STATISTICAL ANALYSIS PLAN

Study: PS0010

Product: BIMEKIZUMAB

A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, PARALLEL-GROUP, DOSE RANGING STUDY TO EVALUATE THE SAFETY, EFFICACY, PHARMACOKINETICS, AND PHARMACODYNAMICS OF BIMEKIZUMAB IN ADULT SUBJECTS WITH MODERATE TO SEVERE CHRONIC PLAQUE PSORIASIS

SAP/Amendment Number	Date
Final SAP	13 Oct 2016
Amendment 1	13 June 2017

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AB	antibody
AbAb	anti-bimekizumab antibody
ADR	adverse drug reaction
AE	adverse event
AESM	adverse events of special monitoring
ATC	Anatomical Therapeutic Chemical
BLQ	below level of quantification
BSA	body surface area
CMH	Cochran-Mantel-Haenszel
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DAP	Data Analysis Plan
DLQI	Dermatology Life Quality Index
DMC	Data Monitoring Committee
EAER	exposure adjusted event rate
EAIR	exposure adjusted incidence rate
ES	Enrolled Set
ECG	electrocardiogram
eCRF	electronic Case Report Form
FAS	Full Analysis Set
Full MSE	Full Missing Score Estimation
HADS	Hospital Anxiety and Depression Scale
HLGT	Higher level group term
IGA	Investigator's Global Assessment
LLOQ	lower level of quantification
LOCF	last observation carried forward
MAR	missing at random
MCID	minimal clinically important difference
MCMC	Markov Chain Monte Carlo
MCS	Mental Component Summary
MI	multiple imputation
MMRM	mixed-effects model of repeated measures
mNAPSI	modified Nail Psoriasis Severity Index
NRI	non-responder imputation
NSAID(s)	non-steroidal anti-inflammatory drug(s)
PASI	Psoriasis Area and Severity Index
PCS	Physical Component Summary
PD	pharmacodynamics
PDILI	Potential drug-induced liver injury
PGADA	Patient's Global Assessment of Disease Activity
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PK	pharmacokinetics

PPS	Per Protocol Set
PSSI	Psoriasis Scalp Severity Index
PT	preferred term
Q4W	every four weeks
QOL	quality of life
RS	Randomized Set
sc	subcutaneously
SD	standard deviation
SAP	statistical analysis plan
SF-36	Short Form 36 item Health Survey
SFU	Safety Follow-Up
SMQ	Standard MedDRA query
SOC	system organ class
SS	Safety Set
TEAE	treatment-emergent adverse event
TFLs	tables, figures, and listings
ToPAS	Toronto Psoriatic Arthritis Screening questionnaire
VAS	visual analog scale
WHO	World Health Organization

1 INTRODUCTION

This statistical analysis plan (SAP) defines the summary tables, figures, and listings (TFLs) to be generated for the clinical study report (CSR) and is based on the final protocol (01 Jun 2016), protocol amendment 1 (08 Jul 2016), and protocol amendment 2 (18 Jul 2016).

2 PROTOCOL SUMMARY

2.1 Study objectives

2.1.1 Primary objective

The primary objective of the study is to evaluate the dose response of bimekizumab administered subcutaneously (sc) every four weeks (Q4W) for 12 weeks in the treatment of subjects with moderate to severe chronic plaque psoriasis.

2.1.2 Secondary objectives

The secondary objectives of the study are to:

- Evaluate the efficacy of individual dose regimens of bimekizumab compared to placebo after 12 weeks of treatment for subjects with moderate to severe chronic plaque psoriasis
- Compare the efficacy of the bimekizumab 160mg treatment group versus the bimekizumab 160mg treatment group with a loading dose
- Assess the pharmacokinetics (PK) of bimekizumab following sc dosing Q4W
- Assess the safety, immunogenicity, and tolerability of bimekizumab
- Assess the exposure response relationship of bimekizumab as it relates to efficacy and safety
- Characterize the population PK of bimekizumab in the moderate to severe psoriatic population following repeat dose administration.

2.1.3 Other objectives

The other objectives of the study are to:

- Assess the improvement of skin-related quality of life (QOL)
- Assess the improvement of general health-related QOL
- Assess psoriatic nail disease in subjects with nail disease at Baseline
- Assess psoriatic scalp disease in subjects with scalp psoriasis at Baseline
- Assess the effect of bimekizumab on gene and protein expression, and explore the relationship between genomic, genetic, and proteomic biomarkers and disease biology, drug treatment and inflammatory and immune responses.

2.2 Study variables

2.2.1 Efficacy variables

2.2.1.1 Primary efficacy variable

The primary efficacy variable is the Psoriasis Area and Severity Index (PASI)90 response defined as a subject that achieves 90% reduction from Baseline in the PASI score at Week 12.

2.2.1.2 Secondary efficacy variables

The secondary efficacy variables are:

- Investigator's Global Assessment (IGA) response (Clear or Almost Clear with at least two category improvement from Baseline) at Week 12
- IGA response (Clear or Almost Clear with at least two category improvement from Baseline) at Week 8
- PASI90 response at Week 8
- PASI75 response at Week 12
- PASI100 response at Week 12

2.2.1.3 Other efficacy variables

The other efficacy variables detailed below will be evaluated at all scheduled visits in accordance with the Schedule of Assessments (see study protocol Table 5-1). This excludes the primary and secondary variables as specified in Section 2.2.1.1 and Section 2.2.1.2.

The other efficacy variables are:

- PASI50, PASI75, PASI90, and PASI100 response
- IGA response (Clear or Almost Clear with at least 2 category improvement from Baseline)
- Absolute and percent change from Baseline in PASI score
- Shift from Baseline in IGA score
- Change from Baseline in Dermatology Life Quality Index (DLQI)
- Percentage of subjects achieving a DLQI score of 0 or 1
- Percentage of subjects achieving a minimal clinically important difference (MCID) (improvement from Baseline of four or more) in the DLQI
- Time to PASI50 response
- Time to PASI75 response
- Time to PASI90 response
- Time to PASI100 response
- Absolute and percent change from Baseline in the body surface area (BSA) affected by psoriasis
- Change from Baseline in modified Nail Psoriasis Severity Index (mNAPSI) score (in the subgroup of subjects with psoriatic nail disease at Baseline)
- Change from Baseline in the Subject's Global Assessment of Disease Activity (PGADA) for the arthritis visual analog scale (VAS)
- Change from Baseline in the Psoriasis Scalp Severity Index (PSSI) (in the subgroup of subjects with scalp psoriasis at Baseline)

- Change from Baseline in Short Form 36 item Health Survey (SF-36) Physical Component Summary (PCS) score, and Mental Component Summary (MCS) score, and individual domains
- Change from Baseline in the Subject Symptom Diary responses (for subjects at sites in selected countries)
- Patient Global Impression of Change (PGIC)
- Change from Baseline in Patient Global Impression of Severity (PGIS)
- Change from Baseline in Hospital Anxiety and Depression Scale (HADS) HADS-A and HADS-D scores
- Percentage of subjects with scores below eight in HADS-A and HADS-D (subjects with normal scores)

2.2.2 Pharmacokinetic/pharmacodynamic variables

2.2.2.1 Pharmacokinetic variables

The PK variables are:

- Plasma concentration of bimekizumab and population PK
- Dose-exposure response characterizing the relationship between PASI score and bimekizumab concentration and dose, including EC50/ED50 and E_{max} to support dose selection for subsequent Phase 3 studies

2.2.2.2 Pharmacodynamic variable

The pharmacodynamic (PD) variable is to determine the plasma concentrations of cytokines of relevance to IL17-A/F signaling pathway and psoriasis biology, including but not limited to serum complement concentrations, mononuclear cell subtypes, and cytokines and other candidate biomarkers.

2.2.3 Pharmacogenomic variables

2.2.3.1 Genomic, genetic, and metabolite variables

Where local regulations permit, additional blood samples will be collected from consenting subjects at specific time points and stored at -80°C for up to 20 years to allow for potential, exploratory analyses of genomic, genetic, and proteomic biomarkers relevant to the inflammatory and immune response processes. The nature and format of these tentative analyses will be determined when the results of the main study are made available.

2.2.4 Immunological variable

The immunological variable is anti-bimekizumab antibody (AbAb) detection prior to and following study treatment.

2.2.5 Safety variables

Safety variables to be assessed are:

- Severity and frequency of adverse events (AEs) (including serious AEs)

- Change from Baseline in clinical laboratory values (chemistry, hematology, and urinalysis)
- Change from Baseline in vital signs
- Change from Baseline in physical examination
- Electrocardiogram (ECG) results

2.3 Study design and conduct

2.3.1 Study description

This is a Phase 2b, multicenter, randomized, double-blind, placebo-controlled, parallel-group, dose-ranging study to evaluate the safety, efficacy, PK, and PD of bimekizumab administered sc to subjects with chronic plaque psoriasis. To be eligible to participate in this study, subjects must be adults with a diagnosis of moderate to severe psoriasis (Baseline PASI ≥12 and BSA affected by psoriasis ≥10% and IGA score ≥3 [on a 5-point scale]) who are a candidate for systemic psoriasis therapy and/or phototherapy and/or chemophototherapy.

2.3.2 Study periods

This study will include three periods which are:

- Screening Period (2 to 4 weeks)
- Treatment Period (12 weeks)
- Safety Follow-Up (SFU) Period (20 weeks after the last dose of study medication for subjects not entering the extension study)

After the 12-week Treatment Period, eligible subjects will be allowed to enroll in an extension study (PS0011).

Screening Period

The Screening Period will last two weeks, but can be extended up to a total of four weeks in the event that applicable laboratory screening tests require retesting if the initial results are in error, borderline, or indeterminate. During this time, laboratory data (hematology, urine, and biochemistry tests) will be obtained, the doses of non-steroidal anti-inflammatory drug (NSAIDs) (if used to treat psoriatic arthritis), will be verified as stable.

Treatment Period

During the Treatment Period subjects will be randomized 1:1:1:1:1:1 to receive the following blinded study treatment regimens:

- bimekizumab 64mg administered sc Q4W
- bimekizumab 160mg administered sc Q4W
- bimekizumab 320mg loading dose administered sc at Baseline followed by 160mg dose administered sc Q4W
- bimekizumab 320mg administered sc Q4W
- bimekizumab 480mg administered sc Q4W
- Placebo administered sc Q4W

Approximately 320 subjects will be screened in order to have 240 subjects randomized in the study. There will be approximately 40 subjects per treatment group and study medication will be administered in the clinic at Baseline, Week 4, and Week 8. Additional non-dosing study visits will occur at Week 1, Week 2, and Week 6.

At Week 12, all subjects enrolling in the extension study (PS0011) will undergo the Week 12 study assessments and then receive their extension study dose of study treatment. All subjects not enrolling in the extension study will have the Week 12 study assessments and will enter the SFU Period.

Subjects who withdraw early from study treatment will be asked to return for study assessments 12 weeks after the first dose (ie, the Week 12 Visit) and for the SFU Visit.

Subjects withdrawing early from the study who are unable to return for the Week 12 Visit will undergo the early Withdrawal Visit assessments and will enter the SFU Period.

Safety Follow-Up Period

All subjects not continuing in the extension study, including those withdrawn from study treatment, will have a SFU Visit 20 weeks after their last dose of study medication.

2.3.3 Study duration per subject

For each subject, the study will last for a maximum of up to 29 weeks. This includes the following study period durations:

- Screening Period: 2 to 4 weeks
- Double-blind, placebo-controlled Treatment Period: 12 weeks
- SFU Period: a SFU Visit is planned 20 weeks after the last dose of study medication

After the 12-week Treatment Period, subjects will be allowed to enroll in an extension study. The SFU Visit will not be required for subjects who enroll in the extension study.

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

2.4 Determination of sample size

A total of 240 subjects (40 in each treatment group) are planned to be randomized in this study. The primary efficacy analysis will be based on the full analysis set (FAS). While some randomized subjects may not be in the FAS, it is expected that this number will not impact the following calculations.

The sample size is calculated based on published PASI90 response data from the secukinumab psoriasis program (ERASURE and FIXTURE studies) and from a Phase 2 bimekizumab study in subjects with psoriatic arthritis (PA0007). As an IL-17A inhibitor, the secukinumab data are important in the estimation of the response of bimekizumab, an IL-17A/F inhibitor. In the ERASURE and FIXTURE studies, PASI90 responses at Week 12 were reported to be 59.2% and 54.2% for the 300mg treatment group and 39.1% and 41.9% for the 150mg treatment group versus 1.2% and 1.5% in the placebo group, respectively (Langley et al, 2014). Limited data from a subset of subjects with psoriasis BSA >3% in the PA0007 study indicate a PASI90 response of 86.7% after eight weeks of treatment for the combined top three doses versus no response in the placebo group.

For the sample size calculation, responder rates of 60%, 50%, and 2% at the end of a 12-week Treatment Period for bimekizumab 480mg Q4W, bimekizumab 320mg Q4W, and placebo have been assumed, respectively. These assumptions are similar to the results observed in the ERASURE and FIXTURE studies and conservative relative to the PA0007 study in order to account for uncertainty in the estimates. The PASI90 responder rates for the 160mg Q4W and 64mg Q4W doses are assumed to decrease, using estimates of 40% and 20%, respectively. Note that the bimekizumab 320mg loading dose followed by bimekizumab 160mg Q4W has been left out of this sample size calculation as it will not be included in the test for dose response, but will rather be compared separately to the bimekizumab 160mg Q4W dose.

The sample size for the primary objective of evaluating the dose response relationship was calculated using a two-sided test for detecting a linear trend across proportions at a significance level of 0.05. With 40 subjects in each treatment group, the test for detecting the overall dose response based on PASI90 response is powered at >99%.

The sample size calculations were performed using the software nQuery Advisor® 7.0.

3 DATA ANALYSIS CONSIDERATIONS

3.1 General presentation of summaries and analyses

Statistical analysis and generation of tables, figures, subject data listings, and statistical output will be performed using SAS Version 9.3 or higher. All tables and listings will use Courier New font size 9.

Descriptive statistics will be displayed to provide an overview of the study results. For categorical variables, the number and percentage of subjects in each category will be presented. Unless otherwise noted, all percentages will be displayed to 1 decimal place. No percentage will be displayed for zero counts, and no decimal will be presented when the percentage is 100%. For continuous variables, descriptive statistics will include number of subjects with available measurements (n), mean, standard deviation (SD), median, minimum, and maximum. For bimekizumab PK concentrations, summary statistics will include geometric mean, geometric coefficient of variation (CV), 95% confidence intervals for geometric mean, arithmetic mean, SD, median, minimum, and maximum. All summaries of PK variables will be based on the observed values. No imputation will be used.

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

- "n" will be an integer
- Mean, SD, and median will use one additional decimal place compared to the original data
- CV[%] will be presented with one decimal place
- Minimum and maximum will have the same number of decimal places as the original value.

Derived variables in general will display the mean, SD and median to 1 more decimal place than the variables used in the derivation. If the number of decimal places reported in the raw data is varied then use either the maximum raw number of reported decimal places or 3, whichever is the lowest, as a guide for the descriptive statistics.

Statistical tests of efficacy variables will be presented as two-sided p-values rounded to three decimal places. P-values less than 0.001 will be presented as "<0.001" and p-values greater than

0.999 will be presented as ">0.999." Statistical comparisons will be two-sided and will be performed at the 0.05 level of significance.

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

3.2 Definition of Baseline values

Unless otherwise specified, the last valid measurement before study medication administration in the double-blind period will be used as the Baseline value. The same Baseline definition will be used for the follow-up period. If a Baseline visit measurement is missing, and a Screening visit measurement is available, the Screening value will be utilized as Baseline. If a scheduled

Baseline assessment is taken on the same day as first administration of study medication and no time is collected, it will be assumed to have been taken prior to study medication.

3.3 Mapping of assessments performed at early study withdrawal visit

Study assessments at an early study withdrawal visit where visit date matches the visit date of a scheduled visit will be summarized at the scheduled visit with the same visit date. Premature study withdrawal visit assessments that do not have a scheduled visit with a matching date will be assigned to the next scheduled site visit following the last visit where assessments were available. The only exception to this rule is for anti-bimekizumab antibody (AbAb) assessments where all premature withdrawal visit assessments will be assigned to the next scheduled visit at which AbAb are assessed. All by-visit summaries will contain nominal visits only. Unscheduled visits will not be mapped to scheduled visits. For subjects who discontinue study treatment early and return for the Week 12 visit as per the protocol, the assessments collected at that visit are summarized as Week 12 assessments in observed case summaries. For summaries based on imputed data, the Week 12 value will be treated as missing and imputed according to the rules for the given variable.

3.4 Protocol deviations

Important protocol deviations are defined as those deviations from the protocol likely to have a meaningful impact on the efficacy, safety, or PK/PD outcomes for an individual subject. The criteria for identifying important protocol deviations will be defined within the appropriate protocol-specific document. All protocol deviations will be reviewed as part of the ongoing data cleaning process and all important deviations will be identified and documented in a separate Protocol Deviation Tracker prior to unblinding to confirm exclusion from analysis sets.

3.5 Analysis sets

The primary efficacy variable will be summarized using the FAS and also the Per Protocol Set (PPS) as a supportive analysis. All other efficacy analyses will be based on the FAS. Safety variables will be summarized using the Safety Set (SS).

3.5.1 Enrolled Set

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

3.5.2 Randomized Set

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

3.5.3 Safety Set

The SS will consist of all subjects who received at least one dose of the study medication.

3.5.4 Full Analysis Set

The FAS will consist of all randomized subjects who received at least one dose of the study medication and have a valid measurement of the primary efficacy variable at Baseline.

3.5.5 Per Protocol Set

The PPS will consist of all subjects in the FAS who had no important protocol deviation affecting the primary efficacy variable. The subjects with important protocol deviations will be pre-defined and evaluated during a data evaluation meeting prior to unblinding of the data.

3.5.6 Other analysis sets

3.5.6.1 Pharmacokinetics Per-Protocol Set

The PK-PPS will consist of all randomized subjects who took at least one dose of the study medication and provided at least one quantifiable plasma concentration post-dose and had no important protocol deviations affecting the pharmacokinetic variables, as confirmed during ongoing data cleaning meetings prior to database lock.

3.5.6.2 Pharmacodynamics Per-Protocol Set

The PD-PPS will consist of all randomized subjects who took at least one dose of the study medication and provided at least one PD measurement post-dose with no important protocol deviations affecting the measurement.

3.6 Treatment assignment and treatment groups

At Baseline, eligible subjects will be randomly assigned as described in Section 2.3.2. Treatment assignment for the FAS and PPS will be according to randomization and not actual treatment received.

It is expected that subjects will receive treatment as randomized. Hence, safety analyses should be based on randomized treatment for the SS. However, if it is determined after unblinding that subjects randomized to placebo received bimekizumab at any time, then these subjects will be reallocated to the appropriate bimekizumab group and will be summarized accordingly from the time point at which the misallocation occurred. If a subject receives different doses of bimekizumab at different time points (not including bimekizumab 160mg subjects who have a 320mg loading dose at Baseline), then the subject will be allocated to the highest dose received. If the highest dose was received earlier than the other(s), then all safety data will be summarized under that dose. If the highest dose was received later than the other(s), then safety data will be summarized under that dose starting from the time point at which it was first received. Since the period of exposure at risk (140 days) is longer than the treatment period (84 days), no consideration is given to the time at which exposure to misallocated doses can be considered complete.

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

3.7 Center pooling strategy

Centers will be pooled into regions for analysis purposes. Due to the small number of subject in the Asia region, this will be combined with Europe. Centers will be grouped in the following regions: North America, Europe & Asia. These regions include the following countries:

- North America: Canada, USA
- Europe & Asia: Czech Republic, Hungary, Poland, Japan

3.8 Coding dictionaries

All prior and concomitant medications other than study drug will be classified by WHO Anatomical Therapeutic Chemical (ATC) Classification, presenting Anatomical Main Group (ATC Level 1), Pharmacological Subgroup (ATC level 3), and preferred term, using version SEP/2015 of the World Health Organization Drug Dictionary (WHO-DD), according to UCB standard operating procedures (SOP).

All AEs will be classified by primary system organ class (SOC), high level term (HLT) and preferred term (PT) using version 19.0 of MedDRA® according to UCB Standard Operating Procedures (SOP). Previous and ongoing medical history will be classified by MedDRA® SOC and PT.

3.9 Analysis time points

3.9.1 Pre-treatment period

The pre-treatment period (Screening period) of the study is the period prior to a subject's first dose of study medication intake. This period starts at the Screening visit (2 to 4 weeks prior to first dose) and ends at the Baseline visit (Week 0) up to the time of first study medication administration (exclusive). Unless specific time information is available to indicate that a Baseline visit assessment was performed after a subject's first study medication administration, all assessments performed at the baseline visit will be attributed to the pre-treatment period.

3.9.2 Double-blind treatment period

The double-blind treatment period begins at the Baseline visit (Week 0) at the time of first study medication administration (inclusive) and ends at the Week 12/WD Visit.

Premature withdrawal visit assessments will be assigned to the next scheduled visit following the last visit where assessments are available. All visit measurements, even violating the visit window, will be utilized for the respective visit as long as they are in the proper sequence.

Subjects will be classified as completing the study if they complete the Week 12 Visit without early withdrawal from the study. This is regardless of whether they attend the Follow-up visit and regardless of whether they consent to the extension study.

3.9.3 Follow-up period

For the subjects who do not consent to the extension study the SFU visit will take place 20 weeks after the last dose of study medication.

AEs that occur during the follow-up period will be reported separately. In addition AEs that occur in subjects who have entered the extension study will be captured within the extension study.

3.10 Relative day

The relative day will be included in different listings and will be calculated as follows:

- If the start (stop) date occurred on or after the first dose, but prior to the double-blind drug stop date, relative day is calculated as start (stop) date minus first dose date + 1
- If the start (stop) date occurred after the last dose of double-blind drug, the relative day to the most recent dose is calculated as start (stop) date minus most recent dose date. The relative day in this situation should be preceded by a '+'
- If the start (stop) date occurred before the first dose, the relative day is calculated as start (stop) date minus first dose date. The relative day in this situation should be preceded by a '-'.

For AEs, relative days for start and stop dates will be calculated as the number of days since the first injection of the medication. For non-treatment emergent AEs, relative day of onset will be negative if the event started and stopped before the first dose. Relative day will only be computed for fully completed dates and will be missing for partial dates.

3.11 Changes to protocol-defined analyses

Section 14.6 of the protocol (Handling of dropouts or missing data) describes the sensitivity analysis to be used in this study. More specifically, the protocol describes a sensitivity analysis using multiple imputation as the method for handling missing data and states that the MCMC method will be used for intermittent missing data and that monotone regression will be used for monotone missing data. This SAP has been modified such that the MCMC method will be used for all missing data whether intermittent or monotone.

Section 14.6 of the protocol also notes that the missing data for all continuous efficacy variables that have not been specifically described elsewhere will be imputed using the last observation carried forward (LOCF) approach. This has been updated in this SAP so that these continuous variables will in fact be imputed using the multiple imputation method. However, in the event that multiple imputation is not possible (eg, Due to small group sizes or insufficient missing data) then the LOCF method will be used.

4 STATISTICAL/ANALYTICAL ISSUES

4.1 Adjustments for covariates

The primary efficacy analyses and selected secondary analyses will be adjusted for the following covariates:

- Prior biologic exposure
- Region

4.2 Handling of dropouts or missing data

Based on previous studies of biologics in subjects with moderate to severe chronic plaque psoriasis, it is expected that the number of subjects who discontinue prior to Week 12 will be low. For the small percentage of subjects for whom primary endpoint data are unavailable at Week 12, this lack of data is suggestive of an ineffective study treatment, thereby supporting the imputation of nonresponse. Achieving the clinical response and making it through 12 weeks of study treatment are both critical components of the primary outcome. Therefore, nonresponder imputation (NRI) will be used as the method for handling missing data in the primary analysis. However, to account for other possible mechanisms of missing data other sensitivity analyses will be performed as described in Section 8.1.4.1.

4.2.1 Handling missing data for the primary efficacy variable

For the primary efficacy analysis based on evaluating dose response, missing data will be imputed using NRI. That is, subjects with missing data at Week 12 or who discontinue double blind study treatment prior to Week 12 will be counted as non-responders for the analysis.

In addition, sensitivity analyses will be performed which will assess the impact of different methods of handling missing data. These methods are described in Section 8.1.4.1.

4.2.2 Handling missing data for the secondary analysis of the primary efficacy variable

For the key secondary analysis of the primary efficacy variable based on performing pairwise comparisons of bimekizumab doses versus placebo, missing data will be imputed using NRI as described in Section 4.2.1.

In addition there will be 4 sensitivity analyses to assess the impact of missing data. These are further described in Section 8.1.4.1.

4.2.3 Handling missing data for other efficacy variables

For the dose/exposure response model, missing data will not be imputed. If a subject drops out, this will be used as a covariate in the model. However, the dose/exposure response model is outside of the scope of this SAP and is described in greater detail in a separate data analysis plan (DAP).

Other categorical efficacy variables will be imputed using NRI (ie. Missing data or subjects who prematurely discontinue study treatment prior to Week 12 will be counted as non-responders) or summarized as observed cases. One exception to this rule is the imputation of missing data when summarizing HADS-A and HADS-D scores below 8 (ie. Normal HADS scores). In these summaries, LOCF will be used. That is, the last available HADS-A/HADS-D score while the subject was on study treatment will be carried forward to the missing time points to determine whether or not the subject had normal HADS scores.

Continuous efficacy variables will be imputed using multiple imputation (MI) via the Markov-Chain Monte Carlo (MCMC) method. Summaries based on observed case data will also be prepared.

The MI procedure for continuous efficacy variables (based on MCMC) will be applied as follows using a starting "seed" value of 17402:

1. Create a dataset, one for each treatment group, of subjects with observed values and those needing estimation by multiple imputation. The missing values in each dataset will be filled in using the MCMC method with a total of 100 sets of imputations being performed. Biologic exposure, geographic region and values at baseline and at each post-baseline visit (in chronological order) will be included in the imputation model. If required, in order to facilitate the imputation model, variables can be omitted in the order of biologic exposure and geographical region. Cut-off values will be applied to the imputed data, as appropriate, to ensure that imputed values do not exceed the range of plausible values for the variable being imputed. The resulting datasets for each treatment arm will be combined into one complete dataset based on each of the 100 imputations
2. The relevant change from Baseline (or percent change from Baseline) values will then be derived based on the datasets which include observed and imputed values for the variable being considered
3. The results from each of the 100 imputed datasets will be combined for the calculation of means and standard errors using Rubin's rules, which account for the uncertainty associated with the imputed values (Rubin, 1987). This will be done using SAS PROC MIANALYZE. For calculation of other descriptive statistics such as the median, min and max, Rubin's rules do not apply. Multiple imputation estimates will be computed by simply averaging the estimates from the $m = 1, \ldots, M$ independent repetitions of the imputation algorithm:

$$\bar{\theta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m$$

where $\hat{\theta}_m$= estimate of θ from the completed dataset $m = 1, \ldots, M$.

There may be cases where the multiple imputation model fails to converge (eg, sparse subgroups), in such situations the LOCF approach will instead be used to impute the missing data. If LOCF is used instead of multiple imputation for this reason, this will be clearly specified in the corresponding table summary. Should there be no missing data for a study variable then only observed case data will be produced.

Summaries for all variables based on observed case data will also be prepared.

4.2.4 Safety variables

For analyses of AEs and concomitant medication usage, a complete date must be established in order to correctly identify the AE or medication as occurring during treatment or not. For purposes of imputing missing components of partially-reported start and stop dates for AEs and for medication use, the algorithms listed below will be followed. Start and stop dates of AEs or concomitant medication will be displayed as reported in the subject data listings (i.e., no imputed values will be displayed in data listings).

Partial AE and concomitant medication start dates will be imputed as follows:

Imputation of Partial Start Dates

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

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

Imputation of Partial Stop Dates

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

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

4.3 Interim analyses and data monitoring

After all enrolled subjects have completed the Week 8 visit, an interim analysis will be performed to analyze the dose-exposure response for PASI, including the PASI90 response, in each dose group to aid in determining the optimal therapeutic dose(s) for subsequent studies. The purpose of performing this analysis at this time is so as to be able to develop a dose-response model in advance of the Week 12 database lock. The final model will include all data up to and including Week 12 and the model will be revised/updated accordingly (if required). The details of this model are outside of the scope of this SAP and are described in a separate DAP.

Decisions related to the conduct of the PS0010 study will not be impacted by these analyses as they will be conducted by individuals that are not part of the clinical study team, and they will not be shared with any clinical study team members until after the database lock, at which point the PS0010 study team will be unblinded.

The database lock to support the development of the PS0010 interim CSR will occur before all subjects have completed the PS0010 study. Specifically, subjects who do not enter the extension study (PS0011) will need to complete the SFU Visit, which occurs 20 weeks after the last dose of study treatment. The Week 12 interim database lock will be performed based on the last subject completing the Week 12 visit. It is anticipated that there will be some subjects still awaiting the SFU at that time. The number of subjects in that position is expected to be low, and the minimal data to be collected from their SFU visits is not considered mandatory to evaluating the key efficacy and safety objectives of the study, therefore it is considered acceptable to proceed with the PS0010 data analysis and interim CSR development in the absence of these data. Once all SFU data have been collected and the full PS0010 study database has been locked, all TLFs will be re-run to include the SFU data that have been added to the database following the Week 12 interim database lick. A final CSR will subsequently be prepared based on this final set of TLFs. Other than the additional data from the pending SFU visits, it is expected that this final CSR will

include biomarker and flow cytometry data summaries that will not be available in the interim CSR. In addition, an independent Data Monitoring Committee (DMC) will periodically review and monitor the safety data from this study and advise UCB.

Details will be provided in the DMC Charter and further details related to this DMC will be outlined in a separate analysis plan.

4.4 Multicenter studies

Not applicable.

4.5 Multiple comparisons/multiplicity

In order to control for the study-wise Type I error rate, the secondary analysis of pairwise comparisons of bimekizumab will be formally evaluated for statistical significance only if the primary efficacy analysis is statistically significant at the 5% 2-sided level. In addition, the pairwise comparisons will follow a sequential testing sequence and the formal evaluation of statistical significance of each comparison is dependent upon the previous comparison achieving statistical significance at the 5% 2-sided level. More details can be found in Section 8.1.3.

4.6 Use of an efficacy subset of subjects

The primary dose response analysis and pairwise comparisons will be repeated using the PPS as a supportive analysis.

4.7 Active-control studies intended to show equivalence

Not applicable.

4.8 Examination of subgroups

Subgroup analyses will be performed on the primary and secondary efficacy variables.

The following variables for subgroup analyses will be used:

- Age (<65 years, ≥65 years)
- Gender (male, female)
- Disease duration (<median, ≥median)
- Region (North America [Canada, USA], Europe & Asia [Czech Republic, Hungary, Poland, Japan]
- Body weight (<100 kg, ≥100 kg)
- BMI (<25 kg/m^2, 25 to <30 kg/m^2, ≥30 kg/m^2)
- Prior systemic phototherapy or chemophototherapy (yes, no)
- Prior nonbiologic systemic therapy (yes, no)
- Prior biologic therapy (yes, no)
- Prior anti-TNF therapy (yes,no)
- Any prior systemic therapy (yes, no)
- IGA score at baseline (3 vs 4)

- PASI score at baseline (<20, ≥20)
- Psoriasis BSA at baseline (<20%, ≥20%)

These summaries will be based on imputed data (NRI) and will include descriptive statistics only.

5 STUDY POPULATION CHARACTERISTICS

5.1 Subject disposition

Summaries of reasons for screen failures (for all subjects screened), disposition of subjects (for all subjects screened) as well as the subjects who discontinued due to AEs (for the RS) will be produced. The disposition of subjects for all subjects screened will include the number of subjects included in each analysis set (ES, RS, FAS, PPS, SS, PK-PPS, and PD-PPS) overall and by site.

In addition, the number and percentage of subjects who discontinued treatment and who discontinued the study, including a breakdown of the main reason for discontinuation will be presented for subjects in the RS.

5.2 Protocol deviations

A summary, using the RS, of the number and percentage of subjects with an important protocol deviation (including a summary of subjects excluded from the PPS due to important protocol deviations) by treatment group and overall and by type of deviation will be provided.

A by-subject listing of important protocol deviations will be provided.

6 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Unless specified otherwise all summaries detailed in this section will be performed on the SS and repeated using the FAS. If the SS and FAS analysis sets are identical the summaries will not be repeated.

6.1 Demographics

Demographic variables will be summarized by treatment group and overall.

The following continuous variables will be summarized using descriptive statistics (number of subjects, mean, SD, minimum, median and maximum).

- Age at the time of study entry (years)

The following categorical variables will be summarized using frequency counts and percentages.

- Age group (≤18, 19-<65, ≥65 years)
- Age group (18-<65, 65-<85, ≥85 years)
- Age group (<40, 40-<65, ≥65 years)
- Gender
- Race

- Ethnicity
- Height (cm)
- Weight (kg)
- BMI (kg/m^2)

By-subject listings of demographics will be provided.

6.2 Other Baseline characteristics

Baseline characteristics (including Baseline clinical measures) will be summarized by treatment group and overall.

- The following continuous variables will be summarized using descriptive statistics (number of subjects, mean, SD, minimum, median and maximum).Psoriasis BSA (%)
- Duration of disease (in years)
- PASI score
- mNAPSI total score
- PGADA VAS score
- PSSI score
- DLQI total score
- HADS-A score
- HADS-D score

BMI (kg/m^2) will be calculated as:

$$BMI = \frac{\text{Weight (kg)}}{(\text{Height (m)})^2} \quad (1)$$

Duration of disease (years) will be calculated as:

$$Disease\ Duration = \frac{(\text{Date of randomization} - Date\ of\ onset\ of\ Plaque\ Psoriasis^{1})}{365.25} \quad (2)$$

[1] If the date of onset of plaque psoriasis is partial, it should be imputed to the most recent feasible date (ie, last day of the month if only day is missing, or the last day of the year if day and month are missing).

Should the date of randomization be missing then the duration of disease will be derived using the date of screening.

The following categorical variables will be summarized using frequency counts and percentages.

- Age (<40 years, 40 to <65 years, ≥65 years)
- Region (North America [Canada, USA], Europe & Asia [Czech Republic, Hungary, Poland, Japan]

- Country
- BMI (<25 kg/m2, 25 to <30 kg/m2, ≥30 kg/m2)
- Duration of disease (<median, ≥median)
- IGA score
- Body weight (<100 kg, ≥100 kg)
- Prior biologic therapy (yes, no)
- Prior anti-TNF therapy (yes, no)
- Prior phototherapy or chemophototherapy (yes, no)
- Prior nonbiologic systemic therapy (yes, no)
- Any prior systemic therapy (yes, no)
- HADS-A score (<8, 8 to 10, 11 to 14, >14)
- HADS-D score (<8, 8 to 10, 11 to 14, >14)
- Tobacco use (never, current, former)

By-subject listings of Baseline characteristics will be provided.

6.3 Medical history and concomitant diseases

All medical history will be listed and the number and percentage of subjects with any medical history will be summarized by treatment group, system organ class (SOC) and preferred term (PT).

6.4 Prior and concomitant medications

Medication start and stop dates will be compared to the date of first dose of treatment to allow medications to be classified as either Prior or Concomitant.

Prior medications include any medications that started prior to the start date of study medication. Past medications are a subset of prior medications, and include prior medications with a stop date before the date of first study medication administration. Concomitant medications are medications taken at least one day in common with the study medication dosing period.

Details of imputation methods for missing or partial dates are described in Section 4.2.4.

The number and percentage of subjects taking Prior medications (excluding past psoriasis medications) will be summarized by treatment group, overall and by ATC class, presenting Anatomical Main Group (ATC Level 1), Pharmacological Subgroup (ATC level 3), and preferred term.

The number and percentage of subjects taking Concomitant medications will be summarized similarly.

Past psoriasis treatment will be captured separately and will also be summarized by treatment group.

A by-subject listing of all Prior and Concomitant medications and of past psoriasis treatment will be provided.

7 MEASUREMENTS OF TREATMENT COMPLICANCE

All summaries detailed in this section will be performed on the SS.

Due to the method of administration of the treatments, compliance will be examined in terms of completed doses, where one dose will be considered as one complete treatment administration (ie, three injections).

Treatment compliance will be calculated as:

$$\frac{total\ number\ of\ completed\ doses}{total\ number\ of\ expected\ doses}\ x\ 100 \quad (3)$$

where the total number of expected doses is derived relative to when the subject finishes treatment. If a subject completes treatment nine doses are expected (three each at Baseline, Week 4 and Week 8). If a subject discontinues early, then the number of expected injections is based on the time of early discontinuation relative to the dosing visits.

A summary of percent treatment compliance categorized as ≤75% and >75% will be provided by treatment group.

A by-subject listing of treatment compliance will be provided.

8 EFFICACY ANALYSES

8.1 Statistical analysis of the primary efficacy variable

8.1.1 Derivations of PASI score and response

PASI scoring of psoriatic plaques is based on three criteria: redness (R), thickness (T), and scaliness (S). Severity is rated for each index (R, S, T) on a 0-4 scale (0 for no involvement up to 4 for very marked involvement). The body is divided into four areas comprising the head (h), upper extremities (u), trunk (t), and lower extremities (l). In each of these areas, the fraction of total surface area affected is graded on a 0-6 scale (0 for no involvement; up to 6 for 90% - 100% involvement).

The various body regions are weighted to reflect their respective proportion of BSA. The composite PASI score is then calculated by multiplying the sum of the individual-severity scores for each area by the weighted area-of-involvement score for that respective area, and then summing the four resulting quantities as follows (note for R, T, and S scores are as follows: 0 = none, 1 = slight, 2 = moderate, 3 = marked, and 4 = very marked):

$$PASI = (0.1 \times (R_h + T_h + S_h) \times A_h) + (0.2 \times (R_u + T_u + S_u) \times A_u) + (0.3 \times (R_t + T_t + S_t) \times A_t) + (0.4 \times (R_l + T_l + S_l) \times A_l) \quad (4)$$

where

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

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

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

A_h, A_u, A_t, A_l = numerical value translation of % area of psoriatic involvement score for the head, upper extremities, trunk, and lower extremities, respectively (where 0 = 0% [clear], 1 = >0% to <10%, 2 = 10% to <30%, 3 = 30% to <50%, 4 = 50% to <70%, 5 = 70% to <90%, and 6 = 90% to 100%).

The highest potential PASI score is 72 for severe disease; the lowest is 0 for no psoriasis lesion. PASI scores are treated as a continuous score, with 0.1 increments within these values. The percent improvement in PASI scores from Baseline will be computed as:

$$\text{Percent improvement from Baseline} = 100 \times \frac{\text{Baseline PASI} - \text{Post Baseline PASI}}{\text{Baseline PASI}} \quad (5)$$

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

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

A categorical variable, PASI90, is defined to be equal to 1 if the percentage improvement from Baseline in the PASI scores is 90% or greater and 0 if the percentage improvement from baseline is less than 90%. This definition is introduced for the purpose of identifying subjects who respond to the treatment (1 = responder, 0 = non-responder). Similarly, the categorical secondary efficacy variables PASI50, PASI75, and PASI100 response are equal to 1 for subjects with improvements of 50% or greater, 75% or greater and 100% from Baseline in PASI score, respectively (and equal to 0 otherwise),

8.1.2 Primary analysis of the primary efficacy variable

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

The number and percentage of subjects achieving a PASI90 response (as defined in Section 8.1.1) at Week 12 will be presented by treatment group.

The dose-response relationship between treatment and PASI90 response will be assessed to test the following hypotheses:

$$H_0: -2\mu_1 - \mu_2 + 0\mu_3 + \mu_4 + 2\mu_5 = 0$$

$$H_1: -2\mu_1 - \mu_2 + 0\mu_3 + \mu_4 + 2\mu_5 \neq 0$$

where 1 = placebo, 2 = bimekizumab 64mg Q4W, 3 = bimekizumab 160mg Q4W, 4 = bimekizumab 320mg Q4W, and 5 = bimekizumab 480mg Q4W.

This will be done by using a logistic regression model which will include fixed effects for treatment, region (see Section 4.8), and prior biologic exposure (yes/no). This will be evaluated using a linear contrast in which equal spacing between each of the doses is assumed.

The coefficients for this linear contrast will be -2, -1, 0, 1 and 2 for placebo, bimekizumab 64mg Q4W, bimekizumab 160mg Q4W, bimekizumab 320mg Q4W, and bimekizumab 480mg Q4W, respectively. While the bimekizumab 320mg loading dose followed by bimekizumab 160mg Q4W has been left out of the linear contrast for evaluating dose response, it will remain a part of the logistic regression model and will be compared to the bimekizumab 160mg Q4W dose as a secondary analysis of the primary efficacy variable.

If the logistic regression model is unable to converge, then prior biologic exposure will be dropped from the model to facilitate convergence. If convergence is not obtained after dropping prior biologic exposure, then region may be dropped from the model. The linear contrast from the logistic regression model will be evaluated at a two-sided significance level of α=0.05.

NRI will be used to account for missing data; ie, subjects with a missing PASI score at Week 12 or who discontinued study treatment prior to the Week 12 Visit will be considered non-responders for the primary analysis.

The results produced in the SAS modelling process for the primary endpoint will be provided as a supportive output.

8.1.3 Secondary analyses of the primary efficacy variable

As a supportive analysis for the primary efficacy variable, the same logistic regression model described in Section 8.1.2 will be used to assess the effect of each individual dose vs placebo on PASI90 response. Comparisons will be made for each dose vs placebo at a two-sided significance level of α=0.05. For each dose, the odds ratio vs placebo, the 95% confidence interval, and the corresponding p-value will be calculated.

The pairwise comparisons of bimekizumab versus placebo will be formally evaluated for statistical significance only if the primary dose response efficacy analysis is statistically significant. Once the dose response has been established, pairwise testing of each bimekizumab dose versus placebo will account for multiplicity by using a fixed sequence testing procedure with each bimekizumab dose being tested sequentially from the highest dose to the lowest dose. Each test will only be conducted if the previous test reaches significance at a two-sided significance level of α=0.05. This procedure will control the overall Type I error rate. While formal evaluation of statistical significance will only proceed as described above, the p-values for all pairwise tests vs placebo will be provided, regardless of whether or not the testing procedure was completed for all bimekizumab doses.

A pairwise comparison will also be made for the bimekizumab 160mg Q4W group versus the bimekizumab 320mg loading dose followed by bimekizumab 160mg Q4W group, but this evaluation will not be part of the sequential testing procedure described above (ie, the p-value will be nominal).

The results produced in the SAS modelling process for the secondary analyses of the primary efficacy variable will be provided as a supportive output.

8.1.4 Supportive and sensitivity analyses of the primary efficacy variable

8.1.4.1 Sensitivity analyses

The primary dose response analysis and pairwise comparisons will be repeated for all subjects in the PPS to evaluate the effect of important protocol deviations on the analysis.

An ordered categorical analysis using a non-parametric correlation statistic of Mantel and Haenszel and modified ridit scores will be performed as a sensitivity analysis to the primary dose-response evaluation based on the logistic regression model. This analysis will include region and prior biologic exposure as stratification factors to adjust for potential differences in the dose response based on these variables, and will be evaluated at a 2-sided significance level of α=0.05.

Sensitivity analyses for the primary dose response analysis of the primary efficacy variable, designed to evaluate the assumptions related to the handling of missing data are as follows:

1. An analysis will be performed in which all available data at Week 12 will be considered. This analysis will be based on the logistic regression model to evaluate dose-response as specified in the primary analysis (see Section 8.1.2). However, in this case, subjects will be analyzed according to their randomized treatment, even if they discontinued prior to Week 12 and were no longer on the randomized study treatment when the assessment was performed at Week 12. Even though efforts will be made to collect the primary outcome data for all subjects at Week 12, there may still be some subjects for whom Week 12 efficacy data cannot be obtained. In this case, NRI will be used for these subjects. For the purposes of analyses these will be referred to as "Week 12 de facto analysis". The estimand here is the difference in outcome improvement at the planned endpoint for all randomized participants (Mallinckrodt, 2012). This is an estimand of effectiveness to evaluate the de facto hypothesis. It should be noted that this measures something different from the primary analysis and could be confounded by placebo subjects who withdraw and are subsequently on another active medication at the time of the Week 12 assessment. Therefore, the results of this analysis should be interpreted in the appropriate context.
2. The treatment by region interaction will be investigated by including the interaction term in the primary model, should this interaction be statistically significant at the 10% level this term will be included into the main model.

The following sensitivity analyses will be performed for the secondary pairwise comparison analysis of the primary efficacy variable in order to assess the missing data assumptions:

1. Missing data will be addressed by using MI based on the Markov-Chain Monte Carlo (MCMC) method for missing data. The actual PASI scores will be imputed and then dichotomized to obtain the PASI90 response status. The multiple imputed data sets will be analyzed using a logistic regression model with factors of treatment group, prior biologic exposure, and region. Finally, SAS PROC MIANALYZE will be used to combine the results into a single inference (see Section 8.1.4.3). This procedure assumes a missing at random (MAR) pattern of missingness and corresponds to an estimand of what has been called the difference in outcome improvement if all subjects tolerated or adhered to treatment (Mallinckrodt, 2012).

2. The logistic regression model will be applied where all available data at Week 12 will be considered, as described in #1 in the sensitivity analysis for the primary dose response in Section 8.1.4.1.

3. The logistic regression model will be applied where only observed data for subjects still on the initially randomized treatment at Week 12 will be considered, as described in #2 in the sensitivity analysis for the primary dose response in Section 8.1.4.1.

8.1.4.2 Supportive analysis

A supporting analysis will be performed based on observed data only for subjects who are still on the initially randomized treatment at Week 12. Subjects with missing data or who have prematurely discontinued study treatment will be treated as missing. For the purposes of analysis these will be referred to as "Observed Values for Randomized Treatment Completers". The same statistical method described for the primary efficacy analysis of dose response will be used.

An additional supportive analysis will be performed to take into account the data for subjects that are withdrawn due to CTCAE Grade 3 and/or repeated CTCAE Grade 2 lab abnormalities highlighted in Withdrawal Criterion #11. These subjects will be identified programmatically as those with a reason for discontinuation of "Other" where the specific description provided in the free text field refers to "Withdrawal Criterion #11". This sensitivity analysis will use LOCF for these subjects where the last observed value of a variable prior to discontinuation (including Baseline) is used to impute the missing data. For all other subjects with missing data or who discontinued study treatment prior to Week 12, NRI will be used.

8.1.4.3 Multiple imputation analysis

The multiple imputation analysis is a sensitivity analysis for the secondary analysis of the primary variable (see Section 8.1.4.1).

Investigators will be given discretion to discontinue study treatment for subjects not responding in order to provide them with alternative therapy to treat the condition. Therefore, in many cases, missing efficacy data due to study treatment discontinuation should be dependent on the observed efficacy scores, but independent of unobserved data. This would be consistent with a missing at random (MAR) pattern of missingness. In order to investigate the efficacy results under the assumption of data being MAR, a multiple imputation method will be applied as follows:

1. Create a data set, one for each treatment group, of subjects with observed values and those needing estimation by multiple imputation. The missing PASI values in each data set will be filled in using the MCMC method with a total of 100 sets of imputations being performed. Biologic exposure, geographic region and values at baseline and at each post-baseline visit (in chronological order) will be included in the imputation model. Note that PASI scores at earlier visits will also be used as predictors for the model of PASI at later visits. The resulting data sets for each treatment arm will be combined into one complete data set based on each of the 100 imputations.

   Note: The imputation model based on the MCMC method will only allow continuous variables in the imputation model. Therefore, prior biologic exposure and region will be re-coded as indicator variables (with values of 0 or 1 for each level of the variable). In order to

achieve model convergence prior biologic exposure may be dropped from the model, if convergence is still not obtained then region may be also be dropped from the model.

2. For each complete imputed data set, the dichotomous responder rate based on the PASI scores will be computed. Each complete imputed data set will then be analyzed based on a logistic regression model with factors of treatment group, region, and prior biologic exposure.

   Note: For derivation of PASI90 response, the PASI value at Week 12 in the imputed data sets will be compared directly to the observed Baseline PASI value to determine whether or not a reduction of at least 90% was achieved. If values outside of the pre-defined range of values for PASI (0-72) are imputed, they will be cut off as appropriate after the multiple imputation procedure but before deriving the responder variable. For example, an imputed PASI value of -0.5 would be changed to 0 before deriving the PASI90 responder variable.

3. The Week 12 results from the logistic regression analysis of each of the 100 imputed data sets will be combined for overall inference using Rubin's rules, which account for the uncertainty associated with the imputed values (Rubin, 1987). This will be done using SAS PROC MIANALYZE.

Some key points to consider relative to the calculation of the odds ratios and corresponding confidence intervals are noted below:

As the estimates of the odds ratios from the logistic regression model in step 3 follow a log-normal distribution, a log transformation is needed to normalize these estimates since the procedures for combining results from multiple imputed datasets are based on the assumption that the statistics estimated from each imputed dataset are normally distributed. Therefore, the log of the odds ratio estimates from the logistic regression model are used when combining into a single inference (the use of PROC MIANALYZE in step 3). Additionally, the standard errors (SE) for the odds ratios are transformed as follows:

$$SE = \frac{\log(UCL) - \log(LCL)}{2Z_{\alpha/2}} \quad (6)$$

Where UCL and LCL are the upper and lower confidence limit, respectively, for the confidence interval of the odds ratio from the logistic regression model, and $Z_{\alpha/2}$ is the relevant critical value from the standard normal distribution (1.96 for a 95% confidence interval). After the use of PROC MIANALYZE in step 4, the estimates of the log odds ratio for each bimekizumab dose relative to placebo and the corresponding upper and lower confidence limits will be provided. The odds ratio is estimated by exponentiating the estimate of the log odds ratio. The confidence limits of the odds ratio are then estimated as follows:

$$LCL = OR * \exp^{-(SE*Z_{\alpha/2})} \quad (7)$$

$$UCL = OR * \exp^{(SE*Z_{\alpha/2})} \quad (8)$$

Where OR is the back-transformed estimate of the odds ratio just described, SE is the standard error of the log odds ratio derived in PROC MIANALYZE and $Z_{\alpha/2}$ is the relevant critical value from the standard normal distribution (1.96 for a 95% confidence interval). These calculations

will be done such that odds ratios and corresponding confidence intervals are calculated for the odds ratio of each bimekizumab dose versus placebo. Note that the p-values presented in the tables will be the ones provided initially by PROC MIANALYZE and are not impacted by the transformations described above.

8.2 Statistical analysis of the secondary efficacy variables

All secondary efficacy variables will be summarized for subjects in the FAS.

Unless specified otherwise, both NRI (or MI for continuous variables) and Observed Case analyses will be performed (see Section 8.1.4). No other sensitivity or supportive analyses will be presented for the secondary efficacy variables.

Also, in addition to any endpoint tabulations, summary tables of each endpoint by visit and treatment will be provided.

All p-values presented for secondary efficacy variables will be nominal.

8.2.1 Derivation of IGA response

A static IGA for psoriasis will be used to assess disease severity in all subjects during the study.

The Investigator will assess the overall severity of psoriasis using the following five-point scale (see Table 8–1):

Table 8–1: Five-point Investigator's Global Assessment

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

IGA response is defined as clear [0] or almost clear [1] with at least a two category improvement from Baseline.

8.2.2 Analysis of Investigator's Global Assessment

The number and percentage of subjects achieving an IGA response (defined as clear [0] or almost clear [1] with at least two category improvement from Baseline) at Week 12 will be presented by treatment group.

The same logistic model as described for the secondary analysis of the primary efficacy variable (see Section 8.1.3) will be used to analyze IGA response at Week 12, performing pairwise

comparisons versus placebo and presenting the odds ratio vs placebo, 95% confidence interval and corresponding p-value for each dose.

Similarly the number and percentage of subjects achieving an IGA response at Week 8 will also be presented by treatment group and will be analyzed in the same way as the Week 12 IGA response.

A line plot of the IGA responder rate over time, by treatment group will be produced.

8.2.3 Analyses of secondary PASI responder variables

8.2.3.1 PASI90 at Week 8

The number and percentage of subjects achieving a PASI90 response (as defined in Section 8.1.1) at Week 8 will be presented by treatment group.

The same logistic regression model as described in Section 8.1.3 will be used to analyze PASI90 at Week 8, performing pairwise comparisons of each bimekizumab dose versus placebo and presenting the odds ratio vs placebo, 95% confidence interval and corresponding p-value for each dose.

A line plot of the PASI90 responder rate over time, by treatment group will be produced.

8.2.3.2 PASI75 at Week 12

The number and percentage of subjects achieving a PASI75 response (as defined in Section 8.1.1) at Week 12 will be presented by treatment group.

The same logistic regression model as described in Section 8.1.3 will be used to analyze PASI75 at Week 12, performing pairwise comparisons of each bimekizumab dose versus placebo and presenting the odds ratio vs placebo, 95% confidence interval and corresponding p-value for each dose.

A line plot of the PASI75 responder rate over time, by treatment group will be produced.

8.2.3.3 PASI100 at Week 12

The number and percentage of subjects achieving a PASI100 response (as defined in Section 8.1.1) at Week 12 will be presented by treatment group.

The same logistic regression model as described in Section 8.1.3 will be used to analyze PASI100 at Week 12, performing pairwise comparisons of each bimekizumab dose versus placebo and presenting the odds ratio vs placebo, 95% confidence interval and corresponding p-value for each dose.

A line plot of the PASI100 responder rate over time, by treatment group will be produced.

8.3 Analysis of other efficacy variables

All other efficacy variables will be summarized for subjects in the FAS.

All other efficacy variables will be summarized based on imputed data (NRI and MI for binary and continuous variables, respectively) and observed case data (ie, subjects with missing data or who have prematurely discontinued study treatment are treated as missing) unless otherwise specified.

8.3.1 Further analysis of Investigator's Global Assessment

The distribution of the IGA score will be summarized in a shift table comparing the change from Baseline versus post-Baseline by treatment group for the FAS. Cross-tabulations of the Baseline vs post-Baseline IGA score will be reported for each time point, including the number of subjects with missing data.

In addition, the IGA responses as described in Section 8.2.1 will be summarized by treatment group across all visits.

8.3.2 Further analysis of PASI

Absolute value, change from baseline, and percentage change from Baseline in PASI score will be summarized by treatment group and visit.

In addition summaries of the PASI50 response, PASI75 response, PASI90 response and PASI100 response will be summarized by treatment group and visit.

A line plot of the percentage change from Baseline in PASI score over time, by treatment group will be produced.

8.3.3 Analysis of BSA

Absolute value, change from baseline, and percentage change from Baseline in BSA affected by psoriasis will be summarized by treatment group and visit.

8.3.4 Derivation of DLQI score and response

The DLQI questionnaire is used for patients with psoriasis and consists of 10 questions. Question 7 consists of a sub-question which is only to be answered following a "No" response to the main question. The questions are scored as shown below in Table 8–2, and the DLQI score is categorized as shown in Table 8–3.

Table 8–2: DLQI Scoring

Question Number	0 points	1 point	2 points	3 points
1,2	Not at all	A little	A lot	Very Much
3, 4, 5, 6, 8, 9, 10	Not at all/ Not Relevant	A little	A lot	Very Much
7	No – Not at all/ Not Relevant	No – A little	No – A lot	Yes

Table 8–3: DLQI - Categorization

Category	Score Range
No Effect	0-1
Small Effect	2-5
Moderate Effect	6-10
Very Large Effect	11-20
Extremely Large Effect	21-30

In the case of one missing value the result imputed for that question will be 0 and the DLQI score created as normal. Should two or more questions be unanswered then the DLQI score will be set to missing. For question 7, if "Not relevant" is selected the score for the question will be 0, if the question is answered "No" but the second half of the question is incomplete then the question will still be scored as 0.

A four-point change in the DLQI score has been reported to be meaningful for the subject (within-Subject MCID); while a DLQI absolute score of 0 or 1 indicates no or small impact of the disease on health related QOL.

A line plot of the percentage changes from baseline in DLQI score over time, by treatment group will be produced.

8.3.5 Analysis of DLQI

Absolute value, change from Baseline, and percentage change from Baseline value for DLQI will be summarized by treatment group and visit.

The number and percentage of subjects achieving a DLQI score of 0 or 1 at each visit will be presented by treatment group.

The number and percentage of subjects achieving DLQI MCID (as defined in Section 8.3.4) at each visit will be presented by treatment group.

8.3.6 Derivation of mNAPSI score

Psoriatic nail disease will be evaluated at the Baseline visit using the mNAPSI. All affected nails will be scored (0 to 3) for onycholysis/oil drop dyschromia, nail plate crumbling, and pitting and will be scored (0 for "no" or 1 for "yes") for leukonychia, nail bed hyperkeratosis, splinter haemorrhages and red spots in the lunula. The score for an individual nail ranges from 0 to 13 with higher scores indicative of more severe nail disease. The total mNAPSI score is the sum of the scores for each individual nail. If a nail is unaffected, it will be recorded as such and will not contribute to the total mNAPSI score. Subjects with nail disease at Baseline are defined as those with a mNAPSI score >0 at Baseline.

If 1 or 2 response items scored on the 0 to 1 scale are missing, the missing response(s) will be imputed by the average of the available responses. Otherwise, the total mNAPSI score will be set to missing.

8.3.7 Analysis of mNAPSI

The change from Baseline in mNAPSI scores will be summarized at each visit by treatment group for the subset of subjects with psoriatic nail disease at Baseline.

8.3.8 Derivation of PGADA for the arthritis VAS

The PGADA for the arthritis VAS will be used to provide an overall evaluation of arthritis disease symptoms. Subjects will respond to the question, "Considering all the ways your arthritis affects you, please mark a vertical line on the scale below to show how you are feeling today", using a VAS where 0 is "very good, no symptoms" and 100 is "very poor, severe symptoms".

8.3.9 Analysis of PGADA for the arthritis VAS

The actual and change from Baseline in PGADA VAS scores will be summarized at each visit by treatment group.

Subjects who have psoriatic arthritis as defined by the Toronto Psoriatic Arthritis Screening questionnaire (ToPAS) questionnaire results at the baseline assessment will also have the PGADA VAS scores summarized separately.

8.3.10 Derivation of PSSI

The PSSI considers both the extent of the scalp area of involvement and the severity based on the scoring scales outlined in Table 8–4 and Table 8–5.

The assessment considers erythema, induration and desquamation on the scalp. For each of these three elements, the scores of the area and severity are multiplied. Then, the score from each element is totaled to obtain the PSSI score. As with the PASI, the PSSI score ranges from 0 to 72 with a higher score indicating increased scalp disease severity.

If a subject is missing 1 severity measurement, the average of the remaining severity measurements will be utilized to substitute for the missing severity measurement. If the area of affected skin is missing or if 2 or more severity measurements are missing, then the PSSI score will be set to missing.

Table 8–4: PSSI assessment of extent of scalp psoriasis

Score	Definition
1	<10%
2	10-29%
3	30-49%
4	50-69%
5	70-89%
6	90-100%

Table 8–5: PSSI assessment of clinical symptoms

Score	Definition
0	Absent
1	Slight
2	Moderate
3	Severe
4	Severest possible

8.3.11 Analysis of PSSI

The analysis of PSSI score will be conducted for subjects who have scalp psoriasis, defined as PSSI>0 at Baseline.

The change from Baseline will be summarized at each visit by treatment group.

8.3.12 Derivation of SF-36

The SF-36v2, standard recall, measures the following 8 health domains as rated by the subjects over the past four weeks: Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional, and Mental Health. The classification of the questionnaire items to the health domains is shown in Section 12.1.

The SF-36 Physical and Mental Component Summary scores (PCS and MCS, respectively) are used to measure the two broad components, or aspects, of health-physical and mental. PCS and MCS are based on the aggregate of 8 health concepts described above and all of the 8 health domain scales are used to score both components summary measures.

One additional item asks respondents about health change over the past year.

The SF-36 will be used using QualityMetric's Health Outcomes™ Scoring Software. The software uses updates 2009 U.S. population norms and applies a Full Missing Score Estimation (Full MSE) method as follows:

- A health domain score (except the PF domain) will be estimated provided that at least one non-missing response is available within that domain
- For the PF domain item response theory will be used to develop a model for estimates of the missing score
- Regression methods are then applied to estimate the PCS and the MCS on the basis of the available domains.

8.3.13 Analysis of SF-36

The change from Baseline in the SF-36 PCS and MCS scores as well as for the individual domain scores will be summarized at each visit by treatment group.

As a sensitivity analysis the SF-36 PCS, MCS and individual domain scores will be also summarized at each visit and by treatment group for the complete cases, where any missing data will result in the affected domains and component scores for that visit also being set to missing.

8.3.14 Definition of Time to PASI Response

Time to PASI50, PASI75, PASI90, and PASI100 response (in days) will each be calculated as:

Date of first PASIx response – Date of Baseline visit +1 (9)

8.3.15 Analysis of Time to PASI Response

Time to PASI50, PASI75, PASI90, and PASI100 response will each be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.

Kaplan-Meier plots of time to PASI50, PASI75, PASI90, and PASI100 response will be presented by treatment group.

Subjects who discontinue study treatment prior to achieving a response will be censored at the date of study treatment discontinuation. Subjects who reach the Week 12 Visit without achieving the given response will be censored at the date of the Week 12 Visit.

The median time to response, including the two-sided 95% confidence interval, will be calculated for each treatment. Comparisons of each bimekizumab dose vs placebo will be analyzed using a log-rank test stratified by region and prior biologic exposure. In addition between group differences for bimekizumab 320mg loading dose + 160mg Q4W vs bimekizumab 160mg Q4W and bimekizumab 320mg loading dose + 160mg Q4W vs bimekizumab 320mg Q4W will be analyzed using the log-rank test stratified by region and prior biologic exposure.

8.3.16 Definition of Subject Symptom Diary responses

UCB is developing a new PRO measure that will be used to assess key symptoms relevant to subjects with moderate to severe chronic plaque psoriasis.

PS0010 will use the draft PRO measure in selected countries to enable psychometric validation of the PRO. Site staff will train the participating subjects on the use of the electronic PRO (ePRO) diary at the Screening visit, following which the device will be dispensed to the subject for home use until the Week 12 Visit.

The ePRO diary will be administered on a daily basis from Baseline to the Week 12 Visit. Item content of the ePRO diary will be finalized prior to study initiation and will include a variety of psoriasis-related characteristics, reported as relevant by subjects in the published literature and planned concept elicitation interviews.

The ePRO diary will also administer appropriate anchor items (eg, subject global impression of change, and subject global impression of concept) at the end of every study week.

The weekly score is the sum of the individual scores within the week. A maximum of 3 missed days are permitted (irrespective of whether these are consecutive or not) for the derivation of the weekly score and will be included in the analysis. The missing days will be counted as missing data for the purposes of the weekly score.

8.3.17 Analysis of Subject Symptom Diary responses

The mean responses for each question on the subject symptom diary will be summarized by treatment group and week, in terms of absolute values and changes from baseline.

8.3.18 Definition of HADS

The HADS was chosen for its well-established psychometric properties and its use in clinical research on biological therapy in subjects with chronic plaque psoriasis (Snaith et al, 1994). The HADS scores for anxiety and for depression range from 0 to 21 with higher scores indicating worse state. A score below 8 is considered to be normal, a score between 8 and 10 is considered mild, scores between 11 and 14 are considered moderate and a score of 15 and above is considered severe (Langley et al, 2009).

8.3.19 Analysis of HADS

Change from Baseline in HADS-A and HADS-D will be summarized at each visit by treatment group.

In addition the percentage of subjects with scores below 8, between 8 and 10, between 11 and 14 and greater than 14 in HADS-A and HADS-D will be summarized by visit and treatment group.

8.3.20 Definition of PGIC

The PGIC will be assessed at Baseline and on a weekly basis following the completion of all the patient symptom diary items. Only subjects in selected countries who participate in the completion of the patient symptom diary will participate in the completion of the PGIC.

The PGIC is a five-item instrument that will ask the participants to think about their psoriasis related symptoms (including itching, pain, scaling, and sleep loss) since they started this study. Each item has seven response categories ranging from “Very much improved,” “Much improved,” “Minimally improved,” “No change,” “Minimally worse,” “Much worse,” to “Very much worse.”

8.3.21 Analysis of PGIC

The response category for each item of the PGIC will be summarized by treatment group and week.

8.3.22 Definition of PGIS

The PGIS will be completed at Baseline and on a weekly basis following the completion of all patient symptom diary items. Only subjects in selected countries who participate in the completion of the patient symptom diary will participate in the completion of the PGIS.

The PGIS is a five-item instrument that will ask the participants to think about their psoriasis related symptoms (including itching, pain, scaling, and sleep loss) over the past week and an additional five items asking about the same set of symptoms at the current point in time. Each item has five response categories ranging from “No (symptoms),” “Mild (symptoms),” “Moderate (symptoms),” “Severe (symptoms),” to “Very severe (symptoms).”

8.3.23 Analysis of PGIS

The shift from Baseline in response category to each relevant visit will be summarized by question, treatment group and week.

9 PHARMACOKINETICS AND PHARMACODYNAMICS

All PK and antibody (AB) data will be listed together.

9.1 Pharmacokinetics

Bimekizumab plasma concentrations will be summarized for each treatment at each scheduled visit as described in Section 3.1 using the PK-PPS analysis set.

If bimekizumab plasma concentration measurements are deemed to be below the level of quantification (BLQ), then for calculation of the derived statistics this sample result will be set to half the LLOQ. The subjects with at least 1 result that is defined as BLQ will also be listed within the respective analysis table. Descriptive statistics will be calculated if at least 2/3 of the values are above the LLOQ. If this is not the case, only median, minimum, and maximum will be presented.

In addition geometric mean bimekizumab plasma concentration time curves will be plotted by treatment group, and by cumulative antibody status for subjects randomized to bimekizumab.

The bimekizumab concentrations will also be listed.

Additional PK analyses will be described in the DAP.

9.2 Pharmacodynamics

IL-17A, IL-17F, IL-6, and IL-23 data will be summarized for each treatment at each scheduled visit using the PD-PPS analysis set.

Additional PD analyses will be described in the DAP.

9.3 Immunogenicity

9.3.1 Anti-bimekizumab antibodies

The AbAb status will be determined for each visit where samples are taken for drug concentration means. The cut point for determining whether the anti-bimekizumab antibody level is sufficiently high to be considered AbAb positive is not yet known. However, this will be determined prior to the database lock of the PK and AbAb data.

The number and percentage of subjects with AbAb levels above the specified cut point will be reported as:

- Number and percentage of subjects with AbAb level above the specified cut point at the time of each visit
- Number and percentage of subjects with AbAb level above the specified cut point at any visit during the treatment period.

  In addition the timepoint of the first occurrence of AbAb positivity during the treatment period (excluding baseline and pre-treatment) will be summarized for each treatment group.

All individual subject-level AbAb results will be listed.

10 SAFETY ANALYSES

All safety summaries and listings will be performed using all subjects in the SS.

10.1 Extent of exposure

A by-subject listing of treatment administration (including exposure duration and total dose) will be provided.

10.1.1 Duration of exposure

The duration of exposure (in days) will be calculated as:

$$\text{Date of last injection} - \text{Date of first injection} + 28 \quad (10)$$

Should the date of the last injection plus 28 days equate to a date that is beyond the Week 12 visit date then the exposure will be calculated as:

$$\text{Week 12 visit date} - \text{Date of first injection} \quad (11)$$

For subjects who have died the exposure will be as follows:

$$\text{Date of Death - Date of first injection} + 1 \quad (12)$$

The duration of exposure will be summarized by treatment group using descriptive statistics.

10.1.2 Number of doses received

The number of injections received will be summarized by treatment group over the 12 week administration period. There is anticipated to be 3 injections per dosing visit.

10.1.3 Exposure days at risk

For subjects who complete the Week 12 Visit and continue to the extension study, exposure days at risk will be calculated as:

$$Date\ of\ Week\ 12\ Visit^{1} - \text{Date of first administration of study medication} + 1 \quad (13)$$

[1] and the extension study administration of study medication.

For subjects who discontinue on or prior to the Week 12 Visit for reasons other than death, exposure days at risk will be fixed as 84 days (which is equivalent to the intended 12-week initial period). This method is used to provide balance for the comparison of the exposure denominator for the initial period.

For subjects who die prior to the Week 12 Visit, exposure days at risk will be calculated as:

$$\text{Date of death} - \text{Date of first administration of study medication} + 1 \quad (14)$$

10.2 Adverse events

10.2.1 Treatment-emergent adverse events

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

medicinal (investigational) product, whether or not related to the medicinal (investigational) product.

Treatment-emergent AEs (TEAEs) are defined as those AEs that have a start date on or following the first administration of study treatment through the final administration of study treatment + 140 days (covering the 20-week SFU period). If it is not possible (due to partial dates) to determine whether or not an AE is treatment-emergent then it will be assumed to be a TEAE.

To allow for a fair comparison across all subjects, exposure at risk for defining treatment-emergence will be cut off at the Week 12 Visit. Subjects who discontinue prior to the Week 12 Visit will also have their treatment-emergence cut off at Week 12 (84 days). All events that occur on the date of the Week 12 visit, with the exception of those with a HLT of injection-site reactions or any event that meets the criteria for anaphylactic reaction as defined in the AEs of Special Monitoring, will be considered treatment emergent and will be captured as part of the PS0011 study.

An overview of TEAEs that emerge following the 12-week treatment period (Post-Treatment Period) but within <140 days of the last dose (while sill in he PS0010 study) will be summarized separately. However, subjects who consent to the PS0011 extension study will have any Post-Treatment Period AEs reported only within the extension study to prevent duplicate reporting.

Adverse drug reactions (ADRs) are defined as a response to a drug which is noxious and unintended and which occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of disease or for modification of physiological function. Any AE that is considered "Related" to study treatment will be classed as an ADR.

If the intensity of an AE is unknown, it is considered as severe.

If the relationship to study drug is missing, it is considered as related.

Subject time at risk represents the time a subject was at risk for having an AE. The definitions for subject time at risk (in days) are outlined in Section 10.1.3. These definitions will be used for exposure-adjusted AE summaries.

Selected AE summaries will include the exposure adjusted incidence rate (EAIR) with associated 95% confidence interval and the exposure adjusted event rate (EAER).

The EAIR is defined as the number of subjects (n) with a specific AE adjusted for the exposure and will be scaled to 100 patient-years:

$$EAIR = 100 \times n / \sum_{i=1}^{N} \left(T_{Exp(i)}\right) \quad (15)$$

If a subject has multiple events, the time of exposure is calculated to the first occurrence of the AE being considered. If a subject has no events, the total time at risk is used.

Exact Poisson 95% confidence intervals for incidence rates are calculated using the relationship between the Poisson and the Chi-square distribution (Ulm, 1990; Fay and Feuer, 1997):

$$LCL = \chi^2_{2n,\alpha/2}/2 \quad (16)$$

$$UCL = \chi^2_{2(n+1),1-\alpha/2}/2 \quad (17)$$

where n is the number of subjects with a specific AE for the incidence rate of interest and is the basis for the number of the degrees of freedom for the chi-square quantile for the upper tail probability χ^2.

The EAER will be the number of AEs including repeat occurrences in individual subjects divided by the total time at risk scaled to 100 patient-years and calculated using:

$$EAER = 100 \times N_{AE} / \sum_{i=1}^{N} \left(T_{Risk(i)}\right) \quad (18)$$

where N_{AE} is the total number of AEs.

No confidence interval will be computed for EAER.

Subject time at risk represents the time a subject was at risk for having an AE as described in Section 10.1.3.

The following summaries will be provided by treatment group and for the 'All bimekizumab' group. In addition all summaries of TEAEs 'per 100 subject years' will display EAIR and EAER:

- Incidence of TEAEs During Treatment Period – Overview
- Incidence of TEAEs During Post-Treatment Period – Overview (subjects not entering the extension study only)
- Incidence of TEAEs per 100 subject years During Treatment Period by SOC, HLT, and PT
- Incidence of TEAEs per 100 subject years During Post-Treatment Period by SOC, HLT, and PT (subjects not entering the extension study only)
- Incidence of TEAEs per 100 subject years During Treatment Period by SOC, HLT, and PT and by Time of Onset Relative to Anti-bimekizumab Antibody Status
- Incidence of TEAEs During Treatment Period by SOC, HLT and PT
- Incidence of TEAEs During Treatment Period by Relationship by SOC, HLT, and PT
- Incidence of TEAEs During Treatment Period by decreasing frequency of PT
- Incidence of TEAEs During Treatment Period by Severity, SOC, HLT, and PT
- Incidence of TEAEs Above Reporting Threshold of 5% During Treatment Period by SOC, and PT
- Incidence of Non-Serious TEAEs During Treatment Period by SOC, HLT, and PT
- Incidence of Non-Serious TEAEs During Treatment Period by Relationship SOC, HLT, and PT

- Incidence of Non-Serious TEAEs Above Reporting Threshold of 5% During Treatment Period by SOC and PT
- Incidence of Non-Serious TEAEs Above Reporting Threshold of 5% During Treatment Period by Relationship SOC and PT
- Incidence of Adverse Drug Reactions During Treatment Period by SOC, HLT, and PT
- Incidence of Adverse Drug Reactions Above Reporting Threshold of 5% During Treatment Period by PT

All AE summaries will be ordered alphabetically for SOC and HLT within SOC and in terms of decreasing frequency for PT within HLT, in the "All bimekizumab" treatment group.

For each subject and each AE, the worst severity recorded will be attributed and used in the by-severity summaries. Similarly, the worst causality (most related to treatment) will be attributed and used in the by-causality summaries. If severity or causality is missing, the worst case will be assumed.

The following are AEs of special monitoring (AESM) that require special statistical analyses:

1. Serious infections, including TB and opportunistic infections

Serious infections will be identified based on MedDRA classification (SOC "Infections and infestations") using the "Any SAE" table. A separate table does not need to be produced to summarize these events.

Fungal infections will be summarized in a stand-alone table which presents EAIR and EAER. The table will include all TEAEs which code into the High Level Group Term (HLGT) of "Fungal infectious disorders".

Opportunistic infections (including tuberculosis) will be summarized in a stand-alone table which presents EAIR and EAER. The table will include all TEAEs identified using UCB-defined search criteria as described in Section 12.2.

These events will be presented in two stand-alone tables which include EAIR and EAER:

2. Malignancies, including lymphoma.

These events will be presented in two stand-alone tables which include EAIR and EAER:

- One table will be based on the criteria SMQ = "Malignant or unspecified tumours (SMQ)"
- One table will be based on the criteria SMQ="Malignant tumours (SMQ)".

Note that the events included in the "Malignant tumours" table will be a subset of the events included in the "Malignant or unspecified tumours" table. The SMQ search should included 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 High Level Term (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)".

3. Major cardiovascular events

Major cardiovascular events will be presented in a stand-alone table which includes EAIR and EAER. The table will include TEAEs that are 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 SMQs:
  - Haemorrhagic central nervous system vascular conditions (SMQ)
  - Ischaemic central nervous system vascular conditions (SMQ)
- All serious TEAEs which code to a PT included in the HLT "Ischaemic coronary artery disorders" except events coding to PT "Chest Pain" or "Chest discomfort"
- All serious TEAEs which code to a PT of "Cardiac failure congestive".

4. Cytopenias

These events will be presented in a stand-alone table that is based on the SMQ = "Haematopoietic cytopenias". 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.

5. Neuropsychiatric events (in particular depression and suicide)

These events will be presented in a stand-alone table including EAIR and EAER. The table will be based on the SMQ = "Depression and suicide/self-injury (SMQ)". The SMQ search will include all TEAEs which code to a PT included in the Scope=Broad and/or Scope=Narrow groups within the SMQ.

6. Inflammatory bowel disease

These events will be presented in a stand-alone table which includes EAIR and EAER. The table will include all TEAEs which code into the HLGT of "Colitis excl infective".

7. Anaphylactic reaction

Anaphylactic reactions will be summarized together in a stand-alone table.

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

The second row within the body of the table will be labeled "Any hypersensitivity reaction" and will represent the overall incidence of subjects who reported at least one hypersensitivity reaction.

The third row within the body of the table will be labeled "Any anaphylactic reaction" and will represent the overall incidence of subjects who reported at least one anaphylactic reaction.

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

Hypersensitivity reactions and anaphylactic reactions will be identified as follows:

a) Hypersensitivity reactions

All TEAEs that either emerged on the same day as when a study medication injection reaction was received, or that emerged one day after a study medication injection was received, which code to a PT which contains the term "hypersensitivity" will be considered to be a hypersensitivity reaction and included in the summary table as such.

b) Anaphylactic reactions

An algorithmic approach will be used to identify TEAEs that are considered to be anaphylactic reactions. PTs are separated into the 4 distinct categories (A, B, C, D) prior to the algorithmic approach being applied.

All TEAEs that either emerged on the same day as when a study medication injection reaction was received, or that emerged one day after a study medication injection was received, and which fulfill any of the 3 criteria described in Section 12.3 will be included in the summary table.

8. Hepatic events and DILI

Although not officially considered to be AEs of special monitoring but hepatic events are nonetheless considered to be interesting enough to be summarized in stand-alone tables.

Hepatic events will be summarized in a stand-alone table that includes all TEAEs in the SMQ "Drug related hepatic disorders - comprehensive search (SMQ)". Note that the following two sub-SMQs are to be excluded: "Liver neoplasms, benign (incl cysts and polyps)" and "Liver neoplasms, malignant and unspecified (SMQ)".

The SMQ search should include all TEAEs (regardless of whether they have been judged as related to study medication or not) which code to a PT included in the Scope=Narrow group within each SMQ.

Hy's Law cases are to be reported separately in the liver function test summary table (with adjudication for "Potential drug-induced liver injury (PDILI)" cases).

By-subject listings of all AEs, all TEAEs during the treatment period, all TEAEs that occurred within 140 days of the last dose during the post treatment period and within the TEAE window and all non-TEAEs will be provided. A by-subject listing of all ADRs will also be provided. In addition listings will be produced for AESM.

10.2.2 Deaths, serious adverse events, and other significant adverse events

The following summaries will be provided:

- Incidence of Serious TEAEs per 100 subject years During Treatment Period by SOC, HLT, and PT
- Incidence of Serious TEAEs per 100 subject years During Post-Treatment Period by SOC, HLT and PT (subjects not entering the extension study only)

- Incidence of Serious TEAEs During Treatment Period by SOC, HLT, and PT
- Incidence of Serious TEAEs During Treatment Period by Relationship SOC, HLT, and PT
- Incidence of TEAEs Leading to Discontinuation During Treatment Period by SOC, HLT, and PT
- Incidence of TEAEs Leading to Death During Treatment Period by SOC, HLT, and PT
- Incidence of TEAEs Leading to Death by Relationship by SOC, HLT, and PT
- Incidence of Serious Adverse Drug Reactions During Treatment Period by SOC, HLT, and PT

The following listings will be provided:

- A by-subject listing of all serious AEs
- A by-subject listing of all AEs leading to discontinuation of study treatment
- A by-subject listing of all death data

10.3 Clinical laboratory evaluations

All laboratory data recorded in the electronic case report form (eCRF) will be summarized. If any additional analytes to those in Table 10–1 are also recorded then these will be listed only.

Table 10–1: Laboratory measurements		
Hematology	**Chemistry**	**Urinalysis**
Basophils	Calcium	Albumin
Eosinophils	Chloride	Bacteria
Lymphocytes	C-reactive protein	Crystals
Atypical lymphocytes	Magnesium	Glucose
Monocytes	Potassium	pH
Neutrophils	Sodium	Red blood cell
Hematocrit	Glucose	White blood cell
Hemoglobin	Blood urea nitrogen	Urine dipstick for pregnancy testing[1]
Mean corpuscular hemoglobin	Creatinine	
Mean corpuscular hemoglobin concentration	Alkaline phosphatase	
Mean corpuscular volume	Aspartate aminotransferase (AST)	
Platelet count	Alanine aminotransferase (ALT)	
Red blood cell count	Gamma glutamyltransferase (GGT)	
White blood cell count	Total bilirubin	
	Lactate dehydrogenase	
	Total cholesterol	
	Serum pregnancy testing[1]	

[1] Pregnancy testing will consist of serum testing at the Screening and SFU visits. The pregnancy test will be urine at all other visits.

For tables where data are summarized by visit, unscheduled and repeat visits will not be summarized, but these data will be included in listings. For tables where multiple measurements over a period of time are considered (as in shift tables), unscheduled and repeat visits will be considered as long as they were collected in the period being summarized. All summaries will be presented in SI units and will be based on observed case values. In the case where laboratory values are below the lower limit of quantification, then these will be set to the midpoint between 0 and the lower limit of quantification for the purpose of summarizing the data. The following summaries will be provided:

- A summary of the absolute and change from Baseline values in each laboratory variable by treatment group and visit
- A summary of the number and percentage of subjects experiencing markedly abnormal values by laboratory variable, treatment group and visit
- A shift table of the number and percentage of subjects experiencing low, normal or high values at Baseline to maximum post-Baseline value (ie, low, normal, high), by laboratory variable and treatment group

- A Shift table of the number and percentage of subjects experiencing low, normal or high values at Baseline to minimum post-Baseline value (ie, low, normal, high), by laboratory variable and treatment group
- A Shift table of the number and percentage of subjects experiencing low, normal or high values at Baseline to the End of Treatment[1] value (ie, low, normal, high), by laboratory variable and treatment group.

[1]The End of Treatment value refers to the value from the last observed non-missing post baseline visit prior to the end of treatment.

A by-subject listing of all laboratory data will be provided. This listing will be presented by treatment group and will include: center, subject identifier, age, sex, race, weight, visit, laboratory variable, result (with abnormal values flagged as "L" or "H" accordingly) and unit.

A summary table highlighting the potential cases of Hy's Law, within each treatment group will be presented. Hy's Law is defined as.

- AST >3xULN or ALT >3xULN and
- Total Bilirubin >2xULN

Markedly abnormal values are defined as those with a severity of Grade 3 and above based on the common terminology criteria for adverse events (CTCAE) criteria (U.S. Department of Heath and Human Services 2010). Definitions of markedly abnormal values using the Grade 3 cutpoints are given in the tables below for age ranges of ≥17 years (Table 10–2 for markedly abnormal liver function tests values, **Error! Reference source not found.** for markedly abnormal hematology values and Table 10–4 for markedly abnormal biochemistry values). The laboratory results classified as Grade 3 or Grade 4 will be listed separately.

Table 10–2: Definitions of Marked Abnormal Liver Function Tests

Parameter name	Conventional		Standard		
	Unit	Criteria	Unit	Criteria	Abnormal Designation
Alkaline Phosphatase		>5.0 x ULN		>5.0 x ULN	AH
ALT	U/L	>5.0 x ULN	U/L	>5.0 x ULN	AH
AST	U/L	>5.0 x ULN	U/L	>5.0 x ULN	AH
Total Bilirubin	mg/dL	>3.0 x ULN	umol/L	>3.0 x ULN	AH
GGT	U/L	>5.0 x ULN	U/L	>5.0 x ULN	AH

Table 10–3: Definitions of Marked Abnormal Clinical Chemistry Values					
Parameter name	**Conventional**		**Standard**		
	Unit	**Criteria**	**Unit**	**Criteria**	**Abnormal Designation**
Creatinine	mg/dL	>3.0 x ULN	mmol/L	>3.0 x ULN	AH
Glucose	mg/dL	<30 >250	mmol/L	<1.7 >13.9	AL AH
Calcium	mg/dL	>12.5 <7.0	mmol/L	>3.1 <1.75	AH AL
Magnesium	mg/dL	>3.0 <0.9	mmol/L	>1.23 <0.4	AH AL
Potassium	mmol/L	>6.0 <3.0	mmol/L	>6.0 <3.0	AH AL
Sodium	mmol/L	>155 <130	mmol/L	>155 <130	AH AL
Total Cholesterol	mg/dL	>400	mmol/L	>10.34	AH

Table 10–4: Definitions of Marked Abnormal Hematology Values					
Parameter name	**Conventional**		**Standard**		
	Unit	**Criteria**	**Unit**	**Criteria**	**Abnormal Designation**
Hemoglobin	g/dL	<8.0 >4.0 above ULN	g/L	<80 >40 above ULN	AL AH
Lymphocytes Absolute	10^9/L	<0.5 >20.0	10^9/L	<0.5 >20.0	AL AH
Neutrophils Absolute	10^9/L	<1.0	10^9/L	<1.0	AL
Platelets	10^9/L	<50	10^9/L	<50	AL
WBC/Leukocytes	10^9/L	<2.0 >100	10^9/L	<2.0 >100	AL AH

Abbreviations: AH=abnormal high; AL=abnormal low; ALT= alanine aminotransferase; AST = aspartate aminotransferase; dL = deciliter; GGT: gamma-glutamyltransferase; L = liter; mg = milligram; mmol = millimoles; µg = microgram; ULN = upper limit of normal.

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

10.4.1 Vital signs

The following vital signs variables will be summarized: systolic blood pressure (mmHg), diastolic blood pressure (mmHg), body temperature (°C) and heart rate (beats/min). The following summaries will be provided:

- A summary of the absolute and change from Baseline value for each vital sign variable by treatment group and visit
- A summary of the number and percentage of subjects experiencing at least one markedly abnormal value for a vital sign variable as defined in Table 10–5, by treatment group and visit.

Table 10–5: Definitions of markedly abnormal blood pressure values

Parameter (unit)	Markedly Abnormal Low	Markedly Abnormal High
Systolic blood pressure (mmHg)	<90 and a decrease from Baseline of ≥20 <90	>180 and an increase from Baseline of ≥20 >140
Diastolic blood pressure (mmHg)	<50 and a decrease from Baseline of ≥15 <60	>105 and an increase from Baseline of ≥15 >90

A by-subject listing of all vital signs data will be provided. This listing will be presented by treatment group and will include: center, subject identifier, age, sex, race, weight, visit, vital sign variable and result (with abnormal values flagged as "L" or "H" accordingly).

10.4.2 Electrocardiograms

A summary of the number and percentage of subjects with normal, abnormal not clinically significant and abnormal clinically significant ECG results at all applicable visits will be presented.

The following ECG variables will be summarized (absolute values and change from baseline) by visit: QTcF, RR, PR, QRS and QT.

QTc outliers are defined as QTcF values following dosing that are greater than 450 ms or are increases from baseline greater than 30 ms. QTcF outliers will be highlighted in the data listings and summarized using the following categories:

- Values >450 ms, >480 ms, >500 ms
- Increase from baseline of >30 ms, Increase from baseline of >60 ms, Increase from baseline of >90 ms
- Values >450 ms and increases of >30 ms. Values >500 ms and increases of >60 ms

The number and percentage of subjects who meet the ECG outlier criteria at any assessment post-date of first dose will be summarized.

A by-subject listing of all 12-lead ECG data will be provided.

10.4.3 Other safety variables

10.4.3.1 Assessment and management of TB and TB risk factors

A by-subject listing of all of the "Evaluation of signs and symptoms of tuberculosis" questionnaire data will be provided.

10.4.3.2 Electronic Columbia Suicide Severity Rating Scale

The results of the Columbia Suicide Severity Rating Scale (CSSRS) will be summarized using the number of subject and percentage with (i) events in suicide behavior, (ii) suicidal ideation, (iii) suicidal behavior and ideation, and (iv) self-injurious behavior without suicidal intent.

Suicidal ideation is defined as an event in any of the following 4 categories:

- [redacted]

Q3 is not supplied for this study.

Suicidal behavior is defined as an event in any of the following 4 categories:

- Actual attempt
- Interrupted attempt
- Aborted attempt
- Preparatory acts or behavior

Self-injurious behavior without suicidal intent is defined as an event in the category non-suicidal self-injurious injuries.

The incidence of subjects with suicidal intent, suicidal behavior and self-injurious behavior will be summarized by treatment group.

A by-subject listing of the electronic Columbia Suicide Severity Rating Scale questionnaire data will be provided.

10.4.3.3 Physical examination

A summary of the number and percentage of subjects experiencing abnormalities including clinically significant abnormalities by treatment group and visit will be presented.

A by-subject listing of all physical examination data will be provided.

11 REFERENCES

Basra, M. K., Fenech, R., Gatt, R. M., Salek, M. S., & Finlay, A. Y., The Dermatology Life Quality Index 1994-2007: a comprehensive review of validation data and clinical results. Br J Dermatol., 2008; 159, 997-1035.

Daudén, E., Griffiths, C. E., Ortonne, J. P., Kragballe, K., Molta, C. T., 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

Langley, R. G., Feldman, S. R., Han, C., Schenkel, B., Szapary, P., Hsu, M. C., 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, doubleblind, placebo controlled phase III trial. J Am Acad Dermatol., 2010; 63(3), 457-65.

Maruish, M. E., User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: Quality Metric Incorporated, 2011.

Mosteller, R. D., Simplified calculation of body-surface area. The New England Journal of Medicine, 1987, October 22; 317(17), 1098.

Ratitch, B., Lipkovich, I., O'Kelly, M., Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG, 2013.

Rubin D.B. 1987. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons.

Snaith, R. P., & Zigmond, A. S., The hospital anxiety and depression scale, with the irritability-depression-anxiety scale and the Leeds situational anxiety scale manual, 1994.

Ulm, K., A simple method to calculate the confidence interval of a standardized mortality ratio. American Journal of Epidemiology, 1990; 131(12), 373-375

U.S. Department of Health and Human Services, Common Terminology Criteria for Adverse Events (CTCAE), Version 4.0, June 2010.

Ware, J. E., Kosinski, M., Bjorner, J. B., Turner-Bowker, D. M., Gandek, B., & Maruish, M. E., User's Manual for the SF-36v2TM Health Survey (2nd Ed). Lincoln, RI: QualityMetric Incorporated, 2007.

12 APPENDICES

12.1 Classification of the SF-36v2 questionnaire

Items	Scales
[redacted]	Physical Functioning
	Role-Physical
	Bodily Pain
	General Health
	Vitality
	Social Functioning
	Role-Emotional
	Mental Health

12.2 AESM Opportunistic infections

Opportunistic infections are identified in two steps:

Step 1: Refer to column B of the spreadsheet which identifies the Preferred Terms (PTs) to be classified as opportunistic infections using either a single 'x' or a double 'xx'.

- TEAEs which code to a PT flagged with a single 'x' need to also be serious 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:

1. 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 case-by-case. Information from the AE dataset to be included in the spreadsheet: Subject ID, AE verbatim term, SOC, High Level Term (HLT), Lower Level Term (LLT), 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.
2. 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'.
3. Study programming team incorporates these decisions into the AE 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 a case-by-case PT reported that has been confirmed by the study physician to be an opportunistic infection will be summarized as such in the stand-alone table, along with all of the events identified in Step 1 of this process.

The timing and frequency of Step 2 should be outlines and agreed to by the study team at the beginning of the study. It is suggested that this process be executed multiple times throughout the course of the study, more frequently in the weeks leading up to database lock, and one final time immediately prior to database lock.

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.

12.3 MedDRA algorithmic approach to anaphylaxis

The SMQ *Anaphylactic reaction* consists of three parts:

- A **narrow search:** 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

- Anaphylactic reaction (SMQ)
  - Anaphylactic reaction
  - Anaphylactic shock
  - Anaphylactic transfusion reaction
  - Anaphylactoid reaction
  - Anaphylactoid shock
  - Circulatory collapse
  - Dialysis membrane reaction
  - Kounis syndrome
  - Shock
  - Shock symptom
  - Type I hypersensitivity

- A **broad search:** 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

- Acute respiratory failure
- Asthma
- Bronchial oedema
- Bronchospasm
- Cardio-respiratory distress
- Chest discomfort
- Choking
- Choking sensation
- Circumoral oedema
- Cough
- Cyanosis
- Dyspnoea
- Hyperventilation
- Irregular breathing
- Laryngeal dyspnoea
- Laryngeal oedema
- Laryngospasm
- Laryngotracheal oedema
- Mouth swelling
- Nasal obstruction
- Oedema mouth
- Oropharyngeal spasm
- Oropharyngeal swelling
- Respiratory arrest
- Respiratory distress
- Respiratory dyskinesia
- Respiratory failure
- Reversible airways obstruction
- Sensation of foreign body
- Sneezing
- Stridor
- Swollen tongue
- Tachypnoea
- Throat tightness
- Tongue oedema
- Tracheal obstruction
- Tracheal oedema
- Upper airway obstruction
- Wheezing

– Category C

- Allergic oedema
- Angioedema
- Erythema
- Eye oedema
- Eye pruritus
- Eye swelling
- Eyelid oedema
- Face oedema
- Flushing
- Generalised erythema
- Injection site urticaria
- Lip oedema
- Lip swelling
- Nodular rash
- Ocular hyperaemia
- Oedema
- Periorbital oedema
- Pruritus
- Pruritus allergic
- Pruritus generalised
- Rash
- Rash erythematous
- Rash generalised
- Rash pruritic
- Skin swelling
- Swelling
- Swelling face
- Urticaria
- Urticaria papular

– Category D

- Blood pressure decreased
- Blood pressure diastolic decreased
- Blood pressure systolic decreased
- Cardiac arrest
- Cardio-respiratory arrest
- Cardiovascular insufficiency
- Diastolic hypotension
- 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.

13 AMENDMENT TO THE STATISTICAL ANALYSIS PLAN

13.1 Amendment 1

Rationale for the amendment

The purpose of this amendment is to apply modifications as a result of conclusions made at Data Evaluation Meeting 2 including changing LOCF methods to MI methods.

Global changes

Various spelling and formatting amendments were made.

Due to the small population size in the Asia region, Asia was combined with Europe in all cases.

Specific changes

Change #1

Page 10, Section 2.2.2.1 Pharmacokinetic variables, removed "CL/F" and "V/F" from the PK variables collected.

Change #2

Page 13, Section 3.1 General Presentation of Summaries and Analyses, added "Derived variables in general will display the mean, SD and median to 1 more decimal place than the variables used in the derivation. If the number of decimal places reported in the raw data is varied then use either the maximum raw number of reported decimal places or 3, whichever is the lowest, as a guide for the descriptive statistics.".

Change #3

Page 14, Section 3.3. Mapping of assessments performed at early stage study withdrawal visit, added in the following two sections of text.

"The only exception to this rule is for anti-bimekizumab antibody (AbAb) assessments where all premature withdrawal visit assessments will be assigned to the next scheduled visit at which AbAb are assessed. All by-visit summaries will contain nominal visits only. Unscheduled visits will not be mapped to scheduled visits."

"in observed case summaries. For summaries based on imputed data, the Week 12 value will be treated as missing and imputed according to the rules for the given variable."

Change #4

Page 15, Section 3.5.6.1 Pharmacokinetic Per Protocol Set, added the following to the end of the sentence "and had no important protocol deviations affecting the pharmacokinetic variables, as confirmed during ongoing data cleaning meetings prior to database lock.".

Change #5

Page 17, Section 3.11 Changes to protocol-defined analyses, added the following text into this section.

"Section 14.6 of the protocol (Handling of dropouts or missing data) describes the sensitivity analysis to be used in this study. More specifically, the protocol describes a sensitivity analysis using multiple imputation as the method for handling missing data and states that the MCMC

method will be used for intermittent missing data and that monotone regression will be used for monotone missing data. This SAP has been modified such that the MCMC method will be used for all missing data whether intermittent or monotone.

Section 14.6 of the protocol also notes that the missing data for all continuous efficacy variables that have not been specifically described elsewhere will be imputed using the LOCF approach. This has been updated in this SAP so that these continuous variables will in fact be imputed using the multiple imputation method. However, in the event that multiple imputation is not possible (eg, Due to small group sizes or insufficient missing data) then the LOCF method will be used. ”

Change #6

Page 18, Section 4.2.3 Handling missing data for other efficacy variables, the text was changed.

From

For the dose/exposure response model, missing data will not be imputed. If a subject drops out, this will be used as a covariate in the model.

Other categorical efficacy variables will be imputed using NRI and observed cases.

To

For the dose/exposure response model, missing data will not be imputed. If a subject drops out, this will be used as a covariate in the model. However, the dose/exposure response model is outside of the scope of this SAP and is described in greater detail in a separate data analysis plan (DAP).

Other categorical efficacy variables will be imputed using NRI (ie. Missing data or subjects who prematurely discontinue study treatment prior to Week 12 will be counted as non-responders) and summarized as observed cases. One exception to this rule is the imputation of missing data when summarizing HADS-A and HADS-D scores below 8 (ie. Normal HADS scores). In these summaries, LOCF will be used. That is, the last available HADS-A/HADS-D score while the subject was on study treatment will be carried forward to the missing time points to determine whether or not the subject had normal HADS scores.

From

Other continuous efficacy variables will be imputed using last observation carried forward (LOCF) where the last observed value of a variable prior to discontinuation (including Baseline) is used to impute the missing data. Summaries based on observed case data will also be prepared.

To

Continuous efficacy variables will be imputed using multiple imputation (MI) via the Markov-Chain Monte Carlo (MCMC) method. Summaries based on observed case data will also be prepared.

The MI procedure for continuous efficacy variables (based on MCMC) will be applied as follows:

1. Create a dataset, one for each treatment group, of subjects with observed values and those needing estimation by multiple imputation. The missing values in each dataset will be

filled in using the MCMC method with a total of 100 sets of imputations being performed. Biologic exposure, geographic region and values at baseline and at each post-baseline visit (in chronological order) will be included in the implementation model, if required, in order to facilitate the imputation model, variables can be omitted in the order of biologic exposure and geographical region. Cut-off values will be applied to the imputed data, as appropriate, to ensure that imputed values do not exceed the range of plausible values for the variable being imputed. The resulting datasets for each treatment arm will be combined into one complete dataset based on each of the 100 imputations.

2. The relevant change from Baseline (or percent change from Baseline) values will then be derived based on the datasets which include observed and imputed values for the variable being considered
3. The results from each of the 100 imputed datasets will be combined for the calculation of means and standard errors using Rubin's rules, which account for the uncertainty associated with the imputed values (Rubin, 1987). This will be done using SAS PROC MIANALYZE. For calculation of other descriptive statistics such as the median, min and max, Rubin's rules do not apply. Multiple imputation estimates will be computed by simply averaging the estimates from the $m = 1, \ldots, M$ independent repetitions of the imputation algorithm:

$$\bar{\theta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m$$

where $\hat{\theta}_m$= estimate of θ from the completed dataset $m = 1, \ldots, M$.

There may be cases where the multiple imputation model fails to converge (eg, sparse subgroups), in such situations the LOCF approach will instead be used to impute the missing data. If LOCF is used instead of multiple imputation for this reason, this will be clearly specified in the corresponding table summary. Should there be no missing data for a study variable then only observed case data will be produced.

Summaries for all variables based on observed case data will also be prepared.

Change #7

Page 20, Section 4.3 Interim analyses and data monitoring, added the following paragraph.

The database lock to support the development of the PS0010 interim CSR will occur before all subjects have completed the PS0010 study. Specifically, subjects who do not enter the extension study (PS0011) will need to complete the SFU Visit, which occurs 20 weeks after the last dose of study treatment. The Week 12 interim database lock will be performed based on the last subject completing the Week 12 visit. It is anticipated that there will be some subjects still awaiting the SFU at that time. The number of subjects in that position is expected to be low, and the minimal data to be collected from their SFU visits is not considered mandatory to evaluating the key efficacy and safety objectives of the study, therefore it is considered acceptable to proceed with the PS0010 data analysis and interim CSR development in the absence of these data. Once all SFU data have been collected and the full PS0010 study database has been locked, all TLFs will be re-run to include the SFU data that have been added to the database following the Week 12 interim database lick. A final CSR will subsequently be prepared based on this final set of TLFs.

Other than the additional data from the pending SFU visits, it is expected that this final CSR will include biomarker and flow cytometry data summaries that will not be available in the interim CSR..

Change #8

Page 21, Section 4.8 Examination of subgroups, removed the subgroup "Anti-bimekizumab antibody positivity at Baseline (Yes, No)".

Change #9

Page 22, Section 6.1 Demographics, amended the first age category by changing "≥65-<85, ≥85" to "≥65" in order to match exactly EudraCT age categorizations.

Change #10

Page 23, Section 6.2 Other Baseline Characteristics, removed "SF-36 Total score" and "Anti-bimekizumab antibody positivity at Baseline (Yes, No)" from the list of variables to be summarized. In addition height, weight and BMI have been moved to Section 6.1 Demographics.

Change #11

Page 26, Section 8.1.2 Primary analysis of the primary efficacy variable, added in the 320mg loading dose followed by 160mg Q4W into the primary model with a linear contrast coefficient of 0. In addition, added the following text to the end of the section "The results produced in the SAS modelling process for the primary endpoint will be provided as a supportive output."

Change #12

Page 27, Section 8.1.3 Secondary analyses of the primary efficacy variable, added the following text to the end of the section "The results produced in the SAS modelling process for the secondary analyses of the primary efficacy variable will be provided as a supportive output.".

Change #13

Page 29, Section 8.1.4.2 Supportive Analysis, included an additional sensitivity analysis to be based on LOCF for the subjects that were withdrawn due to withdrawal criterion #11.

Change #14

Page 40, Section 10.1.1 Duration of exposure – Equation 13, removed "Date of last injection" from the equation.

Change #15

Page 40, Section 10.1.3 Exposure days at risk – Equations 14 and 15 added a "+1" to the deviation.

Change #16

Page 41, Section 10.2.1 Treatment-emergent Averse Events, added to the exclusion criteria for considering AEs that occur on the date of the Week 12 Visit to be treatment emergent to include any AE that met the criteria for anaphylactic reaction

Change #17

Page 43, Section 10.2.1 Treatment-emergent Averse Events, changed the reporting level from 10% to 5% in all cases.

Change #18

Page 47, Section 10.3 Clinical laboratory evaluations Table 10-2 – Definitions of Marked Abnormal Liver Function Tests, the abnormal ranges were amended from ≥5 and ≥3 to >5 and >3 respectively.

Change #19

Page 49, Section 10.3 Clinical laboratory evaluations Table 10-3 – Definitions of Marked Abnormal Clinical Chemistry Tests, the standard units for the chemistry parameters have been changed from umol/L to mmol/L. In addition the abnormal high for calcium was corrected from >12.5 to >3.1. In addition Albumin was removed from the table.

Change #20

Page 50, Section 10.4.1 Vital signs Table 10-5 – Definitions of markedly abnormal blood pressure values. Added in the additional lower and upper limits for systolic blood pressure and diastolic blood pressure.

Change #21

Page 51, Section 10.4.3.2 Electronic Columbia Suicide Severity Rating Scale, section has been re-written to contain more detail.

From

The incidence of subjects with suicidal ideation, behavior and injuries will be summarized by treatment group and visit

To

The results of the Columbia Suicide Severity Rating Scale (CSSRS) will be summarized using the number of subject and percentage with (i) events in suicide behavior, (ii) suicidal ideation, (iii) suicidal behavior and ideation, and (iv) self-injurious behavior without suicidal intent.

Suicidal ideation is defined as an event in any of the following 4 categories:

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

Q3 is not supplied for this study.

Suicidal behavior is defined as an event in any of the following 4 categories:

- Actual attempt
- Interrupted attempt
- Aborted attempt

- Preparatory acts or behavior.

Self-injurious behavior without suicidal intent is defined as an event in the category non-suicidal self-injurious injuries.

The incidence of subjects with suicidal intent, suicidal behavior and self-injurious behavior will be summarized by treatment group.

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.

PS0010 Statistical Analysis Plan

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Approval Date (dd-mon-yyyy (HH:mm))
[redacted]	Clinical Approval	30-Jan-2018 15:54 GMT+01
	Clinical Approval	01-Feb-2018 23:08 GMT+01