of which has a possible value ranging from 0 to 10, and the overall score is calculated as follows:

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

To give each symptom equal weighting, the average of the 2 scores relating to morning stiffness (questions 5 and 6) is taken. The sum of the scores from questions 1 to 4 and the average of the scores from questions 5 and 6 result in a total score ranging from 0 to 50. This is divided by 5 to give a final BASDAI score between 0 and 10, with lower scores indicating lower disease activity.

The individual question responses from the BASDAI together with the derived BASDAI score will be listed by treatment group and visit for each subject, including changes from Baseline for the BASDAI score. Additionally, for each subject the mean result from questions 5 and 6, together with the change from Baseline and percentage change from Baseline in this mean value, will be listed.

The observed BASDAI score and changes from Baseline will be summarized by treatment group and visit.

Additionally, the mean of question 5 and 6, together with the corresponding changes from Baseline and percentage changes from Baseline will be summarized by treatment group and visit.

Observed mean BASDAI score and changes from Baseline, including 95% CI for the means, will be plotted versus time, with both treatment groups overlaid on the same axes.

The full BASDAI questionnaire is shown in Section 14.1. A glossary of the BASDAI questionnaire will be presented.

8.1.5.2 Patient's Global Assessment of Disease Activity

Patient's Global Assessment of Disease Activity will be listed by treatment group, subject and visit, including changes from Baseline and percentage changes from Baseline.

Observed values, changes from Baseline and percentage changes from Baseline will be summarized by treatment group and visit.

Observed mean PGADA and changes from Baseline, including 95% CI for the means, will be plotted versus time, with both treatment groups overlaid on the same axes.

8.1.5.3 High sensitivity C-reactive protein

Measurements of hs-CRP will be listed by treatment group, subject and visit for each subject including changes from Baseline and ratio to Baseline. The latter will be calculated as follows:

$$Ratio\ to\ Baseline\ =\ hs\text{---}CRP\ at\ Post\text{–}Baseline\ Visit/\ hs\text{–}CRP\ at\ Baseline\ Visit$$

Observed values, changes from Baseline and ratio to Baseline will be summarized by treatment group and visit. Summary statistics will include n, arithmetic mean, SD, geometric mean, 95% CI for the means, geoCV, median, Q1, Q3, minimum and maximum.

The geoCV will be calculated as a percentage as:

$$geoCV(\%) = 100 \times \sqrt{\exp(SD_L^2) - 1},$$

where SD_L represents the standard deviation of the natural logarithm transformed value.

Observed mean hs-CRP values, changes from Baseline and ratio to Baseline, including 95% CI for the means, will be plotted versus time, with both treatment groups overlaid on the same axes.

The rules for handling any observed concentration values that are BLQ or ALQ are described in Section 4.2.1.2.

8.1.6 Supportive analysis of the primary efficacy variable

8.1.6.1 Alternative analysis sets

The primary efficacy analysis set is the PPS; however, the primary analysis and all supportive analyses of the primary efficacy variable will be repeated using the FAS. These analyses will be performed in order to evaluate the effect that any IPDs may have had on the results of the analysis. The results will be presented as specified in Section 8.1.3.

8.1.6.2 Alternative prior distributions

The primary analysis will be repeated for each of the following prior distribution alternatives:

- A vague prior distribution for all parameters in the model (including the CZP effect).
- An informative prior for the Baseline effect [$\beta_{bl} \sim N(-0.66, \sigma_{bl}^2 = 25)$] and an informative prior for the CZP effect.
- An informative prior for the Baseline effect [$\beta_{bl} \sim N(-0.66, \sigma_{bl}^2 = 25)$] and a vague prior for the CZP effect [$\beta_{CZP} \sim N(0, \sigma_{CZP}^2 = 10^6)$].

The results will be presented as specified in Section 8.1.3.

8.1.6.3 Frequentist analysis – Analysis of Covariance

A frequentist analysis of the change from Baseline in ASDAS at Week 12 will be performed using an analysis of covariance (ANCOVA). The ANCOVA model will match the Bayesian model, ie, including treatment group (CZP and BKZ) as a fixed effect and Baseline ASDAS as a covariate.

Least squares means (LSMeans), SEs and associated 95% CIs for the changes from Baseline in ASDAS at Week 12 in each treatment group will be presented. In addition, the estimated treatment difference in LSMeans (CZP – BKZ), SE, corresponding 95% CI, and p-value for the treatment comparison will be presented.

8.1.6.4 Frequentist analysis – Mixed Model Repeated Measures

A frequentist analysis of the change from Baseline in ASDAS will be performed using a MMRM. This model will use all available change from Baseline in ASDAS data at all post-Baseline visits during the Treatment Period up to and including Week 12. Treatment group (CZP and BKZ), visit, and the treatment group by visit interaction will be included as fixed effects and Baseline ASDAS will be included as a covariate.

An unstructured covariance matrix will be used to model the within-subject errors. If the model fails to converge, and/or if the estimated unstructured covariance matrix is not positive definite, then simpler alternative covariance structures will be considered (eg, compound symmetry).

Least squares means, SEs and associated 95% CIs for the changes from Baseline in ASDAS in each treatment group will be presented by visit. In addition, the estimated treatment difference in LSMeans (BKZ – CZP), SE, corresponding 95% CI, and p-value for the treatment comparison will be presented for each visit. Week 12 will be the primary time point of interest from this analysis.

The LSMeans for the changes from Baseline, including 95% CI, will be plotted versus time, with both treatment groups overlaid on the same axes.

8.1.6.5 Frequentist analysis – Mixed Model Repeated Measures with Baseline by treatment group interaction

A frequentist analysis of the change from Baseline in ASDAS will be performed using a MMRM as specified in Section 8.1.6.4. In addition, the model will include the Baseline ASDAS by treatment group interaction.

The results presentation, and the covariance matrix structures, as specified in Section 8.1.6.4, will be applied.

8.2 Secondary efficacy variables

Secondary efficacy variables will be listed and summarized by treatment group and visit for the PPS and repeated for the FAS.

8.2.1 ASDAS Disease Activity States and Improvement

8.2.1.1 Derivation of ASDAS Disease Activity States and Improvement

Cut-offs for 4 disease activity states based on ASDAS will be defined for each subject at Baseline and at each post-Baseline visit as follows:

- ASDAS-Inactive Disease (ASDAS-ID): ASDAS < 1.3
- ASDAS-Moderate Disease activity (ASDAS-MD): ASDAS ≥ 1.3 to < 2.1
- ASDAS-High Disease activity (ASDAS-HD): ASDAS ≥ 2.1 to ≤ 3.5
- ASDAS-Very High Disease activity (ASDAS-vHD): ASDAS > 3.5

In addition, the following cut-offs for 2 levels of improvement in ASDAS will be defined for each subject at each post-Baseline visit based on the change from Baseline in ASDAS:

- ASDAS Clinically Important Improvement (ASDAS-CII): a reduction (improvement) from Baseline in ASDAS ≥ 1.1 units
- ASDAS Major Improvement (ASDAS-MI): a reduction (improvement) from Baseline in ASDAS ≥ 2.0 units

ASDAS-Inactive Disease (ASDAS-ID) and ASDAS-Major Improvement (ASDAS-MI) at Week 12 are the secondary efficacy variables defined for this study. Subjects with missing ASDAS at Week 12 will not be classified in terms of their ASDAS disease activity state or improvement according to the above definitions.

8.2.1.2 Analysis of ASDAS Disease Activity States and Improvement

The number and percentage of subjects classified in each of the disease activity state categories (ASDAS-ID, ASDAS-MD, ASDAS-HD and ASDAS-vHD) and each disease improvement category (ASDAS-CII [yes/no] and ASDAS-MI [yes/no]) and will be tabulated by treatment group and visit including the 95% CIs calculated using a Wilson approximation.

Additionally, inferential treatment effects for ASDAS-ID and ASDAS-MI at Week 12 will be derived from the Bayesian analysis as described in Section 8.1.2.1. ASDAS-ID and ASDAS-MI rates at Week 12 will be presented.

8.3 Other efficacy variables

8.3.1 Derivation of Assessment in SpondyloArthritis International Society Response

The following ASAS-based variables will be derived:

- ASAS20 response is defined as an improvement of at least 20% (percentage change from Baseline of -20% or more negative), and an absolute improvement of at least 1 unit (change from Baseline of -1 or more negative) in at least 3 of the 4 following domains:
  - PGADA (Section 8.1.5.2)
  - Function (represented by BASFI, Section 8.3.3.1)
  - Pain assessment (the total spinal pain NRS score, Section 8.3.3.2)
  - Inflammation (the mean of the BASDAI Questions 5 and 6 results, Section 8.1.5.1, concerning morning stiffness intensity and duration)

  In addition, *absence* of deterioration in the potential remaining domain is required. Deterioration is defined as the presence of both of the following:
  - percentage change from Baseline of 20% or more;
  - change from Baseline of 1 unit or more.
- The ASAS40 response is defined as an improvement of at least 40% (percentage change from Baseline of -40% or more negative), and an absolute improvement of at least 2 units (change from Baseline of -2 or more negative) in at least 3 of the 4 domains above. In addition, no worsening (a change from Baseline that is non-positive [ie, negative or zero]) in the remaining domain is required.
- Time to ASAS20 response: defined as the time (in days) from the first dose of IMP to the first instance (if more than one instance) of an ASAS20 response.
  - This will be calculated for the Treatment Period and across the entire study period (Treatment Period + Treatment Extension Period) separately.
  - For time to ASAS20 response during the Treatment Period:
    - Subjects who discontinue early prior to an ASAS20 response will be censored on the date of their last efficacy assessment during the Treatment Period.
    - Subjects who do not have an ASAS20 response during the Treatment Period will be censored on the date of their Week 12 efficacy assessment (ie, end of the Treatment Period).
  - For time to ASAS20 response during the study (Treatment Period + Treatment Extension Period):

  - Subjects who discontinue early prior to an ASAS20 response will be censored on the date of their last efficacy assessment during the study (Treatment Period + Treatment Extension Period).
  - Subjects who do not have an ASAS20 response during the study will be censored on the date of their Week 48 efficacy assessment (ie, end of the Treatment Extension Period).

- Time to ASAS40 response: defined as the time (in days) from the first dose of IMP to the first instance (if more than one instance) of an ASAS40 response. As described above for time to ASAS20 response, time to ASAS40 response will be calculated based on the Treatment Period and also across the study period (Treatment Period + Treatment Extension Period) separately and the same censoring rules will be applied.
- The ASAS partial remission (ASAS-PR) response is defined as a score of ≤ 2 units in all 4 domains listed above.

In order for a subject to be defined as a responder (ASAS20, ASAS40 or ASAS-PR, respectively) at a given visit, the criteria stated must be met all together that visit.

Missing component data for the derivation of the ASAS20, ASAS40 and ASAS-PR will be handled as described in Section 4.2.1.4.

8.3.2 Presentation of Assessment in SpondyloArthritis International Society Response

8.3.2.1 ASAS20 and ASAS40 response

A by-subject listing showing ASAS20 and ASAS40 response status will be presented by treatment group and visit for the FAS.

The number and percentage of subjects classified as ASAS20 and ASAS40 at each post-Baseline visit will be tabulated by treatment group for the FAS including the 95% CI calculated using a Wilson approximation.

8.3.2.2 Time to ASAS20 and ASAS40 response

A by-subject listing showing time to ASAS20 response and time to ASAS40 response will be presented by treatment group and visit for the FAS. This listing will include flags for censored observations.

The time to ASAS20 response and time to ASAS40 response during the Treatment Period and during the study (Treatment Period + Treatment Extension Period) will be summarized for the FAS separately using the following statistics:

- Number and percentage of subjects that are responders
- Number and percentage of subjects that are censored
- Minimum, maximum and Kaplan-Meier estimates of the 25th percentile, median and 75th percentile, with corresponding 95% CIs

Kaplan-Meier response rates for ASAS20 and ASAS40 will be presented for each treatment, including the cumulative number of events and the number of subjects at risk at each study visit.

This will be presented separately for the Treatment period and across the study (Treatment Period + Treatment Extension Period).

Kaplan-Meier curves will be presented for time to ASAS20 response and time to ASAS40 response for the Treatment Period and across the study (Treatment Period + Treatment Extension Period) separately, with both treatment groups overlaid on each plot. The plot will start at zero (ie, no responders at time zero) and be non-decreasing.

8.3.2.3 ASAS-PR

A by-subject listing showing ASAS-PR response status will be presented by treatment group and visit for the FAS.

The number and percentage of subjects classified as ASAS-PR at each post-Baseline visit will be tabulated by treatment group for the FAS including the 95% CI calculated using a Wilson approximation.

8.3.3 Presentation of components of Assessment in SpondyloArthritis International Society

8.3.3.1 Bath Ankylosing Spondylitis Functional Index

The BASFI is a validated index used to determine the degree of functional limitation in subjects with AS. The BASFI questionnaire comprises 10 questions, each of which has a possible value ranging from 0 ("Easy") to 10 ("Impossible"). The first 8 questions evaluate activities related to functional anatomical limitations due to AS and the final 2 questions evaluate a subject's ability to cope with everyday life. The BASFI will be administered at all visits per the schedule of events.

The BASFI score is the arithmetic mean of the 10 individual scores giving a value between 0 and 10, with lower scores indicating better physical function.

Missing data for the BASFI questionnaire will be handled as described in Section 4.2.1.5.

The BASFI results will be listed by treatment group and visit for each subject in the FAS including the results from each of the individual 10 questions, the BASFI score, changes from Baseline and percentage changes from Baseline in the BASFI score.

The full BASFI questionnaire is shown in Section 14.2. A glossary of the BASFI questionnaire will be presented.

Observed values, changes from Baseline and percentage changes from Baseline in the BASFI score will be summarized by treatment group and visit for the FAS.

8.3.3.2 Total and nocturnal spinal pain numeric rating scale

The pain experienced by subjects due to AS will be assessed by the following questions, referring to the average experience over the past week:

- Total pain in the spine: How much pain of your spine due to spondylitis do you have?
- Nocturnal spinal pain: How much pain of your spine due to spondylitis do you have at night?

The response to these questions will be rated by the subjects on a scale from 0 (no pain) to 10 (most severe pain) and the response will be recorded on the eCRF.

The Total and Nocturnal spinal pain scores will be listed by treatment group and visit for each subject in the FAS including changes from Baseline and percentage changes from Baseline.

Observed values, changes from Baseline and percentage changes from Baseline will be summarized by treatment group and visit for the FAS for each of Total and Nocturnal spinal pain scores.

8.3.4 Physician's Global Assessment of Disease Activity

The Investigator will assess the overall status of the subject with respect to their AS signs and symptoms and functional capacity (considering both joint and skin components) using an NRS where 0=very good, asymptomatic and no limitations of normal activities and 10=very poor, very severe symptoms which are intolerable and inability to carry out normal activities.

A by-subject listing will present the PhGADA results by treatment group and visit for the FAS, including changes from Baseline. Observed values and changes from Baseline will be summarized by treatment group and visit for the FAS.

8.3.5 Hospital Anxiety and Depression Scale

The HADS was chosen for its well-established psychometric properties and its use in clinical research on biological therapy in subjects with other inflammatory skin diseases (Dauden et al, 2009; Langley et al, 2010).

The HADS is a commonly used instrument to determine the levels of anxiety and depression that a subject is experiencing. The HADS is a 14-item scale that generates ordinal data. Seven items relate to anxiety (assessed in the HADS-A) and seven relate to depression (assessed in the HADS-D). Individual item scores range from 0 to 3 and are summed to give a total score on each of the individual HADS-A and HADS-D scales. These total scores range from 0 to 21 with higher scores indicating a worse state.

In case of missing data, the rules in Section 4.2.1.6 will apply.

The full HADS questionnaire is shown in Section 14.3. A glossary of the HADS questionnaire will be presented.

Scores for each item on the HADS-A and HADS-D will be listed at each visit by treatment group for each subject in the FAS. This listing will also include the total HADS-A and HADS-D scores together with the changes from Baseline in the total HADS-A and total HADS-D scores at each visit. Summary statistics will be presented by treatment group for the observed total HADS-A and HADS-D scores together with changes from Baseline at each visit for the FAS.

8.3.6 Change in Bone Formation (PET-MRI/PET-CT Substudy)

To evaluate the effect of BKZ or CZP on changes in bone formation, an independent assessment of SIJ radiographs at Screening (to assess eligibility of subjects for the PET-MRI/PET-CT substudy) and PET-MRI/PET-CT scans of the entire spine and SIJ (to assess positivity and efficacy in subjects eligible for the substudy) at Baseline, Week 12 and Week 48 will be conducted.

8.3.6.1 Eligibility assessment

SIJ radiographs at Screening will be independently assessed in a blinded fashion by two reviewers. An image quality assessment will be performed, modified New York (mNY) scores

recorded and a subject's eligibility determined (subjects must be mNY-positive to be randomized into the substudy). If there is disagreement between the two primary reviewers in the assessment of eligibility, then a third reviewer will perform an adjudication based on an independent assessment of the images. The results of the adjudication of eligibility will be considered as the final assessment.

All data recorded for the eligibility assessment, including imaging date, quality of image, right and left SIJ sacroiliitis scoring and radiograph assessment (mNY-positive/negative) will be listed by treatment group for each primary reviewer and for the adjudicator (if applicable). The listing will include all subjects in the SS who were randomized into the PET-MRI/PET-CT substudy.

8.3.6.2 Efficacy assessment

SIJ and spine PET-CT and PET-MRI scans will be independently assessed at each time point (Baseline, Week 12 and Week 48) in a blinded fashion by two reviewers, with an additional adjudication read by a third reviewer, if applicable. Independent reviewers will assess PET-MRI/PET-CT examinations quantitatively using VU University Medical Center (VUMC)'s in-house software (ACCURATE).

Initially, two primary readers will independently assess the images visually to identify PET-positive lesions in the SIJ and spine. The readers will dichotomously score the presence or absence (1 versus 0) of PET-positive lesions in each side of the SIJ (Left and Right) and at each of the 24 spinal levels and 4 spinal regional levels [C2-C7 (Cervical), T1-T12 (Thoracic), L1-L5 (Lumbar), Sacrum] in each of the following 8 spinal locations:

- Process spinosus
- Costo-vertebral (right)
- Costo-vertebral (left)
- Facet joint (right)
- Facet joint (left)
- Anterior vertebra
- Posterior vertebra
- Vertebra (other)

For each subject, PET-positive lesions may be identified in one or both sides of the SIJ and in one or more of the 192 possible anatomical locations (spinal location/level combinations). It is expected that only one lesion for a subject may be separated/identified in each side of the SIJ or individual anatomical location in the spine, therefore, there is a direct mapping of a subject's lesion to a SIJ side or spinal location/level.

If there is a discrepancy between the two primary readers in the presence of PET-positive lesion assessment (Yes versus No for lesions in any side of the SIJ and 1 versus 0 for lesions in any spinal location), a consensus adjudication will take place with a third reader who will also be blinded to the clinical data. This reader will make an independent assessment and the results of the adjudication will be considered as the final assessment of the presence of PET-positive lesions.

Another reader who has not been involved in the visual identification of the PET-positive lesions, either as a primary reader or an adjudicator, will then draw volumes of interest (VOIs) on the pre-identified PET-positive lesions. In the case that no PET-positive lesions have been identified for a subject, then the final review step to draw VOIs will not occur. Efficacy PET-MRI/PET-CT review will occur provided that an acceptable PET-CT or PET-MRI scan is available and that PET-positive lesions were identified on the previous scan (for the Week 12 and Week 48 visits). During the PET-positive lesion VOI review, the independent reviewer will perform a quantitative assessment in ACCURATE and record the standard uptake values of $_{18}$F-fluoride in VOIs, corrected for both body weight and individual integrated whole blood activity concentration (SUV_{auc}) for each PET-positive lesion identified during the PET-positive lesion identification review.

Data from the visual identification of PET-positive lesions and the results of the quantitative VOI review (i.e. SUV_{auc}) will be listed, for each independent reader and adjudicator (if applicable), for each anatomical location in the spine, and for each side of the SIJ, by treatment group and visit for subjects in the SS who were enrolled into the PET-MRI/PET-CT substudy.

No formal statistical modelling of the PET-MRI/PET-CT imaging data will be performed for this substudy; only summaries will be presented together with 95% CIs for mean values, as applicable.

The summaries described below based on the visual identification of PET-positive lesions will include the final assessment data, ie, if adjudication was performed, then the adjudicated data will be summarized, otherwise data from one of the primary reviewers will be summarized.

All summaries will be presented for the PET-PPS.

Since the attenuation correction and quantification is variable in PET-MRI scans, PET-MRI imaging data will be listed but will not be aggregated with the PET-CT imaging data in the summaries.

Due to the predicted sparseness of the number of PET-positive lesions in each of the individual anatomical locations of the spine, data will be pooled as follows for the purpose of some of the summaries:

- In the SIJ and spine overall (across the SIJ and all anatomical locations in the spine);
- In the SIJ overall (across both sides), and split by Left and Right side;
- In the spine overall (across all anatomical locations in the spine), and split by spinal regional level (Cervical, Thoracic, Lumbar, Sacrum), and, depending on the distribution of lesions across the spine, further by spinal levels and spinal locations in which at least 1 PET-positive lesion is identified.

An overview of the total number of PET-positive lesions identified in the spine across all subjects in each of the individual anatomical locations will be presented by visit.

The presence of PET-positive lesions identified in the SIJ and spine will be summarized by visit. The summary will include the number and percentage of subjects at each visit with any PET-positive lesions in the spine or the SIJ, in both the spine and the SIJ, in the SIJ overall (and split by Left and Right side) and in the spine overall [and split by spinal regional level (Cervical, Thoracic, Lumbar and Sacrum) and level/location, as applicable]. This summary will also

include a total count of PET-positive lesions across subjects for each of the pooled locations described above. In the spine, the incidence and the total number of lesions will only be presented for the anatomical locations in which at least 1 PET-positive lesion is identified.

The number of PET-positive lesions per subject in the SIJ and spine (overall and in pooled locations, as described above) will be calculated and summarized by visit using frequency counts and percentages, and/or descriptive statistics, as applicable. In addition, in the spine and SIJ overall, in the spine overall and in each spinal regional level, the changes in the number of PET-positive lesions per subject from Baseline to Week 12 and Week 48, and the changes from Week 12 to Week 48 will be summarized using descriptive statistics.

Bar charts summarizing the total number of PET-positive lesions identified in the SIJ and in the spine will be presented by treatment group and visit for each of the pooled locations described above.

Observed SUV_{auc} values for each lesion will be pooled across subjects and summarized by visit for each of the pooled locations, as described above. In the spine, the observed SUV_{auc} values will only be summarized for the anatomical locations in which at least 1 PET-positive lesion is identified. Descriptive statistics will be presented including 95% CIs for the mean observed SUV_{auc} values. In addition, changes and percentage changes in SUV_{auc} values between visits will be calculated for each subject's PET-positive lesion in the SIJ and spine, i.e. Week 12 - Baseline, Week 48 - Baseline and Week 48 - Week 12. These changes in SUV_{auc} values will be pooled across subjects and summarized by visit for each of the pooled locations described above. Descriptive statistics will be presented including 95% CIs for mean changes and percentage changes in SUV_{auc} values between visits.

Spaghetti plots of observed SUV_{auc} values for lesions in the SIJ and spine will be presented over time for the SIJ overall, the spine overall, and for each spinal regional level (Cervical, Thoracic, Lumbar, Sacrum) by treatment group.

In addition, percentage changes in SUV_{auc} values for lesions in the SIJ and spine will be summarized using waterfall plots. These plots will identify subjects' lesions by treatment group and will be presented by pooled location (as described above, as applicable) and for each visit comparison (Week 12 – Baseline, Week 48 – Baseline and Week 48 – Week 12).

The maximum of a subject's observed SUV_{auc} values will be calculated across the pooled locations described above. This maximum SUV_{auc} value per subject will be summarized at each visit together with changes and percentage changes from Baseline between visits. Descriptive statistics will be presented including 95% CIs for the mean and the mean changes/percentage changes in the maximum SUV_{auc} values.

The mean and 95% CI for the maximum SUV_{auc} values per subject in the SIJ and spine over scheduled time will also be presented by treatment group. The same plot will be repeated for the changes and percentage changes in the maximum SUV_{auc} value per subject with treatment group overlaid on the same plot.

In addition, percentage changes in the maximum of SUV_{auc} values per subject in the SIJ and spine will be summarized using waterfall plots. These plots will identify subjects by treatment group and will be presented for each visit comparison (Week 12 – Baseline, Week 48 – Baseline and Week 48 – Week 12).

Additional exploratory analyses may include examining the relationship between changes from Baseline in bone formation and changes from Baseline in ASDAS at Week 12 (primary efficacy variable) and Week 48, and also whether there are any differences in bone formation in clinical responders according to ASAS20 versus clinical non-responders at Weeks 12 and 48. Statistical modelling which accounts for multiple PET-positive lesions per subject may be used in these analyses.

9 PHARMACOKINETICS AND PHARMACODYNAMICS

9.1 Pharmacokinetics

There will be no calculation of PK parameters for this study. The analysis of PK data for this study will focus on the concentrations of CZP and BKZ only. Any additional PK analyses will be performed only if required.

The actual blood sampling times for BKZ and CZP concentrations will be obtained in days relative to the start time of the first administration of IMP for all sampling time points.

Individual blood sampling times and concentrations of BKZ and CZP will be listed by treatment group for the SS and will include the actual sampling time in days relative to the first.

Individual concentrations will be summarized at each scheduled time point based on the PK-PPS, using n, mean, median, SD, minimum, maximum, geometric mean and 95% CI, and geoCV (assuming log-normally distributed data).

Individual concentration versus time (day) profiles will be presented graphically on linear and semi-logarithmic scales with all subjects overlaid on the same plot (spaghetti plots). Geometric mean profiles, separately for BKZ and CZP, will also be presented on both linear and semi-logarithmic scales respectively with and without the corresponding lower and upper limit of the 95% CI. All semi-logarithmic plots will include the respective LLOQ marked on the y-axis.

The following rules will apply for PK data listings and summaries:

- Values below the LLOQ will be reported as BLQ in the listings
- Descriptive statistics of concentrations will be calculated if at most ⅓ of the individual data points are missing or are not quantifiable (<LLOQ). Values that are BLQ will be replaced by the numerical values of the LLOQ/2 in this instance. However, if n≤3, then only n, minimum and maximum will be presented (where 'n' refers to the number of quantifiable values). The other descriptive statistics will be left blank.
- The 95% CI for the geometric mean will be left blank if the SD is 0

The geoCV (%) will be calculated using the following formula where SD_L is the standard deviation from the log-transformed data:

$$geoCV\ (\%) = 100 \times \sqrt{\exp(SD_L^2) - 1}$$

9.2 Pharmacodynamics

All PD variables will be listed by treatment group, subject and visit including changes from Baseline, based on the SS.

Observed values and changes from Baseline will be summarized by treatment group and visit based on the PD-PPS.

Figures of mean and mean changes from Baseline over time will be presented with separate plots for each variable and both treatment groups overlaid on the same plot.

The rules for handling any values that are BLQ or ALQ are described in Section 4.2.2.

10 IMMUNOLOGICAL ANALYSES

The results for the ADA measurements will be listed by treatment group and time point based on the SS, including the Screening assay, confirmatory assay and titre (if applicable).

A cut point will be determined by the bioanalytical laboratory that will be used to determine the status of ADA as above the cut point (ACP) or below the cut point (BCP). For any ADA levels that are ACP, a further confirmatory assay will be performed, the results of which will be determined as either 'confirmed positive' (CP) or 'not confirmed positive' (NCP). For samples that are CP, a further titre assay will be performed and the ADA titre will be reported. The following definitions will be applied:

- An ADA status for a given time point of positive will be concluded for any subject with an ADA level that is ACP and CP at any time point
- An ADA status of negative for a given time point will be concluded for any subject with an ADA level that is either BCP or ACP and NCP at any time points
- A subject will be classified as having ADA positivity at Baseline if the Visit 2 (Week 0), pre-dose result is ACP and CP
- A subject will be classified as having treatment-induced ADA positivity when meeting 1 of the following criteria:
  - The Baseline result is either BCP or ACP and NCP, and at least 1 post-Baseline time point is ACP and CP
  - The Baseline result is positive (ACP and CP) and at least 1 post-Baseline measurement shows a pre-defined fold increase in titre from the Baseline value (the fold increase from Baseline required to meet this criterion will be defined with the development of the assay and will be included in the TFLs)
- A subject will be classified as overall positive if at least 1 post-Baseline measurement is ACP and CP (this includes subject who have negative results at Baseline)
- A subject will be classified as overall negative if at all post-Baseline visits the ADA status is negative (this includes subjects who have positive [ACP and CP] results at Baseline)

The ADA status (positive/negative) will be summarized as a categorical endpoint (number and percentage of subjects) by treatment group, for all time points and overall, based on the PK-PPS. In addition, the first occurrence of treatment-induced ADA positivity (based on the definitions

above) will be summarized by treatment group (number and percentage of subjects) at each post-Baseline visit, based on the PK-PPS. This tabulation will count the number of subjects at each post-Baseline visit who fulfill at least 1 of the above defined criteria for treatment-induced positivity; subjects will be counted in the numerator based on the earliest visit at which 1 of these criteria is fulfilled. At other visits, subjects will be counted in the denominator (assuming a measurement is available). For all tabulations, percentages will be based on the number of observations at each visit.

Separate listings will be presented (based on the PK-PPS) showing the BKZ and CZP concentrations and ADA measurements in the same output in adjacent columns. The listing will include the BKZ and CZP concentration, ADA status (ACP or BCP) and confirmatory assay results if applicable (NCP or CP), together with the titre for results that are CP. In addition, the time since the previous administration of IMP will be reported (in days).

Finally, individual subject plots (based on the PK-PPS) will be presented displaying ADA titre, BKZ and CZP concentrations overlaid on the same figure using a semi-logarithmic scale.

The rules for handling values that are BLQ in the BKZ and CZP concentration data are described in Section 4.2.4. For the ADA data, any negative results for which there are no titres available at a specific visit will be substituted with 0.001 for the purpose of the figure.

11 SAFETY ANALYSES

All analysis of safety variables will be performed using the SS.

11.1 Extent of exposure

All IMP administration details (including date and time of administration, body site and side of administration) will be listed. Additionally, the time between each dose will be listed.

The duration of exposure to BKZ (days) will be calculated separately for the Treatment Period and the Treatment Extension Period, where the last and first dose in each equation will be the last and first dose of IMP within the Treatment Period and Treatment Extension Period respectively:

$$Exposure = (Date\ of\ Last\ Dose - Date\ of\ First\ Dose) + 140\ days$$

The duration of exposure to CZP (days) will be calculated as follows:

$$Exposure = (Date\ of\ Last\ Dose - Date\ of\ First\ Dose) + 70\ days$$

The number of days added to the BKZ and CZP exposure calculations are 5 times the half-life of each drug, respectively: 5 x 28 days for BKZ and 5 x 14 days for CZP.

Missed doses will not be accounted for in the calculation above.

The number of injections received and duration of exposure will be listed for each subject and summarized by treatment group for the Treatment Period and Treatment Extension Period separately using descriptive statistics. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/CT substudy.

11.2 Adverse events

Adverse events with start date prior to the first dose of IMP are defined as pre-treatment AEs. Adverse events with a start date prior to the first dose of IMP will be defined as pre-treatment AEs. A treatment-emergent AE (TEAE) is defined as any AE with a start date/time at the time of

or after the first dose of IMP up until 140 days after the last dose of IMP. Any AE with onset later than 140 days after the last dose of IMP will not be considered as treatment-emergent and therefore will not be included in the tabulations of TEAEs. Selected summaries will be presented separately for TEAEs during the Treatment Period (any AE with a start date/time at the time or after the first dose of IMP up until the Week 12 Visit) and TEAEs during the study (as per the definition of TEAEs given above).

Where dates are missing or partially missing, AEs will be assumed to be treatment-emergent, unless there is clear evidence to suggest that the AE started prior to dosing or later than 140 days post-last dose. Handling of missing dates/times for classification of AEs as TEAEs is described in Section 4.2.5.

All AEs will be recorded in the eCRF from informed consent until study completion or termination. All AEs will be coded and categorized by intensity (mild/moderate/severe) and relationship to IMP (related/not related).

An overview of the number and percentage of subjects who experience TEAEs will be presented by treatment group, based on the SS. Separate tabulations will be presented for TEAEs during the Treatment Period and TEAEs during the study and will include the number and percentage of subjects with any TEAEs, serious TEAEs, related TEAEs, discontinuation due to TEAEs, severe TEAEs, AEs leading to death and TEAEs leading to death; event counts will also be included. The summary of TEAEs during the study will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy.

In addition, the following summaries will be presented by treatment group, SOC, high level term (HLT) and PT, based on the SS:

- Incidence of TEAEs[a] [b]
- Incidence of TEAEs, stratified by overall ADA status (positive/negative) as defined in Section 10
- Incidence of serious TEAEs[a] [b]
- Incidence of non-serious TEAEs[b]
- Incidence of TEAEs by relationship
- Incidence of TEAEs by maximum relationship
- Incidence of TEAEs by maximum intensity
- Incidence of fatal TEAEs by relationship
- Incidence of non-serious TEAEs by relationship
- Incidence of serious TEAEs by relationship
- Incidence of non-serious TEAEs above threshold of 5% of subjects
- Incidence of non-serious TEAEs above threshold of 5% of subjects by relationship
- Incidence of TEAEs by SOC and PT (including the number and percentage of subjects and individual subject numbers for each PT stratified by intensity, relationship and seriousness)

[a] These summaries will be presented separately for TEAEs during the Treatment Period and TEAEs during the study.

[b] These summaries will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy and will be based on TEAEs during the study

In addition, separate summaries by treatment group, SOC, HLT and PT will be included for the following AEs for special monitoring (AESM):

- Major cardiovascular events
- Serious infections, including opportunistic infections and tuberculosis (TB)
- Malignancies including lymphoma
- Cytopenias
- Neuropsychiatric events
- Inflammatory bowel disease
- Anaphylactic reaction (hypersensitivity and anaphylactic reactions)
- Hepatic events and drug-induced liver injury

The criteria for identifying and additional criteria for reporting AESM are provided in Section 14.5.

All summary tables (including those for AESM) will contain the number and percentage of subjects and the number of events where applicable. A subject who has multiple events in the same SOC, HLT and PT will be counted only once in the subjects counts but all events will be included.

In summaries including relationship to IMP, the following relationship categories will be included:

- Related
- Not related

Subjects who experience the same event multiple times will be included in the most related category for tabulations by maximum relationship. Events with missing relationship will be considered as 'Related' for summary purposes but shown as missing in the data listings.

In summaries including intensity, the following categories will be summarized:

- Mild
- Moderate
- Severe

Subjects who experience the same event multiple times will be included in the most severe category for tabulations by maximum intensity. Events with missing intensity will be considered as 'Severe' for summary purposes. All data will be presented as recorded in the database for the listings.

Adverse event summaries will be ordered by alphabetical SOC, alphabetical HLT within SOC and decreasing frequency of PT in the 'BKZ' column for tables including event counts. For tables including only number and percentage of subjects, summaries will be ordered by alphabetical SOC, alphabetical HLT within SOC and decreasing incidence of PT within HLT and SOC in the 'BKZ' column.

Listings will be presented by treatment group and subject for all AEs, serious adverse events (SAEs), AEs leading to death and AEs leading to withdrawal. Listings will include the onset date/time and outcome date/time of the event (including relative days), the AE duration (derived), days since first dose of IMP, days since most recent dose of IMP, pattern of event, intensity, relationship, action taken and outcome. In addition, the listing will flag AEs that led to discontinuation, TEAEs, AESM and SAEs. All AE listings will be based on all subjects screened.

A glossary of all TEAE terms will be provided including the SOC, HLT, PT and reported term.

11.3 Clinical laboratory evaluations

Laboratory data (chemistry, hematology, and urinalysis) and changes from Baseline (if applicable) for numeric variables will be listed by treatment group and visit. Any laboratory variables that are BLQ or ALQ will be handled as described in Section 4.2.3. Values outside the reference range for the numeric variables will be flagged in the listings and, in addition, will be listed separately. The reference ranges will also be reported in the listings.

Chemistry and hematology variables presented in Table 11–1 will be summarized by treatment group at each visit, for both observed values and changes from Baseline. Figures of mean and mean change from Baseline will be presented by treatment for the absolute neutrophil count (ANC). All treatment groups will be overlaid on the same plot and the plot will include the error bars based on the SD (ie, mean +/-SD).

Laboratory variables will be grouped according to the laboratory function panel and categorized as normal, high or low, if applicable, based on the reference range supplied by the analytical laboratory. For selected variables that are identified in Table 11–1, the change in category from Baseline will be presented in shift tables at all post-Baseline visits.

The number and percentage of subjects with treatment-emergent markedly abnormal (TEMA) values will be tabulated by treatment group and visit. Markedly abnormal laboratory values will generally be defined as those categorized as Grade 3 or Grade 4 as per the Rheumatology Common Toxicity Criteria (RCTC). Communication with the lead author of the RCTC publication has revealed that the reference ranges for absolute lymphocyte counts (ALC) are in error, therefore ALC criteria will follow the ranges defined as per the Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03 (United States Department of Health and Human Services, 2010). Additional laboratory variables (not present in the RCTC publication) will also be categorized as per the CTCAE Version 4.03. The criteria for identifying TEMA values are presented in Table 14–5 and Table 14–6 for hematology and chemistry values respectively.

Laboratory measurements meeting the criteria for TEMA will be listed separately for hematology and chemistry. The listings will include all measurements for any variable with at least one TEMA result for each subject.

Any additional laboratory variables not included in the outputs described previously will be listed separately. These will include the following:

- Serology
- Urine drug screen
- Pregnancy tests (serum and urine)
- Follicle stimulating hormone (FSH) (only for women whose last menstrual cycle occurred between 12 and 24 months prior to the Screening Visit)
- The results of the Quantiferon tuberculosis (TB) Gold test will be listed separately.

Subjects with treatment-emergent liver function test abnormalities at any post-Baseline visit will be summarized (number and percentage of subjects) by treatment group. The criteria for this tabulation are presented in Table 14–7.

Table 11–1: Clinical laboratory measurements

Category	Panel	Variable
Serology	Serology	human immunodeficiency virus (HIV), Hepatitis B and C, human leukocyte antigen (HLA)-B27 (HLA-B27)
Hematology	Red blood cell	red blood cell count[a], hemoglobin[a], hematocrit
	Red blood cell indices	mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, mean corpuscular volume
	Platelet	Platelet
	White blood cell	white blood cell count[a]
	White blood cell differential	Absolute counts: absolute neutrophils[a], basophils, eosinophils, absolute lymphocytes, monocytes; atypical lymphocytes Percentages: neutrophils/leukocytes[a], basophils/leukocytes[a], eosinophils/leukocytes[a], lymphocytes/leukocytes[a], monocytes/leukocytes[a].
Chemistry	Electrolytes	Sodium, chloride, potassium, magnesium, total calcium, bicarbonate

Table 11–1: Clinical laboratory measurements

Category	Panel	Variable
	Kidney function	BUN, creatinine, uric acid, glomerular filtration rate
	Proteins	hs-CRP[bc]
	Liver function	AST[a], ALT[a], GGT[a], ALP[a], LDH[a], total bilirubin, direct bilirubin (if indicated), indirect bilirubin (if indicated)
	Lipids	Total cholesterol[a]
	Hormones	FSH (postmenopausal females only)
	Fertility	Pregnancy test[d]
	Substrates	Glucose
Urinalysis	Dipstick	pH, albumin, glucose, blood (RBC, WBC)
	Urine sediment	RBC, WBC, crystals, bacteria

ALP=alkaline phosphatase; ALT=alanine aminotransferase; AST=aspartate aminotransferase; BUN=blood urea nitrogen; GGT=gamma glutamyl transferase; FSH=follicle stimulating hormone; HIV=human immunodeficiency virus; HLA-B27= human leukocyte antigen B27; hs-CRP=high sensitivity C-reactive protein; LDH=lactate dehydrogenase; RBC=red blood cell; WBC=white blood cell

[a] Shift tables will be presented for these variables.

[b] hs‑CRP will be included in the listings and tabulations for the efficacy data and will not be included in the safety data outputs.

[c] hs-CRP will be tested at specified visits per the study schedule of events.

[d] Serum pregnancy test to be performed at Screening; urine pregnancy test to be performed at subsequent.

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

11.4.1 Vital signs

The following vital signs measurements will be obtained:

- Pulse rate
- Systolic and diastolic blood pressure
- Temperature

A by-subject listing of all vital sign measurements and changes from Baseline will be presented by treatment group and visit. The listing will include a flag for measurements identified as TEMA/potentially clinically significant (PCS) as calculated by the criteria outlined in Table 14–8. Body weight will be included in the listing.

Descriptive statistics will be reported for all vital sign measurements. Observed values and changes from Baseline will be summarized by treatment group for each vital signs variable and visit.

Figures of mean change from Baseline over time will be presented for each variable by treatment group.

The number and percentage of subjects with TEMA/PCS vital sign values will be summarized by treatment group at each visit. The denominator for the percentages will be the number of subjects with a non-missing measurement for the variable at the specific visit.

11.4.2 Electrocardiograms

The following ECG variables will be reported together with the Investigator's interpretation of the ECG profile:

- PR interval
- RR interval
- QRS interval
- QT interval
- Heart rate
- QT corrected for heart rate using Fridericia's formula (QTcF)

Electrocardiogram results will be reported in the by-subject listings. The listing will also include the changes from Baseline and percentage changes from Baseline and will be presented by treatment group and visit.

Observed results, changes and percentage changes from Baseline will be summarized for each ECG variable by treatment group and visit.

The following cut points in QTcF (observed data and change from Baseline) will be summarized categorically by treatment group (number and percentage of subjects) and visit. The denominator for the percentages will be the number of subjects with a non-missing measurement for the variable at the specific visit.

Observed QTcF data:

- <450msec
- ≥450msec to <480msec
- ≥480msec to <500msec
- ≥500msec

Change from Baseline in QTcF:

- <30msec
- ≥30msec to <60msec
- ≥60msec

Electrocardiogram findings including the Investigator's interpretation will be listed separately.

11.4.3 Other safety variables

11.4.3.1 C-SSRS

The C-SSRS is divided into 2 time intervals and further sub divided as shown:

- C-SSRS at Baseline – assessed at the Screening visit
  - Suicidal Ideation – over Lifetime and over Past 6 Months
  - Intensity of Ideation – over Lifetime and over Past 6 Months
  - Suicidal Behavior – over Lifetime and over Past 2 Years
- C-SSRS Since Last Visit – assessed at all visits after Screening
  - Suicidal Ideation – Since Last Visit
  - Intensity of Ideation – Since Last Visit
  - Suicidal Behavior – Since Last Visit

For Suicidal Ideation questions, only those questions where "Yes" has been answered will be listed. For Intensity of Ideation questions and Suicidal Behavior questions, any question for which an answer has been recorded will be listed.

Results, as specified above, will be listed by treatment group, subject and visit.

11.4.3.2 Physical examination

Subjects with abnormalities in the physical examination will be listed including details of the abnormality.

11.4.3.3 Chest X-ray/CT scan and TB questionnaire

The results of the chest X-ray and the TB questionnaire will be listed for each subject.

12 OTHER ANALYSES

Not applicable.

13 REFERENCES

Common Terminology Criteria for Adverse Events (CTCAE); Version 4.03 June 2010.U.S. Department of Health and Human Services

Guidance for Industry: E9 Statistical Principles for Clinical Trials. US Department of Health and Human Services, Food and Drug Administration. ICH 1998.

Mundt JC, Greist JH, Gelenberg AJ, Katzelnick DJ, Jefferson JW, Modell JG. Feasibility and Validation of a Computer Automated Columbia Suicide severity Rating Scale Using Interactive Voice Response Technology. J Psychiatr Res. 2010;44(16):1224 8.

Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings from Three Multisite Studies with Adolescents and Adults. Am J Psychiatry. 2011;168(12):1266 77.

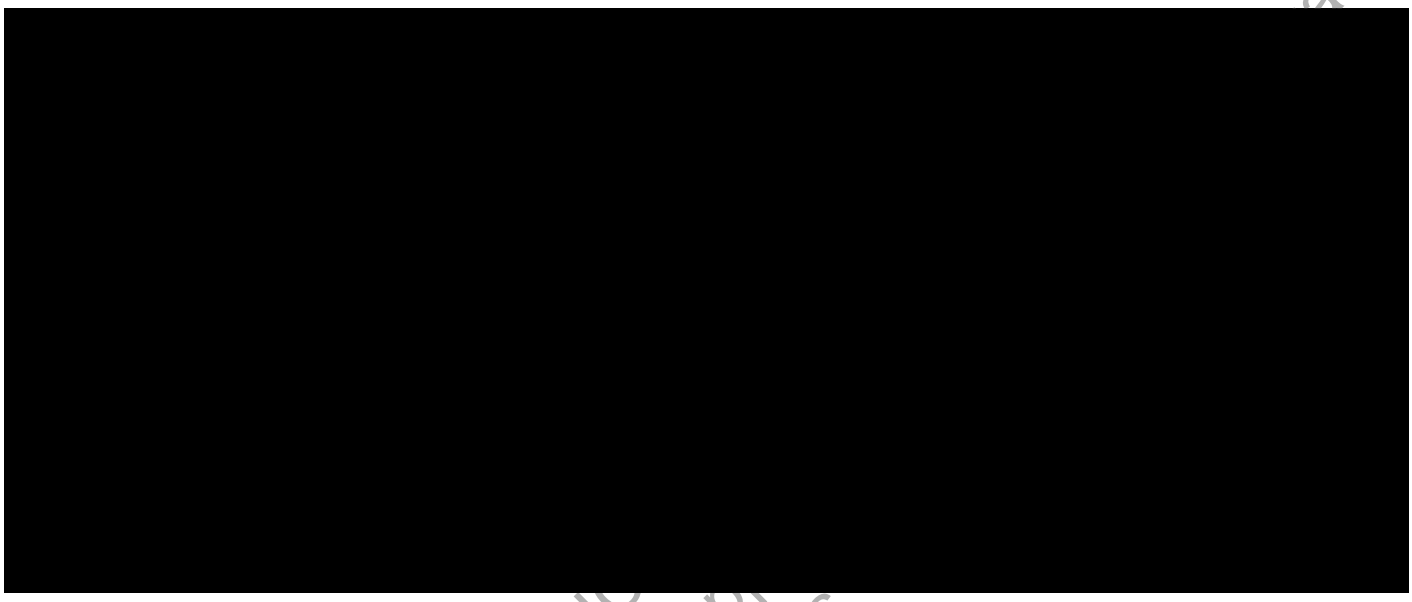
Snaith RP, Zigmond AS. The hospital anxiety and depression scale, with the irritability depression anxiety scale and the Leeds situational anxiety scale manual. 1994.

Van der Heijde D, Sharp J, Wassenberg S, Gladman DD. Psoriatic arthritis imaging: a review of scoring methods. Ann Rheum Dis. 2005;64;ii61-ii64.

14 APPENDICES

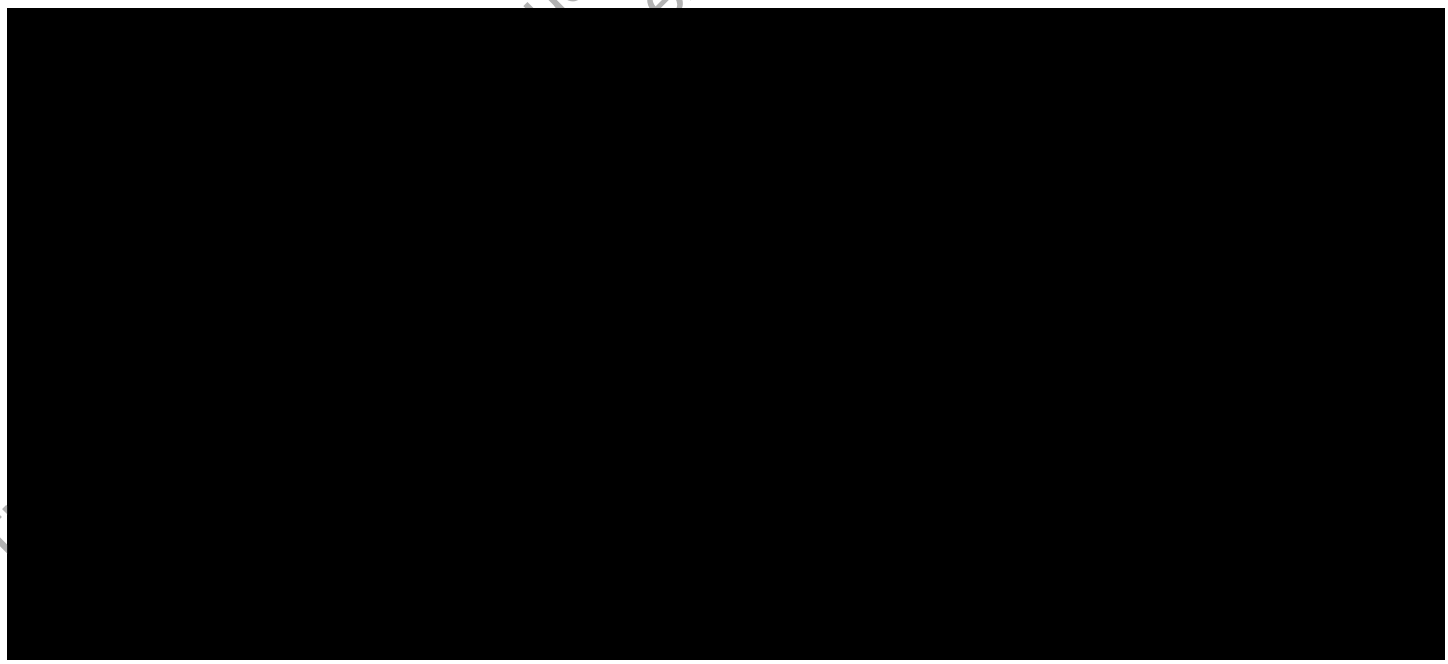
14.1 Bath Ankylosing Spondylitis Disease Activity Index

The BASDAI is shown in Table 14–1. All questions are answered on a scale from 0 to 10; for questions 1 to 5, 0 = none and 10 = very severe; for question 6, 0 = 0 hours, 10 = 2 or more hours.



14.2 Bath Ankylosing Spondylitis Functional Index

The BASFI is shown in Table 14–2. All items are scored on a scale from 0 ("Easy") to 10 ("Impossible"). The first 8 items evaluate activities related to functional anatomical limitations due to ankylosing spondylitis and the final 2 items evaluate a subject's ability to cope with everyday life.



[redacted]

14.3 Hospital Anxiety and Depression Scale

The individual items on the HADS-A and HADS-D scales are listed in Table 14–3.

[redacted]

14.4 Treatment group assignment for the TFLs

Data will be summarized in the TFLs according to the treatment groupings as displayed in Table 14–4.

Table 14–4: Treatment group assignment for TFLs

TFL group	CZP	BKZ	All Subjects
Subject disposition	X	X	X
Protocol deviations	X	X	X
Demographics	X	X	X
AS history and Baseline Disease characteristics	X	X	X
Medical history	X	X	X

Table 14–4: Treatment group assignment for TFLs

TFL group	CZP	BKZ	All Subjects
Prior/concomitant medications	X	X	X
AEs	X	X	
Safety data (including vital signs, ECG, laboratory tests, physical examination, body weight, C-SSRS)	X	X	
Pharmacodynamic data	X	X	
Efficacy data	X	X	
Pharmacokinetic data	X	X	
ADA	X	X	

ADA= anti-drug antibodies; AE=adverse event; C-SSRS=Columbia-Suicide Severity Rating Scale; ECG=electrocardiogram

14.5 Adverse events for special monitoring

The following AESM are defined for BKZ and will be summarized separately as described in Section 11.2:

- Major cardiovascular events
- Serious infections, including opportunistic infections and TB
- Malignancies including lymphoma
- Cytopenias
- Neuropsychiatric events
- Inflammatory bowel disease
- Anaphylactic reaction (hypersensitivity and anaphylactic reactions)
- Hepatic events and drug-induced liver injury

14.5.1 Major cardiovascular events

Major adverse cardiovascular events (MACE) will be tabulated separately and will be identified using the following UCB-defined search criteria:

- All serious TEAEs which code to a PT included in the Scope=Broad and/or Scope=Narrow groups of the following standardized MedDRA queries (SMQ):
  - Haemorrhagic central nervous system vascular disorders
  - Ischaemic central nervous system vascular disorders

- All serious TEAEs which code to a PT included in the HLT = 'Ischaemic coronary artery disorders' except events coding to PT = 'Chest pain' or PT = 'Chest discomfort'
- All serious TEAEs which code to a PT = 'Cardiac failure congestive'

14.5.2 Serious infections, including opportunistic infections and tuberculosis

14.5.2.1 Serious infections

Serious infections are based on MedDRA classification using the SOC = 'Infections and Infestations'. Such events will be included in the tabulations of SAEs and no separate summary tabulations will be presented.

14.5.2.2 Fungal infections

Fungal infections will be summarized separately based on all TEAEs coding to the high level group term (HLGT) = 'Fungal infectious disorders'.

14.5.2.3 Opportunistic infections

Opportunistic infections (including TB) will be summarized in a separate table including all TEAEs identified using search criteria defined by UCB.

Opportunistic infections are identified in 2 steps:

- Step 1: Refer to column B of the spreadsheet which identifies the PTs to be classified as opportunistic infections using either a single 'x' or a double 'xx'
  - All TEAEs which code to a PT flagged with a single 'x' need to also be serious in order 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.

All serious TEAEs in the study database which code to a PT flagged with a single 'x' and all TEAEs in the study database which code to a PT flagged with a double 'xx' will be summarized as an opportunistic infection in the stand-alone table.

- 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 in order to determine whether it is a true opportunistic infection or not. The process for physician review is as follows:
  - Study programming team creates a spreadsheet which lists all of the subjects with a TEAE present in the database which codes to a PT identified as needing case-by-case review. Information from the AE dataset to be included in the spreadsheet: Subject ID, AE verbatim term, SOC, HLT, Low Level Term (LLT), PT, AE start date, AE end date, seriousness, severity, relationship to IMP, action taken. Additionally, a column will be included where the study physician can document their decision on the case
  - Study physician reviews the cases in the spreadsheet and indicates in the additional column which AEs are confirmed to be opportunistic infections via a single 'x'

– Study programming team incorporates these decisions into the ADAE dataset by merging the study physician decisions for individual subjects / PTs and flagging the confirmed opportunistic infections as such in the dataset

All subjects with an AE reported that has been confirmed by the study physician to be an opportunistic infection (based on case-by-case review) will be summarized as such in the stand-alone table, together with all of the events identified in Step 1 of this process.

Following the initial physician review of case-by-case events, subsequent reviews will be based on the cumulative set of case-by-case events present in the database at each time point of spreadsheet creation. Physician decisions from previous runs should be retained in each subsequent run. The final run of the spreadsheet, with all study physician decisions on the full set of case-by-case events, will be archived at the end of the study.

14.5.3 Malignancies including lymphoma

Malignancies will be presented in 2 separate tables based on the following SMQs:

- Malignant or unspecified tumors
- Malignant tumors

Events included in the 'Malignant tumors' tabulation will be a subset of the events in the 'Malignant or unspecified tumors' tabulation. The SMQ search should include all TEAEs which code to a PT included in the Scope=Narrow group within each SMQ.

The output tables will include 2 different overall incidence rows:

- The first overall incidence row will summarize 'Any malignancies (including unspecified)' or 'Any malignancies' (depending on the table) and this row will summarize the incidence of all AEs flagged for inclusion in the table, regardless of the HLT it codes to
- The second overall incidence row will summarize 'Any malignancy (including unspecified, excluding non-melanomic skin cancers)' or 'Any malignancy (excluding non-melanomic skin cancers)' (depending on the table) and this row will summarize the incidence of AEs flagged for inclusion in the table, excluding those which code to an HLT of 'skin neoplasms malignant and unspecified (excl melanoma)'

14.5.4 Cytopenias

Cytopenias will be tabulated separately based on the SMQ = 'Haematopoietic cytopenias'. Only serious TEAEs will be included in the tabulation. The SMQ search should include all serious TEAEs which code to a PT included in the Scope=Broad and/or Scope=Narrow groups within the SMQ.

14.5.5 Neuropsychiatric events

Neuropsychiatric events (in particular depression, anxiety and suicide ideation or behavior) will be tabulated separately based on the SMQ = 'Depression and suicide/self-injury'. The SMQ search should include all TEAEs which code to a PT included in the Scope=Broad and/or Scope=Narrow groups within the SMQ.

14.5.6 Inflammatory bowel disease

Inflammatory bowel disease events will be tabulated separately, based on the HLGT = 'Colitis excl infective'.

14.5.7 Anaphylactic reaction

Anaphylactic reactions will be summarized together in a stand-alone table with the following incidence rows:

- The first row within the body of the table will be labeled 'Any hypersensitivity/anaphylactic reaction' and will represent the overall incidence of subjects who reported at least one hypersensitivity reaction or at least one anaphylactic reaction
- The second row within the body of the table will be labeled 'Any hypersensitivity reaction' and will represent the overall incidence of subjects who reported at least one hypersensitivity reaction
- The third row within the body of the table will be labeled 'Any anaphylactic reaction' and will represent the overall incidence of subjects who reported at least one anaphylactic reaction

Following these three overall incidence rows, all TEAEs that have been identified as either a hypersensitivity reaction or an anaphylactic reaction will be summarized by SOC, HLT and PT (these will not be presented by subcategory).

Hypersensitivity reactions and anaphylactic reactions will be identified as follows:

- Hypersensitivity reactions: all TEAEs with onset (start date/time) within 24 hours after any administration of IMP, which code to a PT which contains the term 'hypersensitivity' will be considered as a hypersensitivity reaction and included in the summary table.
- Anaphylactic reactions: An algorithmic approach will be used to identify TEAEs that are considered to be anaphylactic reactions. Preferred terms are separated into 4 distinct categories (A, B, C, D) prior to the algorithmic approach being applied.

All TEAEs with onset (start date/time) within 24 hours after any administration of IMP, and which fulfill any of the 3 criteria described in Section 14.5.7.1 will be included in the summary table.

Any TEAEs with missing start time will be assumed to have occurred at the time of or after dosing for this purpose, if the event date is on the same day as a dosing date.

14.5.7.1 Anaphylactic reaction algorithm

The SMQ = 'anaphylactic reaction' consists of 3 parts:

- A **narrow search** containing PTs that represent core anaphylactic reaction terms. If a subject reports any TEAE which codes to a PT included in Category A, then the event will be flagged as an anaphylactic reaction and summarized as such in the table
  - **Category A**

1. Anaphylactic reaction

2. Anaphylactic shock
3. Anaphylactic transfusion reaction
4. Anaphylactoid reaction
5. Anaphylactoid shock
6. Circulatory collapse
7. Dialysis membrane reaction
8. Kounis syndrome
9. Shock
10. Shock symptom
11. Type I hypersensitivity

- A **broad search**: If a subject reports any TEAE which codes to a PT included in Category B AND reports any TEAE which codes to a PT included in Category C, and both TEAEs have the same start date, then both events will be flagged as anaphylactic reactions and summarized as such in the table

  – **Category B**

1. Acute respiratory failure
2. Asthma
3. Bronchial oedema
4. Bronchospasm
5. Cardio-respiratory distress
6. Chest discomfort
7. Choking
8. Choking sensation
9. Circumoral oedema
10. Cough
11. Cyanosis
12. Dyspnoea
13. Hyperventilation
14. Irregular breathing
15. Laryngeal dyspnea
16. Laryngeal oedema
17. Laryngospasm
18. Laryngotracheal oedema

19. Mouth swelling
20. Nasal obstruction
21. Oedema mouth
22. Oropharyngeal spasm
23. Oropharyngeal swelling
24. Respiratory arrest
25. Respiratory distress
26. Respiratory dyskinesia
27. Respiratory failure
28. Reversible airways obstruction
29. Sensation of foreign body
30. Sneezing
31. Stridor
32. Swollen tongue
33. Tachypnoea
34. Throat tightness
35. Tongue oedema
36. Tracheal obstruction
37. Tracheal oedema
38. Upper airway obstruction
39. Wheezing

- **Category C**

1. Allergic oedema
2. Angioedema
3. Erythema
4. Eye oedema
5. Eye pruritis
6. Eye swelling
7. Eyelid oedema
8. Face oedema
9. Flushing
10. Generalized erythema

11. Injection site urticarial
12. Lip oedema
13. Lip swelling
14. Nodular rash
15. Ocular hyperaemia
16. Oedema
17. Periorbital oedema
18. Pruritis
19. Pruritis allergic
20. Pruritis generalized
21. Rash
22. Rash erythematous
23. Rash generalized
24. Rash pruritic
25. Skin swelling
26. Swelling
27. Swelling face
28. Urticaria
29. Urticaria popular

– **Category D**

1. Blood pressure decreased
2. Blood pressure diastolic decreased
3. Blood pressure systolic decreased
4. Cardiac arrest
5. Cardio-respiratory arrest
6. Cardiovascular insufficiency
7. Diastolic hypertension
8. Hypotension

- An **algorithmic approach**: If a subject reports any TEAE which codes to a PT included in Category D AND reports (either a TEAE which codes to a PT included in Category B OR a TEAE which codes to a PT included in Category C), and both TEAEs have the same start date, then both events will be flagged as anaphylactic reactions and summarized as such in the table.

14.5.8 Hepatic events and drug-induced liver injury

Hepatic events will include:

- Events based on the SMQ = 'Drug related hepatic disorders - comprehensive search' (excluding sub-SMQs = 'Liver neoplasms, benign [incl cysts and polyps]' and 'Liver neoplasms, malignant and unspecified'). All AEs should be included in the tabulation (including those considered both related and not related to the IMP) which code to a PT included in the Scope=Narrow group within each SMQ
- Hy's Law cases will also be summarized separately in a table of liver function abnormalities as described in Section 11.3 (with adjudication for potential drug-induced liver injury [PDILI] cases)

14.6 Treatment-emergent markedly abnormal laboratory values

The criteria for identifying TEMA laboratory values are provided for hematology in Table 14–5 and for chemistry in Table 14–6.

Table 14–5: Definition of TEMA values for hematology

Variable (SI units)	Source	Markedly abnormal definition			
		Low Grade 4	Low Grade 3	High Grade 3	High Grade 4
Hemoglobin (g/L)	RCTC	<70 OR decrease from Baseline of ≥30	70 to <80 OR value <LLN and decrease from Baseline of 21 to 29	NA	NA
ALC (10^9/L)	CTCAE	<0.2	0.2 to <0.5	NA	NA
ANC (10^9/L)	RCTC	<0.5	0.5 to <1.0	NA	NA
WBC (10^9/L)	RCTC	<1.0	1.0 to <2.0	NA	NA
Platelets (10^9/L)	RCTC	<20	20 to <50	NA	NA

ALC=absolute lymphocyte count; ANC=absolute neutrophil count; CTCAE=Common Terminology Criteria for Adverse Events (CTCAE 2010); LLN=lower limit of normal; NA=not applicable; RCTC=Rheumatology Common Toxicity Criteria; WBC=white blood cell.

Table 14–6: Definition of TEMA values for chemistry

Variable (SI units)	Source	Markedly abnormal definition			
		Low Grade 4	Low Grade 3	High Grade 3	High Grade 4
ALP (U/L)	RCTC	NA	NA	>3.0 to 5.0 x ULN	>5.0 x ULN
ALT (U/L)	RCTC	NA	NA	>3.0 to 8.0 x ULN	>8.0 x ULN
AST (U/L)	RCTC	NA	NA	>3.0 to 8.0 x ULN	>8.0 x ULN

Table 14–6: Definition of TEMA values for chemistry

Variable (SI units)	Source	Markedly abnormal definition			
		Low Grade 4	Low Grade 3	High Grade 3	High Grade 4
Creatinine (µmol/L)	RCTC	NA	NA	>1.8 to 3.0 x ULN	>3.0 x ULN
GGT (U/L)	CTCAE	NA	NA	>5.0 to 20.0 x ULN	>20.0 x ULN
Potassium (mmol/L)	RCTC	<2.5	2.5 to <3.0	>6.4 to 7.0	>7.0
Sodium (mmol/L)[a]	RCTC	<120	120 to <125	>155 to 160	>160
Total bilirubin (µmol/L)	RCTC	NA	NA	≥2.0 to 3.0 x ULN	>3.0 x ULN
Calcium (mmol/L)	RCTC	<1.63	1.63 to <1.75	>3.125 to 3.375	>3.375
Cholesterol (total) (mmol/L)	CTCAE	NA	NA	>10.34 to 12.92	>12.92

ALP=alkaline phosphatase; ALT=alanine aminotransferase; AST=aspartate aminotransferase; CTCAE=Common Terminology Criteria for Adverse Events (CTCAE 2010); GGT=gamma glutamyl transferase; NA=not applicable; RCTC=Rheumatology Common Toxicity Criteria; ULN=upper limit of normal.

[a] For sodium High Grade 3 and 4 are taken from CTCAE.

14.7 Treatment-emergent abnormal liver values

The criteria for identifying treatment-emergent liver function abnormalities are presented in Table 14–7. Subjects will be counted in all applicable categories for the tabulations ie, a subject with ≥5 x upper limit of normal (ULN) in AST will also be counted in the ≥3 x ULN in AST category, the ≥3 x ULN in aspartate aminotransferase (AST) or alanine aminotransferase (ALT) category and the ≥5x ULN in AST or ALT category.

Table 14–7: Definition of treatment-emergent liver function values

Criterion
≥3 x ULN increase for AST
≥5 x ULN increase for AST
≥10 x ULN increase for AST
≥20 x ULN increase for AST
≥3 x ULN increase for ALT
≥5 x ULN increase for ALT
≥10 x ULN increase for ALT

Table 14–7: Definition of treatment-emergent liver function values

Criterion
≥20 x ULN increase for ALT
≥3 x ULN increase for AST or ALT
≥5 x ULN increase for AST or ALT
≥10 x ULN increase for AST or ALT
≥20 x ULN increase for AST or ALT
≥1 x ULN increase for bilirubin
≥1.5 x ULN increase for bilirubin
≥1.5 x ULN increase for ALP
≥1 x ULN increase for bilirubin and 3 x ULN increase of either AST or ALT (at the same visit)
≥2 x ULN increase for bilirubin and 3 x ULN increase of either AST or ALT (at the same visit)

ALP=alkaline phosphatase; ALT=alanine aminotransferase; AST=aspartate aminotransferase; ULN=upper limit of normal.

14.8 Treatment-emergent markedly abnormal vital signs values

The criteria for identifying TEMA/PCS vital signs values are provided in Table 14–8.

Table 14–8: TEMA/PCS criteria for vital signs

Variable	Unit	Low[a]	High[a]
Systolic blood pressure	mmHg	Value ≤90 and ≥30 decrease from Baseline	Value ≥180 and ≥40 increase from Baseline
Diastolic blood pressure	mmHg	Value ≤50 and ≥20 decrease from Baseline	Value ≥105 and ≥30 increase from Baseline
Pulse rate	bpm	Value ≤50 and ≥30 decrease from Baseline	Value ≥120 and ≥30 increase from Baseline

bpm=beats per minute; mmHg=millimeters of mercury; PCS=potentially clinically significant; TEMA=treatment-emergent markedly abnormal.
Note: the change in measurement (increase or decrease) will be calculated relative to the value obtained at Baseline.
[a] Both conditions must be satisfied for a measurement to be considered PCS.

15 AMENDMENTS TO THE SAP

15.1 Amendment 2

15.1.1 Rationale for the amendment

Various updates were suggested with the introduction of the PET-MRI/PET-CT substudy, and to align with the changes made to protocol amendment 3.

15.1.2 List of changes

Change #1

List of Abbreviations

The following abbreviations were added to the list:

- CSR = clinical study report
- mNY = modified New York
- SIJ = sacroiliac joint
- SUV_{auc} = standard uptake value corrected for individual integrated whole blood activity concentration
- VOI = volume of interest
- VUMC = VU University Medical Center

Change #2

Secondary objectives (Section 2.1.2)

The following bullets were removed:-

- To evaluate the effect of BKZ or CZP on changes in osteoblastic activity
- To assess the pharmacokinetics (PK) and immunogenicity of BKZ

Change #3

Other objectives (Section 2.1.3)

The following bullets were added:

- To evaluate the effect of BKZ or CZP on changes in bone formation
- To assess the pharmacokinetics (PK) and immunogenicity of BKZ

Change #4

Other Efficacy Variables (Section 2.2.1.3)

The following bullets were added to the end:

- Change from Baseline in Patient's Global Assessment of Disease Activity (PGADA)
- Changes in bone formation as measured by standardized uptake value corrected for individual integrated whole blood activity concentration (SUV_{auc}) for each PET-positive lesion identified together with the maximum SUV_{auc} value across all lesions within the spine

and sacroiliac joint (SIJ). SUV_{auc} will be derived from positron-emission tomography magnetic resonance imaging (PET MRI) or positron-emission tomography computed tomography (PET CT) scans at Baseline, Week 12, and Week 48

Change #5

Study design and conduct (Section 2.3)

The second paragraph was changed to:

It is planned that at least 60 subjects will be randomized to 1 of 2 treatment arms in a 2:1 ratio and will receive either BKZ or CZP up to Week 44 (final dose of investigational medicinal product [IMP]).

Change #6

Determination of sample size (Section 2.4)

This section was rewritten to:

A sufficient number of subjects will be enrolled to ensure that at least 60 subjects are available in the main study at Week 12 to compare the change from Baseline in ASDAS between BKZ and CZP. This will also provide at least 25 subjects enrolled in the PET-CT or PET-MRI substudy from multiple sites. Subjects will be randomized in a 2:1 ratio to receive BKZ or CZP respectively.

The sample size was calculated to provide at least 80% power to detect a difference between treatment groups in the mean change from Baseline in ASDAS at Week 12 of 0.89 (group mean change from Baseline for CZP and BKZ being -1.78 and -2.67 respectively) with a common standard deviation (SD) of 1.17 using frequentist methods for the comparison of the treatment group mean change from Baseline in ASDAS at Week 12 with Baseline ASDAS as a covariate (Overall and Starbuck, 1979).

The correlation between Baseline and Week 12 raw ASDAS was assumed to be 0.35. Note that the informative prior for the model intercept to be used in the Bayesian modelling of the primary efficacy variable, given by β_{CZP} ~ Normal (-1.78, var=0.0605), contributes an effective sample size of approximately 20 CZP subjects. Prior data conflict will be investigated and a vague prior will be used for the model intercept coefficient if prior data conflict is evident.

Note that a clarification regarding the sample size for this study was provided in protocol amendment 3. As indicated above, the planned sample size of at least 60 subjects provides sufficient power to detect meaningful treatment differences in the primary efficacy variable whilst also providing at least 25 subjects for the PET-MRI or PET-CT substudy. However, due to complexities in the initiation of sites capable of performing PET-MRI or PET-CT scans, more than the planned 35 non-PET-MRI/CT substudy subjects were screened and randomized before it was possible to close enrolment at the sites not participating in the substudy. Owing to the novelty of the PET-MRI/CT outcomes within this disease area, it is considered important to ensure at least 25 evaluable subjects are included in the assessment of this exploratory objective to maximize the possibility of identifying the salient factors or variables that might differ between the treatment groups and, therefore, the total sample size is increased to at least 72 subjects. This sample size increase will not negatively impact the power to detect a difference in the primary efficacy variable at Week 12 (ie, it will remain >80%), but will ensure that the target

number of subjects are included in the exploratory analysis of changes in bone formation at Week 12 and Week 48. Since the analyses of PET-MRI/CT data are exploratory and treatment effect sizes are unknown, formal power calculations were not performed. Clinical expertise does, however, indicate that a sample size of at least 25 subjects in the substudy will characterize broad differences between treatment groups.

Change #7

Determination of sample size (Section 2.4)

The following paragraph was added at the end:

Note that a clarification regarding the sample size for this study was provided in protocol amendment 3. As indicated above, the planned sample size of at least 60 subjects provides sufficient power to detect meaningful treatment differences in the primary efficacy variable whilst also providing at least 25 subjects for the PET-MRI or PET-CT substudy. However, due to complexities in the initiation of sites capable of performing PET-MRI or PET-CT scans, more than the planned 35 non-PET-MRI/CT substudy subjects were screened and randomized before it was possible to close enrolment at the sites not participating in the substudy. Owing to the novelty of the PET-MRI/CT outcomes within this disease area, it is considered important to ensure at least 25 evaluable subjects are included in the assessment of this exploratory objective to maximize the possibility of identifying the salient factors or variables that might differ between the treatment groups and, therefore, the total sample size is increased to at least 72 subjects. This sample size increase will not negatively impact the power to detect a difference in the primary efficacy variable at Week 12 (ie, it will remain ≥80%), but will ensure that the target number of subjects are included in the exploratory analysis of changes in bone formation at Week 12 and Week 48. Since the analyses of PET-MRI/CT data are exploratory and treatment effect sizes are unknown, formal power calculations were not performed. Clinical expertise does, however, indicate that a sample size of at least 25 subjects in the substudy will characterize broad differences between treatment groups.

Change #8

General presentation of summaries and analyses (Section 3.1)

The following paragraph was extended to:

Data listings containing all documented data and all derived data will be generated. Subjects enrolled in the PET-MRI/PET-CT substudy will be flagged in the data listings.

Change #9

The following section was added:

PET Per-Protocol Set (Section 3.5.8)

The PET Per-Protocol Set (PET-PPS) will consist of all randomized subjects who receive at least 1 dose of the IMP and have evaluable PET-MRI or PET-CT scan data at Baseline and at least 1 of the post-Baseline assessments.

Change #10

Subject Disposition (Section 5.1)

The opening paragraph was changed to:

The number and percentage of subjects who were randomized into the study (and into the PET-MRI/PET-CT substudy), subjects who completed or prematurely discontinued the study, as well as the primary reason for discontinuation will be presented by treatment group and for all subjects, based on the RS. A subject who completed the study is defined as a subject who completed all visits up to the last scheduled study visit, ie, Visit 18 (Week 64). An additional summary of subject disposition in the PET-MRI/PET-CT substudy will also be presented separately by treatment group and for all subjects, based on the RS.

Change #11

Subject Disposition (Section 5.1)

The second bullet of the seventh paragraph was changed to:

- for subjects participating in the respective pharmacogenomics, pharmacogenetic and PET MRI/CT substudies: date of respective informed consent

Change #12

Classification of age categories (Section 6.1.2)

The opening paragraph was changed to:

Age will be classified into categories based on requirements for European Union Drug Regulating Authorities Clinical Trials (EudraCT) and clinicaltrials.gov reporting.

Change #13

Demographics (Section 6.2)

The second paragraph was changed to:

All Baseline demographic characteristics (with the exception of year of birth) will be summarized by treatment group and for all subjects based on the SS. The summary of age will include descriptive statistics and categorical summaries, the latter based on requirements for EudraCT and clinicaltrials.gov reporting. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/CT substudy. Childbearing potential will be listed for the SS.

Change #14

Other Baseline characteristics (Section 6.3)

The section was changed to:

The following Baseline disease characteristics will be presented in a by-subject listing for the SS:

- Duration of disease (as calculated in Section 6.1.3), including confirmation of sacroiliac joint X-ray and the date on which the sacroiliac joint X-ray was performed
- hs-CRP
- BASDAI score
- PGADA

- ASDAS
- ASDAS disease activity states (ASDAS-ID, ASDAS-Moderate Disease activity [ASDAS-MD], ASDAS-High Disease activity [ASDAS-HD] and ASDAS-Very High Disease activity [ASDAS-vHD])
- BASFI score
- Total and nocturnal spinal pain NRS
- PhGADA
- Total HADS-A and HADS-D scores

The calculation of Baseline ASDAS and the classification of the ASDAS disease activity states will be based on the Baseline BASDAI, PGADA and hs-CRP values. The Baseline for each of the variables listed above is defined in Table 3–1.

Baseline disease characteristics will be summarized using descriptive statistics by treatment group and for all subjects based on the SS. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy and will also include summary statistics for SUV_{auc} in the spine and in the SIJ at Baseline derived from the PET-MRI/PET-CT scans performed at Screening. SUV_{auc} will be summarized in each SIJ quadrant and in each anatomical location of the spine together with the maximum SUV_{auc} value in the SIJ and spine for each subject.

In addition, the above Baseline disease characteristics will be listed for the SS including SUV_{auc} and the maximum of the SUV_{auc} values in the SIJ and spine.

Subject lifestyle details (alcohol use, smoking history and caffeine use) will be listed, based on the SS.

Change #15

Medical History (Section 6.4)

The opening paragraph was changed to:

Medical history will be listed for the RS and summarized for the SS by treatment group and for all subjects, MedDRA® system organ class (SOC) and preferred term (PT). The reported term will be included in the listing. The summary will include the number and percentage of subjects and will be sorted alphabetically by SOC and by descending incidence of PT within each SOC, based on the 'All Subjects' column. This summary will be repeated for subjects in the RS who were randomized in the PET-MRI/CT substudy. A glossary of all medical history conditions will be presented for the RS including the SOC, PT and reported term.

Change #16

Ankylosing spondylitis history (Section 6.5)

This section was shortened to:

Date of initial diagnosis of AS and the date that AS symptoms first started will be listed.

Change #17

Prior and concomitant medications (Section 6.6)

The opening paragraph was changed to:

Prior and concomitant medications (Section 6.6.1 and Section 6.6.2) will be listed for the RS by treatment group and subject, and summarized for the SS by WHODD Anatomical Main Group (Level 1), Pharmacological Subgroup (Level 3) and PT. The reported term will be included in the listing, together with flags to identify AS, rescue and prohibited medications, respectively. The flags for prohibited, AS and rescue medications will be confirmed and documented at the final DEM.

Change #18

Prior and concomitant medications (Section 6.6)

The second bullet of the fourth paragraph was changed to:

- Rescue medication during study

Change #19

Prior and concomitant medications (Section 6.6)

The following paragraph was added:

These summaries will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy.

Change #20

This section was removed:

Positron-emission tomography-magnetic resonance imaging/positron-emission tomography computed tomography

Change #21

Other efficacy variables (Section 8.3)

This sentence at the start was removed:

Results of the other efficacy variables will be listed and summarized by treatment group and visit for the FAS.

Change #22

The phrase "for the FAS" was added for various requested outputs in the following sections:

- Presentation of Assessment in SpondyloArthritis International Society Response (Section 8.3.2)
- Presentation of components of Assessment in SpondyloArthritis International Society (Section 8.3.3)
- Physician's Global Assessment of Disease Activity (Section 8.3.4)
- Hospital Anxiety and Depression Scale (Section 8.3.5)

Change #23

Change in Bone Formation (PET-MRI/PET-CT Substudy) (Section 8.3.6)

The entire section was added (see above).

Change #24

Extent of Exposure (Section 11.1)

The final paragraph was extended to:

The number of injections received and duration of exposure will be listed for each subject and summarized by treatment group for the Treatment Period and Treatment Extension Period separately using descriptive statistics. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/CT substudy.

Change #25

Adverse Events (Section 11.2)

The fourth and fifth paragraphs were changed to:

An overview of the number and percentage of subjects who experience TEAEs will be presented by treatment group, based on the SS. This tabulation will include the number and percentage of subjects with any TEAEs, serious TEAEs, related TEAEs, discontinuation due to TEAEs, severe TEAEs, AEs leading to death and TEAEs leading to death; event counts will also be included. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy.

In addition, the following summaries will be presented by treatment group, SOC, high level term (HLT) and PT, based on the SS:

- Incidence of TEAEs[a]
- Incidence of TEAEs, stratified by overall ADA status (positive/negative) as defined in Section 10
- Incidence of serious TEAEs[a]
- Incidence of non-serious TEAEs
- Incidence of TEAEs by relationship
- Incidence of TEAEs by maximum relationship
- Incidence of TEAEs by maximum intensity
- Incidence of fatal TEAEs by relationship
- Incidence of non-serious TEAEs by relationship
- Incidence of serious TEAEs by relationship
- Incidence of non-serious TEAEs above threshold of 5% of subjects
- Incidence of non-serious TEAEs above threshold of 5% of subjects by relationship
- Incidence of TEAEs by SOC and PT (including the number and percentage of subjects and individual subject numbers for each PT stratified by intensity, relationship and seriousness)

[a] These summaries will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy.

15.2 Amendment 3

15.2.1 Rationale for the amendment

Updates to the description of the exploratory analysis of the PET-MRI/PET-CT imaging data were required based on a further understanding of the data expected for this study. Other clarifications have also been incorporated into this amendment.

15.2.2 List of changes

Change #1

List of Abbreviations

The SUV_{max} definition has been removed.

Change #2

Protocol deviations (Section 5.2)

The third paragraph in this section has been changed to:

The number and percentage of subjects with IPDs will be summarized by treatment group and for all subjects for each deviation type, based on the RS. This summary will be repeated for the FAS (if this analysis set is different to the RS).

Change #3

Other baseline characteristics (Section 6.2)

The third and fourth paragraphs in this section have been changed to:

Baseline disease characteristics will be summarized using descriptive statistics by treatment group and for all subjects based on the SS. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy and will also include summary statistics for SUV_{auc} in the SIJ and spine at Baseline derived from the PET-MRI/PET-CT scans performed at Screening. SUV_{auc} values for each of the PET-positive lesions identified in the SIJ and spine will be pooled across subjects and across the SIJ and anatomical locations in the spine for the purpose of this summary. In addition, the maximum SUV_{auc} value across the SIJ and spine for each subject at Baseline will be summarized.

The above Baseline disease characteristics will be listed for the SS.

Change #4

Medical history and concomitant diseases (Section 6.4)

This section was amended as follows:

Medical history (except for AS) will be listed for the RS and summarized for the SS by treatment group and for all subjects, MedDRA® system organ class (SOC) and preferred term (PT). The reported term will be included in the listing. The summary will include the number and percentage of subjects and will be sorted alphabetically by SOC and by descending incidence of PT within each SOC, based on the 'All Subjects' column. This summary will be repeated for subjects in the SS who were randomized in the PET-MRI/CT

substudy. A glossary of all medical history conditions will be presented for the RS including the SOC, PT and reported term.

Procedure history will be listed separately for the RS by treatment group for previous procedures relating to AS and previous procedures not relating to AS. Concomitant medical procedures performed during the study will be listed for the RS.

Change #5

Efficacy assessment (Section 8.3.6.2)

The section was amended as follows:

SIJ and spine PET-CT and PET-MRI scans will be independently assessed at each time point (Baseline, Week 12 and Week 48) in a blinded fashion by two reviewers, with an additional adjudication read by a third reviewer, if applicable. Independent reviewers will assess PET-MRI/PET-CT examinations quantitatively using VU University Medical Center (VUMC)'s in-house software (ACCURATE).

Initially, two primary readers will independently assess the images visually to identify PET-positive lesions in the SIJ and spine. The readers will dichotomously score the presence or absence (1 versus 0) of PET-positive lesions in each side of the SIJ (Left and Right) and at each of the 24 spinal levels and 4 spinal regional levels [C2-C7 (Cervical), T1-T12 (Thoracic), L1-L5 (Lumbar), Sacrum] in each of the following 8 spinal locations:

- **Process spinosus**
- **Costo-vertebral (right)**
- **Costo-vertebral (left)**
- **Facet joint (right)**
- **Facet joint (left)**
- **Anterior vertebra**
- **Posterior vertebra**
- **Vertebra (other)**

For each subject, PET-positive lesions may be identified in one or both sides of the SIJ and in one or more of the 192 possible anatomical locations (spinal location/level combinations). It is expected that only one lesion for a subject may be separated/identified in each side of the SIJ or individual anatomical location in the spine, therefore, there is a direct mapping of a subject's lesion to a SIJ side or spinal location/level.

If there is a discrepancy between the two primary readers in the presence of PET-positive lesion assessment (Yes versus No for lesions in any side of the SIJ and 1 versus 0 for lesions in any spinal location), a consensus adjudication will take place with a third reader who will also be blinded to the clinical data. This reader will make an independent assessment and the results of the adjudication will be considered as the final assessment of the presence of PET-positive lesions.

Another reader who has not been involved in the visual identification of the PET-positive lesions, either as a primary reader or an adjudicator, will then draw volumes of interest (VOIs) on the pre-identified PET-positive lesions. In the case that no PET-positive lesions have been identified for a subject, then the final review step to draw VOIs will not occur. Efficacy PET-MRI/PET-CT review will occur provided that an acceptable PET-CT or PET-MRI scan is available and that PET-positive lesions were identified on the previous scan (for the Week 12 and Week 48 visits). During the PET-positive lesion VOI review, the independent reviewer will perform a quantitative assessment in ACCURATE and record the standard uptake values of $_{18}$F-fluoride in VOIs, corrected for both body weight and individual integrated whole blood activity concentration (SUV_{auc}) for each PET-positive lesion identified during the PET-positive lesion identification review.

Data from the visual identification of PET-positive lesions and the results of the quantitative VOI review (i.e. SUV_{auc}) will be listed, for each independent reader and adjudicator (if applicable), for each anatomical location in the spine, and for each side of the SIJ, by treatment group and visit for subjects in the SS who were enrolled into the PET-MRI/PET-CT substudy.

No formal statistical modelling of the PET-MRI/PET-CT imaging data will be performed for this substudy; only summaries will be presented together with 95% CIs for mean values, as applicable.

The summaries described below based on the visual identification of PET-positive lesions will include the final assessment data, ie, if adjudication was performed, then the adjudicated data will be summarized, otherwise data from one of the primary reviewers will be summarized.

All summaries will be presented for the PET-PPS.

Since the attenuation correction and quantification is variable in PET-MRI scans, PET-MRI imaging data will be listed but will not be aggregated with the PET-CT imaging data in the summaries.

Due to the predicted sparseness of the number of PET-positive lesions in each of the individual anatomical locations of the spine, data will be pooled as follows for the purpose of some of the summaries:

- **In the SIJ and spine overall (across the SIJ and all anatomical locations in the spine);**
- **In the SIJ overall (across both sides), and split by Left and Right side;**
- **In the spine overall (across all anatomical locations in the spine), and split by spinal regional level (Cervical, Thoracic, Lumbar, Sacrum), and, depending on the distribution of lesions across the spine, further by spinal levels and spinal locations in which at least 1 PET-positive lesion is identified.**

An overview of the total number of PET-positive lesions identified in the spine across all subjects in each of the individual anatomical locations will be presented by visit.

The presence of PET-positive lesions identified in the SIJ and spine will be summarized by visit. The summary will include the number and percentage of subjects at each visit with

any PET-positive lesions in the spine or the SIJ, in both the spine and the SIJ, in the SIJ overall (and split by Left and Right side) and in the spine overall [and split by spinal regional level (Cervical, Thoracic, Lumbar and Sacrum) and level/location, as applicable]. This summary will also include a total count of PET-positive lesions across subjects for each of the pooled locations described above. In the spine, the incidence and the total number of lesions will only be presented for the anatomical locations in which at least 1 PET-positive lesion is identified.

The number of PET-positive lesions per subject in the SIJ and spine (overall and in pooled locations, as described above) will be calculated and summarized by visit using frequency counts and percentages, and/or descriptive statistics, as applicable. In addition, in the spine and SIJ overall, in the spine overall and in each spinal regional level, the changes in the number of PET-positive lesions per subject from Baseline to Week 12 and Week 48, and the changes from Week 12 to Week 48 will be summarized using descriptive statistics.

Bar charts summarizing the total number of PET-positive lesions identified in the SIJ and in the spine will be presented by treatment group and visit for each of the pooled locations described above.

Observed SUV_{auc} values for each lesion will be pooled across subjects and summarized by visit for each of the pooled locations, as described above. In the spine, the observed SUV_{auc} values will only be summarized for the anatomical locations in which at least 1 PET-positive lesion is identified. Descriptive statistics will be presented including 95% CIs for the mean observed SUV_{auc} values. In addition, changes and percentage changes in SUV_{auc} values between visits will be calculated for each subject's PET-positive lesion in the SIJ and spine, i.e. Week 12 - Baseline, Week 48 - Baseline and Week 48 - Week 12. These changes in SUV_{auc} values will be pooled across subjects and summarized by visit for each of the pooled locations described above. Descriptive statistics will be presented including 95% CIs for mean changes and percentage changes in SUV_{auc} values between visits.

Spaghetti plots of observed SUV_{auc} values for lesions in the SIJ and spine will be presented over time for the SIJ overall, the spine overall, and for each spinal regional level (Cervical, Thoracic, Lumbar, Sacrum) by treatment group.

In addition, percentage changes in SUV_{auc} values for lesions in the SIJ and spine will be summarized using waterfall plots. These plots will identify subjects' lesions by treatment group and will be presented by pooled location (as described above, as applicable) and for each visit comparison (Week 12 – Baseline, Week 48 – Baseline and Week 48 – Week 12).

The maximum of a subject's observed SUV_{auc} values will be calculated across the pooled locations described above. This maximum SUV_{auc} value per subject will be summarized at each visit together with changes and percentage changes from Baseline between visits. Descriptive statistics will be presented including 95% CIs for the mean and the mean changes/percentage changes in the maximum SUV_{auc} values.

The mean and 95% CI for the maximum SUV_{auc} values per subject in the SIJ and spine over scheduled time will also be presented by treatment group. The same plot will be repeated for the changes and percentage changes in the maximum SUV_{auc} value per subject with treatment group overlaid on the same plot.

In addition, percentage changes in the maximum of SUV_{auc} values per subject in the SIJ and spine will be summarized using waterfall plots. These plots will identify subjects by treatment group and will be presented for each visit comparison (Week 12 – Baseline, Week 48 – Baseline and Week 48 – Week 12).

Additional exploratory analyses may include examining the relationship between changes from Baseline in bone formation and changes from Baseline in ASDAS at Week 12 (primary efficacy variable) and Week 48, and also whether there are any differences in bone formation in clinical responders according to ASAS20 versus clinical non-responders at Weeks 12 and 48. Statistical modelling which accounts for multiple PET-positive lesions per subject may be used in these analyses.

Change #6

Adverse events (Section 11.2)

The first paragraph was changed to:

Adverse events with start date prior to the first dose of IMP are defined as pre-treatment AEs. Adverse events with a start date prior to the first dose of IMP will be defined as pre-treatment AEs. A treatment-emergent AE (TEAE) is defined as any AE with a start date/time at the time of or after the first dose of IMP up until 140 days after the last dose of IMP. Any AE with onset later than 140 days after the last dose of IMP will not be considered as treatment-emergent and therefore will not be included in the tabulations of TEAEs. Selected summaries will be presented separately for TEAEs during the Treatment Period (any AE with a start date/time at the time or after the first dose of IMP up until the Week 12 Visit) and TEAEs during the study (as per the definition of TEAEs given above).

The fourth and fifth paragraphs were changed to:

An overview of the number and percentage of subjects who experience TEAEs will be presented by treatment group, based on the SS. Separate tabulations will be presented for TEAEs during the Treatment Period and TEAEs during the study and will include the number and percentage of subjects with any TEAEs, serious TEAEs, related TEAEs, discontinuation due to TEAEs, severe TEAEs, AEs leading to death and TEAEs leading to death; event counts will also be included. The summary of TEAEs during the study will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy.

In addition, the following summaries will be presented by treatment group, SOC, high level term (HLT) and PT, based on the SS:

- **Incidence of TEAEs[a] [b]**
- **Incidence of TEAEs, stratified by overall ADA status (positive/negative) as defined in Section 10**
- **Incidence of serious TEAEs[a] [b]**
- **Incidence of non-serious TEAEs[b]**
- **Incidence of TEAEs by relationship**
- **Incidence of TEAEs by maximum relationship**

- **Incidence of TEAEs by maximum intensity**
- **Incidence of fatal TEAEs by relationship**
- **Incidence of non-serious TEAEs by relationship**
- **Incidence of serious TEAEs by relationship**
- **Incidence of non-serious TEAEs above threshold of 5% of subjects**
- **Incidence of non-serious TEAEs above threshold of 5% of subjects by relationship**
- **Incidence of TEAEs by SOC and PT (including the number and percentage of subjects and individual subject numbers for each PT stratified by intensity, relationship and seriousness)**

[a] These summaries will be presented separately for TEAEs during the Treatment Period and TEAEs during the study.

[b] These summaries will be repeated for subjects in the SS who were randomized in the PET-MRI/PET-CT substudy and will be based on TEAEs during the study

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.