STATISTICAL ANALYSIS PLAN

Study: HS0001

Product: Bimekizumab

A PHASE 2 MULTICENTER, INVESTIGATOR-BLIND, SUBJECT-BLIND, PLACEBO-CONTROLLED STUDY OF THE EFFICACY, SAFETY, AND PHARMACOKINETICS OF BIMEKIZUMAB IN SUBJECTS WITH MODERATE TO SEVERE HIDRADENITIS SUPPURATIVA

SAP/Amendment Number	Date
Final	28 November 2017
Amendment 1	19 February 2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

%ΔAN	percentage change from Baseline in abscess and inflammatory nodule count
ACP	above the cut point
ADA	adalimumab
AE(s)	adverse event(s)
AESM	adverse events of special monitoring
ALC	absolute lymphocyte count
ALP	alkaline phosphatase
ALQ	above the limit of quantification
ALT	alanine aminotransferase
AN	abscess and inflammatory nodule
ANC	absolute neutrophil count
AST	aspartate aminotransferase
BCP	below the cut point
BKZ	bimekizumab
BLQ	below the limit of quantification
BMI	body mass index
bpm	beats per minute
BUN	blood urea nitrogen
CD	cluster of differentiation
CI	confidence interval
CP	confirmed positive
CSR	clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
CT	controlled terminology
CTCAE	Common Terminology Criteria for Adverse Events
DEM	data evaluation meeting
DLQI	Dermatology Life Quality Index
DMC	data monitoring committee
ECG	electrocardiogram
eCRF	electronic case report form
EQ-5D-3L	European Quality of Life-5 Dimensions-3 Level questionnaire

ES	Enrolled Set
EudraCT	European Union Drug Regulating Authorities Clinical Trials
EW	early withdrawal
FAS	Full Analysis Set
FDA	Food and Drug Administration
FSH	follicle stimulating hormone
geoCV	geometric coefficient of variation
GEE	generalized estimating equation
GGT	gamma glutamyltransferase
HADS	Hospital Anxiety and Depression Scale
HADS-A	Hospital Anxiety and Depression Scale Anxiety
HADS-D	Hospital Anxiety and Depression Scale-Depression
HiSCR	Hidradenitis Suppurativa Clinical Response
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HLGT	high level group term
HLT	high level term
HS	hidradenitis suppurativa
hsCRP	high sensitivity C-reactive
ICH	International Council for Harmonization
IHS4	International Hidradenitis Suppurativa Severity Scoring System
IL	interleukin
IMP	investigational medicinal product
IPD	important protocol deviation
LDH	lactate dehydrogenase
LLN	lower limit of normal
LLOQ	lower limit of quantification
mHiSCR25	modified Hidradenitis Suppurativa Clinical Response – 25%
mHiSCR75	modified Hidradenitis Suppurativa Clinical Response – 75%
mHiSCR90	modified Hidradenitis Suppurativa Clinical Response – 90%
MCH	mean corpuscular hemoglobin

MCHC	mean corpuscular hemoglobin concentration
MCV	mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
n	number of subjects
NA	not applicable
NCP	not confirmed positive
NRI	non-responder imputation
NRS	numeric rating scale
PBO	placebo
PCS	potentially clinically significant
PD	pharmacodynamic(s)
PDILI	potential drug-induced liver injury
PD-PPS	Pharmacodynamic Per-Protocol Set
PGA	patient's global assessment
PK	pharmacokinetic(s)
PK-PPS	Pharmacokinetic Per-Protocol Set
PoC	proof of concept
PPS	Per-Protocol Set
PT	preferred term
Q1	first quartile
Q3	third quartile
QoL	quality of life
QTcF	QT corrected for heart rate using Fridericia's formula
RBC	red blood cell
RCTC	Rheumatology Common Toxicity Criteria
RS	Randomized Set
SAE(s)	serious adverse event(s)
SAP	statistical analysis plan
SD	standard deviation
SDTM	Study Data Tabulation Model
SFU	Safety Follow-up
SMQ	standardized MedDRA query
SOC	system organ class

SS	Safety Set
TB	Tuberculosis
TEAE	treatment-emergent adverse event
TEMA	treatment-emergent markedly abnormal
TFLs	tables, figures and listings
ULN	upper limit of normal
VAS	visual analogue scale
WBC	white blood cell
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 HS0001. 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:

- Final protocol amendment 1, dated 08 March 2018

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 must be modified or updated prior to database lock, a further 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 this study is to evaluate the efficacy of bimekizumab in subjects with moderate to severe hidradenitis suppurativa (HS).

2.1.2 Secondary objective

The secondary objective of this study is to assess the safety, tolerability, immunogenicity, and pharmacokinetics (PK) of bimekizumab in subjects with moderate to severe HS.

2.1.3 Exploratory objective

The exploratory objective of the study is to further assess the efficacy of bimekizumab and to explore the effect of bimekizumab on gene and protein expression, and to explore the relationship between pharmacodynamic (PD), nonhereditary pharmacogenomics, pharmacogenetic, and imaging biomarkers and HS disease biology, drug treatment, and inflammatory and immune responses.

2.2 Study variables

2.2.1 Efficacy variables

2.2.1.1 Primary efficacy variable

The primary efficacy variable is the clinical response as measured by Hidradenitis Suppurativa Clinical Response (HiSCR) (defined as at least a 50% reduction from Baseline in the total AN (abscess and inflammatory nodule) count with no increase from Baseline in abscess or draining fistula count) at Week 12.

2.2.1.2 Exploratory efficacy variables

The exploratory efficacy variables are listed below and will be evaluated at scheduled visits in accordance with the schedule of assessments.

- Proportion of subjects with a total abscess and inflammatory nodule count of 0, 1, or 2 at Week 12
- Proportion of subjects with at least a 30% reduction and at least 1-unit reduction from Baseline in the Patient's Global Assessment (PGA) of Skin Pain (based on the subject's worst pain over the previous 24 hours) at Week 12
- Change from Baseline in Modified Sartorius Score at Week 12
- Change and percentage change from Baseline in lesion count (abscess count, inflammatory nodule count, abscess and inflammatory nodule count, and draining fistula count) over time
- Change from Baseline in International Hidradenitis Suppurativa Severity Scoring System (IHS4) score over time
- Change from Baseline in erythema assessment over time
- Change from Baseline in Hurley Stage over time
- Change and percentage change from Baseline in PGA of Skin Pain (pain at its worst and pain on average over the last 24 hours) over time
- Change from Baseline in the HS-Physician's Global Assessment 6-point scale over time
- Change from Baseline in the Dermatology Life Quality Index (DLQI) score over time
- Change from Baseline in European Quality of Life-5 Dimensions-3 Level questionnaire (EQ-5D-3L) (index and visual analogue scale [VAS] scores) over time
- Change from Baseline in Hospital Anxiety and Depression Scale (HADS), HADS-Anxiety (HADS-A) and HADS-Depression (HADS-D) scores over time

2.2.2 Pharmacokinetic, pharmacodynamic, pharmacogenomic, and pharmacogenetic variables

2.2.2.1 Secondary pharmacokinetic variable

Blood samples will be collected for the measurement of plasma bimekizumab concentrations at scheduled visits in accordance with the schedule of assessments.

2.2.2.2 Exploratory pharmacodynamic variables

The PD variables assessed are the blood or blood derivative (eg, plasma) concentrations of cytokines and chemokines of relevance to interleukin (IL) 17A/F signaling pathway and HS biology.

Additional variables may include, but will not be limited to, serum complement concentrations and mononuclear cell subtypes.

2.2.2.3 Exploratory nonhereditary pharmacogenomic variables

Where local regulations permit, blood, urine and skin tissue biopsy samples will be collected and stored for up to 20 years to allow for potential exploratory analyses of ribonucleic acid, proteins, and metabolites biomarkers relevant to HS and the inflammatory and immune response processes. The nature and format of these tentative analyses will be determined at a later stage. The results of any pharmacogenomics analyses will be reported outside the CSR and are not further described in this SAP.

2.2.2.4 Exploratory 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 with response to treatment with bimekizumab or adalimumab, HS disease biology, and inflammatory and immune response processes. The nature and format of these tentative substudy analyses will be determined when the results of the main study are made available. The results of any pharmacogenetic analyses will be reported outside the CSR and are not further described in this SAP.

2.2.3 Safety variables

As part of the secondary objective, the following secondary safety variables will be assessed:

- Frequency and severity of adverse events (AEs) and serious adverse events (SAEs)
- Withdrawal due to AEs
- Change from Baseline in vital signs (blood pressure, pulse rate and body temperature) and body weight
- Change from Baseline in electrocardiogram (ECG) parameters
- Change from Baseline in clinical laboratory parameters (hematology, biochemistry, and urinalysis)
- Change from Baseline in physical examination

2.2.4 Immunological variables

Blood samples will be collected for the measurement of anti-drug (bimekizumab) antibody (anti-bimekizumab antibody).

2.2.5 Exploratory Imaging variables

Imaging biomarkers will be assessed by means of medical infrared imaging (thermography) and photography in specific centers with technological capabilities for subjects who consent to participate in the imaging substudy. Imaging biomarkers may be measured to evaluate the relationship with response to treatment with bimekizumab or adalimumab, HS disease biology, and inflammatory and immune response processes. The thermography and photography results will not be a part of the CSR and are not further described in this SAP.

2.3 Study design and conduct

HS0001 is a Phase 2 multicenter, randomized, Investigator-blind, subject-blind, placebo-controlled, active-reference arm study to assess the efficacy, safety, and PK of bimekizumab in eligible adult subjects with moderate to severe HS.

This study will include 3 periods: a Screening Period (≥2 weeks up to a maximum of 4 weeks prior to randomization), a Treatment Period (12 weeks), and a Safety Follow up (SFU) Period (20 weeks after the last dose of investigational medicinal product [IMP]).

A sufficient number of subjects will be screened in order to have 80 subjects randomized 2:1:1 (bimekizumab to placebo to adalimumab, stratified by Baseline Hurley Stage) to receive blinded IMP (all administered by subcutaneous injection by unblinded study personnel). There will be approximately 40 subjects in the bimekizumab treatment group, 20 subjects in the placebo group, and 20 subjects in the adalimumab group.

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

The predefined success criterion for declaring a positive study requires a posterior probability of at least 97.5% that the HiSCR rate at Week 12 for bimekizumab will be higher than that for placebo.

Additionally, the study has an exploratory objective to informally compare bimekizumab against the current standard of care for HS (adalimumab) using the 60% credible interval for the treatment difference.

Sample size calculations are based on the number of subjects required to meet the above success criterion (comparing against placebo) assuming a true underlying HiSCR rate at Week 12 of 70% for bimekizumab. Published data from the PIONEER I and II studies (Kimball et al, 2016a) indicate that, when antibiotic therapy at entry is allowed, the response rates for placebo and adalimumab at Week 12 are 27.6% and 58.9% respectively.

Using simulation, it is estimated that with data from 40 subjects on bimekizumab, 20 subjects on placebo and 20 subjects on adalimumab (80 subjects total), the study will provide at least a 99% probability of meeting the success criterion and an 80% probability of meeting the exploratory criterion under the assumptions described below.

These simulations were based on Bayesian augmented control models in which the published summary data were used to construct informative priors for the placebo and adalimumab treatment arms in the current study. Based on historic data, the Bayesian model assumes an informative prior distribution for the placebo group HiSCR rate to be β (5.52,14.48). This prior distribution contributes an approximate effective sample size of 20 subjects for the placebo treatment group. Similarly, the Bayesian model assumes an informative prior distribution for the adalimumab group HiSCR rate to be β (11.78, 8.22) which, with the parametrization in the model, is equivalent to an adalimumab response rate of 58.9%. This prior distribution contributes an approximate effective sample size of 20 subjects for the adalimumab group. For the bimekizumab HiSCR rate, a vague prior probability distribution N (0,1E4) was assumed.

3 DATA ANALYSIS CONSIDERATIONS

3.1 General presentation of summaries and analyses

Statistical evaluation will be performed by PAREXEL with oversight and supervision by UCB. The datasets will follow the UCB analysis data model data specifications.

All statistical analyses will be performed using SAS® version 9.4 or later (SAS Institute, Cary NC, USA) or R version 2.10.1 (R Development Core Team) or later, or WinBUGS 1.4.

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

The primary treatment comparison for all analyses of efficacy will be between bimekizumab and placebo.

Continuous variables will be summarized by treatment, visit (where applicable) including number of subjects (n), mean, standard deviation (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 indicated.

Categorical variables will be summarized by treatment, visit (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 data. In all outputs the 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
- For the purpose of the tabulations the lower and upper 95% confidence limits for the percentages will be truncated at 0 and 100% respectively
- Unless otherwise stated, the denominator for the percentages will be based on the number of subjects in the respective analysis set, treatment group, visit and any subgroup defined for the tabulations (eg, Baseline Hurley Stage) with non-missing data for the relevant classification

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 data, for which additional rules are stated below):

- n will be an integer

- Mean (arithmetic and geometric), 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 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, and to 1 additional significant figure for the mean (arithmetic and geometric), median, SD and 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:

- Posterior probabilities associated with the Bayesian analyses of the primary efficacy variable will be presented as percentages to 1 decimal place
- Means, SDs, medians, 95% credible intervals and highest posterior density intervals obtained from the Bayesian analysis of the primary efficacy variable will be presented to 1 decimal place

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

All tabulations will be performed by treatment group and visit (where applicable). In the TFLs the treatment groups will be displayed as follows, where BKZ refers to bimekizumab, ADA refers to adalimumab and PBO refers to placebo:

- PBO
- ADA
- BKZ

Subject data listings containing all documented data and all derived data will be generated and these will be presented by treatment group and subject.

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 Analysis time points

3.2.1.1 Relative day

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

Relative days for an event or measurement occurring before the date of first dose will be calculated as follows:

$$Relative\ Day\ =\ Event\ Date\ -\ Date\ of\ First\ Dose$$

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

$$Relative\ Day\ =\ (Event\ Date\ -\ Date\ of\ First\ Dose)\ +\ 1$$

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

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

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 will be presented as '-' in the subject data listings.

3.2.1.2 Study periods

For each subject completing the study, their maximum study duration will be up to 34 weeks and will comprise the following:

- Screening Period: ≥2 weeks with a maximum of 4 weeks prior to randomization
- Treatment Period: 12 weeks
- Safety Follow-Up Period: 20 weeks after the last dose of IMP (scheduled at Week 10, 2 weeks prior to the end of the Treatment 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 dose of IMP. Scheduled or unscheduled measurements can be used as the Baseline value. If a measurement is repeated at Baseline and is obtained prior to receiving the first dose of IMP, then the last available measurement will be used as the Baseline value.

Measurement-specific Baseline visits (based on the planned visits in the protocol) are presented in Table 3–1.

Table 3–1: Definition of Baseline

Category	Measurement/Derivation	Definition of Baseline
Efficacy	Lesion count (including Hurley Stage)	Visit 2 (Day 1, predose) or if missing, the Screening value

Table 3–1: Definition of Baseline

Category	Measurement/Derivation	Definition of Baseline
	Erythema assessment HADS-A and HADS-D Modified Sartorius Score PGA of skin pain HS-Physician's global assessment DLQI EQ-5D-3L	
Safety	Vital signs 12-lead ECG Safety laboratory tests C-SSRS	Visit 2 (Day 1, predose) or if missing, the Screening value
Pharmacodynamic	Cytokines, candidate biomarkers and flow cytometry	Visit 2 (Day 1, predose)
Immunological	Anti-bimekizumab antibody	Visit 2 (Day 1)

C-SSRS=Columbia-Suicide Severity Rating Scale; DLQI=Dermatology Life Quality Index; ECG=electrocardiogram; EQ-5D-3L=European Quality of Life 5 Dimensions-3 Level questionnaire; HADS= Hospital Anxiety and Depression Scale; HADS-A= Hospital Anxiety and Depression Scale-Anxiety; HADS-D= Hospital Anxiety and Depression Scale-Depression; PGA=Patient's Global Assessment.

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 × (Post-Baseline Visit Value – Baseline Visit Value) / (Baseline Visit Value)

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 as follows:

- Inclusion/exclusion criteria deviations
- Administration of prohibited concomitant medications
- Deviations relating to withdrawal criteria
- Visit schedule deviations

- Study drug 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 3 data evaluation meetings (DEMs) will be performed for this study at the following time points:

- Prior to the informal unblinded interim analysis planned after approximately 45 subjects have completed 4 weeks of the study.
- Prior to the informal unblinded interim analysis planned after the last subject (excluding 2 subjects from site 101 in Australia who entered the study outside of the original planned recruitment period) has completed 12 weeks of the study or has discontinued early.
- Prior to the final analysis after all subjects have completed the SFU period and 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, 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 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 subjects randomized into the study.

3.5.3 Safety Set

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

3.5.4 Full Analysis Set

The Full Analysis Set (FAS) is a subset of the RS and will consist of all subjects who received at least 1 dose (full or partial) of IMP and had a valid Baseline measurement and a post-Baseline measurement for at least 1 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 and for all exploratory efficacy analyses.

3.5.5 Per-Protocol Set

The Per-Protocol Set (PPS) is a subset of the FAS, consisting of those subjects who had no IPDs affecting the primary efficacy variable, as confirmed during a pre-analysis review prior to unblinding of the data (at each of the interim and final analyses). All primary efficacy analyses will be performed for the PPS. The PPS will also be used for the supportive and sensitivity analyses of the primary efficacy variable.

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 formal interim analysis as the primary efficacy variable is assessed at Week 12, whereas the formal interim analysis will be based on only 45 of the 80 randomized subjects completing Week 4 of the study. Therefore, all relevant data will not be available at the time of the interim analysis DEM.

The PPS as agreed for the informal interim analysis may similarly be different from the PPS as agreed for the final analysis due to additional information obtained on subjects during the SFU Period.

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 Pharmacokinetic Per-Protocol Set (PK-PPS) will consist of all randomized subjects who received at least 1 dose of IMP and had at least 1 quantifiable post-dose plasma concentration.

3.5.7 Pharmacodynamics Per-Protocol Set

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

3.6 Treatment assignment and treatment groups

Listings and summaries will be presented by treatment group and for all subjects in the relevant analysis sets, where applicable. Treatment assignment is given in Table 14–11.

The following order will be used in the TFLs (with the treatments displayed accordingly):

- Not randomized (if applicable)
- PBO
- ADA

- BKZ
- All subjects

The dosing schedules and number of injections given at each visit are described in Table 3–2. Due to the differences in the dosing schedule between bimekizumab and adalimumab, placebo injections will be administered along with active drug such that all subjects will receive the same number of injections at each corresponding visit.

Table 3–2: Administration of investigational medicinal product

Visit	Treatment group		
	Adalimumab[a] (Total # injections, # of placebo injections)	**Bimekizumab[b] (Total # injections, # of placebo injections)**	**Placebo**
Baseline	160mg (4 active injections, no placebo injections)	640mg (4 active injections, no placebo injections)	4 placebo injections
Week 2	80mg (2 active injections, no placebo injections)	320mg (2 active injections, no placebo injections)	2 placebo injections
Weeks 4, 6, 8, and 10	40mg (1 active injection, 1 placebo injection)	320mg (2 active injections, no placebo injections)	2 placebo injections
Weeks 5, 7, and 9	40mg (1 active injection, no placebo injection)	NA (no active injection, 1 placebo injection)	1 placebo injection

NA=not applicable
[a] Each individual adalimumab injection will be 40mg.
[b] Each individual bimekizumab injection will be 160mg.

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 rules:

- Subjects randomized to placebo who received at least 1 dose of bimekizumab will be summarized in the 'BKZ' treatment group
- Subjects randomized to bimekizumab who received placebo throughout the Treatment Period will be summarized in the 'PBO' treatment group
- Subjects randomized to bimekizumab who received any doses of placebo during the Treatment Period, but at least 1 dose of bimekizumab, will be summarized in the 'BKZ' treatment group
- Subjects randomized to bimekizumab who received any doses of adalimumab during the Treatment Period, but at least 1 dose of bimekizumab, 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 randomized to placebo who received at least 1 dose of adalimumab will be summarized in the 'ADA' treatment group
- Subjects randomized to adalimumab who received placebo throughout the Treatment Period will be summarized in the 'PBO' treatment group
- Subjects randomized to adalimumab who received any doses of placebo during the Treatment Period, but at least 1 dose of adalimumab, will be summarized in the 'ADA' treatment group
- Subjects randomized to adalimumab who received any doses of bimekizumab during the Treatment Period, but at least 1 dose of adalimumab, 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 randomized to placebo who received at least 1 dose of adalimumab and at least 1 dose bimekizumab will be assessed at the DEM for the treatment group assignment, as it will depend on how many doses of each of the active 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 statistical analyses 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

The data from different sites will be pooled for all analyses. No stratification by site or country will be performed.

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 protocol defines the HiSCR as "at least a 50% reduction from Baseline in the total abscess and inflammatory nodule count, with no increase from Baseline in abscess or draining fistula count". The SAP defines this as "at least a 50% reduction from Baseline in the total abscess and inflammatory nodule count, with no increase from Baseline in abscess **or draining tunnels (fistulas/sinus tracts) count"**

- The section on the determination of the sample size in the protocol states that, with the planned sample size of 80 subjects, the probability of study success is 85%. Subsequent simulations have been performed and the probability found to be 99%. The SAP thus states this 99% probability in Section 2.4 instead.
- 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 these data and to define a separate analysis set accordingly.
- The definition of the FAS was updated to include all subjects with a valid Baseline measurement and a post-Baseline measurement for at least 1 efficacy variable. In the protocol, the definition included only subjects with a valid Baseline measurement and a post-Baseline measurement for the primary efficacy variable. Since the primary efficacy variable is a composite endpoint based on abscess count, inflammatory nodule count and draining tunnel (fistula/sinus tract) count, using this definition, any subjects with at least 1 missing component of the composite endpoint would not be included in any of the efficacy analyses.
- Based on the second interim analysis for this study, additional exploratory efficacy variables based on the lesion count assessment data, HiSCR depth and disease flare status, have been defined in this SAP amendment. Section 8.2.4 includes details on the derivation of different HiSCR depths and Section 8.2.5 describes the planned summaries of this additional variable. Section 8.3.1.2 defines disease flare and includes details on planned summaries of disease flare status and the number of visits at which a subject experienced a disease flare during the Treatment Period.

4 STATISTICAL/ANALYTICAL ISSUES

4.1 Adjustments for covariates

Randomization will be stratified by Baseline Hurley Stage in this study. Accordingly, all statistical models fitted in the analyses of the primary efficacy variable will include the stratification factor for Baseline Hurley Stage as a covariate.

In cases of classification errors at the time of randomization, subjects may be assigned to an incorrect stratum. In these cases, subjects will be included in the statistical modelling according to the observed values of Baseline Hurley Stage regardless of the value used by the site to stratify the subject at randomization.

Concomitant antibiotic use at entry into the study will be included as an additional covariate in the supportive and sensitivity analyses of the primary efficacy variable.

4.2 Handling of dropouts or missing data

4.2.1 Efficacy data

There will be no imputation of missing data for the individual components of the HiSCR (abscess count, inflammatory nodule count and draining tunnel (fistula/sinus tract) count).

For subjects with a missing HiSCR due to missing component data at a specific visit, or due to early discontinuation from the study during the Treatment Period, non-responder imputation

(NRI) will be used for the primary analysis of the primary efficacy variable. These subjects will be defined as non-responders at a specific visit or at all subsequent visits after the time of study discontinuation. In additional sensitivity analyses, an 'observed cases' analysis will also be performed (no imputation of missing HiSCR) and subjects who receive specific rescue therapy will also be considered a non-responder from the time that the rescue therapy was taken to the end of the Treatment Period (Section 8.2.3).

For the hsCRP data, measurements that are below the limit of quantification (BLQ) will be imputed with half of the lower limit of quantification (LLOQ) for the purpose of calculating summary statistics, changes from Baseline and ratio to Baseline.

The handling of any missing data for the other exploratory efficacy data is described or referenced in the respective sections where the analysis of this data is described.

4.2.2 Pharmacodynamic and pharmacokinetic data

For the PD data (cytokines, chemokines, flow cytometry and additional variables as measured), 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.

Measurements of PK concentration that are BLQ will be imputed with half of the LLOQ for the purpose of calculating the geometric mean and its 95% CI, the geoCV, the arithmetic mean and SD for the summaries and figures.

For the individual figures, any concentrations that are BLQ will be regarded as missing, with the exception of predose BLQ measurements on Day 1, which will be imputed with zero for linear scale plots.

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 PK data in Section 4.2.2.

4.2.4 Dates

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

Partially missing dates will be imputed as described above and the duration of each AE will be calculated in days as follows:

$$Duration\ of\ AE\ =\ (Date\ of\ Outcome\ -\ Date\ of\ Onset\ )+\ 1$$

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 early withdrawal visit

Subjects who withdraw early from the study for any reason, including those withdrawn from IMP, will be asked to return for the study assessments 12 weeks after the first dose of IMP (ie, the Week 12 Visit) and will then enter the SFU Period and will undergo the same assessments performed at the SFU Visit (20 weeks after the last dose of IMP).

For subjects who return for the EW visit at any time other than the Week 12 visit, the following rules will apply with regard to the inclusion of the results obtained at the EW visit in the descriptive summaries and in the data listings:

- 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 anti-BKZ antibody assessments, in which all early withdrawal visit assessments will be assigned to the next scheduled visit at which anti-BKZ antibodies are assessed. The rationale for this is that anti-BKZ antibody positivity is summarized over a given study period. As part of that summary, a table indicating the first visit at which anti-BKZ antibody positivity is observed will be presented. In order to match the number of subjects who were anti-BKZ antibody positive at specific visits with the overall positivity for the period, it is necessary to ensure that anti-BKZ antibody positivity is attributed to a visit where anti-BKZ 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 analysis

This study will include two informal unblinded interim analyses, the first after approximately 45 subjects have completed 4 weeks of the study and the second when the last randomized subject for this study has completed the Week 12 Visit at the end of the Treatment Period or this subject's participation has ended prematurely. This second planned interim analysis excludes the 2 subjects from site 101 who entered the study outside of the recruitment period. The purpose of these interim analyses is for sponsor key personnel to review results from the primary efficacy analysis and a subset of the analyses of exploratory efficacy and safety outcomes to facilitate additional Clinical Planning or Portfolio Management decisions. Neither interim analysis will lead to any formal decision to alter or terminate the trial.

Details on the summaries and analyses to be presented at these informal interim analyses are included in the Interim Analysis SAP.

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, 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 12 (Week 30).

In addition, the number and percentage of subjects completing Week 4 and Week 12 of the Treatment Period will be presented. A subject is defined as completing Week 4 and Week 12, 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 summarized separately 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 in each of the analysis sets will be summarized by treatment group and for all subjects based on the ES. Percentages will be calculated based on the RS for the purpose of this summary.

Screen failure reasons will be summarized for the ES. A listing of subjects who did not meet study eligibility criteria (including a glossary) will also be presented for the ES.

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

- Subject disposition (ES)
- Study discontinuation (RS)
- Visit dates (SS)
- Subject analysis sets (ES)

The listing of subject disposition will include the date of informed consent, date of pharmacogenetic informed consent (for subjects participating in the pharmacogenetic substudy), date of randomization, date and time of first and last dose of IMP, date of premature study termination and primary reason (if applicable), and date of final contact. If applicable, the date and reason for premature unblinding will also be included.

The listing of study discontinuation will include the primary 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 deviation 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 (Section 5.2), based on 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 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:

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

In the event that the date of diagnosis is incomplete, it will be imputed to the most recent feasible date:

- 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, the duration of disease will not be calculated

6.2 Demographics and Baseline characteristics

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 BMI for the ES. The age will be entered directly into the study database and will not be re-calculated for the statistical analysis.

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. Childbearing potential (as applicable) and lifestyle (alcohol, tobacco and caffeinated beverage use) will be listed for the SS.

6.3 Other Baseline characteristics

The following Baseline disease characteristics will be listed for the SS, based on the Baseline definitions in Table 3–1 unless otherwise specified:

- Lesion counts by anatomical region and lesion type, total lesion counts across anatomical regions by lesion type, Hurley Stage by anatomical region and worst overall Hurley Stage

across anatomical regions; degree of erythema by anatomical region and worst overall erythema across all anatomical regions (Section 8.1)

- IHS4 score, total HADS-A and HADS-D scores, Modified Sartorius Score, PGA of skin pain (average and worst pain over the last 24 hours), HS-Physician's Global Assessment, total DLQI score and EQ-5D-3L (domain scores, index score and VAS score) (Section 8.3)
- Duration of disease at Screening, based on date of diagnosis as captured in the eCRF
- High sensitivity C-reactive protein (hsCRP)

The above Baseline disease characteristics will be summarized by treatment group and for all subjects in the FAS.

Baseline lesion counts will be summarized by lesion type and anatomical region, and total lesion counts across all anatomical regions will be summarized by lesion type.

Baseline Hurley stage and degree of erythema will be summarized by treatment group and for all subjects by anatomical region. The worst overall Hurley stage and erythema calculated across all anatomical regions will be summarized by treatment group and for all subjects.

In addition, the following Baseline disease characteristics will be summarized by Baseline Hurley Stage and overall, and treatment group, for the FAS.

- IHS4 score
- Total HADS-A and HADS-D scores
- Modified Sartorius Score
- PGA of skin pain (average and worst pain over the last 24 hours)
- HS-Physician's Global Assessment
- Total DLQI score
- EQ-5D-3L (domain scores, index score and VAS score)
- Duration of disease
- hsCRP
- Total lesion counts
- Worst overall erythema

6.4 Medical history and concomitant diseases

Medical history will be listed for the RS and summarized for the SS by treatment group and for all subjects, and by 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. 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 by treatment group for the procedure reported term based on the RS. Concomitant medical procedures performed during the study will be listed for the RS.

6.5 Hidradenitis suppurativa history

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

6.6 Prior and concomitant medications

Prior and concomitant medications (as defined below) 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 antibiotic and rescue medications, respectively. 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 HS medications
- Concomitant antibiotic medications at study entry
- Rescue medication during study

All HS medications, antibiotics 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.

Concomitant antibiotic medication at study entry is defined as any antibiotic with start date on or prior to first dosing, and end date on or after screening.

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.

The listing will include a flag for prohibited prior and concomitant medications. These will be confirmed and documented at the final DEM.

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

Any medications with partially missing dates will be handled as described in Section 4.2.4 in order 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 and volume administered for each injection at each visit will be included in the listing of study drug administration as specified in Section 11.1.

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 is administered on site.

8 EFFICACY ANALYSES

8.1 Lesion count assessment

The primary efficacy variable and some of the exploratory efficacy variables discussed in Section 8.3 (IHS4, Modified Sartorius Score, HS-Physician's Global Assessment, Hurley Stage and Erythema assessment) are based on the assessment and/or counts of different types of lesions in the following main anatomical regions at each visit.

- Inguinal (groin)
- Axillary (armpit)
- Submammary and/or intermammary in women
- Gluteal
- Other

These anatomical regions are further classified into the following locations, for the left and right sides of the body, as applicable:

- Inguinal (groin) excluding genital and pubic area – Left, Right
- Inguinal (groin) including genital and pubic area – Left, Right
- Axillary (armpit) – Left, Right
- Submammary – Left, Right
- Intermammary
- Gluteal – Buttocks – Left, Right
- Gluteal – Perianal/Perineal

All "Other" anatomical regions with lesions present will be specified in free text on the eCRF. Note that investigators will be asked to only select 'Other' regions with more severe lesions present.

The number of each of the following types of lesions (classified as relevant or non-relevant) will be recorded in each anatomical region, and then summed across all anatomical regions:

Relevant Lesions:

- Abscesses
- Draining tunnels (fistulas/sinus tracts)
- Inflammatory nodules

Non-relevant Lesions:

- Non-draining tunnels (fistulas/sinus tracts)
- Non-inflammatory nodules
- Hypertrophic scars

If there are 2 or more relevant lesions present in an anatomical region, the longest distance between 2 relevant lesions will be recorded and an assessment of whether these 2 relevant lesions are clearly separated by normal skin will be made.

If only 1 lesion is present in an anatomical region, then the diameter of the single lesion will be recorded.

In addition to the lesion count data, the degree of erythema will be assessed for each anatomical region and scored as follows:

- 0 = No redness
- 1 = Faint but discernible pink coloration
- 2 = Moderate red coloration
- 3 = Very red or bright red coloration

The worst overall erythema score across all anatomical regions will then be derived automatically in the database and used in the analyses.

Hurley Stage (I, II and III) will also be recorded for each anatomical area and the worst overall Hurley Stage across all anatomical regions will then be derived automatically in the database and used in the analyses.

8.2 Primary efficacy variable

The primary efficacy variable is the clinical response as measured by HiSCR (defined as at least a 50% reduction from Baseline in the total abscess and inflammatory nodule count [AN count], with no increase from Baseline in abscess or draining tunnel [fistula/sinus tract] count) at Week 12.

The HiSCR was developed to address issues with available HS scoring systems and is a validated variable that is responsive to improvement in disease activity, simplifies the scoring process, and increases the sensitivity to detect HS specific lesions (Kimball et al, 2014; Kimball et al, 2016b).

8.2.1 Derivation of the primary efficacy variable

The following algorithm will be applied to derive the HiSCR at each visit, based on total lesion counts across anatomical regions for the 3 relevant lesion types recorded as specified above in Section 8.1:

1. Calculate the AN count at each visit as the total number of abscesses plus the total number of inflammatory nodules, across all anatomical regions
2. Calculate the percentage change from Baseline in AN count (%ΔAN) at each visit as

$$100 \times (AN\ count\ at\ post\text{-}Baseline\ visit\ minus\ Baseline\ AN\ count) \\ / (Baseline\ AN\ count)$$

3. Calculate the change from Baseline in the abscess count by subtracting the Baseline abscess count from the abscess count at each post-Baseline visit
4. Calculate the change from Baseline in the draining tunnel (fistula/sinus tract) count by subtracting the Baseline draining tunnel (fistula/sinus tract) count from the draining tunnel (fistula/sinus tract) count at each post-Baseline visit
5. If the %ΔAN is less than or equal to -50%, and the change from Baseline in the abscess count is zero or negative (ie, non-positive) together with a non-positive change from Baseline in the draining tunnel (fistula/sinus tract) count, then the HiSCR will be assigned a value of 1 (ie, HiSCR is achieved); otherwise, the HiSCR will be assigned a value of 0 (ie, HiSCR is not achieved).

There will not be any imputation of missing lesion count data in the derivation of the HiSCR at any visit. In such cases the HiSCR will be missing and rules for handling missing HiSCR values in the analysis will be applied (Section 4.2.1 and Section 8.2.3).

The primary efficacy variable is attained if the subject has a HiSCR of 1 at Week 12.

8.2.2 Primary analysis of the primary efficacy variable

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

8.2.2.1 Bayesian modeling

The primary Bayesian model (a logistic regression model for the mean binary response $E[y]$) will be:

$$\ln\left(\frac{E(y)}{1-E(y)}\right) = \beta_{PBO}\, X_{PBO} + \beta_{BKZ}\, X_{BKZ} + \beta_{ADA}\, X_{ADA} + \beta_{hII}\, X_{hII} + \beta_{hIII}\, X_{hIII} \qquad (1)$$

where X_{PBO}, X_{BKZ} and X_{ADA} are dummy variables taking the value 1 or 0 according to whether or not each subject is in the PBO, BKZ or ADA treatment groups respectively, X_{hII} and X_{hIII} are dummy variables taking the value 1 or 0 according to whether or not each subject is in the Baseline Hurley Stage II or III groups respectively. Proof-of-concept (PoC) will be declared in this study if the estimated posterior probability, using the above model, of the difference in response rates between the BKZ and PBO treatment groups is 0.975 or greater.

For parameter estimation, the following constraint will be used:

$$\beta_{hII} = -\beta_{hIII}$$

giving the model:

$$\ln\left(\frac{E(y)}{1-E(y)}\right) = \beta_{PBO}\, X_{PBO} + \beta_{BKZ}\, X_{BKZ} + \beta_{ADA}\, X_{ADA} + \beta_{hII}\,(X_{hII} - X_{hIII}) \qquad (2)$$

so that the influence of Baseline Hurley Stage on the model estimated treatment group mean responses are more evident. To see this, note that if $\bar{\eta}_{PBO}$ represents the mean of the logit responses from model (2) for the PBO subjects (the simple "arithmetic" mean) then $logit^{-1}(\bar{\eta}_{PBO})$ is the response probability associated with that mean logit response and is 1 type of estimate of the PBO group response rate. Accordingly:

- $p_{PBO} \mid p_{hII}^{PBO}, p_{hIII}^{PBO} = logit^{-1}(\bar{\eta}_{PBO}) = logit^{-1}\left(\beta_{PBO} + \beta_{hII}\,(p_{hII}^{PBO} - p_{hIII}^{PBO})\right)$ represents the probability of response in the PBO group conditional on p_{hII}^{PBO} and p_{hIII}^{PBO}, which are the Baseline Hurley Stage subject proportions chosen to evaluate the estimated PBO group mean at (here the observed proportions are chosen)
- $p_{BKZ} \mid p_{hII}^{BKZ}, p_{hIII}^{BKZ} = logit^{-1}(\bar{\eta}_{BKZ}) = logit^{-1}\left(\beta_{BKZ} + \beta_{hII}\,(p_{hII}^{BKZ} - p_{hIII}^{BKZ})\right)$ represents the probability of response in the BKZ group conditional on p_{hII}^{BKZ} and p_{hIII}^{BKZ} which are the Baseline Hurley Stage subject proportions chosen to evaluate the estimated BKZ group mean at (here again the observed proportions are chosen)
- $p_{ADA} \mid p_{hII}^{ADA}, p_{hIII}^{ADA} = logit^{-1}\,(\bar{\eta}_{ADA}) = logit^{-1}\left(\beta_{ADA} + \beta_{hII}(p_{hII}^{ADA} - p_{hIII}^{ADA})\right)$ represents the probability of response in the ADA group conditional on p_{hII}^{ADA} and p_{hIII}^{ADA} which are the Baseline Hurley stage subject proportions chosen to evaluate the estimated ADA group mean at (once more here the observed proportions are chosen)

For brevity, and unless otherwise stated the dependence of the above treatment group response rates on the chosen Baseline Hurley Stage subject proportions will be suppressed by denoting them as p_{PBO}, p_{BKZ} and p_{ADA}.

For pragmatic reasons we will ignore the Hurley stage proportions and effects in the equations for the estimated group response rates and use the following relations to map between the probability scale priors for the group response rates and the logit scale treatment group parameters

$$p_{PBO} = logit^{-1}(\beta_{PBO}) \qquad (3)$$

$$p_{ADA} = logit^{-1}(\beta_{ADA}) \qquad (4)$$

The expectations of the prior distributions for the PBO and ADA group response rates will be set to match the observed treatment group response rates in the PII study.

Accordingly, the priors to be used are

$$p_{PBO} \sim \beta(5.52, 14.48) \qquad (5)$$

$$\beta_{BKZ} \sim N(0, 10000)$$

$$p_{ADA} \sim \beta(11.78, 8.22)$$

$$\beta_{hII} \sim N(0, 10000)$$

which have expectations equal to 28% and 59% for the PBO and ADA groups respectively to match the observed treatment group response rates of the PIONEER II study. The implicit assumption using this simple approach is that $E_{\beta_{hII}}[\beta_{hII}(p_{hII}^{BKZ} - p_{hIII}^{BKZ})] \approx 0$. It is important to recognize that even if the proposed study also observes a similar Baseline Hurley Stage effect size and subject proportions as the PIONEER II study the priors (5) and equations (3) and (4), whilst yielding estimated treatment group response rates that should be consistent with those observed in the data, may cause the estimated logit coefficients β_{PBO} and β_{ADA} to be inaccurate.

If the observed Baseline Hurley Stage effects and/or subject proportions in the proposed study are very different from those in the PIONEER II study then, and in addition to the possible invalid estimates of β_{PBO} and β_{ADA}, the estimated treatment group response rates may no longer reflect those in the data since the means of the probability scale priors will no longer match the treatment group response rate means in the observed data.

Both informative priors in (5) have been heavily discounted to account for study to study variation in both the ADA and PBO response rates, and also uncertainty in the agreement between the subject proportions and effect sizes observed in the Baseline Hurley Sage groups in the PIONEER II study and this proposed study.

To protect against severe forms of study to study variation and/or mismatches between Baseline Hurley Stage subject proportions and effect sizes, 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. Investigating any prior data conflict also provides some protection against model misspecification – for example if there turns out to be a Baseline Hurley Stage by treatment group interaction as the PIONEER II data suggest.

To be consistent with models used in previous UCB studies which parameterize PBO as the model intercept with the active treatment being additive effects (on the logit scale) we note that model (2) can be re-parameterized as follows:

$$\begin{aligned} \beta_{PBO} &= \gamma_{PBO} \\ \beta_{ADA} &= \gamma_{PBO} + \gamma_{ADA} \\ \beta_{BKZ} &= \gamma_{PBO} + \gamma_{BKZ} \\ \beta_{hII} &= \gamma_{hII} \end{aligned} \quad (6)$$

giving the model

$$\ln\left(\frac{E(y)}{1-E(y)}\right) = \gamma_{PBO} + \gamma_{BKZ}\, X_{BKZ} + \gamma_{ADA}\, X_{ADA} + \gamma_{hII}(X_{hII} - X_{hIII}) \quad (7)$$

Using this model, the process of mapping the informative priors in (5) to the logit coefficients is:

$$\begin{aligned} \beta_{PLB} &= logit(p_{PBO}) \\ \beta_{ADA} &= logit(p_{ADA}) \\ \gamma_{ADA} &= \beta_{ADA} - \beta_{PBO} \\ \gamma_{BKZ} &= \beta_{BKZ} - \beta_{PB0} \end{aligned} \quad (8)$$

8.2.2.2 Building the prior

From the PIONEER II data, despite the strength of any interaction within treatment groups between antibiotic use and Baseline Hurley Stage being unknown, it is valid to pool the data across Baseline Hurley Stages within these groups to obtain estimates of the overall group response rates and their variability, and to use this information as the basis for building informative priors for the PBO and ADA treatment groups. In this respect, and from the reported event counts and subject numbers within each Baseline Hurley Stage group, we obtained the following Beta distributions that could be used to describe the observed group response rates in that study:

$$p_{PBO} \sim \beta(45,118)$$

$$p_{ADA} \sim \beta(48,67)$$

After heavily discounting the above priors we obtain the priors given in Section 8.2.2.1.

If further information (independent to this study) regarding the PBO and/or ADA treatment group response rates comes to light then the priors for this analysis may be updated prior to study unblinding, and details of this will be supplied in an amendment to this SAP or the CSR. In this case the priors will still be based upon historic data only.

Note that the informative priors for the PBO and ADA treatment group response rates may be updated based upon any observed differences in the proportions of Baseline Hurley Stage II and III subjects as long as no response rate data from the current study, blinded or unblinded, are taken into consideration. For example, if updated, these priors may be fitted by Baseline Hurley Stage if the relative proportions of Baseline Hurley Stage II and III subjects within the current study differ substantially from those reported in the historical data. In this case the priors will still be based only on historic response data, but using the Hurley-stage specific information rather than the average information.

8.2.2.3 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.

8.2.2.4 Presentation of results from 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 1 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 PBO and BKZ group response rates, and of the difference in response rates between the treatment groups, will be summarized using means, medians, SDs, 95% credible intervals, and 95% highest posterior density intervals. The posterior probability that the difference in response rates between the PBO and BKZ groups is greater than zero (ie, BKZ demonstrates a greater response compared with PBO) will also be presented.

8.2.3 Supportive and sensitivity analyses of the primary efficacy variable

The primary efficacy analysis set is the PPS, however the primary analysis and all supportive and sensitivity analyses of the primary efficacy variable will be repeated using the FAS.

The comparison of the BKZ treatment group to the ADA treatment group is described specifically in this section as a sensitivity analysis, however, it must be noted that the results from the BKZ versus ADA comparison will be presented for completeness in each of the sensitivity analyses described in Section 8.2.3.1 to 8.2.3.10 below.

Further analyses of the primary efficacy variable may be performed adjusting for other Baseline covariates. Results from any additional analyses will not be used as a substitute for the planned analyses, but may be used as supplemental information for the CSR. Should such analysis become necessary they will be defined in a further amendment to this SAP.

All sensitivity analyses of the primary efficacy variable are specified in the sections below.

8.2.3.1 Choice of analysis set

The primary analysis of the primary efficacy variable will be repeated for the FAS in order to evaluate the effect that any IPDs may have on the results of the analysis. The analysis and results presentation as specified in Section 8.2.2.4 will be applied

8.2.3.2 Comparison of bimekizumab and adalimumab

In addition to the primary comparison of the BKZ and PBO treatment groups, the comparison between BKZ and ADA in HiSCR at Week 12 will be performed using the methods as described above in Section 8.2.2.4 for the primary efficacy analysis. A 60% credible interval will be presented for the BKZ versus ADA comparison. This will be an informal comparison to assess whether any improvements in response rate observed with BKZ (over PBO) are comparable to the current standard of care (ADA); it will not be part of the evaluation of PoC for this study.

8.2.3.3 Observed case Bayesian analysis

The primary Bayesian analysis of HiSCR at Week 12 will be performed and reported as specified in Section 8.2.2.4 without applying NRI for missing response data, ie, observed case analysis.

8.2.3.4 Vague prior distribution

A sensitivity analysis will be performed to investigate the influence of the choice of prior distribution on the results of the primary analysis. This will be performed and results reported using the same methodology as described in Section 8.2.2.4 but assuming vague priors for all Beta distributed parameters of the model (ie, Beta [0.5, 0.5]).

8.2.3.5 Concomitant antibiotic use – Bayesian analysis

The primary Bayesian analysis of HiSCR at Week 12 will be repeated including treatment group, Baseline Hurley Stage, concomitant antibiotic use at study entry (coded as a binary [yes/no] variable) as an additional covariate, and a concomitant antibiotic use × Baseline Hurley Stage interaction term.

The results of this analysis will be presented as specified in Section 8.2.2.4.

8.2.3.6 Rescue therapy – Bayesian analysis

The primary Bayesian analysis of HiSCR at Week 12 will be performed and reported as specified in Section 8.2.2.4 and will treat subjects who receive specific rescue medication prior to the Week 12 visit as non-responders, from the point of receiving the rescue medication.

8.2.3.7 Frequentist analysis – NRI

An estimate of the treatment differences (BKZ versus PBO and BKZ versus ADA) at Week 12 will be obtained from a longitudinal generalized estimating equation (GEE) model for a binary outcome, controlling for Baseline Hurley Stage and including treatment, visit and a treatment by visit (week) interaction term. An unstructured covariance matrix will be used for this analysis. In the event that the model does not converge, alternate covariance structures will be considered.

For subjects who withdraw from the study, any missing responses from visits after the withdrawal will be imputed as non-response. For any missing response data at any visit that is not due to a withdrawal, the response will not be imputed and the GEE model will analyze these as "missing at random".

At each visit, the number and percentage of responders in each treatment group, together with the estimated treatment differences (odds ratios [BKZ versus PBO and BKZ versus ADA]), will be reported, including 95% CI for the odds ratios.

8.2.3.8 Observed case frequentist analysis

The same model as in Section 8.2.3.7 will be applied to analyze the data. Missing data will not be handled using NRI (ie, observed cases analysis).

The results of the analysis will be presented as described for the analysis in Section 8.2.3.7.

8.2.3.9 Concomitant antibiotic use – Frequentist analysis

An estimate of the treatment differences (BKZ versus PBO and BKZ versus ADA) at Week 12 will be obtained from a longitudinal GEE model for a binary outcome, controlling for Baseline Hurley Stage and including concomitant antibiotic use at study entry (coded as a binary [yes/no] variable), treatment, visit, treatment by visit (week) interaction term and antibiotic use by Baseline Hurley Stage interaction terms. Missing response data will not be handled using NRI (ie, observed cases analysis).

An unstructured covariance matrix will be used for this analysis. In the event that the model does not converge, alternate covariance structures will be considered.

The results of the analysis will be presented as for the analysis in Section 8.2.3.7.

8.2.3.10 Rescue therapy – Frequentist analysis

Subjects who receive specific rescue medication will be considered as non-responders from the point of receiving the rescue medication and the model and presentation of data as specified in Section 8.2.3.7 will be applied. Missing response data will not be handled using NRI (ie, observed cases analysis).

8.2.4 HiSCR depth

As an additional exploratory analysis of HiSCR, the algorithm applied to derive response at each visit (as described in Section 8.2.1) will be adapted to evaluate varying depths of HiSCR during

the Treatment Period. A modified HiSCR will be calculated using different criteria for %ΔAN, namely 25%, 75% and 90%, in order to derive a response at each visit. In the summaries of HiSCR depth, the primary efficacy variable will be referred to as HiSCR and the other depth variables referred to as modified (m) mHiSCR25, mHiSCR75 and mHiSCR90.

8.2.5 Presentation of HiSCR

HiSCR will be listed together with the component efficacy data as specified in Section 8.3.1. This listing will include the different HiSCR depth variables (i.e. mHiSCR25, HiSCR, mHiSCR75 and mHiSCR90).

The number and percentage of subjects who achieved HiSCR will be tabulated overall by treatment group and visit, and also by Baseline Hurley Stage and by concomitant antibiotic use at study entry; and also by Baseline Hurley Stage and rescue medication use at any time during the study. This summary will be presented for the PPS and will include the 95% CI for the percentage of subjects with an HiSCR, calculated using a Wilson approximation. The summary will be presented based on NRI and repeated for observed cases. For the observed cases analysis, the denominator for the percentage calculations will be the number of subjects with non-missing HiSCR data in the PPS for each treatment group, visit and Baseline Hurley Stage or concomitant antibiotic use at study entry (as applicable).

The number and percentage of subjects who achieved mHiSCR25, mHiSCR75 and mHiSCR90, together with 95% confidence intervals, will be presented overall and by Baseline Hurley Stage, and by treatment group and visit based on observed cases for the PPS.

All tabulations described in this section will be repeated for the FAS. For the observed cases analysis, the denominator for the percentage calculations will be the number of subjects in the FAS for each treatment group, visit and Baseline Hurley Stage or concomitant antibiotic use at study entry (as applicable) with non-missing HiSCR data. Listings will be presented for the SS.

A figure will be presented showing the percentage of subjects who achieved HiSCR, together with 95% CIs (calculated using a Wilson approximation), over time. All treatment groups will be overlaid on the same plot. The plot will be generated based on NRI and observed cases for the PPS and the FAS. These plots will be repeated for the percentage of subjects who achieved the different HiSCR depths; mHiSCR25, mHiSCR75 and mHiSCR90.

Additionally, for the PPS and FAS, based on NRI and observed cases, the following bar charts will be presented showing the percentage of subjects who achieved HiSCR during the Treatment Period:

- by treatment group and visit
- by treatment group, visit and Baseline Hurley Stage
- by treatment group and Baseline Hurley Stage at Week 12
- by treatment group, Baseline Hurley Stage and concomitant antibiotic use at study entry, at Week 12
- by treatment group, Baseline Hurley Stage and rescue medication use at any time during the study, at Week 12

In addition to the above bar charts, a stacked bar chart displaying whether the criteria are met (yes/no) for each of the component data used in the calculation of HiSCR (%ΔAN, abscess count and draining tunnel [fistula/sinus tract] count) will be generated by treatment group and visit for the PPS and FAS. This graph will summarize the proportion of subjects in each of the 8 different yes/no binary responses at each visit (2 x 2 x 2 response combination) at each visit.

The following bar charts will be presented for the different HiSCR depths based on observed cases for the PPS and FAS:

- by treatment group and visit
- by treatment group, visit and Baseline Hurley Stage

8.3 Exploratory efficacy variables

Observed values and changes (including percentage changes, where applicable) from Baseline in the continuous exploratory efficacy variables will be listed and summary statistics will be presented by treatment group and visit for the FAS. Changes from Baseline in the categorical exploratory efficacy variables will be summarized in shift tables presented by treatment group and each post-Baseline visit.

The exploratory efficacy variables are listed in Section 2.2.1.2 and defined in more detail in each section below.

8.3.1 Lesion count data

Lesion count data will be listed by treatment group and anatomical region and will show region-specific Hurley Stage and worst overall Hurley Stage for each subject and visit. The total count for each type of lesion, across all anatomical regions at each visit will be listed separately. For the total abscess count and total draining tunnel (fistula/sinus tract) count, the change from Baseline will be listed; for the AN count, the percentage change from Baseline will be listed.

For each anatomical region, the longest distance between 2 relevant lesions (or the diameter of the single lesion if applicable) will be presented, and whether the specified regions were separated by normal skin.

At each visit, lesion counts will be summarized by anatomical region and lesion type. The following lesion types will be summarized:

- Abscesses
- Inflammatory nodules
- Abscesses + inflammatory nodules (AN)
- Draining tunnels (fistulas/sinus tracts)
- Non-draining tunnels (fistulas/sinus tracts)
- Non-inflammatory nodules
- Hypertrophic scars

Total lesion counts (ie, the total across all anatomical regions) will be summarized by visit and lesion type, treatment group, overall and by Baseline Hurley Stage. Summaries will also be

presented for the change and percentage change from Baseline in lesion counts by anatomical region and total lesion counts by lesion type.

8.3.1.1 AN count

The number and percentage of subjects with an AN count of 0, 1, or 2 at Week 12 will be presented by treatment group, overall and by Baseline Hurley Stage. The denominator for the percentage calculations will be the number of subjects in each treatment group in the FAS with non-missing data and, in the summary by Baseline Hurley Stage, percentages will be calculated based on the number of subjects in each treatment group and each level of Baseline Hurley Stage.

In addition, the following plots of AN count and %ΔAN over time will be generated for the PPS and FAS:

- A violin plot of AN count (with box plot inlay) by treatment group and visit
- Box plot of %ΔAN by treatment group and visit
- Jitter plot of Baseline AN count by treatment group and country
- Jitter plot of Baseline AN count by treatment group and site

8.3.1.2 Disease flare

Disease flare is defined when at least a 25% increase in AN count with a minimum increase of 2 AN relative to Baseline is observed. A subject's disease flare status (yes/no) will be determined at each visit using these criteria and will listed with the other lesion count assessment data in the data listings.

Disease flare status will be summarized by treatment group and visit using frequencies and percentages. The denominator for the percentage calculations will be the number of subjects with non-missing data in each treatment group. The number and percentage of subjects who experienced a disease flare at any time during the Treatment Period will also be presented.

A bar chart showing the percentage of subjects with disease flare will be presented by treatment group and visit. A bar chart summarizing the percentage of subjects who experienced a disease flare at any time during the Treatment Period will also be presented by treatment group.

In addition, for each subject, the number of visits at which a disease flare was observed during the Treatment Period will be calculated and summarized by treatment group. Summary statistics will be presented together with a box plot showing the distribution of these data and a histogram summarizing the number of subjects who experienced 0, 1, 2, 3, 4, 5 and 6 visits during the Treatment Period with a disease flare status of 'yes'.

All summaries of disease flare status will be presented for the FAS.

8.3.2 International Hidradenitis Suppurativa Severity Scoring System

The IHS4 is a validated scoring tool to dynamically assess HS severity to be used both in real-life and clinical trials settings (Zouboulis et al, 2017). The determination of the IHS4 score requires counting the inflammatory nodules, abscesses and draining tunnels (fistulas/sinus tracts), making it straightforward to apply in both research and clinical practice and easy to use in conjunction with the HiSCR.

$$IHS4 = (number\ of\ inflammatory\ nodules \times 1) + (number\ of\ abscesses \times 2) + (number\ of\ draining\ tunnels\ (fistulas/sinus\ tracts) \times 4)$$

The IHS4 score will be derived based on observed component total lesion count data; in the case of missing component data, the IHS4 score will be missing.

The observed IHS4 score, together with the change from Baseline and percentage change from Baseline will be listed with the component total lesion count data, by subject and visit, and summarized by treatment group, Baseline Hurley Stage and visit.

Mean observed IHS4 score and percentage changes from Baseline (including 95% CIs) over time will be presented graphically by treatment group.

8.3.3 Patient's Global Assessment of Skin Pain

The PGA of Skin Pain is a unidimensional numeric rating scale (NRS) that allows for rapid (often 1 item) measures of pain that can be administered multiple times with minimal administrative burden. The NRS consists of scores from 0 to 10 with 0 indicating "no skin pain" and 10 indicating "worst skin pain imaginable".

The PGA of Skin Pain NRS will be administered verbally. The pain will be described as "pain at its worst during the last 24 hours" and "pain on average in the last 24 hours".

Both parameters (pain at its worst and pain on average over the last 24 hours) will be listed by treatment group and subject at each visit, including observed values, changes from Baseline and percentage changes from Baseline. For pain at its worst during the last 24 hours, subjects with a 30% reduction and at least 1-unit reduction from Baseline at Week 12 (for example a change from 7 at Baseline to 3 at Week 12) will be flagged.

Summary statistics will be presented for observed values, changes and percentage changes from Baseline in PGA of Skin Pain (pain at its worst and pain on average over the last 24 hours) by treatment group at each visit, overall and by Baseline Hurley Stage.

In addition, the number and percentage of subjects with at least a 30% reduction and at least 1-unit reduction from Baseline in the PGA of Skin Pain (based on the subject's worst pain over the previous 24 hours) at Week 12, will be presented by treatment group. The denominator for the percentage calculations will be the number of subjects in the FAS for each treatment group. This summary will also be presented by Baseline Hurley Stage and percentages will be calculated based on the number of subjects in the FAS for each treatment group and each level of Baseline Hurley Stage.

8.3.4 Modified Sartorius Score

The Sartorius scale was created as a more detailed and dynamic HS severity scale and was modified in order to further develop and simplify this assessment for the clinical setting (Sartorius et al, 2003; Sartorius et al, 2009; Sartorius et al, 2010). The Modified Sartorius Score was the first disease specific instrument for dynamically measuring clinical severity.

The Modified Sartorius Score includes an assessment of the anatomical regions involved, the numbers and types of lesions in each anatomical region, the longest distance between 2 relevant lesions within an anatomical region (or size of a single lesion), and whether all relevant lesions are separated by normal skin (yes or no). Table 14–2 shows the scoring for each criterion.

A Regional Modified Sartorius Score will be derived for each reported anatomical region as:

$$Regional\ Modified\ Sartorius\ Score\ =$$
$$Region\ +4\times DT\ +\ 2\times IN\ +1\ \times\ AB\ +\ 1\times HYS\ +1\ \times\ NIN\ +1\ \times\ NDT\ +\ Distance\ +\ Separate$$

where

- Region = 3 if any lesion is present in the anatomical region, 0 otherwise
- DT = number of draining tunnels (fistulas/sinus tracts) in the region
- IN = number of inflammatory nodules in the region
- AB = number of abscesses in the region
- HYS = number of hypertrophic scars in the region
- NIN = number of non-inflammatory nodules in the region
- NDT = number of non-draining tunnels (fistulas/sinus tracts) in the region
- Distance =
  - 0 if no lesions are present in the anatomical region
  - 2 if longest distance between 2 relevant lesions or diameter of a single lesion is < 50 mm
  - 4 if longest distance between 2 relevant lesions or diameter of a single lesion is ≥ 50 mm and ≤ 100 mm
  - 8 if longest distance between 2 relevant lesions or diameter of a single lesion is > 100 mm
- Separate = 0 if all relevant lesions are clearly separated by normal skin in the region, 6 otherwise

For the purpose of the Regional Modified Sartorius Score derivation, the left and right sides of the body (if applicable) will be treated as 1 anatomical region (ie, the lesion assessment data from both sides will be pooled). For the Region, Distance and Separate variables defined above, the following will apply:

- Region: a value of 3 will be applied if at least 1 lesion is present in either the left or right sides of an anatomical region; if 1 or more lesions are present in both left and right sides of the body, this will still be scored as 3 for the region overall
- Distance: a score will be determined for each side of the body according to the scoring system above and the maximum of these scores will be used
- Separate:
  - if both left and right sides have relevant lesions separated by normal skin then this will be scored as 0;
  - if only one of the left and right sides have relevant lesions separated by normal skin, then this will be scored as 6

- if both left and right sides have no relevant lesions separated by normal skin then this will be scored as 6

The overall Modified Sartorius Score for a subject will then be calculated by summing the Regional Modified Sartorius Scores across all anatomical regions.

The Regional Modified Sartorius Score together with the overall Modified Sartorius Score will be listed by treatment group, subject and visit, including the change from Baseline for the overall Modified Sartorius Score. The observed overall Modified Sartorius Score and the changes from Baseline in the overall Modified Sartorius Score will be summarized by treatment group and visit. This summary will also be produced by Baseline Hurley Stage.

8.3.5 Hidradenitis Suppurativa Physician's Global Assessment – Derived

The HS-Physician's Global Assessment is a validated 6-point scale that is used to measure improvement in inflammatory nodules, abscesses, and draining fistulas (Kimball et al, 2012; Zouboulis et al, 2015). The HS-Physician's Global Assessment scale is defined in Section 14.1.

The scale ranges from "Clear" (no inflammatory or non-inflammatory nodules) to "Very severe" (more than 5 abscesses or draining fistulas). The gradings will be derived on a subject-level basis (ie, across all anatomical regions) based on the following rules:

- Clear:
  - number of abscesses = 0;
  - number of draining tunnels (fistulas/sinus tracts) = 0;
  - number of inflammatory nodules = 0;
  - number of non-inflammatory nodules = 0;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of hypertrophic scars – no limit;
- Minimal:
  - number of abscesses = 0;
  - number of draining tunnels (fistulas/sinus tracts) = 0;
  - number of inflammatory nodules = 0;
  - number of non-inflammatory nodules ≥ 1;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of hypertrophic scars – no limit;
- Mild:
  - number of abscesses = 0;
  - number of draining tunnels (fistulas/sinus tracts) = 0;
  - number of inflammatory nodules ≥ 1 and ≤ 4;
  - number of non-inflammatory nodules – no limit;

  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of hypertrophic scars – no limit;

  OR

  - sum of number of abscesses and number of draining tunnels (fistulas/sinus tracts) = 1;
  - number of inflammatory nodules = 0;
  - number of non-inflammatory nodules – no limit;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of hypertrophic scars – no limit;

- Moderate:
  - number of abscesses = 0;
  - number of draining tunnels (fistulas/sinus tracts) = 0;
  - number of inflammatory nodules ≥ 5;
  - number of non-inflammatory nodules – no limit;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of hypertrophic scars – no limit;

  OR

  - sum number of abscesses and number of draining tunnels (fistulas/sinus tracts) = 1;
  - number of inflammatory nodules ≥1;
  - number of non-inflammatory nodules – no limit;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of hypertrophic scars – no limit;

  OR

  - sum of number of abscesses and number of draining tunnels (fistulas/sinus tracts) ≥ 2 and ≤ 5;
  - number of inflammatory nodules ≤ 10;
  - number of non-inflammatory nodules – no limit;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of hypertrophic scars – no limit;

- Severe:
  - sum of number of abscesses and number of draining tunnels (fistulas/sinus tracts) ≥ 2 and ≤ 5;
  - number of inflammatory nodules > 10;

- number of non-inflammatory nodules – no limit;
- number of non-draining tunnels (fistulas/sinus tracts) – no limit;
- number of hypertrophic scars – no limit;

- Very Severe:
  - sum of number of and number of draining tunnels (fistulas/sinus tracts) > 5;
  - number of inflammatory nodule – no limit;
  - number of non-inflammatory nodules – no limit;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of hypertrophic scars – no limit

The HS-Physician's Global Assessment will be listed by treatment group and subject at each visit. The number and percentage of subjects at each level of the assessment scale (Clear, Minimal, Mild, Moderate, Severe and Very severe) will be presented by treatment group, visit, overall and by Baseline Hurley Stage. The denominator for the percentages will be based on the number of subjects in the given treatment group, Baseline Hurley Stage and visit for which there are no missing data for the HS-Physician's Global Assessment.

Shift tables for the changes from Baseline in this scale will be presented for each post-Baseline visit by treatment group, overall and by Baseline Hurley Stage. The denominator for the percentages will be based on the number of subjects in the given treatment group, Baseline Hurley Stage and visit for which there is no missing data for the change from Baseline in HS-Physician's Global Assessment

8.3.6 Dermatology Life Quality Index

The DLQI is a questionnaire designed for use in adult subjects with inflammatory skin diseases and has been used in patients with HS (Finlay, 1998; Esmann et al, 2010; Basra et al, 2012). The DLQI is a skin disease specific questionnaire aimed at the evaluation of how symptoms and treatment affect subjects' health related quality of life (QoL).

This instrument asks subjects about symptoms and feelings, daily activities, leisure, work and school, personal relationships and treatment. The total DLQI score ranges from 0 to 30 with higher scores indicating lower health related QoL. The complete questionnaire is shown in Section 14.3 together with the scoring and derivation of the total DLQI.

Each question on the DLQI will be listed by treatment group and subject for each visit. The listing will also include the total DLQI score at each visit together with the changes from Baseline in the total DLQI score. Summary statistics will be presented for the observed total DLQI score at each visit together with the changes from Baseline in the total DLQI score by treatment group, overall and by Baseline Hurley Stage.

The number and percentage of subjects with a total DLQI score in each of the following categories will be summarized at each visit, by treatment group, overall and by Baseline Hurley Stage:

- Total DLQI score of 0 or 1

- Change from Baseline in the total DLQI score of -4 or below

For the summary of the change from Baseline in the total DLQI score of -4 or below, the denominator for the percentages will be the number of subjects with non-missing data and who had a total DLQI score of 4 or higher at Baseline.

8.3.7 European Quality of Life - 5 Dimensions-3 Level Questionnaire

The EQ-5D-3L questionnaire is a standardized instrument developed as a measure of health-related QoL that can be used in a wide range of health conditions and treatments. The EQ-5D-3L consists of a descriptive system and the EQ-5D-3L VAS.

The descriptive system comprises 5 questions on the following dimensions: mobility, self care, usual activities, pain/discomfort and psychological status (anxiety/depression) with 3 possible answers for each item (1=no problem, 2=moderate problem, 3=severe problem). The scores on these 5 dimensions (domain scores) represent a health profile for the subject and can be converted to a single summary index with a maximum score of 1 using a pre-defined algorithm and table of scores. The maximum index score of 1 indicates the best health state, by contrast with the scores of individual questions from the descriptive system, where higher scores indicate more severe or frequent problems.

The EQ-5D-3L VAS records the subject's self-rated health on a vertical 0 to 100 scale, with 100 indicating the best health status.

Section 14.4 shows the full questionnaire together with the derivation of the EQ-5D-3L index score (Dolan 1997).

The 5 domain scores, derived EQ-5D-3L index score and VAS scores will be listed by treatment group for each subject at each visit, including changes from Baseline for the EQ-5D-3L index score and VAS score.

The number and percentage of subjects giving each categorical response (1, 2 or 3) to the questions on the EQ-5D-3L domains will be summarized by visit and treatment group, overall and by Baseline Hurley Stage. Shifts from Baseline in the domain scores at each post-Baseline visit will also be presented by treatment group, overall and by Baseline Hurley Stage. Any questions with a missing or ambiguous response (eg, 2 boxes ticked for a single dimension) will be counted in the category "Missing" in the categorical summaries of the EQ-5D-3L domain scores.

The numeric domain scores and changes from Baseline will also be summarized by treatment group, overall and by Baseline Hurley Stage as continuous variables using summary statistics at each visit. Any questions with an ambiguous response (eg, 2 boxes ticked for a single dimension) will be regarded as missing for the tabulations.

The derived EQ-5D-3L index score and VAS scores will be summarized using descriptive statistics together with changes from Baseline, by visit, treatment group, overall and by Baseline Hurley Stage.

8.3.8 Hurley Stage

The Hurley Stage is a severity classification for HS that was developed in 1989 and is widely used for the determination of the severity of HS (Hurley, 1989). Stage I HS is the most common

(68% of patients), while Stage II occurs in about 28% of patients and Stage III only occurs in about 4% of patients.

The Hurley Stage is defined by the following criteria and will be assessed for each anatomical region as part of the lesion count assessment (Section 8.1):

- Stage I: Abscess formation, single or multiple, without sinus tracts and cicatrization (scarring)
- Stage II: Recurrent abscesses with tract formation and cicatrization, single or multiple, widely separated lesions
- Stage III: Diffuse or near diffuse involvement, or multiple interconnected tracts and abscesses across the entire area

The Hurley Stage for each anatomical region will be recorded directly on the eCRF and the overall worst Hurley Stage (ie, the highest Hurley Stage) will be derived automatically in the database across all reported anatomical regions. Hurley Stage will be listed by treatment group, anatomical region and subject at each visit.

A categorical summary of the overall worst Hurley Stage at each visit will be presented by treatment group. The changes from Baseline in the overall worst Hurley Stage will be summarized in shift tables by treatment group and post-Baseline visit. The denominator for the percentages will be the number of subjects in the FAS with non-missing overall worst Hurley Stage in the respective treatment group and visit.

8.3.9 Erythema assessment

The erythema assessment will be recorded as specified in Section 8.1.

Erythema assessment results will be listed by treatment group, subject, visit and anatomical region. In addition, the overall worst erythema (highest score) will be derived automatically in the database across all reported anatomical regions and will be listed by treatment group, subject and visit.

A categorical summary of the overall worst erythema scores, showing the number and percentage of subjects within each category at each visit, will be presented by treatment group, overall and by Baseline Hurley Stage. The denominator for the percentages will be based on the number of subjects in the FAS in the given treatment group, Baseline Hurley Stage and visit with non-missing data for worst erythema.

The changes from Baseline in the degree of erythema will be summarized in shift tables by treatment group, post-Baseline visit, overall and by Baseline Hurley Stage. The denominator for the percentages will be based on the number of subjects in the FAS in the given treatment group, Baseline Hurley Stage and visit with non-missing data for the Baseline worst erythema score at Baseline and at the respective visits.

8.3.10 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.

The full HADS questionnaire is shown in Section 14.5 together with the derivation of the total HADS-A and HADS-D scores.

Scores for each item on the HADS-A and HADS-D will be listed at each visit by treatment group and subject. 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 and Baseline Hurley Stage for the observed total HADS-A and HADS-D scores together with changes from Baseline at each visit.

8.4 High Sensitivity C-Reactive Protein

Concentrations of hsCRP 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 = hsCRP at post-Baseline / hsCRP at Baseline visit

Observed values, changes from Baseline and ratio to Baseline will be summarized by treatment group, Baseline Hurley Stage and visit. For the observed values and changes from Baseline, summary statistics will include n, arithmetic mean, SD, median, Q1, Q3, minimum and maximum. For the ratio to Baseline, summary statistics will include n, geometric mean, geoCV, median, Q1, Q3, minimum and maximum.

The rules for handling any values that are BLQ are described in Section 4.2.1.

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 trough plasma concentrations of BKZ only. Any additional PK analyses will be performed only if required.

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

Individual blood sampling times and concentrations of BKZ will be listed for the SS and will include the actual sampling time in days relative to the first dose of IMP. 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 of BKZ will also be presented on both linear and semi-logarithmic scale respectively with and without the corresponding lower and upper limit of the 95% CI. All semi-logarithmic plots will include the 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
- 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 should be left blank if the SD (or equivalently, the geoCV) is 0
- The geoCV (%) will be calculated using the following formula where SD is the standard deviation from the log transformed data

$$geoCV = 100 \times \sqrt{e^{SD^2} - 1}$$

9.2 Pharmacodynamics

Blood sampling for exploratory PD analyses will be taken for the evaluation of the following:

- Flow cytometry by fluorescence activated cell sorting analysis which might include, but will not be limited to: cluster of differentiation (CD) CD3, CD19, CD4, CD8, and CD69.
- Candidate biomarkers might include, but will not be limited to: IL-17A/IL-17F pathway signaling and HS (eg, IL-17A, IL-17F, IL-23, IL-6, TNF, DC-STAMP, and circulating osteoclast precursors).

The results for all exploratory PD variables will be listed by treatment group, subject and visit, together with the changes from Baseline, based on the SS. Observed values and changes from Baseline will be summarized by treatment group based on the PD-PPS.

Figures of mean and mean changes from Baseline (including 95% CIs for the mean and mean change from Baseline) will be presented with separate plots for each variable and all 3 treatment groups overlaid on the same plot.

The rules for handling any values that are BLQ or ALQ are described in Section 4.2.3.

10 IMMUNOLOGICAL ANALYSES

All immunological analyses will be performed for the PK-PPS. In this study, this will comprise an analysis of anti-BKZ antibody data only which will be listed and summarized at each scheduled visit, and the rate of anti-BKZ antibody positive subjects will be calculated.

The results of the anti-BKZ antibody analysis will be listed by treatment group and visit based on the SS. This will include the screening assay, confirmatory assay, titre (if applicable) and confirmation of whether or not the sample contains BKZ concentration levels above the drug tolerance.

A cut point will be determined by the bioanalytical laboratory that will be used to determine the status of anti-BKZ antibody as above the cut point (ACP) or below the cut point (BCP). For any anti-BKZ antibody 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 anti-BKZ antibody titre will be reported.

The following definitions will be applied:

- An anti-BKZ antibody status of positive will be concluded for any subject with an anti-BKZ antibody level that is ACP and CP at any time point
- An anti-BKZ antibody status of negative will be concluded for any subject with an anti-BKZ antibody level that is either BCP or ACP and NCP at any time point
- A subject will be classified as having anti-BKZ antibody positivity at Baseline if the Day 1, predose result is ACP and CP
- A subject will be classified as having treatment-induced anti-BKZ antibody 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 these criteria 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 anti-BKZ antibody status is negative (this includes subjects who have positive [ACP and CP] results at Baseline)

The number and percentage of subjects with a positive and negative anti-BKZ antibody status will be summarized at each visit and overall. This summary will exclude any samples with BKZ concentrations confirmed to be above the drug tolerance.

In addition, the first occurrence of treatment-induced anti-BKZ antibody positivity (based on the definitions above) will be summarized at each post-Baseline visit. This tabulation will present the number and percentage 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). This summary will exclude any samples with BKZ concentrations confirmed to be above the drug tolerance.

For all tabulations, percentages will be calculated based on the number of subjects with non-missing data and anti-BKZ antibody levels that are not flagged as containing BKZ concentration levels above the drug tolerance at each respective visit.

A separate listing will be presented showing the BKZ concentrations and anti-BKZ antibody measurements in the same output in adjacent columns, based on the PK-PPS. The listing will include the BKZ concentration, anti-BKZ antibody 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 will be presented displaying the anti-BKZ antibody titre and BKZ concentrations overlaid on the same figure. The figure will also show the timing and dose of each administration of BKZ received during the study. The anti-BKZ antibody data will be plotted using a semi-logarithmic scale. This figure will exclude any samples with BKZ concentrations confirmed to be above the drug tolerance.

The rules for handling values that are BLQ in the BKZ concentration data are described in Section 4.2.2. For the anti-BKZ antibody 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, time and location of injection) will be listed.

The percentage of injections administered will be included in the listing based on the number of planned injections. For subjects completing the study this will be 17 injections. The number of planned injections for subjects who withdraw early from the study will be reduced according to the number of injections planned up to EW visit.

All IMP administration details (including date, time and location of injection) will be listed. The percentage of injections administered will be included in the listing based on the number of planned injections. For subjects completing the study this will be 17 injections. The number of planned injections for subjects who withdraw early from the study will be reduced according to the number of injections planned up to EW visit.

The duration of exposure to BKZ and PBO (days) will be calculated as follows:

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

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

$$Exposure = (Date\ of\ Last\ Dose - Date\ of\ First\ Dose) + 100\ days$$

The number of days added to the BKZ and ADA exposure calculations are 5 times the half-life of each drug, respectively: 5 x 28 days for BKZ and 5 x 20 days for ADA. The number of days added to PBO is set to 140 days in order for PBO exposure to be comparable to that of BKZ.

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

Duration of exposure will be listed for each subject and summarized using descriptive statistics.

11.2 Adverse events

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. 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 the first dose of IMP or later than the defined cutoff for TEAEs.

Handling of missing dates/times for classification of AEs as TEAEs is described in Section 4.2.4.

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. 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.

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
- Incidence of TEAEs, stratified by overall anti-BKZ antibody status (positive/negative) as defined in Section 10
- Incidence of serious TEAEs
- 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)

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, opportunistic infections, fungal 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 14.9.

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, 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 the ES, except where otherwise noted.

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

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 for 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) Version 2.0 (Woodworth et al, 2007). 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 14–7 and Table 14–8 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 1 TEMA result for each subject.

Table 11–1: Laboratory measurements

Hematology	Chemistry	Urinalysis	Serology
Basophils	Bicarbonate	Albumin	HIV
Eosinophils	Calcium	Bacteria	HLA-B27
Lymphocytes	Chloride	Crystals	Hepatitis B
Atypical lymphocytes	Creatinine	Glucose	Hepatitis C
Monocytes	hsCRP[a]	pH	
Neutrophils[b]	Magnesium	RBC	
Hematocrit	Potassium	WBC	
Hemoglobin[b]	Sodium	Urine dipstick for pregnancy testing[c]	
MCH	Glucose	Urine drug screen	
MCHC	BUN		

Table 11–1: Laboratory measurements

Hematology	Chemistry	Urinalysis	Serology
MCV	ALP[b]		
Platelet count	AST[b]		
RBC count [b]	ALT[b]		
WBC count[b]	GGT[b]		
	Total bilirubin[d]		
	Uric acid		
	LDH		
	Total cholesterol[b]		
	Serum pregnancy testing		
	Serum FSH[e]		

ALP=alkaline phosphatase; ALT=alanine aminotransferase; AST=aspartate aminotransferase; BUN=blood urea nitrogen; FSH=follicle-stimulating hormone; GGT=gamma glutamyltransferase; hsCRP= high sensitivity C-reactive protein; HLA=human leukocyte antigen; HIV=human immunodeficiency virus; LDH=lactate dehydrogenase; MCH=mean corpuscular hemoglobin; MCHC=mean corpuscular hemoglobin concentration; MCV=mean corpuscular volume; RBC=red blood cell; WBC=white blood cell.

[a] To avoid duplication, hsCRP will be included in the listings and tabulations for the efficacy data and will not be included in the safety laboratory outputs

[b] Shift tables of the change from Baseline in these variables will be presented.

[c] Pregnancy testing will be done in all women of childbearing potential and will consist of serum testing at Screening and at the Safety Follow Up Visit and urine testing at all other visits.

[d] If total bilirubin is >ULN (upper limit of normal), a direct bilirubin estimation (%) will be performed.

[e] The serum FSH test (at Screening only) should only be performed on postmenopausal females who have been postmenopausal for ≥1 year and last menstrual cycle occurred <2 years ago.

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 postmenopausal women)

The results of the TB questionnaire 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 14–9.

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:

- Systolic and diastolic blood pressure
- Pulse rate
- Temperature (oral, axillary, or otic)

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 14–10. Body weight will also be included in the listing.

Descriptive statistics will be reported for all vital sign measurements (with the exception of body weight). 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 percentage calculations will be the number of subjects with a non-missing measurement for the variable at the specific visit.

11.4.2 Electrocardiograms

Standard 12-lead ECG recordings will be obtained after 10 minutes of rest in the supine position. The following ECG variables will be reported together with the Investigator's interpretation of the ECG profile:

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

The results of all ECG variables 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 values, changes and percentage changes from Baseline will be summarized for each variable by treatment group and visit.

The following cut points in QTcF (raw data and change from Baseline) will be summarized categorically by treatment group (number and percentage of subjects) and visit. The denominator

for the percentage calculations 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 QTcF:

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

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

11.4.3 Other safety variables

11.4.3.1 Physical examination

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

11.4.3.2 Columbia-Suicide Severity Rating Scale

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 for each subject and visit.

11.4.3.3 Chest X-ray and tuberculosis questionnaire

The results of the chest X-ray (or computed axial tomography scan if applicable), TB questionnaire and TB test will be listed for each subject.

12 OTHER ANALYSES

A listing of comments will be provided, if applicable. This will be based on the RS.

13 REFERENCES

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

Basra MK, Chowdhury MM, Smith EV, Freemantle N, Piquet V. A review of the use of the dermatology life quality index as a criterion in clinical guidelines and health technology assessments in psoriasis and chronic hand eczema. Dermatol Clin. 2012;30(2):237 44.

Basra MK, Salek MS, Camilleri L, Sturkey R, Finlay AY. Determining the Minimal Clinically Important Difference and Responsiveness of the Dermatology Life Quality Index (DLQI): Further Data. Dermatology. 2015;230(1):27 33.

Dauden E, Griffiths CE, Ortonne JP, Kragballe K, Molta CT, Robertson D, et al. Improvements in patient reported outcomes in moderate to severe psoriasis patients receiving continuous or paused etanercept treatment over 54 weeks: the CRYSTEL study. J Eur Acad Dermatol Venereol. 2009;23(12):1374 82.

Dolan P. Modeling valuations for EuroQol health states. Med Care 1997; 35(11):1095-108.

Esmann S, Dufour DN, Jemec GB. Questionnaire based diagnosis of hidradenitis suppurativa: specificity, sensitivity and positive predictive value of specific diagnostic questions. Br J Dermatol. 2010;163(1):102 6.

Finlay AY. Quality of life assessments in dermatology. Semin Cutan Med Surg. 1998; 17(4):291 6.

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

Hurley H. Axillary hyperhidrosis, apocrine bromhidrosis, hidradenitis suppurativa, and familial benign pemphigus: surgical approach In: Dermatologic surgery. Roenigh R, Roenigh H, editors. Marcel Dekker. 1989:729 39.

Kimball AB, Kerdel F, Adams D, Mrowietz U, Gelfand JM, Gniadecki R. Adalimumab for the treatment of moderate to severe Hidradenitis suppurativa: a parallel randomized trial. Ann Intern Med. 2012;157(12):846 55.

Kimball AB, Jemec GB, Yang M, Kageliery A, Signorovitch JE, Okun MM, et al. Assessing the validity, responsiveness and meaningfulness of the Hidradenitis Suppurativa Clinical Response (HiSCR) as the clinical endpoint for hidradenitis suppurativa treatment. Br J Dermatol. 2014;171(6):1434 42.

Kimball AB, Okun MM, Williams DA, Gottlieb AB, Papp KA, Zouboulis CC, et al. Two Phase 3 Trials of Adalimumab for Hidradenitis Suppurativa. N Engl J Med. 2016a;375(5):422 34.

Kimball AB, Sobell JM, Zouboulis CC, Gu Y, Williams DA, Sundaram M, et al. HiSCR (Hidradenitis Suppurativa Clinical Response): a novel clinical endpoint to evaluate therapeutic outcomes in patients with hidradenitis suppurativa from the placebo controlled portion of a phase 2 adalimumab study. J Eur Acad Dermatol Venereol. 2016b;30(6):989 94.

Langley RG, Feldman SR, Han C, Schenkel B, Szapary P, Hsu MC, et al. Ustekinumab significantly improves symptoms of anxiety, depression, and skin related quality of life in

patients with moderate to severe psoriasis: Results from a randomized, double blind, placebo controlled phase III trial. J Am Acad Dermatol. 2010;63(3):457 65.

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.

Sartorius K, Lapins J, Emtestam L, Jemec GB. Suggestions for uniform outcome variables when reporting treatment effects in hidradenitis suppurativa. Br J Dermatol. 2003;149(1):211 3.

Sartorius K, Emtestam L, Jemec GB, Lapins J. Objective scoring of hidradenitis suppurativa reflecting the role of tobacco smoking and obesity. Br J Dermatol. 2009;161(4):831 9.

Sartorius K, Killasli H, Heilborn J, Jemec GB, Lapins J, Emtestam L. Interobserver variability of clinical scores in hidradenitis suppurativa is low. Br J Dermatol. 2010;162(6):1261 8.

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

UpToDate. 2011. Modified Hidradenitis Suppurativa Score (HSS). [ONLINE] Available at: http://cursoenarm.net/UPTODATE/contents/mobipreview.htm?36/33/37400. [Accessed 27 November 2017].

Zouboulis CC, Desai N, Emtestam L, Hunger RE, Loannides D, Juhász I, et al. European S1 guideline for the treatment of hidradenitis suppurativa/acne inversa. J Eur Acad Dermatol Venereol. 2015; 29(4):619 44.

Zouboulis CC, Kyrgidis A, Jemec GBE, Bechara FG, Giamarellos Bourboulis EJ, Ingram JR, et al. Development and validation of IHS4, a novel dynamic scoring system to assess hidradenitis suppurativa/acne inversa severity. British Journal of Dermatology. 2017 June 21.

14 APPENDICES

14.1 HS-Physician's Global Assessment 6-point scale

The HS-Physician's Global Assessment 6-point scale is scored according to the specifics in Table 14–1.

Table 14–1: HS-Physician's Global Assessment 6-point scale

Score	Rating	Description
0	Clear	No abscesses, no draining tunnels (fistulas/sinus tracts), no nodules
1	Minimal	No abscesses, no draining tunnels (fistulas/sinus tracts), no inflammatory nodules, presence of non-inflammatory nodules
2	Mild	No abscesses or draining tunnels (fistulas/sinus tracts), and less than 5 inflammatory nodules, or Single abscess or draining tunnel (fistula/sinus tract), and no inflammatory nodules
3	Moderate	No abscesses or draining tunnels (fistulas/sinus tracts), and at least 5 inflammatory nodules, or Single abscess or draining tunnel (fistula/sinus tract) in the presence of inflammatory nodules, or Between 2 and 5 abscesses or draining tunnels (fistulas/sinus tracts) with or without inflammatory nodules, up to 10
4	Severe	Between 2 and 5 abscesses and draining tunnels (fistulas/sinus tracts), with inflammatory nodules that are greater than 10
5	Very severe	More than 5 abscesses or draining tunnels (fistulas/sinus tracts)

14.2 Modified Sartorius Score

The scoring system for the presence, counts and measurements of relevant lesions by type are given in Table 14–2.

Table 14–2: Modified Sartorius Score

Characteristics of HS lesions	Points
Anatomical region	
Any	3
None	0
Types of lesions	
Fistula[a]	4
Nodule[b]	2
Abscess	1
Scar[c]	1

Table 14–2: Modified Sartorius Score

Characteristics of HS lesions	Points
Other [d]	1
Total area involved[e]	
< 5 cm	2
5 to 10 cm	4
> 10 cm	8
Are all lesions clearly separated by normal skin?[f]	
Yes	0
No	6

HS=hidradenitis suppurativa
Taken from http://cursoenarm net/UPTODATE/contents/mobipreview htm?36/33/37400 and adapted from:

1. Sartorius, K, Lapins, J, Emtestam, L, Jemec, GB. Suggestions for uniform outcome variables when reporting treatment effects in hidradenitis suppurativa. Br J Dermatol 2003; 149:211.
2. Mahmoud, BH, Tierney, E, Hexsel, CL, Pui, J, Ozog, DM, Hamzavi, IH. Prospective controlled clinical and histopathologic study of hidradenitis suppurativa treated with the long-pulsed neodymium:yttrium-aluminium-garnet laser. J Am Acad Dermatol. 2010; 62:637-45.

[a] Fistula includes draining tunnels (fistulas/sinus tracts)
[b] Nodule refers to inflammatory nodules
[c] Scar refers to hypertrophic scars
[d] Other includes non-inflammatory nodules and non-draining tunnels (fistulas/sinus tracts).
[e] Area is defined as the longest distance between 2 relevant lesions (ie, abscess, draining tunnel (fistula/sinus tract) or inflammatory nodule). If there is only 1 lesion it is the measurement of the single lesion diameter.
[f] This refers to all relevant lesions.

14.3 Dermatology Life Quality Index

The 10 questions and scoring of the DLQI are presented in Table 14–3.

Table 14–3: Dermatology Life Quality Index

Item Nr.	Question	Answers
1	Over the last week, how itchy, sore painful or stinging has your skin been?	Very much
		A lot
		A little
		Not at all
2	Over the last week, how embarrassed or self-conscious have you been because of your skin?	Very much
		A lot
		A little
		Not at all

Table 14–3: Dermatology Life Quality Index

Item Nr.	Question	Answers
3	Over the last week, how much has your skin interfered with you going shopping or looking after your home or garden?	Very much
		A lot
		A little
		Not at all
		Not relevant
4	Over the last week, how much has your skin influenced the clothes you wear?	Very much
		A lot
		A little
		Not at all
		Not relevant
5	Over the last week, how much has your skin affected any social or leisure activities?	Very much
		A lot
		A little
		Not at all
		Not relevant
6	Over the last week, how much has your skin made it difficult for you to do any sport?	Very much
		A lot
		A little
		Not at all
		Not relevant
7	Over the last week, has your skin prevented you from working or studying? If "No", over the last week how much has your skin been a problem at work or studying?	Yes
		No
		Not relevant
		A lot
		A little
		Not at all
8	Over the last week, how much has your skin created problems with your partner or any of your close friends or relatives?	Very much
		A lot
		A little
		Not at all

Table 14–3: Dermatology Life Quality Index

Item Nr.	Question	Answers
		Not relevant
9	Over the last week, how much has your skin caused any sexual difficulties?	Very much
		A lot
		A little
		Not at all
		Not relevant
10	Over the last week, how much of a problem has the treatment for your skin been, for example by making your home messy, or by taking up time?	Very much
		A lot
		A little
		Not at all
		Not relevant

The following rules apply to the scores for the individual questions and the calculation of the total DLQI score:

- Individual responses will be scored as follows, with separate rules for Question 7 below:
  - Very much = 3
  - A lot = 2
  - A little = 1
  - Not at all = 0
- If 1 question is left unanswered this question will be scored as 3 and the scores for each question will be summed to give the total DLQI score.
- If two or more questions are left unanswered a total DLQI will not be calculated.
- If Question 7 is answered "Yes" this will be scored as 3. If question 7 is answered "No" the scoring will be as follows:
  - If "A lot", "A little" or "Not at all" is ticked this will be scored as 2, 1 or 0 respectively.
  - If the second half of the questions is left incomplete, the score will be set to 3.
- If a question is answered as "Not relevant" then this will be scored as 0.
- If two or more response options are ticked, the response option with the highest score will be recorded.
- If there is a response between two tick boxes, the lower of the two score options will be recorded.

- The total DLQI score will be calculated by summing the scores over all the questions.

14.4 European Quality of Life-5 Dimensions-3 Level Questionnaire

The EQ-5D-3L comprises a descriptive aspect and the EQ-5D-3L VAS. Table 14–4 presents the descriptive aspect of the EQ-5D-3L, with the 5 domains (1 question each) of the EQ-5D-3L questionnaire. There are 3 levels of response (a score of 1, 2 or 3) to each question.

Table 14–4: European Quality of Life-5 Dimensions-3 Level Questionnaire

The EQ-5D-3L index score (the EuroQoL Group, 2015) will be derived using the UK time trade off (TTO) value set as shown in Table 14–5.

The EQ-5D-3L index score will be derived as:

$$EQ\text{-}5D\text{-}3L\ Index\ Score = 1 - (0.081 + M02 + MO3 + SC2 + SC3 + UA2 + UA3 + PD2 + PD3 + AD2 + AD3 + N3)$$

where the scores associated with each abbreviation above are explained in Table 14–5 below.

Table 14–5: UK TTO Value Set

Condition	Score if condition is true; else zero
Full Health (all questions have score 1)	1
At least one 2 or 3	0.081
At least one 3 (N3)	0.269

Table 14–5: UK TTO Value Set

Condition	Score if condition is true; else zero
Mobility = 2 (MO2)	0.069
Mobility = 3 (MO3)	0.314
Self-care = 2 (SC2)	0.104
Self-care = 3 (SC3)	0.214
Usual activities = 2 (UA2)	0.036
Usual activities = 3 (UA3)	0.094
Pain/discomfort = 2 (PD2)	0.123
Pain/discomfort = 3 (PD3)	0.386
Anxiety/depression = 2 (AD2)	0.071
Anxiety/depression = 3 (AD3)	0.236

14.5 Hospital Anxiety and Depression Scale

The individual items on the HADS-A and HADS-D scales are listed in Table 14–6.

Table 14–6: Hospital Anxiety and Depression Scale

[redacted]

Each item is scored from 0 to 3, with higher scores indicating more severe anxiety or depression. The total score for HADS-A and HADS-D will be derived by summing the scores over the 7 items and will range from 0 to 21.

In case of missing data, the following rules will be applied:

- 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 HADS-A or HADS-D score will not be calculated.

14.6 Treatment-emergent markedly abnormal laboratory values

The criteria for identifying TEMA laboratory values are provided for hematology in Table 14–7 and for chemistry in Table 14–8.

Table 14–7: 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 14–8: 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
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
Glucose (mmol/L)	RCTC	<1.67	1.67 to <2.22	>13.89 to 27.78	>27.78

Table 14–8: 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
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 glutamyltransferase; 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 14–9. Subjects will be counted in all applicable categories for the tabulations ie, a subject with ≥5 x ULN in AST will also be counted in the ≥3 x ULN in AST category, the ≥3 x ULN in AST or ALT category and the ≥5x ULN in AST or ALT category.

Table 14–9: 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
≥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

Table 14–9: Definition of treatment-emergent liver function values

Criterion
≥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 14–10.

Table 14–10: 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.
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.

14.9 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

14.9.1 Major cardiovascular events

Major adverse cardiovascular events 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'

14.9.2 Serious infections, including opportunistic infections and tuberculosis

14.9.2.1 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.

14.9.2.2 Fungal infections

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

14.9.2.3 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, Lower Level Term, PT, AE start date, AE end date, seriousness, severity, relationship to study medication, 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.

The timing and frequency of Step 2 should be outlined and agreed to by the study team at the beginning of the study.

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 conclusion of the study.

14.9.3 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)'

14.9.4 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.

14.9.5 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.

14.9.6 Inflammatory bowel disease

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

14.9.7 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 1 hypersensitivity reaction or at least 1 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 1 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 1 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 to be 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 14.9.7.1 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.

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

– **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

14.9.8 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 (included 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 PDILI cases)

14.10 Treatment group assignment for the TFLs

The treatment group labels and tabulation requirements for each data type are displayed in Table 14–11.

Table 14–11: Treatment group assignment for the TFLs

TFL group	Placebo	Adalimumab	Bimekizumab	All Subjects
Subject disposition	X	X	X	X
Protocol deviations	X	X	X	X
Demographics	X	X	X	X
HS history and Baseline characteristics	X	X	X	X
Medical history	X	X	X	X
Prior/concomitant medications	X	X	X	X
AEs	X	X	X	
Safety data (including vital signs, ECG and laboratory tests)	X	X	X	
Pharmacodynamic data	X	X	X	
Efficacy data	X	X	X	
Pharmacokinetic data			X	
Anti-BKZ antibody			X	

AEs=adverse events; ECG=electrocardiogram.

15 AMENDMENTS TO THE SAP

15.1.1 Rationale for the amendment

Various updates were suggested prior to and during the interim analysis in late 2018.

15.1.2 List of changes

Change #1

List of Abbreviations

The following abbreviations have been added:

mHiSCR25	modified Hidradenitis Suppurativa Clinical Response – 25%
mHiSCR75	modified Hidradenitis Suppurativa Clinical Response – 75%
mHiSCR90	modified Hidradenitis Suppurativa Clinical Response – 90%
PDILI	potential drug-induced liver injury
PoC	proof of concept

Change #2

Introduction (Section 1), 2nd and 3rd paragraphs

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

- Protocol Amendment 1, dated 08 March 2018
- Primary Analysis Document, dated 19 September 2017

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 IA SAP will be amended accordingly. In addition, if analysis definitions have to be modified or updated prior to the cut-off for IA, an IA SAP amendment will be required. However, if any additional analyses are required to supplement the planned analyses after the cut-off, these analyses will be described in a separate analysis plan.

Has been changed to:

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

- Final protocol amendment 1, dated 08 March 2018

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 must be modified or updated prior to database lock, a further 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.

Change #3

Determination of sample size (Section 2.4), 2nd paragraph

Additionally, the study has an exploratory objective to informally compare BKZ against the current standard of care for HS (ADA) using the 60% credible interval for the treatment difference.

Has been changed to:

Additionally, the study has an exploratory objective to informally compare bimekizumab against the current standard of care for HS (adalimumab) using the 60% credible interval for the treatment difference.

Change #4

Determination of sample size (Section 2.4), 5th paragraph

These simulations were based on Bayesian augmented control models in which the published summary data were used to construct informative priors for the PBO and ADA treatment arms in the current study. Based on historic data, the Bayesian model assumes an informative prior distribution for the PBO group HiSCR rate to be β (5.52,14.48). This prior distribution contributes an approximate effective sample size of 20 subjects for the PBO treatment group. Similarly, the Bayesian model assumes an informative prior distribution for the ADA group HiSCR rate to be β (11.78, 8.22), which, with the parametrization in the model, is equivalent to an ADA response rate of 58.9%. This prior distribution contributes an approximate effective sample size of 20 subjects for the ADA group. For the BKZ HiSCR rate, a vague prior probability distribution N (0,1E4) was assumed.

Has been changed to:

These simulations were based on Bayesian augmented control models in which the published summary data were used to construct informative priors for the placebo and adalimumab treatment arms in the current study. Based on historic data, the Bayesian model assumes an informative prior distribution for the placebo group HiSCR rate to be β (5.52,14.48). This prior distribution contributes an approximate effective sample size of 20 subjects for the placebo treatment group. Similarly, the Bayesian model assumes an informative prior distribution for the adalimumab group HiSCR rate to be β (11.78, 8.22) which, with the parametrization in the model, is equivalent to an adalimumab response rate of 58.9%. This prior distribution contributes an approximate effective sample size of 20 subjects for the adalimumab group. For the bimekizumab HiSCR rate, a vague prior probability distribution N (0,1E4) was assumed.

Change #5

Protocol Deviations (Section 3.4), 3rd paragraph

At least 3 data evaluation meetings (DEMs) will be performed for this study at the following time points:

- Prior to the formal unblinded interim analysis planned after approximately 45 subjects have completed 4 weeks of the study.
- Prior to the informal unblinded interim analysis planned after the last subject has completed 12 weeks of the study or has discontinued early.
- Prior to the final analysis after all subjects have completed the SFU period and data have been verified/coded/entered into the database

Has been changed to:

At least 3 data evaluation meetings (DEMs) will be performed for this study at the following time points:

- Prior to the informal unblinded interim analysis planned after approximately 45 subjects have completed 4 weeks of the study.
- Prior to the informal unblinded interim analysis planned after the last subject (excluding 2 subjects from site 101 in Australia who entered the study outside of the original planned recruitment period) has completed 12 weeks of the study or has discontinued early.
- Prior to the final analysis after all subjects have completed the SFU period and data have been verified/coded/entered into the database

Change #6

Changes to protocol-defined analyses (Section 3.9)

This final bullet has been added:-

- Based on the second interim analysis for this study, additional exploratory efficacy variables based on the lesion count assessment data, HiSCR depth and disease flare status, have been defined in this SAP amendment. Section 8.2.4 includes details on the derivation of different HiSCR depths and Section 8.2.5 describes the planned summaries of this additional variable. Section 8.3.1.2 defines disease flare and includes details on planned summaries of disease flare status and the number of visits at which a subject experienced a disease flare during the Treatment Period.

Change #7

Dates (Section 4.2.4)

This final paragraph has been added:-

Partially missing dates will be imputed as described above and the duration of each AE will be calculated in days as follows:

$$Duration\ of\ AE\ =\ (Date\ of\ Outcome\ -\ Date\ of\ Onset\ )+\ 1$$

Change #8

Calculation rules for duration of adverse events (Table 4-1)

This table has been removed.

Change #9

Interim analysis (Section 4.5.2)

A formal unblinded non-binding interim analysis will be conducted when approximately 45 subjects have completed 4 weeks of the study. During this interim analysis, the sample size assumptions will be checked and the final sample size may be modified if the sample size assumptions at study planning are no longer accurate. Also, at this interim analysis, the study may be stopped for futility. Sample size re-estimation criteria and methods along with futility

rules and operating characteristics will be detailed in the Interim Analysis SAP. The futility rule will be chosen such that the overall power of the study remains high.

An informal, unblinded interim analysis will be performed when the last randomized subject for this study has completed the Week 12 Visit at the end of the Treatment Period or this subject's participation has ended prematurely. The purpose of this informal interim analysis is for the Sponsor key personnel to review the results of the primary efficacy analysis and safety outcomes to facilitate additional Clinical Planning or Portfolio Management decisions. Details on the summaries and analyses to be presented at this informal interim analysis will be included in the Interim Analysis SAP.

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

Has been changed to:

This study will include two informal unblinded interim analyses, the first after approximately 45 subjects have completed 4 weeks of the study and the second when the last randomized subject for this study has completed the Week 12 Visit at the end of the Treatment Period or this subject's participation has ended prematurely. This second planned interim analysis excludes the 2 subjects from site 101 who entered the study outside of the recruitment period. The purpose of these interim analyses is for sponsor key personnel to review results from the primary efficacy analysis and a subset of the analyses of exploratory efficacy and safety outcomes to facilitate additional Clinical Planning or Portfolio Management decisions. Neither interim analysis will lead to any formal decision to alter or terminate the trial.

Details on the summaries and analyses to be presented at these informal interim analyses are included in the Interim Analysis SAP.

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

Change #10

Hidradenitis suppurativa history (Section 6.5)

Detailed information of each subject's HS history will be collected at Screening and will include, but will not be limited to, the following:

- Date of Initial Diagnosis of HS
- Date HS Symptoms First Started
- The date, onset, and duration of past treatments for HS
- The date of any prior surgeries for HS

The HS history will be listed based on the RS.

Has been changed to:

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

Change #11

Prior and concomitant medications (Section 6.6), 3rd and 4th paragraphs

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 HS medications
- Concomitant antibiotic medications at study entry
- Rescue therapy medication during study

All HS medications will be confirmed at the DEMs prior to unblinding.

Has been changed to:

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 HS medications
- Concomitant antibiotic medications at study entry
- Rescue medication during study

All HS medications, antibiotics 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.

Concomitant antibiotic medication at study entry is defined as any antibiotic that was ongoing or started on the Screening visit.

Change #12

Bayesian Modelling (Section 8.2.2.1), 9th paragraph

To protect against severe forms of study to study variation and/or miss-matches between Baseline Hurley Stage subject proportions and effect sizes, a prior data conflict test will be conducted using the above priors as detailed in Section 8.2.2.3 and vague priors for the PBO and ADA treatment groups will be used if either of the two null hypotheses of no prior data conflict are rejected at the 95% confidence level. The prior data conflict test also provides some protection against model miss-specification – for example if there turns out to be a Baseline Hurley Stage by treatment group interaction as the PIONEER II data suggest.

Has been changed to:

To protect against severe forms of study to study variation and/or mismatches between Baseline Hurley Stage subject proportions and effect sizes, 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. Investigating any prior data conflict also provides some protection against model misspecification – for example if there turns out to be a Baseline Hurley Stage by treatment group interaction as the PIONEER II data suggest.

Change #13

Building the Prior (S8.2.2.2), final paragraph

If further information (independent to this study) regarding the PBO group response comes to light then the priors for this analysis may be updated prior to study unblinding, and details of this will be supplied in an amendment to this SAP or the CSR. In this case the priors will still be based upon historic data only.

Has been changed to:

If further information (independent to this study) regarding the PBO and/or ADA treatment group response rates comes to light then the priors for this analysis may be updated prior to study unblinding, and details of this will be supplied in an amendment to this SAP or the CSR. In this case the priors will still be based upon historic data only.

Note that the informative priors for the PBO and ADA treatment group response rates may be updated based upon any observed differences in the proportions of Baseline Hurley Stage II and III subjects as long as no response rate data from the current study, blinded or unblinded, are taken into consideration. For example, if updated, these priors may be fitted by Baseline Hurley Stage if the relative proportions of Baseline Hurley Stage II and III subjects within the current study differ substantially from those reported in the h storical data. In this case the priors will still be based only on historic response data, but using the Hurley-stage specific information rather than the average information.

Change #14

Prior data conflict (Section 8.2.2.3)

For each of the PBO and ADA groups separately, the prior data conflict test will consist of comparing the observed minimal sufficient statistic for the Binomial distribution (the number of HS events) to the prior predictive distribution of this minimal sufficient statistic.

Has been changed to:

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.

Change #15

Supportive and sensitivity analyses of the primary efficacy variable (Section 8.2.3), 1st paragraph

The primary efficacy analysis set is the PPS, however the primary analysis and all supportive and sensitivity analyses of the primary efficacy variable will be repeated using the FAS. Additionally, the comparison of the BKZ treatment group to the ADA treatment group is reported under the following sections.

Has been changed to:

The primary efficacy analysis set is the PPS, however the primary analysis and all supportive and sensitivity analyses of the primary efficacy variable will be repeated using the FAS.

The comparison of the BKZ treatment group to the ADA treatment group is described specifically in this section as a sensitivity analysis, however, it must be noted that the results

from the BKZ versus ADA comparison will be presented for completeness in each of the sensitivity analyses described in Section 8.2.3.1 to 8.2.3.10 below.

Change #16

Comparison of bimekizumab and adalimumab (Section 8.2.3.2)

In addition to the primary comparison of the BKZ and PBO treatment groups, the comparison between BKZ and ADA in HiSCR at Week 12 will be performed using the methods as described above in Section 8.2.2.4 for the primary efficacy analysis. This will be an informal comparison to assess whether any improvements in response rate observed with BKZ (over PBO) are comparable to the current standard of care (ADA).

Has been changed to:

In addition to the primary comparison of the BKZ and PBO treatment groups, the comparison between BKZ and ADA in HiSCR at Week 12 will be performed using the methods as described above in Section 8.2.2.4 for the primary efficacy analysis. A 60% credible interval will be presented for the BKZ versus ADA comparison. This will be an informal comparison to assess whether any improvements in response rate observed with BKZ (over PBO) are comparable to the current standard of care (ADA); it will not be part of the evaluation of PoC for this study.

Change #17

Frequentist analysis – NRI (Section 8.2.3.7), 1st paragraph

An estimate of the treatment difference (BKZ versus PBO) at Week 12 will be obtained from a longitudinal generalized estimating equation (GEE) model for a binary outcome, controlling for Baseline Hurley Stage and including treatment, visit and a treatment by visit (week) interaction term. An unstructured covariance matrix will be used for this analysis. In the event that the model does not converge, alternate covariance structures will be considered.

Has been changed to:

An estimate of the treatment differences (BKZ versus PBO and BKZ versus ADA) at Week 12 will be obtained from a longitudinal generalized estimating equation (GEE) model for a binary outcome, controlling for Baseline Hurley Stage and including treatment, visit and a treatment by visit (week) interaction term. An unstructured covariance matrix will be used for this analysis. In the event that the model does not converge, alternate covariance structures will be considered.

Change #18

Concomitant antibiotic use – Frequentist analysis (Section 8.2.3.9), 1st paragraph

An estimate of the treatment difference (BKZ versus PBO) at Week 12 will be obtained from a longitudinal GEE model for a binary outcome, controlling for Baseline Hurley Stage and including concomitant antibiotic use at study entry (coded as a binary [yes/no] variable), treatment, visit, treatment by visit (week) interaction term and antibiotic use by Baseline Hurley Stage interaction terms. Missing response data will not be handled using NRI (ie, observed cases analysis).

Has been changed to:

An estimate of the treatment differences (BKZ versus PBO and BKZ versus ADA) at Week 12 will be obtained from a longitudinal GEE model for a binary outcome, controlling for Baseline

Hurley Stage and including concomitant antibiotic use at study entry (coded as a binary [yes/no] variable), treatment, visit, treatment by visit (week) interaction term and antibiotic use by Baseline Hurley Stage interaction terms. Missing response data will not be handled using NRI (ie, observed cases analysis).

Change #19

HiSCR depth (Section 8.2.4)

This section has been added:

As an additional exploratory analysis of HiSCR, the algorithm applied to derive response at each visit (as described in Section 8.2.1) will be adapted to evaluate varying depths of HiSCR during the Treatment Period. A modified HiSCR will be calculated using different criteria for %ΔAN, namely 25%, 75% and 90%, in order to derive a response at each visit. In the summaries of HiSCR depth, the primary efficacy variable will be referred to as HiSCR and the other depth variables referred to as modified (m)HiSCR25, mHiSCR75 and mHiSCR90.

Change #20

Presentation of HiSCR (Section 8.2.5)

The HiSCR will be listed together with the component efficacy data as specified in Section 8.3.1.

The number and percentage of subjects who achieved HiSCR will be tabulated by treatment group and visit, based on the PPS. A summary will also be presented by Baseline Hurley Stage. All summaries will include the 95% CI for the percentage of subjects with an HiSCR, calculated using a Wilson approximation. The summary will be presented based on NRI and repeated for observed cases. For the observed cases analysis, the denominator for the percentage calculations will be the number of subjects with non-missing HiSCR data in the PPS for each treatment group, visit and Baseline Hurley Stage (as applicable).

All tabulations described in this section will be repeated for the FAS. For the observed cases analysis, the denominator for the percentage calculations will be the number of subjects in the FAS for each treatment group, visit and Baseline Hurley Stage (as applicable) with non-missing HiSCR data. Listings will be presented for the SS.

A figure will be presented showing the percentage of subjects who achieved HiSCR, together with 95% CIs, over time. All treatment groups will be overlaid on the same plot. The plot will be generated based on the PPS and the FAS.

Additionally, the following bar charts will be presented showing the percentage of subjects who achieved HiSCR at Week 12:

- by treatment group and Baseline Hurley Stage
- by treatment group, Baseline Hurley Stage and antibiotic use at study entry
- by treatment group, Baseline Hurley Stage and rescue therapy use at any time during the study

Has been changed to:

HiSCR will be listed together with the component efficacy data as specified in Section 8.3.1. This listing will include the different HiSCR depth variables (i.e. mHiSCR25, HiSCR, mHiSCR75 and mHiSCR90).

The number and percentage of subjects who achieved HiSCR will be tabulated overall by treatment group and visit, and also by Baseline Hurley Stage and by concomitant antibiotic use at study entry; and also by Baseline Hurley Stage and rescue medication use at any time during the study. This summary will be presented for the PPS and will include the 95% CI for the percentage of subjects with an HiSCR, calculated using a Wilson approximation. The summary will be presented based on NRI and repeated for observed cases. For the observed cases analysis, the denominator for the percentage calculations will be the number of subjects with non-missing HiSCR data in the PPS for each treatment group, visit and Baseline Hurley Stage or concomitant antibiotic use at study entry (as applicable).

The number and percentage of subjects who achieved mHiSCR25, mHiSCR75 and mHiSCR90, together with 95% confidence intervals, will be presented overall and by Baseline Hurley Stage, and by treatment group and visit based on observed cases for the PPS.

All tabulations described in this section will be repeated for the FAS. For the observed cases analysis, the denominator for the percentage calculations will be the number of subjects in the FAS for each treatment group, visit and Baseline Hurley Stage or concomitant antibiotic use at study entry (as applicable) with non-missing HiSCR data. Listings will be presented for the SS.

A figure will be presented showing the percentage of subjects who achieved HiSCR, together with 95% CIs (calculated using a Wilson approximation), over time. All treatment groups will be overlaid on the same plot. The plot will be generated based on NRI and observed cases for the PPS and the FAS. These plots will be repeated for the percentage of subjects who achieved the different HiSCR depths; mHiSCR25, mHiSCR75 and mHiSCR90.

Additionally, for the PPS and FAS, based on NRI and observed cases, the following bar charts will be presented showing the percentage of subjects who achieved HiSCR during the Treatment Period:

- by treatment group and visit
- by treatment group, visit and Baseline Hurley Stage
- by treatment group and Baseline Hurley Stage at Week 12
- by treatment group, Baseline Hurley Stage and concomitant antibiotic use at study entry, at Week 12
- by treatment group, Baseline Hurley Stage and rescue medication use at any time during the study, at Week 12

In addition to the above bar charts, a stacked bar chart displaying whether the criteria are met (yes/no) for each of the component data used in the calculation of HiSCR (%ΔAN, abscess count and draining tunnel [fistula/sinus tract] count) will be generated by treatment group and visit for the PPS and FAS. This graph will summarize the proportion of subjects in each of the 8 different yes/no binary responses at each visit (2 x 2 x 2 response combination) at each visit.

The following bar charts will be presented for the different HiSCR depths based on observed cases for the PPS and FAS:

- by treatment group and visit
- by treatment group, visit and Baseline Hurley Stage

Change #21

Lesion Count Data (Section 8.3.1)

Lesion count data will be listed by treatment group and anatomical region and will show region-specific Hurley Stage and worst overall Hurley Stage for each subject and visit. The total count for each type of lesion, across all anatomical regions at each visit will be listed separately. For the total abscess count and total draining tunnel (fistula/sinus tract) count, the change from Baseline will be listed; for the AN count, the percentage change from Baseline will be listed.

For each anatomical region, the longest distance between 2 relevant lesions (or the diameter of the single lesion if applicable) will be presented, and whether the specified regions were separated by normal skin.

Total lesion counts (ie, the total across all anatomical regions) will be summarized by lesion type, treatment group, overall and by Baseline Hurley Stage. The following lesion types will be summarized:

- Abscesses
- Inflammatory nodules
- Abscesses + inflammatory nodules (AN)
- Draining tunnels (fistulas/sinus tracts)
- Non-draining tunnels (fistulas/sinus tracts)
- Non-inflammatory nodules
- Hypertrophic scars

Summaries will also be presented for the change from Baseline in the total abscess count and total draining tunnel (fistula/sinus tract) count, and for the percentage change from Baseline in the AN count.

The number and percentage of subjects with an AN count of 0, 1, or 2 at Week 12 will be presented by treatment group, overall and by Baseline Hurley Stage. The denominator for the percentage calculations will be the number of subjects in each treatment group in the FAS with non-missing data and, in the summary by Baseline Hurley Stage, percentages will be calculated based on the number of subjects in each treatment group and each level of Baseline Hurley Stage.

Has been changed to:

Lesion count data will be listed by treatment group and anatomical region and will show region-specific Hurley Stage and worst overall Hurley Stage for each subject and visit. The total count for each type of lesion, across all anatomical regions at each visit will be listed separately.

For the total abscess count and total draining tunnel (fistula/sinus tract) count, the change from Baseline will be listed; for the AN count, the percentage change from Baseline will be listed.

For each anatomical region, the longest distance between 2 relevant lesions (or the diameter of the single lesion if applicable) will be presented, and whether the specified regions were separated by normal skin.

At each visit, lesion counts will be summarized by anatomical region and lesion type. The following lesion types will be summarized:

- Abscesses
- Inflammatory nodules
- Abscesses + inflammatory nodules (AN)
- Draining tunnels (fistulas/sinus tracts)
- Non-draining tunnels (fistulas/sinus tracts)
- Non-inflammatory nodules
- Hypertrophic scars

Total lesion counts (ie, the total across all anatomical regions) will be summarized by visit and lesion type, treatment group, overall and by Baseline Hurley Stage. Summaries will also be presented for the change and percentage change from Baseline in lesion counts by anatomical region and total lesion counts by lesion type.

Change #22

AN count (Section 8.3.1.1)

This section has been added:

The number and percentage of subjects with an AN count of 0, 1, or 2 at Week 12 will be presented by treatment group, overall and by Baseline Hurley Stage. The denominator for the percentage calculations will be the number of subjects in each treatment group in the FAS with non-missing data and, in the summary by Baseline Hurley Stage, percentages will be calculated based on the number of subjects in each treatment group and each level of Baseline Hurley Stage.

In addition, the following plots of AN count and %ΔAN over time will be generated for the PPS and FAS:

- A violin plot of AN count (with box plot inlay) by treatment group and visit
- Box plot of %ΔAN by treatment group and visit
- Jitter plot of Baseline AN count by treatment group and country
- Jitter plot of Baseline AN count by treatment group and site

Change #23

Disease flare (Section 8.3.1.2)

This section has been added:

Disease flare is defined when at least a 25% increase in AN count with a minimum increase of 2 AN relative to Baseline is observed. A subject's disease flare status (yes/no) will be determined at each visit using these criteria and will listed with the other lesion count assessment data in the data listings.

Disease flare status will be summarized by treatment group and visit using frequencies and percentages. The denominator for the percentage calculations will be the number of subjects with non-missing data in each treatment group. The number and percentage of subjects who experienced a disease flare at any time during the Treatment Period will also be presented.

A bar chart showing the percentage of subjects with disease flare will be presented by treatment group and visit. A bar chart summarizing the percentage of subjects who experienced a disease flare at any time during the Treatment Period will also be presented by treatment group.

In addition, for each subject, the number of visits at which a disease flare was observed during the Treatment Period will be calculated and summarized by treatment group. Summary statistics will be presented together with a box plot showing the distribution of these data and a histogram summarizing the number of subjects who experienced 0, 1, 2, 3, 4, 5 and 6 visits during the Treatment Period with a disease flare status of 'yes'.

All summaries of disease flare status will be presented for the FAS

Change #24

Definition of TEMA values for chemistry (Table 14–8)

This row has been added:-

Variable (SI units)	**Source**	**Markedly abnormal definition**			
		Low Grade 4	**Low Grade 3**	**High Grade 3**	**High Grade 4**
Glucose (mmol/L)	RCTC	<1.67	1.67 to <2.22	>13.89 to 27.78	>27.78

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

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

Approval Signatures

Name: hs0001-sap-amendment1

Version: 2.0

Document Number: CLIN-000130784

Title: HS0001 Statistical Analysis Plan Amendment 1

Approved Date: 22 Feb 2019

Document Approvals	
Approval Verdict: Approved	Name: [redacted] Capacity: Clinical Date of Signature: 19-Feb-2019 19:45:21 GMT+0000
Approval Verdict: Approved	Name: [redacted] Capacity: Clinical Date of Signature: 22-Feb-2019 09:49:41 GMT+0000