STATISTICAL ANALYSIS PLAN

Study: HS0003

Product: Bimekizumab

A PHASE 3, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDY EVALUATING THE EFFICACY AND SAFETY OF BIMEKIZUMAB IN STUDY PARTICIPANTS WITH MODERATE TO SEVERE HIDRADENITIS SUPPURATIVA

SAP/Amendment Number	Date
Final SAP	3 Apr 2020
SAP Amendment 1	9 Dec 2021
SAP Amendment 2	11 Jul 2022
SAP Amendment 3	07 Oct 2022

Confidentiality Statement

TABLE OF CONTENTS

LIST OF ABBREVIATIONS 9
1 INTRODUCTION 13
2 PROTOCOL SUMMARY 13
2.1 Study objectives 13
2.1.1 Primary objective 13
2.1.2 Secondary objectives 13
2.1.3 Other objectives 13
2.1.4 Exploratory objective 13
2.2 Study endpoints 14
2.2.1 Efficacy endpoints 14
2.2.1.1 Primary efficacy endpoint 14
2.2.1.2 Secondary efficacy endpoints 14
2.2.1.3 Other efficacy endpoints 14
2.2.2 Pharmacokinetic and pharmacogenomic endpoints 16
2.2.2.1 Pharmacokinetic endpoints 16
2.2.2.2 Exploratory pharmacogenomic endpoints 16
2.2.3 Safety endpoints 16
2.2.3.1 Secondary safety endpoints 16
2.2.3.2 Other safety endpoints 16
2.2.4 Immunological endpoints 17
2.3 Study design and conduct 17
2.3.1 Study description 17
2.3.2 Study periods 18
2.3.2.1 Screening Period 18
2.3.2.2 Initial Treatment Period (Weeks 0-16) and Maintenance Treatment Period (Weeks 16-48) 18
2.3.2.3 Safety Follow-up Visit 18
2.3.3 Study duration per participant 18
2.3.4 Planned number of participants and sites 19
2.3.5 Anticipated regions and countries 19
2.4 Determination of sample size 19
3 DATA ANALYSIS CONSIDERATIONS 20
3.1 General presentation of summaries and analyses 20
3.2 General study level definitions 22
3.2.1 Relative day 22
3.2.2 Mapping of data from Premature End of Treatment visits 23
3.3 Definition of Baseline values 23

3.4 Protocol deviations 24
3.5 Analysis sets 24
3.5.1 Enrolled Set 24
3.5.2 Randomized Set 24
3.5.3 Safety Set 24
3.5.4 Full Analysis Set 24
3.5.5 Active Medication Set 24
3.5.6 Maintenance Set 24
3.5.7 Per-Protocol Set 24
3.5.8 Pharmacokinetics Per-Protocol Set 25
3.5.9 COVID-19 Free Set 25
3.6 Treatment assignment and treatment groups 25
3.7 Center pooling strategy 25
3.8 Coding dictionaries 26
3.9 Definition of an intercurrent event 26
3.10 Changes to protocol-defined analyses 27
4 STATISTICAL/ANALYTICAL ISSUES 27
4.1 Adjustments for covariates 27
4.2 Handling of dropouts or missing data 28
4.2.1 Efficacy data 28
4.2.1.1 Handling missing data for the primary efficacy endpoint 28
4.2.1.2 Handling missing data for the secondary efficacy endpoints 28
4.2.1.3 Handling missing data for the other efficacy endpoints 28
4.2.1.4 Missing Data Overview and Summary 29
4.2.2 Missing data algorithms for efficacy analyses 30
4.2.2.1 MI – MCMC / Monotone Regression 31
4.2.2.2 MI – MCMC / Reference-based imputation 36
4.2.2.3 Tipping Point Analysis 37
4.2.3 Rationale for estimand 39
4.2.3.1 Composite estimand 39
4.2.3.2 Hypothetical estimand 39
4.2.4 Dates and times 40
4.3 Interim analyses and data monitoring 41
4.3.1 Data monitoring committee 41
4.3.2 Interim analysis 42
4.4 Multicenter studies 42
4.5 Multiple comparisons/multiplicity 42
4.6 Use of an efficacy subset of participants 44

4.7 Active-control studies intended to show equivalence 44
4.8 Examination of subgroups 44
5 STUDY POPULATION CHARACTERISTICS 45
5.1 Study participant disposition 45
5.2 Impact of COVID-19 46
5.3 Protocol deviations 46
6 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS 46
6.1 Demographics 46
6.2 Other Baseline characteristics 47
6.3 Medical history and concomitant diseases 48
6.4 Prior and concomitant medications 48
6.4.1 Assignment of medications to study period 49
6.4.2 Classification of participants as analgesic, antihistamine users 50
7 MEASUREMENTS OF TREATMENT COMPLIANCE 50
8 EFFICACY ANALYSES 51
8.1 Lesion count assessment 51
8.2 Primary efficacy endpoint 52
8.2.1 Derivation of $HiSCR_{50}$ at Week 16 56
8.2.2 Primary analysis of the primary efficacy endpoint 57
8.2.3 Sensitivity analyses of the primary efficacy endpoint 58
8.2.3.1 Nonresponse imputation 58
8.2.3.2 MI-MCMC / Reference-based imputation 58
8.2.3.3 Tipping point analysis 58
8.2.3.4 Treatment policy 59
8.2.3.5 Analysis on observed cases 59
8.2.3.6 Analysis on FAS 59
8.2.3.7 Analysis on PPS 59
8.2.3.8 Analysis on CFS 59
8.2.3.9 Analysis including COVID-19 impact as intercurrent event 59
8.2.3.10 Cochran-Mantel-Haenszel test 60
8.2.3.11 Center-by-Treatment Interaction 60
8.3 Secondary efficacy endpoints 61
8.3.1 $HiSCR_{75}$ at Week 16 65
8.3.2 Change from Baseline in DLQI Total Score at Week 16 65
8.3.2.1 Primary analysis of change from Baseline in DLQI Total score at Week 16 67
8.3.2.2 Sensitivity analysis of change from Baseline in DLQI Total score at Week 16 67

8.3.3 Change from Baseline in Skin Pain score at Week 16, as assessed by the "worst skin pain" item in the HSSDD 67
8.3.3.1 Primary analysis of change from Baseline in skin pain score at Week 16 68
8.3.3.2 Sensitivity analysis of change from Baseline in skin pain score at Week 16 68
8.3.4 HSSDD skin pain response at Week 16 68
8.3.4.1 Primary analysis of skin pain response at Week 16 68
8.3.4.2 Sensitivity analyses of Skin Pain Response at Week 16 69
8.4 Other efficacy endpoints 69
8.4.1 HiSCR endpoints 69
8.4.1.1 $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ 69
8.4.1.2 Time to response of $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ ..... **Error! Bookmark not defined.**
8.4.1.3 $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response at both Weeks 16 and 48 71
8.4.1.4 $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ among Week 16 Responders 71
8.4.1.5 Time to loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ in Week 16 responders 71
8.4.1.6 Partial response 72
8.4.2 Lesion count 72
8.4.2.1 Change from Baseline in lesion count 72
8.4.2.2 AN count of 0, 1, or 2 73
8.4.2.3 AN25, AN50, AN75, AN90, AN100 73
8.4.3 Flare by Week 16 73
8.4.4 Flare relative to Baseline 74
8.4.5 Time to flare by Week 16 74
8.4.6 Time to flare by Week 48 74
8.4.7 International Hidradenitis Suppurativa Severity Score System (IHS4) 75
8.4.8 HS-Physician's Global Assessment 6-point scale 75
8.4.9 High Sensitivity C-Reactive Protein (hs-CRP) 78
8.4.10 Initiation of systemic antibiotic rescue therapy 79
8.4.11 Time to initiation of systemic rescue therapy in the Initial Treatment Period 79
8.4.12 Time to an intercurrent event in the Initial Treatment Period 79
8.4.13 Hidradenitis Suppurativa Symptom Daily Diary (HSSDD) 80
8.4.14 Hidradenitis Suppurativa Symptom Questionnaire (HSSQ) 81
8.4.15 DLQI 82
8.4.16 Hidradenitis Suppurativa Quality of Life (HiSQOL) 82
8.4.17 Patient Global Impression of HS Severity (PGI-S-HS) 84
8.4.18 Patient Global Impression of Change in HS Severity (PGI-C-HS) 84

8.4.19 Patient Global Impression of Severity of Skin Pain (PGI-S-SP) 84
8.4.20 Patient Global Impression of Change in Severity of Skin Pain (PGI-C-SP) 84
8.4.21 Euro-Quality of Life 5-Dimensions, 3 levels (EQ-5D-3L) 85
8.4.22 Work Productivity and Activity Impairment Questionnaire–Specific Health Problem (WPAI-SHP) v2.0 adapted to HS scores 85
8.4.23 Treatment Satisfaction Questionnaire – Medication-9 86
8.4.24 Lesion intervention 87
8.5 Additional statistical analyses of other efficacy endpoints 87
9 PHARMACOKINETICS AND PHARMACODYNAMICS 88
9.1 Pharmacokinetics 88
9.2 Pharmacodynamics 89
9.3 Immunogenicity 89
9.3.1 Autoantibodies 89
9.3.2 Anti-bimekizumab antibodies 89
10 SAFETY ANALYSES 93
10.1 Extent of exposure 93
10.1.1 Exposure during the Initial Treatment Period 94
10.1.1.1 Study medication duration (days) 94
10.1.1.2 Time at risk (days) 94
10.1.2 Exposure during the Maintenance Treatment Period 95
10.1.2.1 Study medication duration (days) 95
10.1.2.2 Time at risk (days) 95
10.1.3 Exposure during the Initial and Maintenance Treatment Period 96
10.1.3.1 Study medication duration (days) 96
10.1.3.2 Time at risk (days) 96
10.2 Adverse events 97
10.2.1 Data considerations 97
10.2.1.1 COVID-19 related considerations 99
10.2.2 AE summaries 99
10.2.3 Other Safety topics of interest 101
10.2.3.1 Infections (serious, opportunistic, fungal and TB) 101
10.2.3.2 Malignancies 104
10.2.3.3 Major adverse cardiac event 104
10.2.3.4 Neutropenia 106
10.2.3.5 Suicidal Ideation and Behavior 106
10.2.3.6 Inflammatory bowel disease 107
10.2.3.7 Hypersensitivity (including anaphylaxis) 108
10.2.3.8 Hepatic events and PDILI 109

10.3 Clinical laboratory evaluations 109
10.4 Vital signs, physical findings, and other observations related to safety 112
10.4.1 Vital signs 112
10.4.2 Electrocardiograms 113
10.4.3 Other safety endpoints 113
10.4.3.1 Physical examination 114
10.4.3.2 Columbia-Suicide Severity Rating Scale (C-SSRS) 114
10.4.3.3 Assessment and management of TB and TB risk factors 114
10.4.3.4 Pregnancy testing 115
10.4.3.5 Patient Health Questionnaire (PHQ)-9 scores 115
11 REFERENCES 116
12 APPENDICES 118
12.1 Appendix A: MedDRA algorithmic approach to anaphylaxis 118
12.2 Appendix B: Definition of CTCAE grades 120
13 AMENDMENT TO THE STATISTICAL ANALYSIS PLAN 122
13.1 Amendment 1 122
13.2 Amendment 2 175
13.3 Amendment 3 175
STATISTICAL ANALYSIS PLAN SIGNATURE PAGE 216

LIST OF TABLES

Table 2–1: Power calculation assumptions and methods 20
Table 3–1: Geographic regions and corresponding countries 25
Table 4–1: Missing data handling approach by endpoint priority and type 30
Table 4–2: Imputation allowable ranges by variable 32
Table 8–1: Estimand Details and Attributes for Primary Endpoint 52
Table 8–2: Estimand Details and Attributes for Secondary Analyses 61
Table 8–3: Dermatology Life Quality Index 66
Table 8–4: HS-Physician's Global Assessment 6-point scale 76
Table 8–5: Hidradenitis Suppurative Quality of Life 83
Table 10–1: Cardiovascular event classifications 105
Table 10–2: Neuropsychiatric event classifications 107
Table 10–3: IBD event classifications 108
Table 10–4: Definitions of Markedly Abnormal Liver Function Values 110
Table 10–5: Definitions of Markedly Abnormal Biochemistry Values 111

Table 10–6: Definitions of Markedly Abnormal Hematology Values 111
Table 10–7: Definitions of Markedly Abnormal Blood Pressure Values 113
Table 12–1: Definition of CTCAE grades by biochemistry parameters 120
Table 12–2: Definitions of CTCAE grades by hematology parameter 120

LIST OF FIGURES

Figure 2-1: Study Schematic 17
Figure 4-1: Sequence of testing 43

LIST OF ABBREVIATIONS

List of Abbreviations

%ΔAN	percentage change from Baseline in abscess and inflammatory nodule count
ADAb	anti-bimekizumab antibodies
AE(s)	adverse event(s)
ALP	alkaline phosphatase
ALQ	above the limit of quantification
ALT	alanine aminotransferase
AMS	Active Medication Set
AN	abscess and inflammatory nodule
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
BKZ	bimekizumab
BLQ	below the limit of quantification
CFB	change from Baseline
CFS	COVID-19 Free Set
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
COVID-19	coronavirus disease 2019
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
CV-CAC	Cardiovascular Event Adjudication Committee
DLQI	Dermatology Life Quality Index
DMC	data monitoring committee
eCDF	empirical cumulative distribution function
ECG	electrocardiogram
eCRF	electronic case report form
eC-SSRS	electronic Columbia-Suicide Severity Rating Scale
EAER	exposure adjusted event rate
EAIR	exposure adjusted incident rate

List of Abbreviations

EQ-5D-3L	European Quality of Life-5 Dimensions-3 Level questionnaire
ES	Enrolled Set
FAS	Full Analysis Set
FDA	Food and Drug Administration
geoCV	geometric coefficient of variation
GGT	gamma-glutamyltransferase
HiSCR	Hidradenitis Suppurativa Clinical Response
HiSQOL	Hidradenitis Suppurativa Quality of Life
HLT	high level term
HS	hidradenitis suppurativa
hs-CRP	high sensitivity C-reactive protein
HSSDD	Hidradenitis Suppurativa Symptom Daily Diary
HSSQ	Hidradenitis Suppurativa Symptom Questionnaire
IBD	Inflammatory bowel disease
IBD-CAC	Inflammatory Bowel Disease Adjudication Committee
ICF	Informed Consent Form
ICH	International Council for Harmonisation
IDC	Infectious Disease Committee
IGRA	interferon gamma release assay
IHS4	International Hidradenitis Suppurativa Severity Scoring System
IMP	investigational medicinal product
LFT	liver function tests
LLOQ	lower limit of quantification
LOCF	last observation carried forward
LSM	least square mean
MACE	major cardiovascular events
MAR	missing at random
MCID	minimal clinically important difference
MCMC	Markov-Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MI-MCMC	multiple imputation Markov-Chain Monte Carlo

List of Abbreviations

MS	Maintenance Set
MSR	minimum significant ratio
n	number of study participants
NAb	neutralizing antibody
NI	Negative Immunodepletion
nR	New Ratio
NRI	nonresponder imputation
NRS	numeric rating scale
NS	Negative Screen
OC	observed case
PD	pharmacodynamic(s)
pDILI	potential drug induced liver injury
PEOT	premature end of treatment
PGI-C-HS	Patient Global Impression of Change in Hidradenitis Suppurativa Severity
PGI-C-SP	Patient Global Impression of Change in Severity of Skin Pain
PGI-S-HS	Patient Global Impression of Hidradenitis Suppurativa Severity
PGI-S-SP	Patient Global Impression of Severity of Skin Pain
PHQ-9	Patient Health Questionnaire 9
PI	Positive Immunodepletion
PK	pharmacokinetic(s)
PK-PPS	Pharmacokinetic Per-Protocol Set
PPS	Per-Protocol Set
PS	Positive Screen
PT	preferred term
Q2W	every 2 weeks
Q4W	every 4 weeks
QTcF	QT corrected for heart rate using Fridericia's formula
RS	Randomized Set
SAP	statistical analysis plan
SD	standard deviation
SE	standard error

List of Abbreviations

SFU	Safety Follow-up
SIB	suicidal ideation and behavior
SMQ	standardized MedDRA query
SOC	system organ class
SS	Safety Set
SSD	Safety Signal Detection
TEAE	treatment-emergent adverse event
TEMA	treatment-emergent markedly abnormal
TSQM-9	Treatment Satisfaction Questionnaire – Medication 9
ULN	upper limit of normal
VAS	visual analogue scale
WHODD	World Health Organization Drug Dictionary
WPAI-SHP	Work Productivity and Activity Impairment Questionnaire-Specific Health Problem

1 INTRODUCTION

This statistical analysis plan (SAP) defines the scope of statistical analyses and provides a detailed description of statistical methodology to support the final clinical study report (CSR).

The SAP is based on the Protocol Amendment 5, 27 September 2022. If a future protocol amendment necessitates a substantial change to the statistical analysis of the study data, this SAP may be amended accordingly. The content of this SAP is compatible with the International Council for Harmonisation (ICH)/Food and Drug Administration (FDA) E9 Guidance documents.

2 PROTOCOL SUMMARY

2.1 Study objectives

2.1.1 Primary objective

The primary objective of this study is to evaluate the efficacy of bimekizumab in study participants with moderate to severe hidradenitis suppurativa (HS).

2.1.2 Secondary objectives

The secondary objectives of this study are to:

- Evaluate the efficacy of bimekizumab on other measures of disease activity in study participants with moderate to severe HS
- Evaluate the safety of bimekizumab in study participants with moderate to severe HS

2.1.3 Other objectives

The other objectives of this study are to:

- Evaluate the efficacy of bimekizumab on Hidradenitis Suppurativa Clinical Response (HiSCR), other HS Scores, and other clinical measures of disease activity at various timepoints in study participants with moderate to severe HS
- Evaluate the efficacy of bimekizumab on abscesses, nodules, and draining tunnels at various timepoints in study participants with moderate to severe HS
- Evaluate the efficacy of bimekizumab on patient-reported outcome measures at various timepoints in study participants with moderate to severe HS
- Evaluate the effect of bimekizumab on other safety measures at various timepoints in study participants with moderate to severe HS
- Evaluate the pharmacokinetics (PK) of bimekizumab in study participants with moderate to severe HS
- Evaluate the immunogenicity of bimekizumab (antidrug antibodies) in study participants with moderate to severe HS

2.1.4 Exploratory objective

The exploratory objective of the study is to evaluate biomarkers in study participants with moderate to severe HS.

2.2 Study endpoints

2.2.1 Efficacy endpoints

2.2.1.1 Primary efficacy endpoint

The primary efficacy endpoint is the $HiSCR_{50}$ (defined as at least a 50% reduction from Baseline in the total abscess and inflammatory nodule [AN] count with no increase from Baseline in abscess or draining tunnel count) at Week 16.

2.2.1.2 Secondary efficacy endpoints

The secondary efficacy endpoints are defined as:

- $HiSCR_{75}$ response (defined as at least a 75% reduction from Baseline in the total AN count with no increase from Baseline in abscess or draining tunnel count) at Week 16
- Absolute change from Baseline in Dermatology Life Quality Index (DLQI) Total Score at Week 16
- Absolute change from Baseline (CFB) in Skin Pain score at Week 16, as assessed by the "worst skin pain" item (11-point numeric rating scale) in the HS Symptom Daily Diary (HSSDD)
- Skin pain response based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) at Week 16 among study participants with a score of ≥3 at Baseline

2.2.1.3 Other efficacy endpoints

The other efficacy endpoints are defined as:

- Time to response of $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$
- $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$
- Absolute change from Baseline in International Hidradenitis Suppurativa Severity Score System (IHS4)
- Change from Baseline in the HS-Physician's Global Assessment 6-point scale
- Absolute and percentage change from Baseline in high-sensitivity C-reactive protein (hs-CRP)
- Initiation of systemic antibiotic rescue therapy
- $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ at both Weeks 16 and 48
- Time to loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ in Week 16 responders
- Partial responders (defined as a ≥25% reduction in AN count from Baseline [Week 0]) at Week 16 who progress to $HiSCR_{50}$ during the Maintenance Treatment Period
- Partial responders (defined as a ≥25% reduction in AN count from Baseline [Week 0]) at Week 16 who progress to $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response during the Maintenance Treatment Period

- Change and percentage change from Baseline in lesion counts (abscess count, inflammatory nodule count, AN count, and draining tunnel count)
- AN count of 0, 1, or 2
- AN_{25}, AN_{50}, AN_{75}, AN_{90}, AN_{100} (defined as a 25%, 50%, 75%, 90%, 100% reduction in the total AN count relative to Baseline)
- Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Baseline) by Week 16
- Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Week 16) by Week 48
- Time to flare from Weeks 0 to 16
- Time to flare from Week 16 to 48
- Absolute and percentage change (worst and average skin pain) from Baseline in HS Skin Pain score (11-point numeric rating scale)
- Skin pain response based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) among study participants with a score of ≥3 at Baseline
- Skin pain response (at least a 30% reduction and at least a 1-point reduction from Baseline in HSSDD weekly worst skin pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline
- Skin pain response (at least a 30% reduction and at least a 1-point reduction from Baseline in HSSDD weekly average skin pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline
- Skin pain response (at least a 30% reduction and at least a 1-point reduction from Baseline in HS Skin Pain score [11-point numeric rating scale]) assessed by the HSSQ among study participants with a score of ≥3 at Baseline
- Absolute change from Baseline in DLQI Total Score
- DLQI Total Score of 0 or 1
- Minimum clinically important difference (MCID) (improvement from Baseline of 4 or more) in the DLQI Total Score among study participants with a Baseline score of at least 4
- Absolute change from Baseline in Hidradenitis Suppurativa Quality of Life (HiSQOL) domain scores (symptoms, psychosocial, activities and adaptations) and Total score
- Patient Global Impression of HS Severity (PGI-S-HS)
- Patient Global Impression of Change of HS Severity (PGI-C-HS)
- Patient Global Impression of Severity of Skin Pain (PGI-S-SP)
- Patient Global Impression of Change of Skin Pain (PGI-C-SP)

- Absolute change from Baseline in each of the other HS Symptoms - itch, drainage or oozing of HS lesions, and smell or odor
- Response on other HS Symptoms (11-point numeric rating scale) - itch, drainage or oozing of HS lesions, and smell or odor
- Responses to the European Quality of Life-5 Dimensions-3 Level questionnaire (EQ-5D-3L), absolute and changes from Baseline in EQ-5D-3L visual analog scale (VAS) scores
- Absolute change from Baseline in Work Productivity and Activity Impairment Questionnaire-Specific Health Problem (WPAI-SHP) v2.0 adapted to HS scores
- Domain Scores (effectiveness, convenience and global satisfaction) on the Treatment Satisfaction Questionnaire – Medication-9 (TSQM-9)

2.2.2 Pharmacokinetic and pharmacogenomic endpoints

2.2.2.1 Pharmacokinetic endpoints

The PK endpoint is the plasma bimekizumab concentrations over the study duration.

2.2.2.2 Exploratory pharmacogenomic endpoints

[redacted]

A specific SAP will be written to describe the analysis methods for those endpoints, as the results will not be summarized in the CSR. The nature and format of these analyses will be detailed in this SAP.

2.2.3 Safety endpoints

2.2.3.1 Secondary safety endpoints

The secondary safety endpoints are

- Treatment-emergent Adverse Events (TEAEs)
- Serious TEAEs
- TEAEs leading to withdrawal from study

2.2.3.2 Other safety endpoints

The other safety endpoints are

- Adverse events of special interest (Hy's Law)
- Other safety topics of interest: infections (serious, opportunistic, fungal, and TB), neutropenia, hypersensitivity (including anaphylaxis), suicidal ideation and behavior, major adverse cardiovascular events, hepatic events and potential drug-induced liver injury (PDILI), malignancies, and inflammatory bowel disease.

- Absolute change from Baseline in the Patient Health Questionnaire (PHQ-9) score
- Absolute change from Baseline in vital signs
- Absolute change from Baseline in clinical laboratory values (chemistry and hematology)
- Electrocardiogram (ECG) results

2.2.4 Immunological endpoints

The immunological endpoints are

- Bimekizumab antidrug antibodies
- Bimekizumab neutralizing antibodies

The results of the bimekizumab neutralizing antibody analysis will not be summarized in the CSR for this study. All neutralizing antibody analyses will be detailed in the integrated immunogenicity SAP.

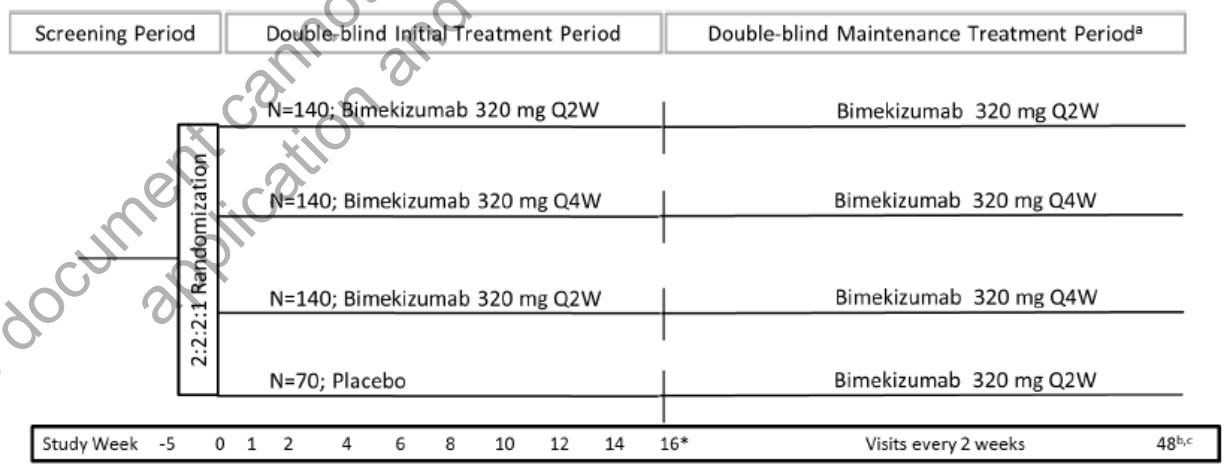
2.3 Study design and conduct

2.3.1 Study description

HS0003 is a Phase 3, randomized, double-blind, placebo-controlled, multicenter, pivotal study evaluating the efficacy and safety of bimekizumab in study participants with moderate to severe HS. Study participants meeting the inclusion criteria who do not meet any exclusion criteria will complete a Screening Period of 14 days to up to 5 weeks; a double-blind, 48-week Treatment Period comprising a 16-week Initial Treatment Period and 32-week Maintenance Treatment Period; and a 20-week Safety Follow-up (SFU) Period following the final injection of investigational medicinal product (IMP) if study participants do not enter a subsequent extension study (HS0005) or withdraw prematurely from treatment.

Study participants will be randomized in a 2:2:2:1 ratio (stratified by Hurley Stage and current antibiotic use) to 1 of 3 dose regimens of bimekizumab or placebo as shown in the schematic (Figure 2-1).

Figure 2-1: Study Schematic



$HiSCR_{50}$=a 50% reduction in the total abscess and inflammatory nodule count with no increase from Baseline in abscess or draining tunnel count;

IMP=investigational medicinal product; Q2W=every 2 weeks; Q4W=every 4 weeks
*Week 16 = primary endpoint ($HiSCR_{50}$ bimekizumab versus placebo)
a Study participant should discontinue from the study from Week 32 on if no partial response is achieved (partial response is defined as ≥25% improvement in abscess and inflammatory nodule count relative to Baseline [Week 0] lesion values,)
b Study participants achieving an improvement of at least 25 % in abscess and inflammatory nodule count continue in HS0005 (Extension Study).
c 20-week Safety Follow-up (from last IMP injection) for any study participant who discontinues from study prior to Week 48, or who does not continue in HS0005.

2.3.2 Study periods

2.3.2.1 Screening Period

The Screening Period will last a minimum of 14 days up to a maximum of 5 weeks prior to randomization.

2.3.2.2 Initial Treatment Period (Weeks 0-16) and Maintenance Treatment Period (Weeks 16-48)

Eligible study participants will be randomized in a 2:2:2:1 ratio as noted in the Study Schema to:

- Bimekizumab 320mg Q2W from Weeks 0 to 48
- Bimekizumab 320mg Q4W from Weeks 0 to 48
- Bimekizumab 320mg Q2W to Week 16, continuing on 320mg Q4W from Weeks 16 to 48
- Placebo to Week 16, continuing on bimekizumab 320mg Q2W from Weeks 16 to 48

2.3.2.3 Safety Follow-up Visit

All study participants not continuing in the extension study, including those withdrawn from IMP as well as those completing all study visits, will have an SFU Visit 20 weeks after their final dose of IMP.

2.3.3 Study duration per participant

The total duration of study participation in HS0003 will be 68 to 71 weeks for those who complete HS0003 and do not participate in the extension study HS0005 and 50 to 53 weeks for those who participate in HS0005 and, thus, do not participate in the 20-week SFU Period. The study is comprised of the following: periods:

- Screening Period: 14 days up to a maximum of 5 weeks prior to randomization
- Initial Treatment Period: 16 weeks
- Maintenance Treatment Period: 32 weeks
- Safety Follow-Up Period: 20 weeks after the last dose of IMP

A study participant will be considered to have completed the study if he or she completed the Week 48 visit.

The end of the study is defined as the date of the last scheduled procedure for the last study participant in the study globally, including the SFU, as applicable.

2.3.4 Planned number of participants and sites

A total of approximately 490 study participants will be randomized into the study. The planned number of study sites is approximately 100.

2.3.5 Anticipated regions and countries

The regions planned for study conduct are Western Europe, Central/Eastern Europe, North America and Asia/Australia, with possible extension to other regions and countries.

2.4 Determination of sample size

A total of 490 study participants will be randomly assigned in a 2:2:2:1 ratio (stratified by Hurley Stage and current antibiotic use) to the following treatment arms:

- Bimekizumab 320mg Q2W during Initial Treatment Period (Weeks 0-16) and Maintenance Treatment (Weeks 16-48) Period, N=140
- Bimekizumab 320mg Q2W during Initial Treatment Period (Weeks 0-16), and Bimekizumab 320mg Q4W during Maintenance Treatment Period (Weeks 16-48), N=140
- Bimekizumab 320mg Q4W during Initial Treatment (Weeks 0-16) and Maintenance Treatment Periods (Weeks 16-48), N=140
- Placebo during Initial Treatment Period (Weeks 0-16), and Bimekizumab 320mg Q2W during Maintenance Treatment Period (Weeks 16-48), N=70

The analysis of the primary efficacy endpoint and secondary efficacy endpoints are based on a comparison of bimekizumab versus placebo at Week 16, with alpha adjustment strategy as indicated in Section 4.5.

The power to detect a statistically significant difference for each of the endpoints are shown in Table 2–1. Notably, with a 2-sided significance level of 0.025, the sample size of 140:70 provides 73% power for detecting at least a difference of 1.5 (bimekizumab Q4W vs placebo) for the Worst Skin Pain change from Baseline (CFB) endpoint.

Given the high level of power for each of the primary and key secondary endpoints at the 0.025 significance level for the Q2W comparison (power≥0.89), and per the alpha spending strategy, there is a high likelihood that the Q4W comparison of Worst Skin Pain CFB vs placebo will be allowed to be tested against the 0.05 level of significance. The power for this latter test is 81%. The sample size is thus ultimately driven by the Worst Skin Pain CFB endpoint. Furthermore, the randomization ratio of 2:2:2:1 has been chosen to provide study participants with a high probability (6/7 ~ 86%) of being randomized to active study drug.

After randomization for this study was complete, an additional endpoint to assess Worst Skin Pain response was included in the sequential testing procedure. This additional endpoint is based on the threshold for clinically meaningful change and is defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score at Week 16 among study participants with a score of ≥3 at Baseline. Note that the power calculations reported in Table 2–1for this endpoint are based on the sample size that was initially driven by the Worst Pain CFB endpoint as described above. With a 2-sided significance level of 0.025, the sample size of 104:52 in the subset of participants reporting Baseline HSSDD worst skin pain score at or beyond the threshold for clinically meaningful change (ie, Baseline HSSDD ≥3) provides 53% power for

detecting a statistically significant difference between bimekizumab Q4W and placebo in the proportion of Worst Pain responders.

Note that the power at the 0.025 level of significance associated with this endpoint for the comparison between bimekizumab Q2W and placebo is 95%. The Q4W comparison of Worst Skin Pain response vs placebo against the 0.05 level of significance is therefore likely, and the power at this significance level is 65%. Given the strength of the power for the bimekizumab Q2W arm vs placebo across endpoints, it is considered acceptable to have a relatively low power for this final endpoint in the testing sequence for the bimekizumab Q4W treatment arm vs placebo.

Table 2–1: Power calculation assumptions and methods

	Power α = 0.025, 2-sided		Assumptions		
Endpoint	**Q2W**	**Q4W**	**Week 16 Bimekizumab Q2W N=280[a]**	**Week 16 Bimekizumab Q4W N=140**	**Week 16 Placebo N=70**
$HiSCR_{50}$	0.99	0.90	Proportion responders=0.60	Proportion responders=0.50	Proportion responders=0.25
$HiSCR_{75}$	0.99	0.98	Proportion responders=0.45	Proportion responders=0.35	Proportion responders=0.10
DLQI	0.99	0.96	Mean CFB=-5.4; SD=6.8	Mean CFB=-4.8; SD=6.8	Mean CFB=-0.8; SD=6.6
Worst Skin Pain CFB[b]	0.89	0.73	Mean CFB=-2.2; SD=3.2	Mean CFB=-2.0; SD=3.2	Mean CFB=-0.5; SD=3.7
Worst Skin Pain Response[c]	0.95	0.53	Proportion responders=0.53	Proportion responders=0.43	Proportion responders=0.23

CFB=change from Baseline; Q2W=every 2 weeks; Q4W=every 4 weeks; SD=standard deviation
Note: Estimates for HS0003 are based on Week 12 data from the HS0001 study.
[a] Pooled Q2W at Week 16 from Q2W/Q2W and Q2W/Q4W arms
[b] Within-participant average of Worst Skin Pain according to 24-hour recall.
[c] Assumes N=208, 104, 52 in Q2W, Q4W, and placebo, respectively, to account for Worst Skin Pain score at or above 3 (ie, the threshold for clinically meaningful change from Baseline).

3 DATA ANALYSIS CONSIDERATIONS

3.1 General presentation of summaries and analyses

Statistical analysis and generation of tables, figures, study participants data listings, and statistical output will be performed using SAS Version 9.4 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 continuous variables, descriptive statistics will include number of study participants with available measurements (n), mean, standard deviation (SD), median, minimum, and maximum.

For PRO continuous variables, descriptive statistics will also include variable score, absolute and percentage changes from baseline, Q1 and Q3, 10th, and 90th percentiles.

If no participants have data at a given time point, then only n=0 will be presented. The other descriptive statistics will be left blank. If n < 3 then the n, minimum, and maximum only will be presented. The other descriptive statistics will be left blank. If n = 3 n, mean, median, minimum and maximum will be presented only. The other descriptive statistics will be left blank.

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

- For summaries of demographics and Baseline characteristics: percentages will be based on all study participants in the analysis set and a "Missing" category (corresponding to study participants with missing data for the variable being summarized) will be included as the last row in the list of categories being summarized.
- For summaries of efficacy and safety endpoints, unless otherwise specified: percentages will be based only on those study participants with observed data for the variable being summarized. As the denominator may be different from the number of study participants in the analysis set being considered, the denominator will be displayed in the table. The general format for displaying this will be "n/Nsub (%)."

Percentages will be presented to 1 decimal place. If the percentage is 100%, a decimal will not be presented. If the count is 0, the percentage will not be presented. Typically, the % sign will be presented in the column header, but not with each individual value.

For the purpose of the tabulations the lower and upper confidence limits for the percentages will be truncated at 0 and 100% respectively. Confidence intervals (CIs) for the response rates in efficacy summaries based on nonresponder imputation (NRI) will be computed using the Wilson approximation.

For bimekizumab PK concentrations, summary statistics will include geometric mean, geometric coefficient of variation (CV), 95% CIs 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 be subject to the following rules:

- "n" will be an integer
- Mean, SD, and median will use 1 additional decimal place compared to the original data
- CV [%] will be presented with 1 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.

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

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

Statistical tests of efficacy variables will be presented as 2-sided p-values rounded to 3 decimal places. P-values less than 0.001 will be presented as "<0.001" and p-values greater than 0.999 will be presented as ">0.999." Statistical comparisons will be 2-sided and will be performed at the 0.05 level of significance unless specified otherwise. The significance levels used as part of the multiple testing procedure are detailed in Section 4.5.

Per protocol, visit windows are ±3 days from the date of first dose. The 20-week SFU Visit window is ±7 days from the date of the final dose. All by-visit summaries will contain nominal (ie, scheduled) visits only. Unscheduled visits will not be mapped to scheduled visits except for assessments that occur within a 3-day time window of a scheduled visit. In that case, the assessment will be mapped to the corresponding scheduled visit and will be used for the analysis. This will only occur for selected vendor data. The only exception to this rule is for unscheduled assessments that occur up to 3 days after the Baseline visit. These unscheduled visits will remain as unscheduled as the Baseline assessment cannot be after the first dose of study drug administration. See Section 3.3 for more details on the definition of Baseline values.

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

3.2 General study level definitions

3.2.1 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 will 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 will 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.

Relative day will be calculated from first dose of IMP for all treatment groups, and additionally from first dose of bimekizumab for the Placebo/BKZ 320mg Q2W arm.

3.2.2 Mapping of data from Premature End of Treatment visits

If the Premature end of treatment (PEOT) visit occurs at a scheduled visit as outlined in the schedule of assessments, then no mapping is necessary, and any early withdrawal assessments will correspond to that scheduled visit. The PEOT assessments that occur on a date after a scheduled visit will be assigned to the next scheduled site visit per the protocol following the last visit where assessments were available. This approach means that there is a chance that data will be mapped to a visit where a given assessment was not actually collected per the protocol schedule of assessments. Such data will not be summarized in by-visit tables (though it will be available in the listings).

The only exception to the above rule is for anti-bimekizumab antibody assessments, in which all PEOT assessments will be assigned to the next scheduled visit at which anti-bimekizumab antibodies are assessed. All by-visit summaries will contain nominal visits only. Unscheduled visits will not be mapped to scheduled visits. Note that based on the early withdrawal mapping conventions described above, a mapped PEOT visit is considered as observed at that visit and will be summarized as such in the tables.

3.3 Definition of Baseline values

Section 8.3.3 details the derivation of the Baseline value for the HSSDD assessment. For all other assessments, the below applies.

A Baseline value for a participant is defined as the latest non-missing measurement for that participant up to and including the day of administration of first study medication, unless otherwise stated. If a Baseline assessment is taken on the same day as first administration of study medication, it is eligible to be used as the Baseline value, even in the case that the time of the assessment is recorded as taking place after the time of first study medication administration. This is considered acceptable as this measurement is still the best representation of the Baseline value of the given assessment since it is highly unlikely that the study medication could have an impact on any measurement in such a short period of time. However, such cases should be rare as study center personnel are instructed to do all assessments at the Baseline visit prior to administering study medication. One exception to this rule is plasma concentration of bimekizumab. If Baseline plasma concentration is measured at a time after the first administration of study medication, then it will not be eligible to be considered as a Baseline plasma concentration. Such cases will be discussed with the quantitative clinical pharmacologist.

For randomized participants for whom no start date of treatment is available, the Baseline value will be considered as the last available value on or before the randomization date.

If a Baseline measurement is missing or not collected, and a Screening value is available, the Screening value will be utilized as Baseline instead.

Baseline values for component scores should be computed using components from the same visit where the relevant measurements were recorded prior to dosing. For example, if the Screening visit has all of the components, but the Baseline visit is missing 1 or more components, the Baseline value for the component score should be calculated using the Screening visit values.

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

3.4 Protocol deviations

Important protocol deviations are deviations from the protocol that could potentially have a meaningful impact on study conduct or on the primary and key secondary efficacy, key safety, or PK outcomes for an individual participant. The criteria for identifying important protocol deviations will be defined within the appropriate protocol-specific document at study start. Important protocol deviations will be reviewed as part of the ongoing data cleaning process. Important protocol deviations including those that lead to exclusion from the analysis sets will be identified and documented prior to unblinding.

Deviations related to the Coronavirus Disease 2019 (COVID-19) global pandemic are unavoidable deviations from the protocol due to confirmed COVID-19 infection, suspected COVID-19 infection, general circumstances around COVID-19 without infection or any other deviation from the protocol due to COVID-19. COVID-19 protocol deviations will also be reviewed separately as part of the ongoing data cleaning process.

3.5 Analysis sets

3.5.1 Enrolled Set

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

3.5.2 Randomized Set

The Randomized Set (RS) will consist of all participants randomized into the study.

3.5.3 Safety Set

The Safety Set (SS) will consist of all participants who received at least 1 dose (full or partial) of IMP. The SS will be used for the demographic, safety, and immunogenicity analyses.

3.5.4 Full Analysis Set

The Full Analysis Set (FAS) will consist of all study participants who received at least 1 dose (full or partial) of IMP and had a valid Baseline measurement and a post-Baseline measurement for abscess, inflammatory nodules, and draining tunnel counts.

3.5.5 Active Medication Set

The Active Medication Set (AMS) will consist of all participants who have received at least 1 dose (full or partial) of bimekizumab. The AMS will be used for summaries of safety that include all data from the Initial Treatment Period and/or Maintenance Treatment Period.

3.5.6 Maintenance Set

The Maintenance Set (MS) will consist of all participants who have received at least 1 dose (full or partial) of bimekizumab in the Maintenance Treatment Period.

3.5.7 Per-Protocol Set

The Per-Protocol Set (PPS) will consist of all study participants in the FAS who had no important protocol deviations affecting the primary efficacy variable. Important protocol

deviations will be predefined and study participants with important protocol deviations will be evaluated during ongoing data cleaning meetings prior to unblinding of the data.

3.5.8 Pharmacokinetics Per-Protocol Set

The Pharmacokinetics Per-Protocol Set (PK-PPS) will consist of study participants who received at least 1 full dose of bimekizumab and provided at least 1 quantifiable plasma concentration post-dose without important protocol deviations that would affect the PK.

3.5.9 COVID-19 Free Set

The COVID-19 Free Set (CFS) will consist of all study participants randomized into the study and who have no COVID-19 impact through Week 16. This analysis set will be used for sensitivity analysis of the primary efficacy endpoint.

3.6 Treatment assignment and treatment groups

It is expected that participants receive treatment as randomized and hence safety analyses will be based on the SS, as randomized. However, if after unblinding it is determined that participants randomized to placebo in the Initial Treatment Period received bimekizumab at any time within the first 16 weeks, then for safety analyses these participants will be reallocated to the appropriate bimekizumab treatment group, unless otherwise specified. Participants randomized to bimekizumab will only be reallocated to the placebo treatment group if they never received bimekizumab. Efficacy analyses will be according to randomized treatment and not actual treatment received.

For the purposes of Initial Treatment Period analyses for the 320mg Q2W dosing regimen, the bimekizumab treatment arms of 320mg Q2W/Q2W and bimekizumab 320mg Q2W/Q4W treatment groups will be pooled.

3.7 Center pooling strategy

Geographic regions have been categorized as North America, Western Europe, Central/Eastern Europe, and Asia/Australia. Below is a table of geographic regions with corresponding countries.

Table 3–1: Geographic regions and corresponding countries

Region	Countries
North America	Canada, United States
Western Europe	Belgium, France, Germany, Italy, Norway, Spain, Switzerland, Denmark, Netherlands
Central/Eastern Europe	Greece
Asia/Australia	Australia, Israel, Turkey

The following center pooling algorithm will be used for each geographic region:

- If a center has 21 or more participants, then no pooling will be done for that center.
- Centers with fewer than 21 participants will be ordered from largest to smallest with pooling proceeding in the following manner:
  - Two or more centers will be combined until the cumulative participant total is at least 21.

- Once a pooled center has at least 21 participants, the process will continue in an iterative fashion for the subsequent centers in the ordered list, where a new pooled center begins each time at least 21 participants has been reached in the previous pool.
- If this iterative process reaches the end of the ordered list of centers where the final pooled center has fewer than 21 participants, then the participants from the centers in that pool will be combined with the pooled center formed in the previous iteration.

This procedure is only to be performed within a geographic region – there will be no pooling of centers across regions.

In the event that the percentage of randomized participants is less than 10% in either of the Asia/Australia or Central/Eastern Europe regions, the two regions will be combined as a geographic region stratum for efficacy modeling, so that there are no modeling convergence issues across efficacy variables.

3.8 Coding dictionaries

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

Medications will be coded according to the World Health Organization Drug Dictionary (WHODD) version MAR2021 B3 or later. Medical procedures will not be coded.

3.9 Definition of an intercurrent event

Handling of intercurrent events is one of the key elements for the analysis of efficacy endpoints.

An intercurrent event is defined as receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy (See Section 8.2.2).

Receipt of systemic antibiotic rescue medication is defined as initiating any systemic antibiotic on or after Baseline for any reason (including in response to an AE). The only exception to this rule is if a participant randomized to the antibiotic stratum on a tetracycline antibiotic interrupts their stable dose of tetracycline antibiotic during the study and subsequently restarts the same tetracycline antibiotic as confirmed using the coded preferred term. The restarted dose and frequency of the antibiotic must be the same or lower than the regimen prior to the interruption.

The dates of an intercurrent event are as follows:

- For receipt of systemic antibiotic rescue medication: start date of the antibiotic
- For discontinuation of study treatment due to an AE or lack of efficacy: Last study treatment date + 17 days. Note: study treatment discontinuation includes study discontinuation.

The choice of 17 days is intended to capture the interval between dosing and lesion assessments (14 days), as well as the visit window (3 days).

An additional sensitivity analysis will be conducted where missing data due to COVID-19 will be considered an intercurrent event and will be imputed as a nonresponse at that particular visit. This will be identified when there are missing data at a visit that has been impacted by COVID-19 according the COVID-19 impact CRF page. The date of this intercurrent event will be the date of the impacted visit.

3.10 Changes to protocol-defined analyses

The MS and AMS were added as analysis sets.

The endpoints for PGI-S-HS, PGI-S-SP, PGI-C-HS, and PGI-C-SP were clarified in Section 2.2.1.3 to indicate that absolute change from Baseline will not be calculated, as these are categorical endpoints.

The HiSQOL endpoint was clarified to show that there are only 3 domains: symptoms, psychosocial, activities and adaptations and to add total score.

In Protocol Amendment 4, the secondary efficacy endpoint for skin pain response based on the worst skin pain HSSDD score is defined as a decrease from Baseline in HSSDD weekly worst skin pain score at or beyond the threshold for clinically meaningful change at Week 16. In the HS0003 SAP, this endpoint is defined to include the exact value of the clinically meaningful threshold of 3, so that the skin pain response based on the HSSDD worst skin pain score is defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score at Week 16 among study participants with a score of ≥3 at Baseline.

The following other efficacy endpoints are included in the protocol but will not be included as part of the analysis:

- Skin pain response based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HS Symptom Questionnaire (HSSQ) weekly worst skin pain score) among study participants with a score of ≥3 at Baseline
- Response on other HS Symptoms (11-point numeric rating scale) - itch, drainage or oozing of HS lesions, and smell or odor
- Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Week 16) by Week 48

The calculation of nominal p-values has been added for selected efficacy endpoints. These nominal p-values are not controlled for multiplicity and should not be used to declare statistical significance.

The protocol defines the PK-PPS separately by period, but there will only be one PK-PPS for the overall study.

4 STATISTICAL/ANALYTICAL ISSUES

4.1 Adjustments for covariates

The primary efficacy analyses and selected secondary analyses will be adjusted for the 2 randomization stratification variables:

- Hurley Stage at Baseline (II or III)
- Baseline antibiotic use (Yes or No)

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

The continuous secondary endpoints will also include the Baseline value as a covariate.

The Worst Skin Pain secondary endpoints (change from Baseline continuous endpoint and pain response binary endpoint) will also include analgesic use as a covariate.

4.2 Handling of dropouts or missing data

4.2.1 Efficacy data

Different approaches will be used to handle missing data including how intercurrent events (defined as receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to the given visit) will be considered. A composite strategy will be implemented in which a positive clinical outcome is defined as the study participant achieving $HiSCR_{50}$ at the given visit and not receiving systemic antibiotic rescue medication, and not discontinuing study treatment due to an AE or lack of efficacy through that visit.

4.2.1.1 Handling missing data for the primary efficacy endpoint

If study participants have an intercurrent event as defined in Section 3.9, then the primary efficacy variable at that timepoint and all subsequent timepoints (whether the data were observed or not) will be set to "nonresponse" as the study participant has not met the criteria for response based on the composite estimand defined in Section 8.2. All remaining missing data for the endpoint will be imputed using multiple imputation Markov-Chain Monte Carlo method (MI-MCMC)/monotone regression for the primary analysis.

In addition, sensitivity analyses using NRI, MI-MCMC/reference-based methods, tipping point analysis, and observed case (OC) methods will be performed, which will assess the impact of different methods of handling missing data.

4.2.1.2 Handling missing data for the secondary efficacy endpoints

For secondary binary efficacy endpoints, intercurrent events will be handled, and missing data will be imputed, using the same methods as for the primary efficacy endpoint. NRI and OC methods will be performed as sensitivity analyses.

For secondary continuous efficacy endpoints, MI-MCMC/monotone regression is the primary method for imputing missing data, regardless of whether the missing data are preceded by an intercurrent event. That is, if an intercurrent event occurs on or before a visit, the result for that visit will be treated as missing and then imputed. If the imputation model cannot converge, last observation carried forward (LOCF) will be used. The OC method will be performed as a sensitivity analysis.

4.2.1.3 Handling missing data for the other efficacy endpoints

For other binary efficacy endpoints, missing data will be imputed using the same method as the primary efficacy endpoint. NRI and OC methods will be performed as sensitivity analyses of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$.

For other continuous efficacy endpoints, the MI-MCMC/monotone regression method will be used to impute missing data as the primary method, regardless of whether the missing data are preceded by an intercurrent event. That is, if an intercurrent event occurs on or before a visit, the result for that visit will be treated as missing and imputed with the missing data. If the imputation model cannot converge, LOCF will be used.

For other ordinal endpoints (EQ-5D-3L, PGI-S-HS, PGI-C-HS, PGI-S-SP, PGI-C-SP), the OC method will be applied as the primary analysis method. No imputation is applied.

4.2.1.4 Missing Data Overview and Summary

In summary, the approaches listed below will be used in this study for handling missing data for efficacy endpoints as appropriate:

- **NRI**: Participants who have missing data at the timepoint of interest are treated as though they did not respond to the treatment. This approach is also referred to as Composite Estimand (NRI).
- **Multiple Imputation (MI) – MCMC / Monotone Regression**: Using multiple imputation methodology, intermittent missing data are imputed based on the MCMC method, and monotone missing data are imputed using monotone regression.
- **MI–MCMC / Reference-based imputation**: Using multiple imputation methodology, intermittent missing data are imputed based on the MCMC method, and monotone missing data are imputed using an imputation model based on placebo (reference) data.
- **LOCF**: Post-Baseline missing data are imputed by carrying forward the last available observation (including Baseline).
- **Tipping point analyses**: Assumptions will be made about average outcomes among the subsets of participants who prematurely discontinued study treatment. Various "delta adjustments" will be made to the assumed responses among missing data in each treatment arm with varying degrees of plausibility in order to identify assumptions about the missing data under which the conclusions change (O'Kelly, 2014). Then, the plausibility of such assumptions is discussed.
- **Observed case (OC)**: Missing data are not imputed. Only participants with available data who have not discontinued study treatment at the given timepoint are considered. Study participants with missing data or who have prematurely discontinued study treatment will be treated as missing.
- **Treatment policy strategy**: All available data observed at the time point of interest will be considered, regardless of the occurrence of intercurrent events. This means the analysis includes on- and off-treatment values collected after study participants prematurely discontinued study treatment but agreed to remain on the study and continued to attend visits and provide assessments at those visits as well as values from study participants who received rescue antibiotic medication. Those observed values will be analyzed according to the study participant's randomized treatment. Study participants for whom efficacy data cannot be obtained at the week of interest, despite attempts to retain them in the study, will have their data imputed using MI – MCMC / monotone.

The following table depicts which missing data handling approaches will be used based on endpoint priority (primary, secondary, other) and endpoint type (responder, continuous, ordinal).

Table 4–1: Missing data handling approach by endpoint priority and type

End-point Priority	Endpoint Type	Com-posite Esti-mand (NRI)	Modified Composite Estimand (MI)	MI (MCMC/ Reference -based)	Tipping Point	Treat-ment Policy	Hypo-thetical Esti-mand	OC
Primary	Responder	S[a]	P	S[a]	S	S[a]		S
Secondary included in the statistical testing procedure	Responder	S[a]	P					S
	Continuous						P	S
Secondary not included in statistical testing procedure	Binary	X	X					X
	Continuous						X	X
Other	Responder	X[d]	X					X[d]
	Continuous						X	X[b]
	Ordinal						X[c]	X[c]

MI=multiple imputation, NRI=Nonresponder imputation, OC=Observed case, P=Primary method, S=Sensitivity method, X=Method to be used (no priority designated).
Note: Composite estimand (NRI) refers to the approach in which data preceded by the intercurrent event of study treatment discontinuation due to AE or lack of efficacy or receipt of rescue antibiotic medication are imputed as nonresponse, and other missing data are also imputed as nonresponse.
Note: Modified Composite Estimand (MI) refers to the approach in which data preceded by the intercurrent event of study treatment discontinuation due to AE or lack of efficacy or receipt of rescue antibiotic medication are imputed as nonresponse, and other missing data are imputed via a multiple imputation model.
Note: Hypothetical Estimand (MI) refers to the approach where outcomes for study participants without an intercurrent event of study treatment discontinuation are as observed, and outcomes for study participants with the intercurrent event are imputed via a multiple imputation model.
[a] Imputation method is applied on continuous data, and responder endpoint is derived from the continuous endpoint based on complete data set where applicable.
[b] Required only for by-visit summaries of variables whose value at Week 16 is part of the hierarchical testing procedure.
[c] For variables with multiple categories, data will be summarized as observed with an additional missing row to capture missing data at a given visit.
[d] NRI/OC sensitivity analysis will be performed only for $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, $HiSCR_{100}$ summaries.

4.2.2 Missing data algorithms for efficacy analyses

These descriptions focus on the missing data procedures themselves and do not specifically account for dealing with intercurrent events, which is addressed in their respective sections.

4.2.2.1 MI – MCMC / Monotone Regression

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. To investigate the efficacy results under the assumption of data being MAR, a multiple imputation method will be applied.

Binary endpoint

For a binary endpoint (eg, $HiSCR_{50}$), the procedure is as follows:

1. Create a data set, one for each treatment group of participants with observed values and those needing estimation by multiple imputation. For the imputation step, a distinction is made between non-monotone missing values (ie, intermittent missing values between completed assessments) and monotone missing values (ie, where all participants data are missing after a given time point).

a. For the intermittent missing values, the missing values in each data set will be filled in using the MCMC method with multiple chain, monotone missing data imputing pattern, and non-informative prior for all parameters. Unless specified differently, the first 200 iterations will not be used (the "Burn-in" option). A total of 100 sets of imputations will be performed. The seed used for these imputations will be 762 and all other multiple imputation procedures described in this SAP will use this same seed as well. The resulting 100 imputed data sets will have a monotone missing pattern and will be imputed using a method for monotone missingness. Note that a separate imputation procedure must be invoked in SAS for each treatment group as the seed cannot be set for by groups beyond the first when using a by statement in PROC MI.

b. For monotone missing data, monotone regression will be used to impute missing data. A separate regression model is estimated for each variable with missing values (ie, measurement at each time point). Based on the resulting model, a new regression model is then drawn and is used to impute the missing values for the variable. Since the data set has a monotone missing data pattern, the process is repeated sequentially for variables with missing values. This procedure will be based on the100 imputed datasets generated from the MCMC procedure and will be performed by imputation. The SAS® PROC MI procedure will be used for the imputation.

   In both cases, Hurley Stage at Baseline, Baseline antibiotic use, and value of the variable of interest at Baseline and at each post-Baseline visit (prior to the time point of interest) will be included in the imputation model. The post- Baseline values will need to be specified in chronological order in the imputation model so that SAS® PROC MI imputes variables from left to right (eg, the Week 2 value will be first imputed using regression based on the Baseline value, and then Week 4 value will be imputed using regression based on Baseline and Week 2 values, etc). 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 joint multivariate normal variables. Therefore, Hurley Stage at Baseline and Baseline antibiotic use will be re-coded as indicator variables. For Baseline antibiotic use, this will simply be 0 for Baseline antibiotic non-users and 1 for Baseline antibiotic users. For Hurley Stage at Baseline, this will be 0 for Hurley Stage II participants and 1 for Hurley Stage III participants. In order to

achieve model convergence, Baseline antibiotic use may be dropped from the model. If convergence is still not obtained, then Hurley Stage at Baseline may also be dropped from the model. Additionally, if a variable is dropped in order to allow convergence for one model in a study, that variable does not have to be dropped from other models in the study if the model converges without dropping the variable. In other words, model convergence should be evaluated for each efficacy variable independently.

Note: The imputation of each lesion type (inflammatory nodule, abscess, draining tunnel, etc) will be performed separately. The 100 data sets obtained for each type will be merged by imputation number and subject number.

2. For each complete imputed data set, the dichotomous responder variable (eg, HiSCR 0 or 1) will be computed. Each complete imputed data set will then be analyzed based on the logistic regression model.

   Note: For derivation of HiSCR response, the AN, inflammatory nodule, abscess, and draining tunnel (fistula/sinus tract) counts at Week 16 in the imputed data sets will be compared directly to the observed Baseline counts to determine response. If values outside of the pre-defined range of values for lesion count (<0) are imputed, they will be cut off as appropriate after the multiple imputation procedure but before deriving the responder variable. For example, an imputed draining tunnel (fistula/sinus tract) count of -1 would be changed to 0 before deriving the HiSCR responder variable. Additional ranges for values for secondary and other endpoints are defined in Table 4-2.

   Note: Standard rounding rules will also be applied to the imputed values of endpoints that can only take integer values (eg, abscess count). For example, if a study participant has an abscess count imputed as 2.4, this imputed value would be rounded down to 2. This rounding step is performed after the multiple imputation but before deriving the responder variable.

Table 4–2: Imputation allowable ranges by variable

Variable	Minimum Value	Maximum Value	Integer Values Only
Lesion count[a]	0	--	Yes
DLQI total score	0	30	Yes
hs-CRP	LLOQ/2	--	No
HSSDD item score	0	10	No
HSSQ item score	0	10	Yes
HiSQOL symptom status score	0	16	Yes
HiSQOL psychosocial impact score	0	20	Yes
HiSQOL impact on physical activities score	0	32	Yes
EQ-5D-3L VAS	0	100	Yes

Table 4–2: Imputation allowable ranges by variable

Variable	Minimum Value	Maximum Value	Integer Values Only
WPAI dimension scores	0	100	No for variables: "Percent work time missed due to problem" and "Percent overall work impairment due to problem". Yes for variables: "Percent impairment while working due to problem" and "Percent activity impairment due to problem". These two variables can only take values that are multiples of 10.

[a] Lesion counts will be imputed separately for each lesion type (abscesses, draining tunnels [fistulas/sinus tracts], inflammatory nodules, non-draining tunnels [fistulas/sinus tracts], non-inflammatory nodules, HS scars). The imputed lesion counts will be used to derive the endpoints that are dependent on the lesion count data (eg, $HiSCR_{50}$).

3. Estimates of the adjusted responder rate for each treatment group and the associated SE are obtained from the logistic regression of each of the 100 imputed data sets. These estimates will be combined for overall inference using Rubin's rules, which account for the uncertainty associated with the imputed values (Rubin, 1987), and the combined estimates and SEs will be used to construct 95% CIs using the logit scale. This will be done using SAS PROC MIANALYZE. The combined estimates and 95% CIs on the logit scale will be back-transformed using the inverse logit link function to obtain the adjusted responder rates (%) and associated 95% CIs.

   Note: The (unadjusted) proportion of responders will be calculated at each time point by treatment group from the imputed datasets using SAS PROC FREQ. These results will also be combined into an overall inference 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 models in Step 3 follow a log-normal distribution, a log transformation is needed to normalize these 100 odds ratio estimates. That is because the procedures for combining results from multiple imputed datasets assume that the statistics estimated from each imputed dataset are normally distributed. Therefore, the log of the odds ratio estimates from the logistic regression model are used when combining into a single inference (Step 3). Additionally, the SE for the odds ratios are transformed as follows:

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

Where UCL and LCL are the upper and lower confidence limit, respectively, for the CI of the odds ratio from the logistic regression model, and $Z_{\alpha/2}$ is the relevant critical value from the standard normal distribution (2.24 for a 97.5% CI). The estimates of the log odds ratio for Bimekizumab relative to placebo and the corresponding upper and lower CLs will be provided. The odds ratio will be then estimated by exponentiating the estimate of the log odds ratio. The odds ratio and the confidence limits of the odds ratio will be estimated as follows:

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

$$LCL = \ OR * \exp(-SE * Z\alpha_{/2})$$

$$UCL = \ OR * \exp(SE * Z\alpha_{/2})$$

Where OR is the back-transformed estimate of the odds ratio just described, SE is the SE of the log odds ratio and $Z_{\alpha/2}$ is the relevant critical value from the standard normal distribution (2.24 for a 97.5% CI). These calculations will be done such that odds ratios and corresponding CIs are calculated for the odds ratio of bimekizumab vs. placebo.

Note: If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose with a corresponding $Z_{\alpha/2}$ of 1.96.

In addition to calculating the odds ratio, associated CIs, and p-values for the pairwise comparisons of bimekizumab and placebo, the estimated proportion of responders (ie, estimated responder rate) and the difference in the proportion of responders between each bimekizumab treatment group and placebo will be estimated, and 2-sided 95% CIs will be created for each difference. The creation of the estimates of the differences will be completed for each bimekizumab treatment group using the process detailed below:

1. Use the logistic regression model to calculate:

   Least squares mean estimates of the log odds of bimekizumab (G_B) and placebo (G_P), as well as their corresponding standard errors (S_B and S_P, respectively).

   Standard error of the least squares mean estimate of the log odds ratio (S_R)

2. Compute estimates for predicted proportions using the following transformations:

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

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

   The difference in proportions is then given by:

   $D = P_B - P_P$

3. Estimate the standard error of D by:

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

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

Note that this procedure indicates that the imputation model will be applied for each treatment group separately in PROC MI, which will be the default method. However, in the event there are computational challenges with the imputation model (eg, due to a standard deviation of 0 for responses of a given imputation), it is acceptable to modify the imputation model to include treatment as a variable in the model rather than running a separate model for each treatment group. It should be noted that doing so assumes that treatment does not interact with any of the other variables in the imputation model.

Continuous endpoint

For continuous endpoints (eg, Change from Baseline in DLQI total score at Week 16), the MI method will be applied as follows:

1. The MCMC/monotone regression method described above in Step 1 for binary endpoints will be performed.
2. Based on the multiply imputed data sets obtained for the given variable, the change from Baseline will be derived for each of the 100 complete imputed data sets based on the observed Baseline value and the observed/imputed post-Baseline values. Note that if the value itself is being summarized, no additional derivation is needed.
3. If a statistical model is being used for the analysis of the variable, then that will be performed for each imputation in this step. If no statistical model is being used, then simple descriptive statistics will be calculated.
4. For data excluding hs-CRP, the following rules apply. The results of the 100 imputed data sets (based on the statistical model or descriptive statistics) are combined with means and standard errors calculated using Rubin's rules (via PROC MIANALYZE). Note that for the calculation of other descriptive statistics such as the median, min, and max, Rubin's rules do not apply. MI estimates will be computed by simply averaging the estimates from the multiple repetitions of the imputation algorithm. Thus, for median, Q1, Q3, minimum and maximum, the following approach will apply:
   - The data will be summarized by treatment, visit, and imputation, and the summary statistics will be computed.
   - Results will be summarized by taking the mean value of each summary statistic at each visit across all imputations.
   - The number of decimal places will remain the same as the original for display purposes (ie, if the mean was presented to 1 decimal place, then the mean of the means will also be presented to 1 decimal place).

For hs-CRP only, the following rules apply. The hs-CRP data will be presented using the geometric mean, 95% CI for the geometric mean, median, Q1, Q3, minimum and maximum. The change from Baseline will be expressed as the ratio to Baseline in the summaries. The following approach will be applied:

- Following the MI procedure, the ratio to Baseline will be calculated for any of the imputed values
- The natural logarithm of the absolute values and of the ratios to Baseline will be calculated

- The logged values will be summarized (using PROC MEANS) by treatment, visit and imputation
- The datasets will be combined using PROC MIANALYZE in order to get the mean and 95% CI estimates from the absolute values and ratios to Baseline (based on logged data) across imputations
- The estimates of the mean and 95% CI will be back-transformed to obtain the geometric mean and 95% CI on the original scale
- For the median, Q1, Q3, minimum and maximum the procedure outlined above for the other endpoints will be followed

If the imputation model cannot converge, LOCF will be used.

4.2.2.2 MI – MCMC / Reference-based imputation

MI-MCMC / Reference-based imputation will be implemented as a supportive analysis for the primary efficacy endpoint (through Week 16).

In this case, placebo will be described as the reference arm.

This procedure will use an imputation model based on data from the placebo group only (Mallinckrodt, 2013). Reference-based MI assumes that the statistical behavior of the bimekizumab and placebo-treated participants after discontinuing study medication becomes that of the placebo-treated participants. All timepoints after discontinuation of the double-blind study treatment for both the bimekizumab and placebo groups will be considered missing. Multiple imputations are used to replace missing outcomes for bimekizumab- and placebo-treated participants who discontinued using multiple draws from the posterior predictive distribution estimated from the placebo arm. For binary efficacy endpoints (eg, $HiSCR_{50}$ at Week 16), imputation will be done on the lesion counts before assessing the imputed results for $HiSCR_{50}$ response.

The steps for the procedure are as follows:

1. For non-monotone (intermittent) missing data, MCMC will be used to impute lesion count data, with Baseline antibiotic use, Hurley Stage at Baseline, and lesion count at Baseline and at each post-Baseline visit (in chronological order) being included in the imputation model. This will be done only once for each participant in order to provide a dataset with monotone missing data.
2. Data will be processed sequentially by repeatedly calling SAS® PROC MI to impute missing outcome data at visits t=1, … T, where T is Week 16 for $HiSCR_{50}$.

a. *Initialization*. Set t=1 (Baseline visit)

b. *Iteration*. Set t=t+1. Create a data set combining records from bimekizumab- and placebo-treated participants with columns for covariates (Hurley Stage at Baseline and Baseline antibiotic use) and outcomes at visits 1 to t. Outcomes for all bimekizumab-treated participants are set to missing at visit t and set to observed or previously imputed values at visits 1 to t-1. Outcomes for placebo-treated participants are set to observed at visit t or observed or previously imputed values at visits 1 to t-1. The outcomes should be sorted in chronological order in the model.

c. *Imputation*. Impute missing values for visit t using previous outcomes for visits 1 to t-1, Baseline antibiotic use, and Hurley Stage at Baseline. Note that only placebo data will be used to estimate the imputation model since no outcome is available for bimekizumab-treated participants at visit t. Consequently, the input dataset should include all study participants from placebo but only study participants from the bimekizumab arm that have values at timepoint t missing.

d. Repeat steps 2a-2c, 100 times with different seed values (seeds ranging from 853 to 952) to create 100 imputed complete data sets. Study participants whose missing values were imputed in the last PROC MI call will be included in the input dataset for the next PROC MI call. Standard rounding rules will be applied to the imputed values. If the MI procedure yields a value outside of the pre-defined range for the given variable, the value will be updated to be within the predefined range of values for the endpoint of interest. For example, the imputed value for inflammatory nodules will be updated to 0 in the case of an imputed value less than 0.

e. *Analysis*. For each completed data set, the outcome (response or change from Baseline) will be calculated using the complete datasets (with no missing data).

3. Each complete imputed data set will then be analyzed based on the statistical model specified in this study (logistic regression). The Week 16 results from logistic regression 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.

4.2.2.3 Tipping Point Analysis

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

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

For tipping point analyses, data for participants after the intercurrent event date (See Section 3.9) will be changed to missing prior to imputation and, for the bimekizumab treated participants, will be changed to non-response after imputation.

The worst-case scenario will be evaluated first. All missing primary endpoint values for study participants randomized to bimekizumab (where missing values include observations after the intercurrent event date and any other missing values) will be imputed as non-responders, while all missing values for placebo-randomized study participants will be imputed as responders. While there is little justification for such an approach, it makes the most putative assumption possible against a bimekizumab treatment effect. After applying this imputation approach, a logistic regression model consistent with the one described for the primary analysis will be applied. If the p-value for the odds ratio of bimekizumab versus placebo remains significant, then no further tipping point analyses are needed.

If this analysis based on the worst-case scenario results in a p-value that is not significant (eg, greater than 0.025), then additional tipping point analyses will be performed to identify the point at which results switch or "tip" from significant to non-significant. Note that each bimekizumab dose of 320mg Q2W and 320mg Q4W will be compared to placebo independently for these analyses. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the p-value in this analysis method will be 0.05 instead of 0.025 throughout for that dose. In the tipping point analysis, a shift parameter or delta adjustment is applied to missing, and subsequently imported primary endpoint values (where missing values include observations after the intercurrent event and any other missing values). These delta implemented on the primary endpoint as follows:

1. Data after intercurrent event date (See Section 3.9) will be set to missing.
2. The same MCMC method described in Section 4.2.2.1 (Step 1a) will be implemented for non-monotone (intermittent) missing pattern values, using the same imputation model. This will be based on 100 sets of imputations.
3. Based on the 100 datasets obtained in Step 2, a monotone regression model will be applied (using the same imputation model as in Step 2) as described in Section 4.2.2.1 (Step 1b). This will be based on 1 imputation.
4. Delta adjustments will be made to imputed lesion count values at Week 16, independently in each treatment group as described below.
5. Delta adjusted imputed values will be truncated so that they are within the range of allowable values for each component.
6. Following the delta adjustments for the lesion counts, $HiSCR_{50}$ will then be derived based on the delta-adjusted multiply imputed data sets obtained for each component.
7. Each of the 100 imputed datasets will then be analyzed using a logistic regression model with factors of treatment group, Baseline Hurley Stage, and Baseline antibiotic use.
8. The results obtained from the 100 logistic regression analyses in Step 7 will be combined for overall inference using Rubin's rules, and the results obtained for each shift parameter will be presented in a single table.
9. Steps 4 to 8 will be repeated so that, at each iteration, missing values are adjusted with a larger delta than at the previous iteration. The process will go on until the p-value for the odds ratio between bimekizumab and placebo is no longer statistically significant (eg, ≥ 0.025). The odds ratio, 97.5% CI (or 95% depending on the significance level being used for testing), and p-values obtained for each value of delta will be combined in one single table.

The delta adjustments result in study participants randomized to bimekizumab with missing data having a lower probability of response compared to study participants randomized to placebo with missing data. Since $HiSCR_{50}$ response is an endpoint for which high lesion counts are associated with a less favorable outcome:

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

To start, imputed values within each lesion type, will be adjusted by the same value in each treatment arm. This adjustment will be 5% of the observed range within that lesion type. Depending on the results obtained, this adjustment will be multiplied for step 9 above (2 times, 3 times the initial adjustment) until the p-value is no longer statistically significant.

Additionally, study participants randomized to bimekizumab with an intercurrent event should be set to non-response, after applying the delta adjustment outlined in Step 6 above. This ensures study participants randomized to bimekizumab do not have a higher probability of response in the tipping point analyses compared to the primary analysis (ie, a study participant randomized to bimekizumab who is non-responder in the primary analysis cannot become a responder in the tipping point analyses).

4.2.3 Rationale for estimand

Intercurrent events have been identified within the estimands for this study because of their potential to impact efficacy assessments linked with the primary and secondary study objectives. In order to account for the effect of any observed post-randomization intercurrent events on the efficacy analyses, the following estimand strategies will be implemented when evaluating the primary and secondary efficacy endpoints:

- A composite estimand strategy will be used for the primary analysis of the primary and binary secondary endpoints ($HiSCR_{50}$, $HiSCR_{75}$, HS worst skin pain response).
- A hypothetical estimand will be used for the primary analysis of the continuous secondary endpoints (CFB in DLQI total score and in "worst skin pain" item for the HSSDD).

4.2.3.1 Composite estimand

A composite estimand strategy as defined in Section 8.2.2 allows incorporation of the two intercurrent events (eg, receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy) within the definition of the endpoint. These intercurrent events are considered meaningful to the efficacy outcome following receipt of study medication. For example, within the proposed composite estimand framework, a randomized study participant who discontinues from study treatment due to lack of efficacy prior to Week 16 will be considered a treatment failure at Week 16 regardless of the lesion count assessment performed at that visit.

The assumptions and robustness of the primary analysis (modified composite estimand as defined in Section 8.2.2) will be assessed through the sensitivity analyses defined in Section 8.2.3. The impact of intercurrent event handling and data imputation methods on endpoint derivation will also be assessed via the analyses of lesion counts and derived HiSCR variables as specified in Section 8.4.2.1 and Section 8.4.1.1, respectively.

4.2.3.2 Hypothetical estimand

The hypothetical estimand is defined in Section 8.3 and involves a data-driven approach to account for the potential impact of intercurrent events (eg, receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy) on the analysis of continuous efficacy endpoints. Under this framework, outcomes for study participants without an intercurrent event are analyzed as observed. Conversely, outcomes for study participants with an intercurrent event are imputed via a multiple imputation model, ie any

recorded data on or after the intercurrent event will be set to missing and imputed via multiple imputation following the strategy established in Section 4.2.2.1.

4.2.4 Dates and times

For analyses of AEs and concomitant medication usage, a complete date is required in order to correctly identify the AE or medication as occurring during treatment or not, and for correctly assigning an AE or concomitant medication to the Initial Treatment Period or Maintenance Treatment Period.

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 participant data listings (ie, 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:
    - If the month and year of first dose of study medication is the same as the month and year of the partial start date, then use the date of first dose of study medication,
    - Else, if the month and year of the partial start date are the same as the month and year of a study medication switch date, then use the date of study medication switch,
    - Otherwise, use the 1st of the month of the partial start date;
  - If only the year is specified:
    - If the year of first dose of study medication is the same as the year of the partial start date, then use the date of first dose of study medication,
    - Else, if the year of the partial date is the same as the year of a study medication switch date, then use the date of study medication switch,
    - Otherwise, use the 1st of January of the year of the partial start date;
  - If the start date is completely unknown:
    - If the stop date is unknown or not prior to the date of first dose of study medication, then use the date of first dose of study medication,
    - If the stop date is prior to the date of first dose of study medication, then use the 1st of January of the year of the stop date.
- Imputation of Partial Stop Dates
  - If only the month and year are specified, :
    - Use the last day of the month of the partial stop date;
  - If only the year is specified
    - use December 31st of the year of the partial stop date;
  - If the stop date is completely unknown,

- Do not impute the stop date.

Note that if the stop date or the imputed stop date is prior to the imputed start date, then follow the procedure outlined below:

- If only the year of the start date is specified:
  - If the year of start date is the same as the year of first dose of study medication and the imputed stop date is after the date of first dose of study medication, then set the start date to the date of first dose of study medication,
  - Otherwise, set the 1st January of the year of the start date;
- If only the month and year of start date are specified:
  - If the month and year of the start date is the same as the month and year of first dose of study medication and the imputed stop date is on or after the date of first dose of study medication then set the start date to the date of first dose of study medication,
  - If the month and year of the start date is the same as the month and year of first dose of study medication and the imputed stop date is before the date of first dose of study medication then set the start date to the 1st of the month of partial start date.

Missing start times for medications will be imputed as 00:00h or with the time of dosing for events occurring on the date of IMP administration in case of missing hour and minute. Otherwise start times with only missing minutes will be imputed with :00 or with the minutes of dosing for events occurring on the date and hour of IMP administration.

In the event of ambiguity or incomplete data that makes it impossible to determine whether a medication was concomitant or an AE was treatment emergent, the medication will be considered as concomitant or the AE will be considered treatment emergent. Similarly, in the event of ambiguity or incomplete data which makes it impossible to determine whether a medication or AE is to be assigned to the Initial Treatment Period or to the Maintenance Treatment Period (or both, for medications), then the medication will be assigned to both Treatment Periods, and the AE will be assigned to the Initial Treatment Period.

4.3 Interim analyses and data monitoring

4.3.1 Data monitoring committee

An independent data monitoring committee (DMC) will periodically review unblinded efficacy and safety data to assess the benefit/risk of bimekizumab in study participants with moderate to severe HS. Efficacy data summaries and individual study participant-level data listings may be provided to the DMC to put the safety review in the context of risk/benefit. Any data to be provided is specified per the DMC charter. [redacted]

[redacted]

4.3.2 Interim analysis

[redacted]

4.4 Multicenter studies

The center-by-treatment interaction will be tested by adding center and a center-by-treatment interaction term (Section 8.2.3.11). In the model, center will be based on the original centers prior to pooling (Section 3.7). However, if the model is unable to converge due to a low number of participants at a given center, a pooling by center will be applied in order to allow the model to converge. If convergence is still not achieved, a pooling by region will be applied. If convergence still cannot be achieved, this analysis will not be performed. Detailed strategy in Section 3.7 will be applied.

4.5 Multiple comparisons/multiplicity

To control the overall type I error rate at 0.05 for the multiple comparisons in the primary and secondary efficacy endpoints, a closed testing procedure under a parallel gatekeeping framework will be applied (Sun, 2018).

Under this framework, each bimekizumab dose of 320mg Q2W and 320mg Q4W will be compared to placebo in the first instance at a familywise error rate of 0.025 (α/2). Simultaneously within each dose, closed testing for the primary and secondary efficacy endpoints will be performed as follows:

1. Step 1: Test $HiSCR_{50}$ at significance level 0.025.
2. Steps 2 to 5 – If Step 1 is significant at 0.025 then test each secondary endpoint sequentially in the order shown in Figure 4-1, moving to the next step only if significance achieved at 0.025.
3. In the event that Step 5 is significant at 0.025 for a given dose, then Steps 1 to 5 will be repeated for the other dose using a significance level of 0.05.

The secondary efficacy variables supporting the primary efficacy variable are listed below, and will be included in the multiplicity adjustment using the analysis methods specified in Section 8.3:

1. Proportion of study participants who achieve $HiSCR_{75}$ at Week 16.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

2. Absolute CFB in DLQI Total Score at Week 16.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

3. Absolute CFB in Skin Pain Score at Week 16, as assessed by the "worst skin pain" item (11-point numeric rating scale) in the HSSDD.

a. bimekizumab 320mg Q2W vs placebo

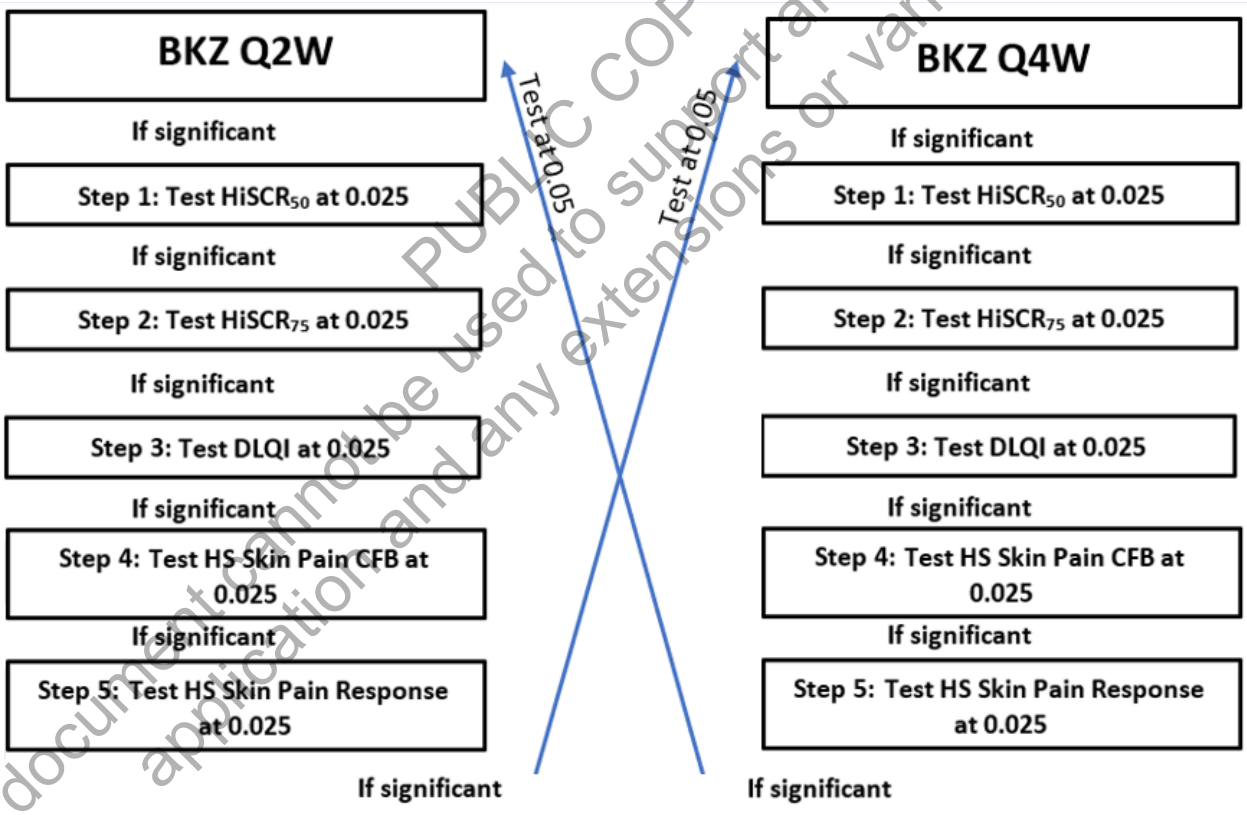
b. bimekizumab 320mg Q4W vs placebo

4. Skin pain response at Week 16, based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) at Week 16 among study participants with a score of ≥3 at Baseline.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

Figure 4-1: Sequence of testing



AN=abscess and inflammatory nodule; DLQI=Dermatology Life Quality Index; $HiSCR_{50}$=a 50% reduction in the total AN count with no increase from Baseline in abscess or draining tunnel count; $HiSCR_{75}$=a 75% reduction in the total AN count with no increase from Baseline in abscess or draining tunnel count; HS=hidradenitis Suppurativa; Q2W=every 2 weeks; Q4W=every 4 weeks

HS skin pain response is tested among study participants with a score of ≥3 at Baseline.

4.6 Use of an efficacy subset of participants

A sensitivity analysis of the primary endpoint will be performed based on the FAS, the PPS, and the CFS.

4.7 Active-control studies intended to show equivalence

Not applicable.

4.8 Examination of subgroups

Subgroup analyses will be performed on the $HiSCR_{50}$, $HiSCR_{75}$, and worst skin pain response endpoints by visit for the Initial Treatment Period and Maintenance Treatment Period. Additional subgroup analyses will be performed on the CFB in the worst skin pain score as measured by HSSDD and in the DLQI total score through Week 16 as described below.

Along with the tables described, there will be tables for $HiSCR_{50}$, $HiSCR_{75}$, and skin pain response endpoints which display the response difference and 95% CIs between each bimekizumab dose regimen versus placebo for each of the subgroups at Week 16. Corresponding forest plots will be prepared.

The following subgroup variables will be determined using Baseline data, except for analgesic use, lesion intervention, and antibody positivity:

- Age (<40 years, 40 to <65 years, ≥65 years)
- Gender (male, female)
- Disease duration (<median, ≥median)

The median disease duration will be calculated based on all participants in the analysis set used for analysis.

- Region (North America [Canada, USA], Western Europe [Belgium, France, Germany, Italy, Norway, Spain, Switzerland, Denmark, Netherlands], Central/Eastern Europe [Greece], Asia/Australia [Australia, Israel, Turkey])
- Weight (≤100 kg, >100 kg)
- BMI (<25 kg/m^2, 25 to <30 kg/m^2, ≥30 kg/m^2)
- Race (Black or African American, White, All Other Races [American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other/Mixed])
- Systemic antibiotic therapy at randomization (yes, no)
- Prior biologic therapy for any indication (yes, no)
- Prior biologic therapy for HS (yes, no)
- Hurley Stage at Baseline (II or III)
- Analgesic users (yes, no), separately for the Initial Treatment Period and the Maintenance Treatment Period (Section 6.4.2 specifies how participants are classified as analgesic users)
- Lesion intervention (yes, no), separately for the Initial Treatment Period and the Maintenance Treatment Period

- Antibody positivity (confirmatory assay: negative or positive; see Section 9.3.2)
- Antihistamine users during the Initial Treatment Period (yes, no) (Section 6.4.2 specifies how participants are classified as antihistamine users) (applicable only to the skin pain response endpoint)

The following subgroups for analysis on the CFB in the worst skin pain score as measured by HSSDD and in the DLQI total score will be determined based on medication use during the Initial Treatment Period:

- Antihistamines users during the Initial Treatment Period (yes, no) (Section 6.4.2 specifies how participants are classified as antihistamine users)
- Analgesics users during the Initial Treatment Period (yes, no) (Section 6.4.2 specifies how participants are classified as analgesic users)
- Systemic antibiotic therapy start/increase after randomization during the Initial Treatment Period (yes, no)

All summaries will be based on imputed data as appropriate and will include descriptive statistics only.

5 STUDY POPULATION CHARACTERISTICS

5.1 Study participant disposition

Summaries of reasons for screen failures (for ES), disposition of participants (for ES), disposition of analysis sets (for RS), disposition and discontinuation reasons in the Initial Treatment Period (for RS) and the Maintenance Treatment Period (for MS), as well as the participants who discontinued due to AEs in the Initial Treatment Period (for RS) and Maintenance Treatment Period (for MS) will be produced. The disposition of participants for all participants screened will include the number of participants included in each analysis set (ES, RS, SS, FAS, AMS, MS, PPS, and PK-PPS) overall and by site.

Participants are defined as completing the Initial Treatment Period if they have a Week 16 visit, or if they fail to attend the Week 16 visit but attend at least one visit in the Maintenance Treatment Period.

The following listings for participant disposition will be provided: participants who did not meet study eligibility criteria (for ES), participant disposition (for ES), participant discontinuation (RS), visit dates (for RS), participant analysis sets (for ES), participants excluded from efficacy analysis (for RS).

To assess participant disposition (entry and periods in the study) during the COVID-19 pandemic, study participant disposition will also be assessed by period of the COVID-19 pandemic (pre – during – post), by comparing the dates of visits (or events) to the dates of the COVID-19 pandemic period. The dates to categorize the periods of the COVID-19 pandemic (pre/during/post) are defined below:

- Pre-COVID-19 pandemic period: Period prior to COVID-19 pandemic start date defined as 11-Mar-2020

- COVID-19 pandemic period: Period from 11-Mar-2020 though the COVID-19 pandemic end date which is currently not defined at the time of approval of the SAP
- Post-COVID-19 pandemic period: Period after the declaration of the end of the pandemic

5.2 Impact of COVID-19

A listing of visits affected by COVID-19 will be presented based on the ES including the visit, date of visit, relationship to COVID-19, impact category and a narrative (short description) of the event. These data will be summarized for non-randomized participants and by treatment group and overall, for enrolled participants.

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

In addition, in order to assess the potential impact of COVID-19 on the collection and reporting of efficacy data, a separate summary on the RS will be presented to display missing data as well as data collected via an alternative modality (e.g.: phone, video call) for efficacy endpoints included in the hierarchy (Section 4.5). For these displays, missing data will be presented only for visits affected by COVID-19, as reported on the dedicated eCRF page. Missing data at other visits and for other reasons will not be included. Note that the remote contingencies for COVID-19 or other exceptional circumstances are not applicable to efficacy assessments and documentation (eg, lesion-based assessments, photography) that require direct face-to-face physician/participant interaction.

5.3 Protocol deviations

Summaries, based on the RS and the MS, displaying the number and percentage of participants with an important protocol deviation (including a summary of participants excluded from the PPS or PK-PPS due to important protocol deviations) by treatment group in the Initial Treatment Period and in the Maintenance Treatment Period, respectively, will be provided. A separate summary of participants with protocol deviations related to COVID-19 will be provided.

A by-participant listing of protocol deviations will be provided. Protocol deviations (eg, missing assessments or visits) related to COVID-19 will be listed separately.

6 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

All summaries detailed in this section will be performed on the RS by treatment group. Summaries for demographics and other baseline characteristics will also be repeated on SS and MS. If the RS and SS are identical, the SS summaries will not be created.

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 study participants, mean, SD, minimum, median, and maximum).

- Age (years)
- Height (cm)

- Weight (kg)
- BMI (kg/m^2)

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

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

The following categorical variables will be summarized generally 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)
- Body weight (≤100 kg, >100 kg)
- Gender
- Race
- Ethnicity
- Ethnic subgroup
- BMI (<25 kg/m2, 25 to <30 kg/m2, ≥30 kg/m2)
- Region
- Smoking history
- Country

By-participant listing of demographics for all study participants screened will be provided.

Childbearing potential and lifestyle will be collected at Screening.

6.2 Other Baseline characteristics

The following Baseline disease characteristics will be summarized by treatment group:

- Lesion counts by anatomical region and lesion type, total lesion counts across anatomical regions by lesion type, Hurley Stage by anatomical region and worst overall Hurley Stage across anatomical regions
- IHS4 score, individual items of the HSSDD, HS-Physician's Global Assessment, DLQI total score and HiSQOL domain and total scores
- hs-CRP
- Duration of disease (years)

Duration of disease (years) will be calculated as:

$$\text{Disease Duration} = \frac{(\text{Date of randomization} - \text{Date of HS Diagnosis}^1)}{365.25}.$$

[1]If the date of HS diagnosis is partial, it will 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). Note that if the date of randomization is missing then the duration of disease will be derived using the date of screening. If the imputed date results in a participant having a disease duration of less than 6 months and the inclusion criterion related to having HS for at least 6 months is confirmed to not have been violated, then the participant's duration of disease will be set to 6 months. If that criterion has been violated, then the participant's duration of disease will be the imputed value of less than 6 months.

- Duration of disease (<median, ≥median)

The median disease duration will be calculated based on all participants in the analysis set used for the summary.

- Baseline antibiotic use (yes, no) (According to the randomization strata)
- Baseline antibiotic use (yes, no) (Derived)
- Hurley Stage at Baseline (According to the randomization strata)
- Hurley Stage at Baseline (Derived)

In addition, the following Baseline disease characteristics will be summarized by the derived Baseline Hurley Stage and by the derived Baseline antibiotic use and treatment group for the RS:

- IHS4 score
- "worst skin pain score" and "average skin pain score" in the HSSDD
- HS-Physician's Global Assessment
- DLQI total score
- hs-CRP
- Duration of disease (years)
- Total lesion counts

6.3 Medical history and concomitant diseases

Previous and ongoing medical history will be summarized by treatment groups, system organ class (SOC), and preferred term (PT) using MedDRA®. Medical procedures are not coded.

The following listings for medical history and concomitant diseases will be provided: medical history, HS history, concomitant medical procedures, previous and ongoing medical history glossary, previous and ongoing medical history conditions, and procedure history.

6.4 Prior and concomitant medications

Prior medications include any medications that started before the start date of study medication. Concomitant medications are any medication that has a start date on or after the start date of study medication, or any medication that has a start date on or before the last dose of study medication + 28 days (whether placebo or bimekizumab).

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

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

The number and percentage of participants taking prior 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 PT. Prior antibiotic medications will be summarized similarly.

The number and percentage of participants taking concomitant medications will be summarized similarly for the Initial Treatment Period and Maintenance Treatment separately. The number and percentage of participants taking concomitant antibiotic medications, antihistamines, and analgesics will be summarized separately by treatment group, overall, and by ATC class, presenting Anatomical Main Group (ATC Level 1), Pharmacological Subgroup (ATC level 3), and PT for the Initial Treatment Period and Maintenance Treatment Period separately.

Separate summaries will be presented for participants taking rescue medication for the Initial Treatment Period and Maintenance Treatment Period separately, identified by a 'yes' response to the 'Is this a rescue medication' question on the electronic case report form (eCRF). This summary will be performed separately for analgesic use and antibiotic use.

Additional summaries for the Initial Treatment Period and Maintenance Treatment Period will be presented for participants taking systemic antibiotic medications that qualify as intercurrent events as described in Section 3.9.

The number and percentage of study participants with concomitant vaccines for COVID-19 will be summarized by treatment group, overall and by World Health Organization Drug Dictionary Standardized Drug Grouping (SDG), presenting SDG subgroup, and preferred term. The SDG subgroup Vaccines for COVID-19 will be used to identify vaccines for COVID-19 using the narrow scope; this subgroup is divided further into separate subgroups which is the level that will be presented. The number of individual occurrences of the vaccine for COVID-19 will also be summarized.

A listing of concomitant vaccines for COVID-19 will be provided.

6.4.1 Assignment of medications to study period

The following rules will be used to assign a concomitant medication to a study period:

- **Initial Treatment Period:** a medication will be assigned to the Initial Treatment Period if it has been taken at least once between the first administration of IMP on Day 1 up to Week 16. This includes medications that started prior to the Initial Treatment Period and those that continued into the Maintenance Treatment Period.
- **Maintenance Treatment Period:** a medication will be assigned to the Maintenance Treatment Period if it has been taken at least once between Week 16 and the final visit. This includes medications that started prior to the Maintenance Treatment Period.

Thus, a medication taken from the time of the first drug administration in the Initial Treatment Period to any timepoint after Week 16 will be assigned to both the Initial Treatment Period and the Maintenance Treatment Period.

Methods for dealing with partial dates are specified in Section 4.2.4.

6.4.2 Classification of participants as analgesic, antihistamine users

If a participant has taken a new analgesic/increased regimen of analgesic, or taken an antihistamine, on 1 or more days (need not be consecutive) in a study period (Initial Treatment Period or Maintenance Treatment Period), then for that period the participant will be classified as an analgesic or antihistamine user, respectively. The period under consideration is to match the period as defined for the HSSDD for the Initial Treatment Period or HSSQ for the Maintenance Treatment Period, based on dates/times of the medications taken.

New analgesic/increased regimen of analgesic use, regardless of indication, is defined as an analgesic medication with start date on or after the first dose of study medication. Stable analgesics (ie, analgesics which were taken already before randomization) will not be included in this category of analgesic user. This classification will be used for selected subgroup analyses.

Antihistamine use is identified by considering the ATC classification. This classification is used for analyzing the Worst Itch endpoint and for selected subgroup analyses, by visit, for the Initial Treatment Period and Maintenance Period as applicable.

Additionally, if a participant has taken a new analgesic/increased regimen of analgesic on 1 or more days (need not be consecutive) prior to the Week 16 visit, then for that week the participant will be classified as an analgesic user. This classification will be used to adjust the formal analysis of the Worst Skin Pain secondary endpoints. If there is a visit date but no HSSDD available at the visit, then the analgesic/antihistamine user status for that week will be derived based on the visit date. If there is no visit available, then the weekly analgesic/antihistamine user status will default to the analgesic/antihistamine status for the overall study period.

7 MEASUREMENTS OF TREATMENT COMPLIANCE

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

Treatment compliance will be calculated as:

$$\frac{N_{actual}}{N_{expected}} \times 100\%$$

where N_{actual} is the total number of actual (completed) injections, and $N_{expected}$ is the total number of expected injections. In this study, dosing occurs every 2 weeks from Week 0 to Week 46, where 2 injections are administered at each given visit either with active dosing or placebo injection. It is expected that a participant should complete a total of 48 injections by the end of study. If a participant discontinues early, then the number of expected injections is based on the time of early discontinuation relative to the dosing visits. For example, if a participant discontinues after Week 8 visit and prior to Week 10 visit, the total number of expected injections will be 10.

A summary of percent treatment compliance categorized as <75% and ≥75% will be provided by treatment group for each study period (Initial Treatment Period for the RS, Maintenance Treatment Period for the MS, and the combined Initial and Maintenance Treatment Period for the AMS).

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

8 EFFICACY ANALYSES

All efficacy analyses of primary, secondary, and other variables will be performed on the RS unless otherwise specified. All efficacy summary tables will be displayed by treatment sequence unless otherwise specified. The primary and secondary endpoints, and their components, will also be summarized by the derived Hurley Stage at Baseline (grouping each stage and overall) and treatment sequence and by the derived Baseline antibiotic use (yes/no and overall) and treatment group.

8.1 Lesion count assessment

The primary efficacy endpoint and some of the secondary and other efficacy endpoints discussed in Section 8.3 and Section 8.4 are based on the assessment and/or counts of different types of lesions in the following main anatomical regions at each visit.

- Inguinal (groin)
- Axillary (armpit)
- Chest/breast
- Gluteal
- Abdomen including supra pubic
- Back
- Head
- Neck
- Leg
- Other

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

- Inguinal excluding genital and pubic area
- Inguinal including genital and pubic area
- Submammary
- Intermammary
- Chest
- Breast
- Gluteal – Buttocks
- Gluteal – Perianal/Perineal
- Scalp
- Face

All "Other" anatomical regions with lesions present will be specified in free text on the eCRF.

The number of each of the following types of lesions will be recorded in each anatomical region, and then summed across all anatomical regions:

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

If participants undergo lesion interventions as specified in the study protocol, the affected lesions will be counted by the Investigator as permanently present, thus accounting for potential bias due to the intervention. Section 8.4.24 specifies how the intervention data will be presented.

8.2 Primary efficacy endpoint

The primary and sensitivity analyses of $HiSCR_{50}$ response at Week 16 are summarized in Table 8–1.

Table 8–1: Estimand Details and Attributes for Primary Endpoint

		Estimands for Primary Endpoint			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
Primary Objective: To evaluate the efficacy of bimekizumab in study participants with moderate to severe HS					
$HiSCR_{50}$	Primary (Section 8.2.2)	$HiSCR_{50}$ response at Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. A **composite strategy** will be used, ie, the occurrence of an intercurrent event will be handled by evaluating the corresponding participants as treatment failures (nonresponders).	The odds ratio versus placebo based on a logistic regression. Missing values will be imputed using MI – MCMC/Monotone Regression under a missing at random assumption.

Table 8–1: Estimand Details and Attributes for Primary Endpoint

		Estimands for Primary Endpoint			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
$HiSCR_{50}$	Sensitivity (Section 8.2.3.1)	$HiSCR_{50}$ response at Week 16	RS	**Composite strategy**, as for the primary analysis where the occurrence of an intercurrent event will be handled by evaluating the corresponding participants as nonresponders.	The odds ratio versus placebo based on a logistic regression. Missing values for any other reason will also be imputed as nonresponders.
$HiSCR_{50}$	Sensitivity (Section 8.2.3.2)	$HiSCR_{50}$ response at Week 16	RS	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo based on a logistic regression. Missing values will be imputed using MI – Reference-Based Regression under a missing not at random assumption.

Table 8–1: Estimand Details and Attributes for Primary Endpoint

		Estimands for Primary Endpoint			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
$HiSCR_{50}$	Sensitivity (Section 8.2.3.3)	$HiSCR_{50}$ response at Week 16	RS	**Composite strategy**[a], as for the primary analysis.	A tipping point analysis will be used where various delta adjustments will be made to the assumed responses on the monotone missing data in each treatment group independently with various degrees of plausibility. The odds ratio versus placebo is based on a logistic regression for each value of delta.
$HiSCR_{50}$	Sensitivity (Section 8.2.3.4)	$HiSCR_{50}$ response at Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **treatment policy strategy**, whereby the data from the Initial Treatment Period are used regardless of whether the intercurrent event occurred.	The odds ratio versus placebo is based on a logistic regression, as for the primary analysis.

Table 8–1: Estimand Details and Attributes for Primary Endpoint

		Estimands for Primary Endpoint			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
$HiSCR_{50}$	Sensitivity (Section 8.2.3.5)	$HiSCR_{50}$ response at Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a lesion count assessment at Week 16 who have not had an intercurrent event on or before Week 16 are included.	The odds ratio versus placebo is based on a logistic regression. Missing values will not be imputed.
$HiSCR_{50}$	Sensitivity (Section 8.2.3.6)	$HiSCR_{50}$ response at Week 16	FAS	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo is based on a logistic regression, as for the primary analysis.
$HiSCR_{50}$	Sensitivity (Section 8.2.3.7)	$HiSCR_{50}$ response at Week 16	PPS	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo is based on a logistic regression, as for the primary analysis.
$HiSCR_{50}$	Sensitivity (Section 8.2.3.8)	$HiSCR_{50}$ response at Week 16	CFS	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo is based on a logistic regression, as for the primary analysis.

Table 8–1: Estimand Details and Attributes for Primary Endpoint

		Estimands for Primary Endpoint			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
$HiSCR_{50}$	Sensitivity (Section 8.2.3.9)	$HiSCR_{50}$ response at Week 16	RS	The same two intercurrent events used for the primary analysis will be used. Any missing data due to COVID-19 will also be considered an intercurrent event. A **composite strategy** will be used, ie, the occurrence of an intercurrent event will be handled by evaluating the corresponding participants as treatment failures (nonresponders).	The odds ratio versus placebo is based on a logistic regression, as for the primary analysis.
$HiSCR_{50}$	Sensitivity (Section 8.2.3.10)	$HiSCR_{50}$ response at Week 16	RS	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo based on a stratified Cochran-Mantel-Haenszel (CMH) test. Missing values not preceded by an intercurrent event will be imputed using MI – MCMC/Monotone Regression under a missing at random assumption.

CFS=Covid-19 Free Set; CMH=Cochran-Mantel-Haenszel; COVID-19=coronavirus disease 2019; FAS=Full Analysis Set; HiSCR=Hidradenitis Suppurativa Clinical Response; IES=intercurrent event(s) strategy; MCMC=Markov Chain Monte Carlo; MI= multiple imputation; PLS=Population-level summary; Pop=Population; PPS=Per-Protocol Set; RS=Randomized Set

[a] The composite estimand strategy will be modified in the tipping point analysis such that participants with intercurrent events will be treated as nonresponders only in the bimekizumab treatment groups.

8.2.1 Derivation of $HiSCR_{50}$ at Week 16

The following algorithm will be applied to derive $HiSCR_{50}$ at each visit, based on total lesion counts across anatomical regions for the 3 relevant lesion types recorded as specified above in Section 8.1:

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

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

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

In cases where the inflammatory nodule, abscess or draining tunnel (fistula/sinus tract) count is missing and will not allow for the $HiSCR_{50}$ calculation, the rules for handling missing values in the analysis will be applied (Section 4.2.1 and Section 8.2.3).

The primary efficacy endpoint is attained if the participant has a $HiSCR_{50}$ of 1 at Week 16.

8.2.2 Primary analysis of the primary efficacy endpoint

The primary endpoint is the $HiSCR_{50}$ response at Week 16 and corresponding analyses are based on the RS. The primary efficacy analysis will evaluate the composite estimand in the RS as described in Table 8–1. The composite estimand combines the clinically meaningful improvement from Baseline based on the $HiSCR_{50}$ response and completion of study treatment through Week 16 without receiving systemic antibiotic rescue medication or discontinuing IMP due to an AE or lack of efficacy.

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

1. Population=Study participants meeting the protocol-specified inclusion/exclusion criteria.
2. Study participant-level outcome=$HiSCR_{50}$ at Week 16.
3. Intercurrent event handling=An intercurrent event is defined as receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. More information is provided in Section 3.9. A composite strategy will be implemented in which a positive clinical outcome is defined as achieving $HiSCR_{50}$ at Week 16 and not receiving systemic antibiotic rescue medication, and not discontinuing study treatment due to an AE or lack of efficacy through Week 16. Any missing data at Week 16 that are not preceded by an intercurrent event will be imputed using multiple imputation as defined in Section 4.2.1. The rationale for this composite estimand is provided in Section 4.2.3.1.

4. Population-level summary measure=conditional odds ratio comparing bimekizumab to placebo adjusted for stratification factors used in the randomization.

The statistical hypothesis for the $HiSCR_{50}$ response at Week 16 is that the conditional odds ratio for the $HiSCR_{50}$ response in the bimekizumab treatment group relative to the placebo group is equal to 1.

The primary analysis will be based on a logistic regression model including a fixed effect for treatment, Hurley stage at Baseline, and Baseline antibiotic use. The odds ratio versus placebo, p-value (from Wald test), and 97.5% CI will be calculated. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose with a corresponding $Z_{\alpha/2}$ of 1.96.

The number and percentage of participants who are $HiSCR_{50}$ responders at Week 16 will be summarized.

By-participant listings of HiSCR responder endpoints will be provided.

8.2.3 Sensitivity analyses of the primary efficacy endpoint

The following sensitivity analyses for the primary efficacy endpoint will be performed to evaluate the assumptions related to the handling of missing data. Details of the estimands for each analysis are described in Table 8–1.

8.2.3.1 Nonresponse imputation

As a sensitivity analysis, any missing data at Week 16 that are not preceded by an intercurrent event (ie, receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy) will be imputed as nonresponse. That is, participants who experience an intercurrent event will be imputed as nonresponder at the timepoint of the event and all subsequent timepoints (including any recorded data after the event), and all missing data will also be imputed as nonresponse.

The same analysis model as in the primary efficacy analyses will then be used on the imputed data set.

8.2.3.2 MI-MCMC / Reference-based imputation

Deviations from the missing at random pattern will be evaluated using a reference-based MI approach (see Section 4.2.2.2). Intermittent missing data will be imputed using the MCMC method. The remaining monotone missing data will be assumed to follow a missing not at random pattern. These data will be imputed using a reference-based approach in which the MI model is based on data from the placebo group, thereby assuming that monotone missing data follow a trajectory similar to the placebo group.

The same analysis model as in the primary efficacy analyses will then be used on the imputed data set.

8.2.3.3 Tipping point analysis

Tipping point analyses will be performed to evaluate missingness assumptions. Various delta adjustments will be made to the assumed responses on the monotone missing data in each treatment group independently with various degrees of plausibility. It will include scenarios where study participants who have missing data and are randomized to bimekizumab have a

lower probability of response compared to study participants who have missing data and were randomized to placebo. This includes the worst-case scenario where study participants who have missing data and are randomized to bimekizumab are considered nonresponders, while study participants who have missing data and were randomized to placebo are considered responders. The goal of the tipping point analysis is to systematically vary assumptions until there is no evidence of a treatment effect (if possible). The plausibility of such required delta adjustment will then be assessed.

Refer to Section 4.2.2 for more details on the methodology.

8.2.3.4 Treatment policy

The treatment policy strategy for addressing intercurrent events will be considered. This will be based on an analysis of all available data at Week 16 regardless of the occurrence of intercurrent events. This analysis will use the same models specified for the primary analysis, where study participants are analyzed according to their randomized treatment, even if they discontinued study treatment for any reason prior to Week 16. Even though efforts will be made to collect the primary outcome data for all study participants at Week 16, there may still be some study participants for whom Week 16 efficacy data cannot be obtained. In this case, missing data will be imputed using multiple imputation under the assumption of MAR (see Section 4.2.2). The same analysis model as in the primary efficacy analyses will then be used on the imputed data set and the resulting inferential statistics will then be combined into a single inference using Rubin's rule.

8.2.3.5 Analysis on observed cases

An additional supportive analysis will be based on observed data only for study participants with a lesion count assessment at Week 16 who have not had an intercurrent event on or before Week 16. Study participants with missing data at Week 16 will be treated as missing (see Section 4.2.2).

The same analysis model as in the primary efficacy analyses will then be used on the imputed data set.

8.2.3.6 Analysis on FAS

The primary efficacy analyses from Section 8.2.2 will be repeated based on the FAS.

8.2.3.7 Analysis on PPS

The primary efficacy analyses from Section 8.2.2 will be repeated based on the PPS.

8.2.3.8 Analysis on CFS

The primary efficacy analyses from Section 8.2.2 will be repeated based on the CFS.

8.2.3.9 Analysis including COVID-19 impact as intercurrent event

An additional sensitivity analysis will include an additional intercurrent event. The composite estimand combines the clinically meaningful improvement from Baseline based on the $HiSCR_{50}$ response and completion of study treatment through Week 16 without receiving systemic antibiotic rescue medication or discontinuing IMP due to an AE or lack of efficacy.

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

1. Population=Study participants meeting the protocol-specified inclusion/exclusion criteria.
2. Study participant-level outcome=$HiSCR_{50}$ at Week 16.
3. Intercurrent event handling=An intercurrent event is defined as receipt of systemic antibiotic rescue medication, discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16, or missing data due to COVID-19. More information is provided in Section 3.9. A composite strategy will be implemented in which a positive clinical outcome is defined as achieving $HiSCR_{50}$ at Week 16 and not receiving systemic antibiotic rescue medication, not discontinuing study treatment due to an AE or lack of efficacy through Week 16, and not having missing data due to COVID-19. Any missing data at Week 16 that are not preceded by an intercurrent event will be imputed using multiple imputation as defined in Section 4.2.1.
4. Population-level summary measure=conditional odds ratio comparing bimekizumab to placebo adjusted for stratification factors used in the randomization.

The same imputation techniques and analysis model as in the primary efficacy analyses will then be used.

8.2.3.10 Cochran-Mantel-Haenszel test

The primary efficacy analyses from Section 8.2.2 will be repeated where the CMH test with fixed effects for treatment, Hurley stage at Baseline, and Baseline antibiotic use will be used as stratification variables. Pairwise treatment comparisons will be made based on the CMH test using the p-value for the general association. The odds ratio and associated confidence interval (CI) based on the Wald test will be provided.

8.2.3.11 Center-by-Treatment Interaction

The center-by-treatment interaction will be tested by adding center and a center-by-treatment interaction term in the logistic regression model described in Section 8.2.2. In the model, center will be based on the original centers prior to pooling. However, if the model is unable to converge due to a low number of participants at a given center, a pooling (see Section 3.7) will be described in order to allow the model to converge. In order to obtain reasonable estimates of variability for a treatment arm at a given center, a minimum of 21 participants will be considered acceptable for a center to be included in the model without pooling. Given the 2:2:2:1 randomization allocation scheme, this should provide a minimum of about 12 participants in the bimekizumab 320mg Q2W treatment group, 6 participants in the bimekizumab 320mg Q4W treatment group, and 3 participants in the placebo treatment group. Centers with fewer than 21 participants will be eligible for pooling. The pooling algorithm used is described in Section 3.7.

In order to achieve model convergence, other explanatory variables eg, Baseline Hurley Stage and Baseline antibiotic use may be dropped from the model. If model convergence is still not achieved, region and a region-by-treatment interaction term will be added to the model instead. Regions are defined in Section 3.7.

If the center-by-treatment interaction is not found to be significant (α=0.05), then no further analyses will be performed. On the other hand, if the interaction is significant, further analyses

will be conducted to determine which center or centers may be the source of interaction. This will be done by running the logistic regression model (including the interaction term) where each center will be systematically removed from the model. This impact of a given center will be based on the change in the interaction p-value when that center is removed. The center or centers that appear to be driving the significant interaction effect will then be removed from the model to verify that conclusions remain the same with or without the influential center(s). This sensitivity analysis will be based on RS with MI/MCMC Monotone Regression for missing data.

8.3 Secondary efficacy endpoints

The secondary efficacy analyses will be performed based on the RS. Sensitivity analyses of the secondary endpoints will be performed on the CFS.

Missing data handling and sensitivity analyses of the secondary efficacy endpoints are described in Section 4.2.1.2.

The analyses of the secondary endpoints are summarized in Table 8–2 .

Table 8–2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
Secondary Objective: Evaluate the efficacy of bimekizumab on other measures of disease activity in study participants with moderate to severe HS					
$HiSCR_{75}$	Secondary (Section 8.3.1)	$HiSCR_{75}$ response at Week 16	RS	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo based on a logistic regression, as for the primary analysis.

Table 8–2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
DLQI	Secondary (Section 8.3.2. 1)	Change from Baseline in DLQI total score to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **hypothetical strategy**, whereby all data at and after the intercurrent event will be treated as missing.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the DLQI total score for participants receiving bimekizumab versus placebo. Missing values will be imputed using MI – MCMC/Monotone Regression under a missing at random assumption.
DLQI	Secondary - Sensitivity (Section 8.3.2. 2)	Change from Baseline in DLQI total score to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a DLQI total score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the DLQI total score for participants receiving bimekizumab versus placebo. Missing values will not be imputed.

Table 8–2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
HSSDD	Secondary (Section 8.3.3.1)	Change from Baseline in worst skin pain score, as assessed by "worst skin pain" item in HSSDD to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **hypothetical strategy**, whereby all data at and after the intercurrent event will be treated as missing.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the worst skin pain score for participants receiving bimekizumab versus placebo. Missing values will be imputed using MI – MCMC/Monotone Regression under a missing at random assumption.
HSSDD	Secondary (Section 8.3.3.2)	Change from Baseline in worst skin pain score, as assessed by "worst skin pain" item in HSSDD to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a HSSDD worst skin pain score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the worst skin pain score for participants receiving bimekizumab versus placebo. Missing values will not be imputed.

Table 8–2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
HSSDD	Secondary (Section 8.3.4.1)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improvement in the weekly worst skin pain score of at least 3 points	RS[a]	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo based on a logistic regression, as for the primary analysis.
HSSDD	Secondary Sensitivity (Section 8.3.4.2.1)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improvement in the weekly worst skin pain score of at least 3 points	RS[a]	**Composite strategy**, as for the primary analysis where the occurrence of an intercurrent event will be handled by evaluating the corresponding participants as nonresponders.	The odds ratio versus placebo based on a logistic regression. Missing values for any other reason will also be imputed as nonresponders.

Table 8–2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
HSSDD	Secondary Sensitivity (Section 8.3.4.2.2)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improve-ment in the weekly worst skin pain score of at least 3 points	RS[a]	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a HSSDD worst skin pain score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	The odds ratio versus placebo is based on a logistic regression. Missing values will not be imputed.

AE=adverse event; ANCOVA=analysis of covariance; DLQI=Dermatology Life Quality Index; HiSCR=Hidradenitis Suppurativa Clinical Response; HSSDD=Hidradenitis Suppurativa Symptom Daily Diary; IES=intercurrent event(s) strategy; LSMD=Least Squares Mean Difference; MCMC=Markov Chain Monte Carlo; MI=multiple imputation; PLS=Population-level summary; Pop=Population; RS=Randomized Set

[a] Analysis includes all study participants in the RS with a Baseline HSSDD Worst Skin Pain score of 3 or higher.

8.3.1 HiSCR$_{75}$ at Week 16

A categorical response variable, HiSCR$_{75}$ at Week 16 is defined to be equal to 1 if %ΔAN is less than or equal to -75%, and the change from Baseline in the abscess count is zero or negative (ie, non-positive) together with a non-positive change from Baseline in the draining tunnel (fistula/sinus tract) count, and 0 otherwise. This definition is introduced for identifying participants who respond to the treatment (1 = responder, 0 = nonresponder). The definition of percentage improvement from Baseline is given in Section 8.2.1.

For HiSCR$_{75}$ at Week 16, logistic regression as specified for the primary analysis will be implemented to test for superiority. The same analysis approach as outlined for the primary efficacy endpoint will be applied.

8.3.2 Change from Baseline in DLQI Total Score at Week 16

The DLQI is a questionnaire designed for use in adult participants with skin diseases and has been used in patients with HS. This is a validated, quality-of-life questionnaire that covers 6 domains including symptoms and feelings, daily activities, leisure, work and school, personal relationships, and treatment, as assessed over the past week.

The scoring of each answer for the DLQI is as follows:

Table 8–3: Dermatology Life Quality Index

DLQI Scoring	
Response	Score
Very much	3
A lot	2
A little	1
Not at all	0
Not relevant	0
Question unanswered	0
Q7: 'prevented work or studying' = yes	3

The DLQI total score is calculated by adding the score of each question. The maximum score is 30, and the minimum score is 0. The higher the score, the more quality of life is impaired.

Meaning of DLQI Total Score

0-1 = no effect at all on patient's life

2-5 = small effect on patient's life

6-10 = moderate effect on patient's life

11-20 = very large effect on patient's life

21-30 = extremely large effect on patient's life

This categorization will not be utilized in the analysis.

Because Q7 has a sub-question (referred to as Q7a here) after the leading yes/no question, some clarifying rules for scoring are provided:

- If Q7 is marked as "yes", a score of 3 is given regardless of the responses to Q7a.
- If Q7 is marked as "no", "not relevant", or is missing and Q7a is "A lot", a score of 2 is given.
- If Q7 is marked as "no", "not relevant", or is missing and Q7a is "A little", a score of 1 is given.
- If Q7 is marked as "no", "not relevant", or is missing and Q7a is "Not at all", a score of 0 is given.
- If Q7 is marked as "no" or "not relevant" and Q7a is missing, a score of 0 is given.
- If Q7 is missing and Q7a is missing, Q7 is considered unanswered (see below for details on how this impacts the DLQI total score).

If 1 question is left unanswered, this is scored 0 and the scores are summed and expressed as usual out of a maximum of 30. If 2 or more questions are left unanswered, the questionnaire is not scored.

Change from Baseline in DLQI total score is defined as Week 16 DLQI total score minus Baseline DLQI total score.

8.3.2.1 Primary analysis of change from Baseline in DLQI Total score at Week 16

Missing data imputation described in Section 4.2.1.2 will be applied.

Change from Baseline in DLQI total score will be presented by treatment group. The analysis model will be based on an ANCOVA with fixed effects of treatment, Hurley Stage at Baseline, Baseline antibiotic use and Baseline value as a covariate. The least square mean (LSM), standard error (SE), and 95% CI for the LSM will be presented by treatment group. For the comparison between placebo and bimekizumab: the difference between the LSM, the associated 97.5% CI for the contrasts, and the corresponding p-value will be presented. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose with a corresponding $Z_{\alpha/2}$ of 1.96. Estimand and intercurrent event details are specified in Table 8–2.

8.3.2.2 Sensitivity analysis of change from Baseline in DLQI Total score at Week 16

A sensitivity analysis using the same analysis model as in Section 8.3.2.1 will be used, based on observed data. Estimand and intercurrent event details are specified in Table 8–2.

8.3.3 Change from Baseline in Skin Pain score at Week 16, as assessed by the "worst skin pain" item in the HSSDD

The items on the HSSDD assess patients' perception of the core symptoms of HS experienced in the past 24 hours: pain, smell or odor, drainage or oozing from HS lesions, and itch on an 11-point numeric rating scale (NRS). Two items assess skin pain: worst skin pain and average skin pain.

Weekly averages will be derived for each of the items of the HSSDD for weeks matching the post-Baseline dosing weeks up to Week 16. A weekly average is the sum of the scored item over the course of the study week divided by the number of days on which the item was completed. The weekly averages will be relative to the respective visit date except for Baseline, which will be anchored to the first dose of study drug. A weekly average will only be calculated if at least 4 non-missing values (not necessarily consecutive) are available. Otherwise, the HSSDD weekly average for the given question will be set to missing.

Baseline will be computed as the average from the first 7 consecutive day period in which there are at least 4 non-missing entries. That is, first consider the first 7 consecutive days prior to the Baseline visit, but not including the Baseline visit day itself. If there are at least 4 non-missing values (not necessarily consecutive), then the Baseline average will be calculated. If there are less than 4 values, the 7 consecutive day period will move one day earlier. If there are at least 4 non-missing values (not necessarily consecutive) in that period, then the Baseline average will be calculated. This will continue until there are at least 4 non-missing values in a 7 consecutive day

period in the 14 days prior to Baseline. If there is no period in which there are at least 4 non-missing entries, then the Baseline value will be set to missing.

Change from Baseline in worst skin pain score is defined as the average Week 16 worst skin pain score minus the Baseline worst skin pain score. Missing data imputation described in Section 4.2.1.2 will be applied to the weekly averages and not to the individual daily PRO data.

8.3.3.1 Primary analysis of change from Baseline in skin pain score at Week 16

Change from Baseline in worst skin pain score will be presented by treatment group. The analysis model will be based on an ANCOVA with fixed effects of treatment, Hurley Stage at Baseline, Baseline antibiotic use, analgesic use (Section 6.4.2) and Baseline value as a covariate.

The LSM, SE, and 95% CI for the LSM will be presented by treatment group. For the comparison between placebo and bimekizumab, the difference between the LSM, the associated 97.5% CI for the contrasts, and the corresponding p-value will be presented. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose with a corresponding $Z_{\alpha/2}$ of 1.96.

8.3.3.2 Sensitivity analysis of change from Baseline in skin pain score at Week 16

A sensitivity analysis using the same analysis model as in Section 8.3.3.1 will be used, based on observed data. Estimand and intercurrent event details are specified in Table 8–2.

8.3.4 HSSDD skin pain response at Week 16

The analysis set for the analyses of the skin pain response will be restricted to those study participants in the RS with a Baseline worst skin pain score of 3 or higher. The weekly scores and Baseline score are derived as specified in Section 8.3.3.

8.3.4.1 Primary analysis of skin pain response at Week 16

Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, is defined as an improvement in the weekly worst skin pain score of at least 3 points versus Baseline.

The primary analysis will be based on a logistic regression model including a fixed effect for treatment, Hurley stage at Baseline, Baseline antibiotic use, and analgesic use (Section 6.4.2).

The odds ratio versus placebo, p-value (from Wald test), and 97.5% CI will be calculated. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose. Missing data will be handled as specified in Section 4.2.1.2. Estimand and intercurrent event details are specified in Table 8–2.

The number and percentage of participants who are pain responders at Week 16 will be summarized by treatment group.

By-participant listings of pain response status will be provided.

8.3.4.2 Sensitivity analyses of Skin Pain Response at Week 16

8.3.4.2.1 Nonresponse imputation

As a sensitivity analysis, any missing data at Week 16 that are not preceded by an intercurrent event (Table 8–2) will be imputed as nonresponse. That is, participants who experience an intercurrent event will be imputed as nonresponder at the timepoint of the event and all subsequent timepoints (including any recorded data after the event), and all missing data will also be imputed as nonresponse.

The same analysis model as Section 8.3.4.1 will then be used on the imputed data set.

8.3.4.2.2 Analysis on observed cases

An additional supportive analysis will be based on observed data only for study participants with a worst skin pain score at Week 16 who have not had an intercurrent event on or before Week 16. Study participants with missing data or who have prematurely discontinued study treatment will be treated as missing (see Section 4.2.2).

The same analysis model as in Section 8.3.4.1 will then be used on the imputed data set.

8.4 Other efficacy endpoints

The other efficacy endpoints are listed below and will be evaluated according to the planned assessments in the protocol. This excludes the timepoints for the primary and secondary endpoints specified above in Section 8.2.1 and Section 8.3.

Missing data handling for these endpoints is described in Section 4.2.1.3.

8.4.1 HiSCR endpoints

8.4.1.1 $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$

Categorical response variables $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ are defined to be equal to 1 if %ΔAN is less than or equal to -25%, -50%, -75%, -90%, and -100%, respectively, and the change from Baseline in the abscess count is zero or negative (ie, non-positive) together with a non-positive change from Baseline in the draining tunnel (fistula/sinus tract) count, and 0 otherwise. This definition is introduced for identifying study participants who respond to the treatment (1 = responder, 0 = nonresponder). The definition of percentage improvement from Baseline is given in Section 8.2.1.

$HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response will be summarized using frequency tables by treatment group and visit.

A line plot of the HiSCR responder ($HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$) rate over time, by treatment group, will be produced.

In order to investigate the effect of intercurrent event handling and missing data handling on the binary response variables, the following iterations of each of the aforementioned plots ($HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$, respectively) will be presented, with corresponding summary statistics tables:

- Observed data only (non-imputation for either intercurrent events or missing data)

- Non-response imputation to reflect intercurrent events, non-imputation for missing data
- Full imputation: non-response imputation to reflect intercurrent events, imputation for missing data per Section 4.2.1.3

Bar charts of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ will be produced by visit and Hurley Stage at Baseline for the Initial Treatment Period and the combined Initial and Maintenance Treatment Period. These bar charts will be repeated for HiSCR by visit and Baseline antibiotic use.

Another bar chart of HiSCR rate at Week 16 will be produced by Hurley Stage at Baseline and Baseline antibiotic use.

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

8.4.1.2 Time to response of $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$

See Section 8.4.1.1 for the derivation of HiSCR.

Initial Treatment Period

Time to $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ (in days) during the Initial Treatment Period will each be calculated as:

> Min (Date of first HiSCRxx response, Date of Week 16 visit) – Date of first dose of study medication + 1, here xx represents 25, 50, 75, 90, 100 respectively. All visits including unscheduled visits are considered.

Participants who discontinue study treatment without achieving a given HiSCR response prior to Week 16 visit will be censored at the date of last lesion count assessment. Participants who reach the Week 16 Visit without achieving the given response will be censored at the date of the Week 16 Visit. Participants who experience an intercurrent event prior to achieving a HiSCR response will be censored at the date of the intercurrent event. Participants will be censored at Baseline if there is no Baseline lesion count assessment or no Post-Baseline lesion count assessment.

Time to $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response during Initial Treatment Period will each be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.

Kaplan-Meier plots of time to HiSCR responses will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 (since there are no responders at Week 0) and will increase over time, representing time to achieving the response.

The median time to response, including the 2-sided 95% confidence interval, will be calculated for each treatment. Comparisons of bimekizumab vs placebo will be analyzed using a log-rank test stratified by Hurley Stage at Baseline and Baseline antibiotic use.

Combined Initial and Maintenance Treatment Period

An additional time to $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ (in days) during the combined Initial and Maintenance Treatment Period will be calculated as above, where the Week 48 visit is considered instead of Week 16.

Time to $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response during the combined Initial and Maintenance Treatment Period will each be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.

Kaplan-Meier plots of time to HiSCR responses will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 (since there are no responders at Week 0) and will increase over time, representing time to achieving the response.

The median time to response, including the 2-sided 95% confidence interval, will be calculated for each treatment.

8.4.1.3 $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response at both Weeks 16 and 48

See Section 8.4.1.1 for the derivation of HiSCR response.

The number and percentage of $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ responders at both Weeks 16 and 48 will be summarized based on the RS and MS.

Missing data for the above summaries will be handled using NRI. That is, participants are counted as responders only if they have an observed HiSCR at both Weeks 16 and 48 and have no intercurrent events through Week 48. Otherwise, they are treated as not responding.

8.4.1.4 $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ among Week 16 Responders

See Section 8.4.1.1 for the derivation of HiSCR response.

Summaries of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ responders at each visit from Week 16 through Week 48 will be summarized based on a subset of participants in the MS who achieve response at Week 16. The summaries will be as follows:

- $HiSCR_{50}$ responder rate based on participants who achieved $HiSCR_{50}$ response at Week 16
- $HiSCR_{75}$ responder rate based on participants who achieved $HiSCR_{75}$ response at Week 16
- $HiSCR_{90}$ responder rate based on participants who achieved $HiSCR_{90}$ response at Week 16
- $HiSCR_{100}$ responder rate based on participants who achieved $HiSCR_{100}$ response at Week 16

Line plots of the above HiSCR responder rate categories over time (from Week 16 to Week 48), by treatment group, will be produced.

8.4.1.5 Time to loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ in Week 16 responders

See Section 8.4.1.1 for the derivation of HiSCR response.

Time to loss of response will be based on the MS and include only participants who had the corresponding HiSCR response at Week 16 (considering intercurrent event handling from the composite estimand described in Section 8.2.2).

Time to loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ (in days) is defined as: Date of loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ - Date of Week 16 treatment administration + 1.

Time to loss of response will be estimated and presented using the Kaplan-Meier product-limit method for each treatment. Participants who experience an intercurrent event prior to loss of response will be considered as having lost response on the date of intercurrent event. Participants who reach the Week 48 Visit without loss of response will be censored at the date of the Week 48 Visit. Participants who discontinue treatment or study, for reasons other than those already defined for an intercurrent event, and who have not yet displayed loss of response by the time of withdrawal, will be censored at the date of the last lesion count assessment.

8.4.1.6 Partial response

See Section 8.2.1 for the derivation of AN count.

A partial response is defined as a ≥25% reduction in AN count from Baseline (Week 0) at a particular timepoint.

The number and percentage of participants who are partial responders at Week 16 and become $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ responders in the Maintenance Treatment Period will be summarized by treatment group and visit. These analyses will be based on the subset of participants in the MS that are partial responders but not $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ responders, respectively, at Week 16. These summaries will be based on observed case data and will not consider the occurrence of intercurrent events.

8.4.2 Lesion count

8.4.2.1 Change from Baseline in lesion count

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

- Abscesses
- Inflammatory nodules
- ANs
- Non-inflammatory nodules
- Draining tunnels (fistulas/sinus tracts)
- Non-draining tunnels (fistulas/sinus tracts)
- HS scars

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

A line plot of the percentage change from Baseline in AN count over time by treatment group will be produced. Separate plots will also be produced for the percentage change from Baseline

in the abscess count, inflammatory nodules count, and draining tunnel count, respectively, over time.

In order to investigate the effect of intercurrent events and missing data handling on lesion count data, the following **additional** iterations of each of the aforementioned plots (percentage change in AN count, abscess count, inflammatory nodule count and draining tunnel count, respectively) will be presented, with corresponding summary statistics tables:

- Intercurrent events:
  - Participants who have experienced intercurrent events
  - Participants who have not experienced intercurrent events
- Missing data handling:
  - Observed lesion counts (i.e., non-imputation)
  - Imputed lesion counts

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

8.4.2.2 AN count of 0, 1, or 2

The number and percentage of participants with an AN count of 0, 1, or 2 will be presented by treatment group and visit. The denominator for the percentage calculations will be the number of participants in each treatment group in the RS with non-missing data at each visit.

8.4.2.3 AN25, AN50, AN75, AN90, AN100

Categorical response variables AN_{25}, AN_{50}, AN_{75}, AN_{90}, and AN_{100} are defined to be equal to 1 if %ΔAN is less than or equal to -25%, -50%, -75%, -90%, and -100%, respectively, and 0 otherwise. This definition is introduced for identifying participants who respond to the treatment (1 = responder, 0 = nonresponder). The definition of percentage change from Baseline is given in Section 8.2.1. AN25, AN50, AN75, AN90, and AN100 response will be summarized using frequency tables by treatment group for each visit.

8.4.3 Flare by Week 16

See Section 8.2.1 for the derivation of AN count.

Disease flare by Week 16 is defined as at least a 25% increase in AN count with an absolute increase of ≥2 AN relative to Baseline is observed by Week 16. A participant's disease flare status (yes/no) will be determined at each visit.

The number of participants who experience at least 1 disease flare by Week 16 will be analyzed using a logistic regression model including a fixed effect for treatment, Hurley stage at Baseline, and Baseline antibiotic use. The odds ratio versus placebo, p-value (from Wald test), and CI will be calculated. Missing data will be handled as described in Section 4.2.1.2.

8.4.4 Flare relative to Baseline

See Section 8.4.3 for the derivation of flare.

Disease flare status will be summarized by treatment group and visit using frequencies and percentages. The denominator for the percentage calculations will be the number of participants with non-missing data in each treatment group. This summary will also include the number of participants with any flare in the Initial Period, Maintenance Period, and the combined Initial and Maintenance Period. A bar chart of percentage of participants with flare, by visit and treatment, will be presented.

In addition, for each participant, the number of flares during the Initial Treatment Period will be calculated and summarized by treatment group. A corresponding histogram summarizing the number of flares during the Initial Treatment Period will be presented.

8.4.5 Time to flare by Week 16

See Section 8.4.3 for the derivation of flare.

Time to flare (in days) during the Initial Treatment Period will each be calculated as:

> Min (Date of first flare, Date of Week 16 visit) – Date of first dose of study medication + 1.
> All visits in the Initial Treatment Period including unscheduled visits are considered.

Participants who discontinue study treatment without experiencing a flare prior to Week 16 Visit will be censored at the date of last lesion count assessment. Participants who reach the Week 16 Visit without experiencing a flare will be censored at the date of the Week 16 Visit. Participants who experience an intercurrent event prior to experiencing a flare will be treated as experiencing a flare at the date of the intercurrent event. Participants will be censored at Baseline if there is no Baseline lesion count assessment or no Post-Baseline lesion count assessment.

Time to flare will be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.

Kaplan-Meier plots of time to flare will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 and will increase over time, representing time to achieving the response.

The median time to flare, including the 2-sided 95% confidence interval, will be calculated for each treatment. Comparisons of bimekizumab vs placebo will be analyzed using a log-rank test stratified by Hurley Stage at Baseline and Baseline antibiotic use.

8.4.6 Time to flare by Week 48

See Section 8.4.3 for the derivation of flare relative to Baseline.

Time to flare (in days) during the combined Initial and Maintenance Treatment Period will be calculated as:

> Min (Date of first flare, Date of Week 48 visit) – Date of first dose of study medication + 1.
> All visits in the up to Week 48 including unscheduled visits are considered.

Flare will be defined relative to the Baseline visit. Participants who discontinue study treatment without experiencing a flare prior to Week 48 visit will be censored at the date of last lesion count assessment. Participants who reach the Week 48 Visit without experiencing a flare will be

censored at the date of the Week 48 Visit. Participants who experience an intercurrent event prior to experiencing a flare will be treated as experiencing a flare at the date of the intercurrent event. Participants will be censored at Baseline if there is no Baseline lesion count assessment or no Post-Baseline lesion count assessment.

Time to flare will each be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.

Kaplan-Meier plots of time to flare will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 and will increase over time, representing time to achieving the response.

The median time to flare, including the 2-sided 95% confidence interval, will be calculated for each treatment.

8.4.7 International Hidradenitis Suppurativa Severity Score System (IHS4)

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

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

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

The observed IHS4 score, change and percentage change from Baseline will be summarized by treatment group and visit. Missing IHS4 scores will be imputed using the multiple imputation procedure specified in Section 4.2.2.1, where IHS4 scores will be derived based on the imputed lesion counts.

The IHS4 scores will be categorized into 3 HS categories (mild HS: ≤3, moderate HS: 4-10, severe HS: ≥11).

The number and percentage of participants in each category (mild, moderate, severe) will be presented by treatment group and visit. The denominator for the percentages will be based on the number of participants in the given treatment group and visit for which there are no missing data (OC).

Shift tables for the changes from Baseline in this scale will be presented for each post-Baseline visit by treatment group. The denominator for the percentages will be based on the number of participants in the given treatment group and visit for which there is no missing data for the change (OC).

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

The HS-Physician's Global Assessment is a validated 6-point scale that is used to measure improvement in inflammatory nodules, abscesses, and draining tunnels. The HS-Physician's Global Assessment scale is defined in Table 8–4.

Table 8–4: HS-Physician's Global Assessment 6-point scale

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

The gradings will be derived on a participant-level basis (ie, across all anatomical regions) based on the following rules:

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

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

OR

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

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

  OR

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

  OR

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

- Severe:
  - sum of number of abscesses and number of draining tunnels (fistulas/sinus tracts) ≥ 2 and ≤ 5;
  - number of inflammatory nodules > 10;
  - number of non-inflammatory nodules – no limit;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of HS scars – no limit;
- Very Severe:
  - sum of number of and number of draining tunnels (fistulas/sinus tracts) > 5;
  - number of inflammatory nodule – no limit;
  - number of non-inflammatory nodules – no limit;
  - number of non-draining tunnels (fistulas/sinus tracts) – no limit;
  - number of HS scars – no limit

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

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

The HS-Physician's Global Assessment will be listed by treatment group and participant at each visit.

8.4.9 High Sensitivity C-Reactive Protein (hs-CRP)

Concentrations of hs-CRP, changes from Baseline, and percent change from Baseline will be summarized by treatment group and visit, where percent change is calculated as

$$\text{Percent change from Baseline} = 100 \times \frac{\text{Post Baseline hs-CRP} - \text{Baseline hs-CRP}}{\text{Baseline hs-CRP}}$$

Summary statistics will include n, arithmetic mean, SD, median, Q1, Q3, minimum and maximum. For the ratio to Baseline, summary statistics will include n, geometric mean, geoCV, median, Q1, Q3, minimum and maximum.

The ratio to Baseline will be calculated as follows:

Ratio to Baseline = hs-CRP at post-Baseline / hs-CRP at Baseline visit

The ratio to Baseline will also be summarized by treatment group and visit.

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

Concentrations of hs-CRP, changes from Baseline, and ratio to Baseline will be listed.

8.4.10 Initiation of systemic antibiotic rescue therapy

See Section 3.9 for the definition of a systemic antibiotic rescue therapy.

The number of participants that use rescue antibiotic therapy will be summarized by treatment group for each period.

8.4.11 Time to initiation of systemic rescue therapy in the Initial Treatment Period

See Section 3.9 for the definition of a systemic antibiotic rescue therapy.

Time to initiation of systemic rescue therapy (in days) during the Initial Treatment Period will be calculated as:

> Min (Date of initiation of rescue therapy, Date of change in the dose/type of current antibiotic, Date of Week 16 visit) – Date of first dose of study medication + 1.

Participants who discontinue the study without initiating systemic rescue therapy prior to Week 16 visit will be censored at the date of discontinuation. Participants who reach the Week 16 Visit without initiating systemic rescue therapy will be censored at the date of the Week 16 Visit. Participants will be censored at Baseline if there is no Post-Baseline visit.

Time to initiation of systemic rescue therapy will be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.

Kaplan-Meier plots of time to initiation of systemic rescue therapy will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 and will increase over time, representing time to requiring rescue therapy.

The median time to initiation of systemic rescue therapy, including the 2-sided 95% confidence interval, will be calculated for each treatment. Comparisons of bimekizumab vs placebo will be analyzed using a log-rank test stratified by Hurley Stage at Baseline and Baseline antibiotic use.

8.4.12 Time to an intercurrent event in the Initial Treatment Period

See Section 3.9 for the definition of an intercurrent event.

Time to an intercurrent event (in days) during the Initial Treatment Period will be calculated as:

> Min (Date of intercurrent event, Date of Week 16 visit) – Date of first dose of study medication + 1.

Participants who discontinue the study without experiencing an intercurrent event prior to Week 16 visit will be censored at the date of discontinuation. That includes participants who discontinue from the study for reasons other than Adverse Event and Lack of Efficacy. Participants who reach the Week 16 Visit without experiencing an intercurrent event will be censored at the date of the Week 16 Visit. Participants will be censored at Baseline if there is no Post-Baseline visit.

Time to an intercurrent event will be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.

Kaplan-Meier plots of time to initiation of systemic rescue therapy will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 and will increase over time, representing time to intercurrent event.

The median time to an intercurrent event, including the 2-sided 95% confidence interval, will be calculated for each treatment. Comparisons of bimekizumab vs placebo will be analyzed using a log-rank test stratified by Hurley Stage at Baseline and Baseline antibiotic use.

8.4.13 Hidradenitis Suppurativa Symptom Daily Diary (HSSDD)

See Section 8.3.3 for details on HSSDD Baseline and weekly average definitions and derivations.

Percent change from Baseline in HSSDD responses for worst and average skin pain score is defined as

$$\text{Percent change from Baseline } = 100 \times \frac{\text{Post Baseline HSSDD score} - \text{Baseline HSSDD score}}{\text{Baseline HSSDD score}}$$

Change from Baseline in each HSSDD item (worst skin pain, average skin pain, smell or odor, itch at its worst, and amount of drainage or oozing) score will be summarized using descriptive statistics by treatment group and visit, based on weekly averages. The table will display descriptive statistics for the Baseline, followed by descriptive statistics for the change from Baseline for all visits. Percentage change will be summarized for the worst and average skin pain items.

Additionally, change from Baseline in each HSSDD item will be evaluated by treatment group at Week 16 via continuous empirical cumulative distribution function (eCDF) plots showing the absolute change from Baseline on the horizontal axis and the cumulative percent of participants experiencing that change on the vertical axis.

Change from Baseline in Worst Skin Pain score and Worst Itch score will additionally be summarized by visit and by analgesic and antihistamine use status (Section 6.4.2), respectively.

HSSDD response based on clinically meaningful change for the worst skin pain item is defined as at least a 3-point reduction from Baseline in HSSDD among study participants with a score of ≥3 at Baseline, based on weekly averages.

The number and percentage of responders based on clinically meaningful change for the worst skin pain item will be summarized by treatment group and visit.

The number and percentage of participants who were responders based on clinically meaningful change at any timepoint in the Initial Treatment Period will be summarized by treatment group for the worst skin pain item.

HSSDD response for the worst skin pain and average skin pain items is defined as at least a 30% reduction and at least a 1-point reduction from Baseline among study participants with a score of ≥3 at Baseline. The number and percentage of responders for each item will be summarized by treatment group and visit.

The number and percentage of participants who were responders (based on the 30% improvement and 1 point improvement definition) at any timepoint in the Initial Treatment Period will be summarized by treatment group for the worst skin pain and average skin pain items.

The number and percentage of participants that complete the HSSDD will be calculated for each visit by treatment group. A participant will be counted as completing the HSSDD at a visit if the minimum number of daily entries is present to calculate the weekly average (see Section 8.3.3). The percentage will be based on the number of participants in the RS. A participant will be considered a completer at a visit if the weekly average can be calculated for that visit.

8.4.14 Hidradenitis Suppurativa Symptom Questionnaire (HSSQ)

The 4 items on the HS Symptom Questionnaire (HSSQ) assesses participants' perception of the core symptoms of HS experienced in the past 7 days - skin pain, smell or odor, drainage or oozing from HS lesions, and itch on an 11-point NRS.

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

Summary statistics of the actual values and change and percentage change from Baseline values will be used to summarize each HSSQ item for each visit by treatment group. The table will display descriptive statistics for the Baseline, followed by descriptive statistics for the change from Baseline for all visits.

Additionally, change from Baseline in each HSSQ item will be evaluated by treatment group at Week 16 and at Week 48 via continuous eCDF plots showing the absolute change from Baseline on the horizontal axis and the cumulative percent of participants experiencing that change on the vertical axis.

Missing data for the continuous change from Baseline will be handled by using MI via the MCMC and monotone regression method specified in Section 4.2.1.3.

HSSQ response for skin pain item is defined as at least a 30% reduction and at least a 1-point reduction from Baseline in HS Skin Pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline.

The number and percentage of responders for skin pain item will be summarized by treatment group and visit based on the MS.

The number and percentage of participants who were responders at any timepoint in the Maintenance Treatment Period will be summarized by treatment group for the skin pain score based on the MS.

Change from Baseline in skin pain score and itch score will additionally be summarized by visit and by analgesic and antihistamine use status (Section 6.4.2), respectively.

The number and percentage of participants that complete the HSSQ will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the MS. The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if each of the items are completed at that visit.

8.4.15 DLQI

See Section 8.3.2 for the derivation of DLQI total score.

A DLQI total score of 0 or 1 indicates no impact of the skin disease on health-related quality of life and will be summarized.

A participant is considered to have achieved the minimally clinical important difference (MCID) if their individual improvement (ie, decrease) from Baseline in total score is >=4. A 4-point improvement in the DLQI total score (DLQI response) has been reported to be meaningful for the participant (within-participant MCID). The summary of MCID will be restricted to participants with a DLQI total score of at least 4 at Baseline to ensure that it is possible for the participant to achieve the MCID.

The DLQI related efficacy variables are defined as follows:

- Change from Baseline in DLQI total score is defined as Post-Baseline DLQI total score minus Baseline DLQI total score.
- Percent of study participants achieving a DLQI total score of 0 or 1 is defined as the number of study participants with DLQI total score of 0 or 1 divided by the number of study participants in RS.
- Percent of study participants achieving a MCID in DLQI total score is defined as the number of study participants with improvement from Baseline in total score of 4 or more divided by the number of study participants in RS that have a Baseline DLQI total score of at least 4.

Missing data for the DLQI total score will be handled by using MI via the MCMC and monotone regression method specified in Section 4.2.1.3.

Change from Baseline in DLQI total score will be summarized using descriptive statistics by treatment group and visit. The table will display descriptive statistics for the Baseline, followed by descriptive statistics for the change from Baseline for all visits.

Frequency tables will be produced to show the number and percentage of DLQI responders for MCID for each visit by treatment groups.

The number and percentage of participants achieving a DLQI total score of 0 or 1 at each visit will be summarized descriptively using counts and percentages by treatment group and visit.

The number and percentage of participants that complete the DLQI total score will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS. The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if the total score is calculated at that visit.

A by-participant listing of the DLQI questionnaire, DLQI total score, change from Baseline and DLQI response for MCID and 0 or 1 data will be provided by treatment group.

8.4.16 Hidradenitis Suppurativa Quality of Life (HiSQOL)

The HiSQOL includes 17 items assessed using a 7-day recall period, grouped in 3 subscales: symptoms, psychosocial, activities and adaptation.

The assessment of each item of the HiSQOL is as follows:

Table 8–5: Hidradenitis Suppurative Quality of Life

HiSQOL Scoring	
Response	Score
Unable to do, due to my HS	4
Extremely	4
Very Much	3
Moderately	2
Slightly	1
Not at all	0
I normally do not do this, HS did not influence	0
I am not sexually active	0
I do not work or study	0
Unanswered	0

The HiSQOL total score is calculated by adding the score of each question. The maximum score is 68, and the minimum score is 0.

Subscale scores will be summarized for each of the 3 subscales. The maximum scores for the subscales are 16 (symptoms), 20 (psychosocial), and 32 (activities and adaptations), and the minimum score is 0 for all subscales.

For all scores, the higher the score, the more quality of life is impaired.

Summary statistics of the actual values and change from Baseline values will be used to summarize HiSQOL domain and total scores for each visit by treatment group. The table will display descriptive statistics for the Baseline, followed by descriptive statistics for the change from Baseline for all visits.

Additionally, change from Baseline in each HiSQOL subscale will be evaluated by treatment group at Week 16 and at Week 48 via continuous eCDF plots showing the absolute change from Baseline on the horizontal axis and the cumulative percent of participants experiencing that change on the vertical axis.

Missing data for the continuous change from Baseline will be handled by using MI via the MCMC and monotone regression method specified in Section 4.2.1.3. The imputed HiSQOL total score will be derived based on the imputed subscales.

The number and percentage of participants that complete the HiSQOL will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS (or MS, as appropriate). The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if the total score is calculated at that visit.

A by-participant listing of the HiSQOL questionnaire, HiSQOL responses, domain and total scores and change from Baseline will be provided.

8.4.17 Patient Global Impression of HS Severity (PGI-S-HS)

The PGI-S-HS is a single item to assess study participants perceptions of the overall severity of HS over the past 7 days (none, mild, moderate, severe, very severe).

The number and percentage of participants with each response will be summarized for each visit by treatment group based on OC data.

The number and percentage of participants that complete the PGI-S-HS will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS. The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if the response is nonmissing at that visit.

8.4.18 Patient Global Impression of Change in HS Severity (PGI-C-HS)

The PGI-C-HS is a single item to assess study participants perception of the change in HS since they started taking the study medication (much better, a little better, no change, a little worse, much worse).

The number and percentage of participants with each response will be summarized for each visit by treatment group based on OC data.

The number and percentage of participants that complete the PGI-C-HS will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS (or MS, as appropriate). The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if the response is nonmissing at that visit.

8.4.19 Patient Global Impression of Severity of Skin Pain (PGI-S-SP)

The PGI-S-SP is a single item to assess study participants' perceptions of the severity of their skin pain from their HS lesions, over the past 7 days (none, mild, moderate, severe, very severe).

The number and percentage of participants with each response will be summarized for each visit by treatment group based on OC data.

The number and percentage of participants that complete the PGI-S-SP will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS (or MS, as appropriate). The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if the response is nonmissing at that visit.

8.4.20 Patient Global Impression of Change in Severity of Skin Pain (PGI-C-SP)

The PGI-C-SP is a single item to assess study participants' perceptions of change in their skin pain from their HS lesions, since they started taking the study medication (much better, a little better, no change, a little worse, much worse).

The number and percentage of participants with each response will be summarized for each visit by treatment group based on OC data.

The number and percentage of participants that complete the PGI-C-SP will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS. The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if the response is nonmissing at that visit.

8.4.21 Euro-Quality of Life 5-Dimensions, 3 levels (EQ-5D-3L)

The EQ-5D-3L comprises 5 questions on mobility, self-care, pain, usual activities, and psychological status with 3 possible answers for each item (1=no problem, 2=moderate problem, 3=severe problem). In addition, there is a VAS to indicate the general health status with 100 indicating the best health status.

Change from Baseline in EQ-5D-3L VAS scores is defined as Post-Baseline EQ-5D-3L VAS score minus Baseline EQ-5D-3L VAS score.

Responses to EQ-5D-3L will be summarized based on OC only as primary analysis. No imputation is applied to responses to EQ-5D-3L but is applied to EQ-5D-3L VAS scores.

Changes from Baseline in EQ-5D-3L VAS will be summarized using descriptive statistics by treatment group and visit. The table will display descriptive statistics for the Baseline, followed by descriptive statistics for the change from Baseline for all visits.

The number and percentage of participants with each response in the EQ-5D-3L will be summarized for each visit by treatment group based on OC data.

The number and percentage of participants that complete the EQ-5D-3L will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS. The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if each of the domains and VAS are completed at that visit.

8.4.22 Work Productivity and Activity Impairment Questionnaire–Specific Health Problem (WPAI-SHP) v2.0 adapted to HS scores

The WPAI-SHP V2.0 is a patient-reported questionnaire that assesses study participant's employment status, work absenteeism, work impairment while working (presenteeism), overall work, and daily activity impairment attributable to a specific health problem. It has been used in several clinical studies of biologic therapy in participants with plaque PSO.

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

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

Questions:

- 1 = currently employed
- 2 = hours missed due to specified problem
- 3 = hours missed other reasons
- 4 = hours actually worked

- 5 = degree problem affected productivity while working
- 6 = degree problem affected regular activities

Scores:

- Percent work time missed due to problem: [Q2 hours/(Q2 hours +Q4 hours)]*100
- Percent impairment while working due to problem: [Q5 score/10]*100
- Percent overall work impairment due to problem: [Q2 hours/(Q2 hours +Q4 hours)+[(1-(Q2 hours/(Q2 hours +Q4 hours))x(Q5 score/10)]]*100
- Percent activity impairment due to problem: [Q6 score/10]*100

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

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

Summary statistics of the actual values and change from Baseline values will be used to summarize WPAI-SHP for each visit by treatment group. The table will display descriptive statistics for the Baseline, followed by descriptive statistics for the change from Baseline for all visits.

Missing data for the continuous change from Baseline will be handled by using MI via the MCMC and monotone regression method specified in Section 4.2.1.3.

The number and percentage of participants that complete the WPAI-SHP will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS. The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if the percentages in each dimension are calculated at that visit.

A by-participant listing of the WPAI-SHP questionnaire, WPAI-SHP domains and change from Baseline will be provided.

8.4.23 Treatment Satisfaction Questionnaire – Medication-9

The TSQM-9 is an abbreviated 9-item version of the TSQM, excluding the side effects of medication domain. The domains included in the TSQM-9 include effectiveness (3 items), convenience (3 items) and global satisfaction (3 items). The TSQM-9 domain scores range from 0 to 100 with higher scores representing higher satisfaction.

The scores for each measure are as follows:

- Global Satisfaction:
  - If no items are missing: ([Sum(Item 7 to Item 9) – 3]/14)*100
  - If either Item 7 or 8 is missing: ([Sum(the two completed items)) – 2]/10)*100
  - If Item 9 is missing: ([Sum(Item 7 and Item 8))-2]/8)*100
- Effectiveness
  - If no items are missing: ([Item 1 + Item 2 + Item 3) – 3]/18)*100

  - If one item is missing: ([(Sum(the two completed items) – 2]/12)*100

- Convenience
  - If no items are missing: ([Sum(Item 4 to Item 6) – 3]/18)*100
  - If one item is missing: ([Sum(the two completed items) – 2]/12*100

Frequency tables will be produced to summarize answers provided to each of the 9 items of the TSQM-9 at Weeks 16 and 48 by treatment group. Responses to TSQM-9 will be summarized based on OC. No imputation will be applied.

The number and percentage of participants that complete the TSQM-9 will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS. The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit. A participant will be considered a completer at a visit if each of the domains are completed at that visit.

A by-participant listing of TSQM-9 will be provided.

8.4.24 Lesion intervention

Investigators will have the option to perform interventions in the event an acutely painful lesion occurs that requires immediate intervention.

The following rules will be used to assign a lesion intervention to a study period:

- **Initial Treatment Period:** a lesion intervention will be assigned to the Initial Treatment Period if it has been performed between the first administration of IMP on Day 1 up to and including Week 16.
- **Maintenance Treatment Period:** a lesion intervention will be assigned to the Maintenance Treatment Period if it has been performed between Week 16 through the Week 48 visit.

Methods for dealing with partial dates are specified in Section 4.2.4.

A listing of participants who receive any lesion intervention will be provided.

The number and percentage of participants who receive at least 1 lesion intervention will be summarized by treatment group for the Initial Treatment Period, Maintenance Treatment Period, and the combined Initial and Maintenance Treatment Period.

The number and percentage of participants with 2, 3, and 4 or more lesion interventions will also be summarized by treatment group for the Initial Treatment Period, Maintenance Treatment Period, and the combined Initial and Maintenance Treatment Period.

The number and percentage of participants with 2 or more lesion interventions performed on the same lesion will be summarized by treatment group for the Initial Treatment Period, Maintenance Treatment Period, and the combined Initial and Maintenance Treatment Period.

8.5 Additional statistical analyses of other efficacy endpoints

For selected other efficacy variables, it is of interest to perform statistical tests and to calculate inferential statistics. As these tests are not part of the multiplicity-controlled procedure, the associated p-values are considered nominal and are not controlled for multiplicity.

For binary variables, the analysis will follow what was specified for the primary analysis of the primary endpoint and the corresponding p-value reported. Missing values will be imputed as for the primary analysis. For continuous variables, the MI – MCMC / Monotone Regression approach used for the secondary continuous endpoint will be applied for the imputation model. The analysis model will be as for the corresponding secondary continuous endpoint analysis, unless otherwise indicated.

Below is a list of variables for which these nominal p-values will be calculated. The results of these inferential tests will be presented in a single table summarizing the testing performed outside of the multiplicity-controlled testing procedure.

All tests will be for both 320mg Q2W vs Placebo and 320mg Q4W vs Placebo (tested separately) and will be performed for the Week 12 and Week 16 visits only. If the Week 16 visit test is already part of the controlled testing procedure in the primary or secondary analyses, only Week 12 is indicated here.

- $HiSCR_{50}$ at Week 12
- $HiSCR_{75}$ at Week 12
- $HiSCR_{90}$
- $HiSCR_{100}$
- Flare by Week 12
- Flare by Week 16
- Time to flare by Week 12 (based on time-to-event analysis per Section 8.4.5 and adjusted appropriately)
- IHS4 change from Baseline
- IHS4 percentage change from Baseline
- HS Physician's Global Assessment: rate of participants who are Clear or Mild
- DLQI total score change from Baseline at Week 12
- Worst Skin Pain per HSSDD change from Baseline at Week 12
- Skin Pain response per HSSDD at Week 12

9 PHARMACOKINETICS AND PHARMACODYNAMICS

9.1 Pharmacokinetics

Pharmacokinetic variables will be analyzed for all participants in the PK-PPS. Bimekizumab plasma concentrations will be summarized for each treatment at each scheduled visit.

PK summaries will be based on observed values. No imputation will be used. However, if plasma concentration measurements are below the level of quantification (BLQ), then for calculation of the derived statistics the result will be set to ½ of the lower level of quantification (LLOQ). Descriptive statistics including geometric mean, geometric coefficient of variation, and geometric mean 95% CI if applicable will be calculated if at least 2/3 of the values of interest are above the LLOQ. If this is not the case, only median, minimum, and maximum will be presented.

Geometric mean plasma concentration will be plotted by treatment group, and by cumulative antibody status for participants randomized to bimekizumab on linear and log linear scale. In addition, $HiSCR_{50}$ response (0=not achieved, 1=achieved) assessed at each PK visit (excluding Week 1) will be plotted against the visit's bimekizumab plasma concentration, separated by treatment group. Spaghetti plots of bimekizumab plasma concentrations by week from bimekizumab first dosing separated by treatment group and antibody status will be presented for participants with and without $HiSCR_{50}$ response at Week 16.

If the dosing for a visit is +/- 7 days out of window, then the plasma concentration from that visit and all subsequent visits will be excluded from the PK summary. In addition, if the PK sampling date is >1 day after the dosing date, then the plasma concentration from that visit will be excluded from the PK summary.

All PK concentrations collected will be listed irrespective of the dosing or sampling occurring out of window.

9.2 Pharmacodynamics

Not applicable.

9.3 Immunogenicity

9.3.1 Autoantibodies

Not applicable.

9.3.2 Anti-bimekizumab antibodies

Anti-bimekizumab antibodies (ADAb) will be measured using a 3-tiered assay approach: screening assay, confirmatory assay, and titration assay. Samples confirmed as positive within the confirmatory assay will be further evaluated for the presence of neutralizing anti-bimekizumab antibodies specific to IL-17AA, IL-17FF or both. Samples will be taken at Baseline, then at study Weeks 4, 8, 12, 16, 20, 24, 36 and 48, and at PEOT and SFU timepoints.

ADAb samples are not analyzed when study participants are on a treatment other than bimekizumab. For study participants who switch from placebo to bimekizumab, samples are analyzed starting at the visit when the switch to bimekizumab occurs (Week 16). The sample at Week 16 will act as the Baseline for that treatment group.

The screening cut point will be used to determine the status of anti-bimekizumab antibodies in the test sample as Positive Screen (PS) or Negative Screen (NS). For samples presenting anti-bimekizumab antibody levels that are PS, a further confirmatory assay will be performed, and the result of which will be reported as either Positive Immunodepletion (PI) or Negative Immunodepletion (NI).

ADAb status for each sample will be derived as follows:

- Sample values that are either NS, or PS and NI and where the bimekizumab concentration is less than the validated ADAb assay drug tolerance limit will be defined as anti-bimekizumab antibody negative.
- Sample values that are either NS, or PS and NI and where the bimekizumab concentration exceeds the validated ADAb assay drug tolerance limit will be defined as inconclusive.

- Sample values that are PS and PI will be defined as ADAb positive (regardless of availability of a titer value)
- Missing or non-evaluable samples will be defined as missing

Positive immunodepletion samples will be titrated, and the ADAb titer (reciprocal dilution factor including minimum required dilution) will be reported. Subsequently, PI samples will also be subject to a neutralizing assay to evaluate the potential of ADAb to neutralize the target binding of bimekizumab (IL-17AA or IL-17FF or both) in vitro.

Cumulative ADAb status will be derived as follows:

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

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

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

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

Overall ADAb status will be derived as follows:

A study participant will be classified as:

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

ADAb categories will be derived as follows:

- **Pre ADAb negative – treatment-emergent ADAb negative (Category 1)**: Includes study participants who are anti-bimekizumab antibody negative at Baseline and anti-bimekizumab antibody negative at all sampling points during the period of interest (one post-Baseline missing/inconclusive sample is allowed for subjects with pre- anti-bimekizumab antibody negative sample). This group also includes study participants who have a missing or inconclusive sample (either missing or inconclusive or insufficient volume) at Baseline (ie, pre-treatment) with all post-Baseline samples as ADAb negative.
- **Pre ADAb negative – treatment-emergent ADAb positive (Category 2)**: Includes study participants who are ADAb negative at Baseline and ADAb positive at any sampling points post-Baseline during the period of interest. This group also includes study participants who

have a missing sample (either missing or insufficient volume) at Baseline (ie, pre-treatment) with 1 or more post-Baseline samples as ADAb positive.

- **Pre ADAb positive – treatment-emergent reduced ADAb (Category 3)**: Includes study participants who are ADAb positive at Baseline, and ADAb negative at all sampling points post-Baseline during the period of interest.
- **Pre ADAb positive – treatment-emergent unaffected ADAb positive (Category 4)**: Includes study participants who are ADAb positive at Baseline and are ADAb positive at any sampling point post-Baseline during the period of interest with titer values of the same magnitude as Baseline (ie, less than a predefined fold difference from the Baseline titer).
  - For this analysis, this is set at an increase of less than the validated Minimum Significant Ratio (MSR) of 2.07-fold from Baseline.
- **Pre ADAb positive – treatment-emergent ADAb boosted positive (Category 5)**: Includes study participants who ADAb positive at Baseline and are ADAb positive at any sampling point post-Baseline during the period of interest with increased titer values compared to Baseline (equal to or greater than a predefined fold difference increase from Baseline titer which will be defined within the validation of the assay).
  - For this analysis, this is set at an increase equal to or greater than the validated MSR of 2.07-fold from Baseline.
  - Note: for any study participant who is ADAb positive at Baseline and ADAb positive at a post-Baseline time point during the period of interest, but for whom titers are not available to determine treatment unaffected or treatment boosted status, the study participant will be considered as treatment boosted, assuming no other samples are available.
- **ADAb Inconclusive (Category 6)**: Includes study participants who have an ADAb positive Baseline (pre-treatment) sample and some post-Baseline samples during the period of interest are missing or inconclusive, while other post-Baseline samples are ADAb negative.
- **Total treatment-emergent ADAb positivity** (**Category 7** [Categories 2 and 5 combined]): Includes study participants who are pre ADAb negative – treatment-emergent ADAb positive (Category 2) and pre ADAb positive – treatment boosted ADAb positive (Category 5).
- **Total prevalence of pre- ADAb positivity (Category 8** [Categories 3, 4, 5 and 6 combined]): Study participants that are tested ADAb positive at Baseline.
- **Missing:** Includes study participants who are ADAb negative, missing, or inconclusive at Baseline with some post-Baseline samples as missing or inconclusive, and other samples as ADAb negative.

For purposes of efficacy subgroup analyses based on anti-bimekizumab antibody status, the following categories can also be used:

- ADAb positive – This is defined as study participants who are anti-bimekizumab antibody positive on at least 2 time points while on treatment (ie, excluding Baseline, excluding SFU).
- ADAb negative – Study participants for who either:

  - All samples (including Baseline) are ADAb negative and there are no missing or inconclusive samples
  - Only 1 sample is ADAb positive and all other samples (including Baseline) are ADAb negative or missing or inconclusive
  - Only 1 sample is missing or inconclusive and the remaining ADAb samples are negative.
- ADAb missing - Defined as study participants who do not fulfil the criteria for one of the 2 groups listed above.

The rationale for requiring at least 2 time points in which ADAb levels are above the specified cut point is to exclude those study participants who have only one occurrence of ADAb levels during the course of treatment. Including such study participants would increase the number of ADAb positive study participants with potentially no impact on efficacy.

In the case that a sample is collected 1 or more days following the scheduled visit date in which the drug was administered, the ADAb results for that sample will be associated with the scheduled visit and summarized accordingly. Such samples will also be considered when ADAb results are summarized over a given study period.

Analysis

Immunogenicity will be assessed through summary tables and figures, and listing of individual results by participants. All analyses will be run on the AMS, unless specified otherwise.

- Summary of ADAb status overall and by each visit separated by treatment group
- Summary of the time-point of the first occurrence of ADAb positivity during the treatment period by treatment group. A plot of the titer by time to first ADAb positivity will be prepared.
- All individual participant-level ADAb results will be listed.
- The number and percentage of participants in each of the 8 ADAb categories during the treatment period by treatment group.
- The prevalence of immunogenicity, separated by treatment group, and defined subcategory, will be reported by visit, defined as (cumulative) proportion of participants having confirmed positive ADAb samples at any visit up to and including that visit. Missing samples will not be included in the denominator.
- The time to achieving treatment-emergent ADAb positivity, separated by treatment group and defined subcategory, will be analyzed based on Kaplan-Meier methods. This will be shown only for Categories 2 and 8 above. Participants will be considered to have an event at the time point at which treatment emergent ADAb positive is first achieved (taking the MSR into consideration for sub-category 5). Participants classified as treatment-emergent ADAb negative will be censored at the time of the last available ADAb result.
- A summary of $HiSCR_{50}$ responders at Week 16, separated by treatment group, as a function of ADAb titer will be presented graphically.
- Individual plots of plasma bimekizumab concentrations/ ADAb titer both plotted on the Y-axes by visit (x-axis) for the full treatment period (excluding SFU for interim analyses and

including SFU for final analyses) will be presented for participants with and without $HiSCR_{50}$ response at Week 16.

- Spaghetti plots of ADAb titer (y-axis) by visit (x-axis), separated by treatment group for all ADAb positive participants, including Baseline positive participants.
- Box plots of ADAb titer (logscale) by time to first ADAb positivity by treatment group.

The groups for defining ADAb status for safety subgroup analyses are as follows:

- AEs starting before first ADAb positive result
- AEs starting on or after first ADAb positive result
- AEs for participants who were always ADAb negative

This is further explained in Section 10.2.2.

10 SAFETY ANALYSES

All analysis of safety variables will be performed using the SS, MS, and AMS.

The AMS will be used for summaries of safety that include data from the Initial Treatment Period and Maintenance Treatment Period.

Summaries of safety will be presented for the Initial Treatment Period, Maintenance Treatment Period, and combined Initial and Maintenance Treatment Period unless specified otherwise.

10.1 Extent of exposure

Summaries for exposure will be provided. This consists of a descriptive summary of study medication duration in days. In addition, total study medication duration and time at risk will be summarized in years by treatment group and treatment period (ie, the Initial Treatment Period, the Maintenance Treatment Period, and the Initial and Maintenance Treatment Period). Summary of exposure in Maintenance Treatment Period will be on MS. The cumulative study medication duration will be summarized for study participants exposed for given durations of time. For the cumulative duration through Week 48 the following categories for duration will be used:

- >0 weeks
- ≥4 weeks
- ≥8 weeks
- ≥12 weeks
- ≥16 weeks
- ≥20 weeks
- ≥24 weeks
- ≥28 weeks
- ≥32 weeks
- ≥40 weeks
- ≥48 weeks

Definitions for study medication duration and time at risk in days are provided below for each period. Time at risk will be summarized in years. Time at risk in years is calculated by dividing the time at risk in days by 365.25.

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

10.1.1 Exposure during the Initial Treatment Period

Definitions for study medication duration (days) and time at risk (days) during the Initial Treatment Period are provided as follows:

10.1.1.1 Study medication duration (days)

Definitions for study medication duration (days) are provided as follows:

- Date of last dose in the Initial Treatment Period – Date of first dose in the Initial Period + 14 days.

Note: The use of 14 days assumes a Q2W dosing interval (bimekizumab 320mg Q2W and placebo). For participants randomized to bimekizumab 320mg Q4W, this will be adjusted based on the dosing interval (eg, use Date of last bimekizumab dose in the Initial Treatment Period – Date of first dose in the Initial Period + 28 days).

Note: If date of last dose in the Initial Treatment Period + 14 days (or + 28 days in the case of Q4W dosing) extends to a date beyond the date of first dose in the Maintenance Treatment Period, then this calculation reverts to:

  - Date of first dose in the Maintenance Treatment Period – Date of first dose in the Initial Treatment Period + 1.

- For participants who die during the Initial Period, if date of last dose in the Initial Period + 14 days (or date of last bimekizumab dose in the Initial Treatment Period + 28 days in the case of Q4W dosing) extends to a date beyond the date of death, then this calculation reverts to:

  - Date of death – Date of first dose in the Initial Period + 1.

10.1.1.2 Time at risk (days)

Definitions for time at risk (days) are provided as follows:

- For participants who complete the Week 16 visit and continue to the Maintenance Treatment Period:

  - Date of first dose in the Maintenance Treatment Period – Date of first dose in the Initial Period + 1.

- For participants who discontinue on or prior to the final visit of the Initial Period, use the minimum of the following:

  - Date of last dose in the Initial Treatment Period – Date of first dose in the Initial Treatment Period + 141

  - The total number of days in the Initial Treatment Period (112 days). For AEs that emerged after 112 days but still within the 140 days window, those AEs would be

classified as TEAE, but will be excluded from the output based on the Initial Treatment Period. However, these AEs will be included in the AE summaries for Maintenance Treatment Period.

  - Date of last clinical contact – Date of first dose in the Initial Treatment Period + 1.

- For participants who die prior to the final visit of the Initial Treatment Period: Date of death – date of first dose in the Initial Period +1.

10.1.2 Exposure during the Maintenance Treatment Period

Definitions for study medication duration (days) and time at risk (days) during the Maintenance Treatment Period are provided as follows:

10.1.2.1 Study medication duration (days)

Definitions for study medication duration (days) are provided as follows:

- Date of last dose in the Maintenance Treatment Period – date of first dose in the Maintenance Treatment Period + 14 days.

The use of 14 days assumes a Q2W dosing interval. For participants randomized to bimekizumab 320mg Q4W in the Maintenance Treatment Period, this will be adjusted based on the dosing interval (eg, use Date of last bimekizumab dose in the Maintenance Treatment Period – date of first dose in the Maintenance Treatment Period + 28 days).

Note: If date of last dose in the Maintenance Treatment Period + 14 days (or + 28 days in the case of Q4W dosing) extends to a date beyond the final visit date of the Maintenance Treatment Period (not including SFU), then this calculation reverts to:

  - Final visit date of the Maintenance Treatment Period (not including SFU) – date of first dose in the Maintenance Treatment Period + 1.

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

  - Date of death – Date of first dose in the Maintenance Treatment Period + 1.

10.1.2.2 Time at risk (days)

Definitions for time at risk (days) are provided as follows:

- For participants who complete the Maintenance Treatment Period as planned and continue into an extension study (and, therefore, do not have the SFU visit in the feeder study):

  - Date of last visit of the Maintenance Treatment Period – Date of first dose in the Maintenance Treatment Period + 1.

- For participants who die prior to the final visit of the Maintenance Treatment Period:

  - Date of death – Date of first dose in the Maintenance Period + 1.

- For all other participants, use the minimum of the following:

    - Date of last dose in the Maintenance Treatment Period – Date of first dose in the Maintenance Treatment Period + 141 days.

- Date of last clinical contact – Date of first dose in the Maintenance Treatment Period + 1.

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

10.1.3 Exposure during the Initial and Maintenance Treatment Period

Definitions for study medication duration (days) and time at risk (days) during the Initial and the Maintenance Treatment Period are provided as follows:

10.1.3.1 Study medication duration (days)

Definitions for study medication duration (days) are provided as follows:

- Sum of study medication durations from the individual Initial and Maintenance Treatment Period.

Note: The algorithms for calculating these durations are specified in Section 10.1.1.1 and Section 10.1.2.1.

Note: If date of last dose in the Initial Treatment Period + 14 days (or + 28 days in the case of Q4W dosing) extends to a date beyond the date of first dose in the Maintenance Treatment Period, then this calculation reverts to:

- Sum of study medication durations from the individual Initial and Maintenance Treatment Periods - 1.

10.1.3.2 Time at risk (days)

Definitions for time at risk (days) are provided as follows:

- For participants who complete the Maintenance Treatment Period as planned and continue into an extension study (and, therefore, do not have the SFU visit in the feeder study):
  - Final visit date – Date of first dose + 1.
- For participants who die prior to the final visit:
  - Date of death – Date of first dose in the + 1.
- For all other participants, use the minimum of the following:
  - Date of last dose – Date of first dose + 141 days.
  - Date of last clinical contact – Date of first dose + 1.

  Note: This group could include participants who discontinue early, participants who complete the Maintenance Treatment Period as scheduled but choose not to continue into an extension study, or participants who are ongoing in the SFU period at the time of the data snapshot (in the case of the interim analysis).

10.2 Adverse events

An AE is any untoward medical occurrence in a patient or clinical study participant, temporally associated with the use of IMP, whether or not considered related to the IMP. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease (new or exacerbated) temporally associated with the use of IMP.

10.2.1 Data considerations

Treatment-emergent AEs are defined as those AEs that have a start date on or following the first dose of study treatment through the final dose of study treatment + 140 days (covering the 20-week SFU period). 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.

An AE will be assigned to the Initial Treatment Period if it started between the first administration of IMP on Day 1 up to Week 16. An AE will be assigned to the Maintenance Treatment Period if it started between the Week 16 study drug administration and Week 48.

If an AE occurs on the date of a treatment switch, the event is attributed to the original treatment. The only exception to this is if the AE fulfills any of the criteria specified below:

- Events that fulfill the anaphylaxis criteria for acute events (refer to Appendix A in Section 12.1)
- Hypersensitivity events identified by the SMQ "Hypersensitivity (SMQ)" (see Section 12.1 Appendix A)
- Events with an high level term (HLT) of "Administration site reactions NEC"
- Events with an HLT of "Injection site reactions"

The rules for imputing partial start or stop dates are outlined in Section 4.2.4.

Any TEAEs that occur during the SFU Period will be attributed to the period in which the participant was before initiating the SFU Period.

Duration of AEs will not be calculated if there is missing stop date information.

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. Note that if the seriousness of an adverse event is unknown, every attempt should be made to resolve this prior to a snapshot for an interim analysis or database lock; in the exceptional case that the seriousness of an adverse event is still missing then no imputation should be applied for this characteristic.

Adverse events will be presented as "number of participants (percentage of participants) [number of events]". In this style of output, "[number of events]" will include all cases of an AE including repeat occurrences in individual participants, while "number of participants" will count each participant only once.

Subject time at risk represents the time a participant is at risk for having an AE. The definitions for subject time at risk (in days) are outlined in Section 10.1. These definitions will be used for exposure-adjusted AE summaries.

Selected AE summaries will include the exposure-adjusted incident rate (EAIR) with associated 95% CI and the exposure adjusted event rate (EAER).

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

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

Where $T_{Exp(i)}$ is the exposure time and N is the number of participants at risk.

If a participant has multiple events, the time of exposure is calculated to the first occurrence of the AE being considered. If a participant 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):

$$LCL = \chi^2_{2n,\alpha/2} / 2$$

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

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

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

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

where N_{AE} is the total number of AEs, T_{Risk} is the time at risk for each participant, and N is the total number of participants at risk.

No confidence interval will be computed for EAER.

Selected summaries, as specified in Section 10.2.2, will include the risk difference between bimekizumab and placebo. The risk difference is calculated as:

$$RD = IP_{BKZ} - IP_{PBO}$$

where IP_{BKZ} is the incidence proportion for the bimekizumab-treated group and IP_{PBO} is the incidence proportion for the placebo group. Note that incidence proportion simply refers to the percentage of participants within the specified treatment group that experienced a given adverse event.

The standard error for the risk difference is calculated as follows:

$$SE_{RD} = \sqrt{\left(IP_{BKZ} \times \left(\frac{1 - IP_{BKZ}}{n_{BKZ}}\right)\right) + \left(IP_{PBO} \times \left(\frac{1 - IP_{PBO}}{n_{PBO}}\right)\right)}$$

where n_{BKZ} is the number of participants in the bimekizumab-treated group and n_{PBO} is the number of participants in the placebo group.

The corresponding confidence interval for the risk difference is as follows:

$$CI_{RD} = RD \pm Z_{1-\alpha/2} \times (SE_{RD})$$

where $Z_{1-\alpha/2}$ is the Z statistic for the corresponding level of alpha. For the risk difference confidence intervals calculated in this SAP, 1.96 will be used (corresponding to a two-sided alpha of 0.05 and 95% confidence interval). The risk difference and corresponding CI will be displayed as percentage.

10.2.1.1 COVID-19 related considerations

To assess the impact of COVID-19 mass vaccination on TEAEs, a sensitivity analysis will present all TEAEs excluding TEAEs assessed as exclusively related to COVID-19 vaccine by the investigator. TEAEs recorded as related to both study medication and COVID-19 vaccination should not be excluded. A complementary table and listing of TEAEs related to COVID-19 vaccine will be presented.

Another sensitivity analysis will present all TEAEs excluding TEAEs with start date on or up to 5 days after date of COVID-19 vaccine. Note that study participants may receive more than one administration of COVID-19 vaccine. A complementary table and listing of TEAEs with start date on or up to 5 days after date of COVID-19 vaccine will also be presented.

10.2.2 AE summaries

The following summaries will be provided by treatment group for the Initial Treatment Period, and the Initial and Maintenance Treatment Period combined based on the SS, and AMS respectively. In addition, all summaries of TEAEs based on "100 subject years" will include EAIR (with 95% confidence interval) and EAER. For AEs that emerged after 112 days but still within the 140 days window, those AEs would be classified as TEAE. These AEs will be excluded from the outputs based on the Initial Period but included in the AE summaries for Initial and Maintenance Treatment Period.

- Incidence of TEAEs – Overview
- Incidence of TEAEs per 100 subject years by SOC, HLT, and PT
- Incidence of Serious TEAEs per 100 subject years by SOC, HLT, and PT
- Incidence of TEAEs Leading to Study Participant Discontinuation per 100 subject years by SOC, HLT, and PT
- Incidence of TEAEs Leading to Death by SOC, HLT, and PT
- Incidence of TEAEs by Maximum Relationship by SOC, HLT, and PT
- Incidence of Serious TEAEs by Relationship SOC, HLT, and PT – Note: For EudraCT reporting purposes
- Incidence of Related Serious TEAE by SOC, HLT, and PT
- Incidence of Severe TEAE per 100 subject years by SOC, HLT, and PT

- Incidence of TEAEs Leading to Death by Relationship by SOC, HLT, and PT – Note: For EudraCT reporting purposes
- Incidence of TEAEs by Maximum Severity, SOC, HLT, and PT
- Incidence of TEAEs by decreasing frequency of PT
- Incidence of TEAEs Above Reporting Threshold of 5% by SOC and PT
- Incidence of Non-Serious TEAEs by SOC, HLT, and PT
- Incidence of Non-Serious TEAEs by Maximum Relationship SOC, HLT, and PT
- Incidence of Non-Serious TEAEs Above Reporting Threshold of 5% by SOC and PT
- Incidence of Non-Serious TEAEs Above Reporting Threshold of 5% by Relationship SOC and PT
- Incidence of Related TEAEs per 100 subject years by SOC, HLT, and PT
- Incidence of Related TEAEs Above Reporting Threshold of 5% by SOC and PT
- Incidence of TEAEs – Suspected and Confirmed COVID-19 cases by SOC, HLT and PT
- Incidence of TEAEs Excluding TEAEs Exclusively Related to COVID-19 Vaccine by SOC, HLT, and PT
- Incidence of TEAEs Exclusively Related to COVID-19 Vaccine by SOC, HLT, and PT
- Incidence of COVID-19 Vaccine Interval Censored TEAEs by SOC, HLT, and PT
- Incidence of COVID-19 Vaccine Interval TEAEs by SOC, HLT, and PT

Suspected and confirmed COVID-19 cases will be identified with the preferred terms “Corona virus infection” or “Corona virus test positive”.

The following subset of tables will also be presented for the Maintenance Treatment Period using the MS:

- Incidence of TEAEs – Overview
- Incidence of TEAEs per 100 subject years by SOC, HLT, and PT
- Incidence of Serious TEAEs per 100 subject years by SOC, HLT, and PT
- Incidence of TEAEs Leading to Discontinuation per 100 subject years by SOC, HLT, and PT
- Incidence of TEAEs Leading to Death by SOC, HLT, and PT

The following tables will be presented for the Initial Treatment Period:

- Incidence of TEAEs Above Reporting Threshold of 5% with Risk Differences by SOC and PT
- Incidence of Serious TEAEs and Risk Differences by SOC and PT

The following table will be presented for the combined Initial and Maintenance Treatment Period. This summary will include only AEs that occur while a participant is on bimekizumab.

Any AEs in the Initial Treatment Period that begin while a participant is on placebo will be excluded.

- Incidence of TEAEs per 100 subject years by SOC, HLT, and PT and by Time of Onset Relative to Anti-bimekizumab Antibody Status. This will include columns for the following:
  - TEAEs starting before the first ADAb positive result (includes ADAb categories 2 and 5) where TEAEs have occurred before the following events: a) the first positive ADAb result for subjects in category 2 and b) the first post-Baseline boosted ADAb titer result for subjects with titer results and the first post-Baseline positive ADAb result for subjects with positive ADAb at Baseline with no other samples with titer available for subjects in category 5
  - TEAEs starting on the same date or after the first ADAb positive result (includes ADAb Categories 2, 3, 4, 5 and 6) where TEAEs have occurred on or after the following events: a) the first positive ADAb results for subjects in categories 2, 3, 4 and 6, and b) the first post-Baseline boosted ADAb titer result for subjects with titer results and the first post-Baseline positive ADAb result for subjects with positive ADAb at Baseline with no other samples with titer available for subjects in category 5
  - TEAEs for subjects who are ADAb negative at all timepoints (includes ADAb Category 1)

The tables with risk differences will also be accompanied by figures (dot plots) which show the incidence of the adverse events and corresponding 95% risk difference confidence intervals. These will be ordered by descending order of risk difference (bimekizumab vs placebo).

10.2.3 Other Safety topics of interest

The following are AEs considered to be other safety topics of interest that require special statistical analyses. Along with the tables described, there will be a table which displays the risk difference and 95% confidence intervals for each of the topics of interest in the Initial Treatment Period. A corresponding figure (with dot plots) will be prepared.

A by-participant listing of all AEs of safety topics of interest will be presented by type of safety topics of interest.

10.2.3.1 Infections (serious, opportunistic, fungal and TB)

- **Incidence of Serious Infection TEAEs per 100 subject years by SOC, HLT and PT**

  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.

- **Incidence of Fungal Infection TEAEs per 100 subject years by SOC, HLT and PT**

  Fungal infections will be summarized in a stand-alone table. The table will include all TEAEs (serious and non-serious) that code into the High Level Group Term (HLGT) "Fungal infectious disorders"

- **Incidence of Opportunistic Infection TEAEs per 100 subject years by SOC, HLT and PT**

Opportunistic infections (including tuberculosis) will be summarized in a stand-alone table. The table will include all TEAEs identified using UCB-defined search criteria .

The following steps will be followed for identifying and reviewing opportunistic infections:

Identification Process

The steps below outline 2 ways in which opportunistic infections (or potential opportunistic infections) can be identified:

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

- All TEAEs that code to a PT flagged with a single 'x' need to also be serious to be considered an opportunistic infection.
- All TEAEs that code to a PT flagged with a double 'xx' are considered to be an opportunistic infection, regardless of seriousness.

Step 2: Refer to column C of the spreadsheet, which identifies the PTs that need to be evaluated on a case-by-case basis by the study physician to determine whether or not it is an opportunistic infection. If column C has a single 'x', then the corresponding preferred term will be flagged for case-by-case review by the study physician.

Review Process

Opportunistic infections for a given study will be reviewed on the following occasions:

- At quarterly Infectious Disease Committee (IDC) Meetings, listings will be produced for each study (see details below) and reviewed by the corresponding study physician ahead of the IDC Meeting. These listings will be posted as part of the broader Safety Signal Detection (SSD) deliverable to a folder named for the given quarter (eg, 2018Q4) on the SharePoint. They will be based on the same data cut as the one used for SSD and will be delivered at the same time as the SSD outputs. The IDC will then agree on the final adjudication for each potential opportunistic infection.
- For each study, a final listing for opportunistic infections (in the format described below) will be produced and agreed upon between the study physician and the IDC prior to finalizing the database.

In each of the circumstances described above, the study programming team will produce an Excel listing that will be provided to the project lead statistician, project lead programmer, and to the study physician (who will subsequently provide it to the IDC). The Excel listing will contain the following columns (using the descriptions below as the column headings in the Excel listing):

- Study ID
- Unique Participant ID
- AE Term (Verbatim)
- AE Preferred Term
- AE System Organ Class
- AE High Level Term

- AE Low Level Term
- Date of Onset
- Outcome of Adverse Event
- Date of Outcome
- TEAE Flag
- Serious Adverse Event?
- Relationship to Study Medication
- Intensity
- Action Taken with IMP
- Opportunistic Infection – Automatic
- Opportunistic Infection – Manual Review
- Flag
- Data Cut Date
- Opportunistic Infection – Final Adjudication

Note the following about the final 5 variables in this listing:

- *Opportunistic Infection – Automatic*: This is flagged as "Y" if the criteria for automatic selection as described in "Step 1" of the identification process are met.
- *Opportunistic Infection – Manual Review*: This is flagged as "Y" if the criteria for case-by-case selection as described in "Step 2" of the identification process are met.
- *Flag* – This has a value of either "NEW" or "OLD". It is marked as "NEW" if the event is appearing for the first time in that run of the listing. Otherwise, if it has appeared previously, it is marked as "OLD". Unique records are determined by USUBJID AESPID for purposes of identifying whether an event has been modified from a previous run.
- *Date* – Only for cases where Flag is "NEW", this field will be populated with the data cut date for that particular run of the listing.
- *Opportunistic Infection – Final Adjudication* – For new events, this is always left blank by the programmers. It will be completed by the study physician/IDC for every event that appears in the listing. For events adjudicated as opportunistic, the field will be populated with a "Y".

Following each review by the study physician and IDC, the Opportunistic Infection – Final Adjudication column will be completed (as described above), and the spreadsheets for each study will be returned to the study programming team via e-mail (coordinated by the IDC secretary). Then, for subsequent runs of the listing, the study programming teams will incorporate adjudications from previous runs.

10.2.3.2 Malignancies

- **Incidence of Malignant or Unspecified Tumours TEAEs per 100 subject years by SOC, HLT and PT**

These events will be presented in the following tables:

  – One table will be based on the criteria Standardized MedDRA Query (SMQ) = "Malignant or unspecified tumours (SMQ)"

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

SMQ search will include all TEAEs which code to a PT included in the Scope=Narrow group within each SMQ.

Note that the events included in the "Malignancies" table will be a subset of the events included in the "Malignancies (including unspecified)" table. While the "Malignant tumours (SMQ)" is most relevant, "Malignant or unspecified tumours (SMQ)" must be reviewed for potential malignancies.

The output tables will include 2 different overall incidence rows:

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

10.2.3.3 Major adverse cardiac event

- **Incidence of Adjudicated Major Cardiac Event (MACE) TEAEs per 100 subject years by SOC, HLT and PT**

Potential cardiovascular events are adjudicated by the independent Cardiovascular Event Adjudication Committee (CV-CAC) according to the CV-CAC Charter (version 6.0). Adjudicated events are classified by the CV-CAC to one of the event types as defined in Table 10–1. The classification of an event as a Major Adverse Cardiac Event (MACE) is also determined by the CV-CAC. Events which are classified by the CV-CAC as any of the event types identified in the third column of Table 10–1 will be considered an extended MACE. Note that extended MACE is determined programmatically and includes a broader scope definition of MACE.

MACE as determined by the CV-CAC will be presented in a table and listing. Extended MACE will be presented separately in a table and listing.

Another table and listing will present the adjudicated cardiovascular events by type. For each cardiovascular event type, the individual PTs which fall within each event type will be summarized. This listing will indicate whether each event was determined to be a MACE and/or an extended MACE.

Additionally, a listing of all events identified for potential review by the CV-CAC will be produced. This listing will indicate whether each event was identified by the CV-CAC Chair for full committee review.

Table 10–1: Cardiovascular event classifications

Event Type Code	Event Type	Extended MACE
1	Non-Fatal Myocardial Infarction (MI)	Yes
2	Non-Fatal Stroke: hemorrhagic	Yes
3	Non-Fatal Stroke: ischemic	Yes
4	Non-Fatal Stroke: embolic	Yes
5	Non-Fatal Stroke: undeterminable	Yes
6	Hospitalization or ER for Unstable Angina with urgent revascularization	Yes
7	Hospitalization or ER for Unstable Angina without urgent revascularization	No
8	Hospitalization for Heart Failure	Yes
9	Transient Ischemic Attack (TIA)	No
10	Coronary Revascularization Procedures (e.g. percutaneous coronary intervention, coronary artery bypass grafting)	Yes
11	Urgent Revascularization Procedures (i.e. due to symptoms of brain ischemia or pending infarction)	Yes
12	Arrhythmia (not associated with ischemia)	No
13	Peripheral Arterial Event	No
14	Venous Thromboembolic Event: pulmonary embolism (PE)	No
15	Venous Thromboembolic Event: deep vein thrombosis (DVT)	No
16	Venous Thromboembolic Event: PE and DVT	No
17	Other CV Event	No
18	Death due to Myocardial Infarction (MI)	Yes
19	Death due to Stroke	Yes
20	Sudden Cardiac Death	Yes

Table 10–1: Cardiovascular event classifications

Event Type Code	Event Type	Extended MACE
21	Other CV Death (e.g. heart failure, pulmonary embolism, cardiovascular procedure-related)	Yes
22	Cardiovascular: Undetermined Cause of Death (i.e. cause of death unknown)	Yes
23	Non-Cardiovascular Death	No
24	Non-Cardiovascular Event	No
99	Inadequate information to adjudicate	No

CV=Cardiovascular; DVT=Deep Vein Thrombosis; ER=Emergency Room; MACE=Major Adverse Cardiac Event; MI=Myocardial Infarction; PE=Pulmonary Embolism; TIA=Transient Ischemic Attack.
MACE is determined by the adjudication committee and is not identified programmatically based on event type.

10.2.3.4 Neutropenia

- **Incidence of Neutropenia TEAEs per 100 subject years by SOC, HLT and PT**

This table will be based on the following PTs (regardless of seriousness):

- Autoimmune neutropenia
- Band neutrophil count decreased
- Cyclic neutropenia
- Febrile neutropenia
- Idiopathic neutropenia
- Neutropenia
- Neutropenic infection
- Neutropenic sepsis
- Neutrophil count decreased

10.2.3.5 Suicidal Ideation and Behavior

- **Incidence of Suicidal Ideation or Behavior TEAEs per 100 subject years by SOC, HLT and PT**

Potential neuropsychiatric events are adjudicated by the independent Neuropsychiatric Adjudication Committee according to the Neuropsychiatric Adjudication Committee (version 8.0). Adjudicated events are classified by the Committee as Suicidal or Non-suicidal. Adjudicated events are also further classified by the Committee to one of the event types as defined in Table 10–2. Suicidal Ideation and Behavior (SIB) is defined as events classified by the Committee as Suicidal.

A table and listing will present SIB events.

Another table and listing will present the adjudicated neuropsychiatric events by type. For each neuropsychiatric event type, the individual PTs which fall within each event type will be summarized. This listing will indicate whether each event was determined to be Suicidal or Non-Suicidal. For event type suicidal ideation, the listing will also indicate if intent was present and if the suicidal ideation was clinically significant.

Additionally, a listing of all events identified for potential review by the Committee will be produced. This listing will indicate whether each event was identified by the Neuropsychiatric Event Adjudication Committee Chair for full committee review.

Table 10–2: Neuropsychiatric event classifications

Event Type Code	Event Classification	Event Type
1	Suicidal	Suicidal events/completed suicide
2	Suicidal	Suicide attempt
3	Suicidal	Preparatory acts toward imminent suicidal behavior
4	Suicidal/Non-suicidal[a]	Suicidal ideation
7	Non-suicidal	Nonsuicidal Self-injurious behavior
8	Non-suicidal	Nonsuicidal Other
99	Not applicable	Inadequate information to adjudicate

[a] Suicidal ideation event types can be classified by the Neuropsychiatric Adjudication Committee as Suicidal or Non-suicidal depending on whether intent to die was present.

10.2.3.6 Inflammatory bowel disease

- **Incidence of Inflammatory Bowel Disease TEAEs per 100 subject years by SOC, HLT and PT**

Selected gastrointestinal events are adjudicated by the independent Inflammatory Bowel Disease (IBD) Adjudication Committee (IBD-CAC) according to the IBD-CAC Charter (version 3.0). Adjudicated events are classified by the IBD-CAC into one of the diagnostic types as defined in Table 10–3. The events will further be classified as definite, probable or possible IBD.

An overview of adjudicated IBD events will be stratified by subjects with or without a previous medical history of IBD. Previous medical history of IBD will be determined using the information recorded on the History of IBD CRF page ("Does subject have a history of IBD?"). This overview table will present events adjudicated by the IBD-CAC as either possible, probable or definite IBD. Definite and probable IBD will also be aggregated and summarized in this table. In addition, this table will summarize each IBD event classification (possible, probable or definite) separately.

Another table and listing will present the adjudicated IBD events by type. For each IBD event type, the individual PTs which fall within each event type will be summarized.

Additionally, a listing of all events identified for potential review by the IBD-CAC will be produced. This listing will indicate whether each event was identified by the IBD-CAC Chair for full committee review.

A further supportive listing will present the individual diagnostic criteria met for each adjudicated IBD event.

Table 10–3: IBD event classifications

Event Type Code	Event Type (Classification and diagnosis)	Classification
1	Possible Inflammatory Bowel Disease – Crohn's Disease	Possible
2	Probable Inflammatory Bowel Disease – Crohn's Disease	Probable
3	Definite Inflammatory Bowel Disease – Crohn's Disease	Definite
4	Possible Inflammatory Bowel Disease – Ulcerative Colitis	Possible
5	Probable Inflammatory Bowel Disease – Ulcerative Colitis	Probable
6	Definite Inflammatory Bowel Disease – Ulcerative Colitis	Definite
7	Possible Inflammatory Bowel Disease – type unclassified	Possible
8	Probable Inflammatory Bowel Disease – type unclassified	Probable
9	Definite Inflammatory Bowel Disease – type unclassified	Definite
10	Symptoms not consistent with Inflammatory Bowel Disease	Not applicable
11	Possible Inflammatory Bowel Disease – Microscopic Colitis	Possible
12	Probable Inflammatory Bowel Disease – Microscopic Colitis	Probable
13	Definite Inflammatory Bowel Disease – Microscopic Colitis	Definite
14	Possible Inflammatory Bowel Disease – no further differentiation possible	Possible
15	Probable Inflammatory Bowel Disease – no further differentiation possible	Probable
16	Definite Inflammatory Bowel Disease – no further differentiation possible	Definite
99	Not enough information to adjudicate	Not applicable

IBD=inflammatory bowel disease.
Note: IBD diagnoses of "microscopic colitis" and "no further differentiation possible" were added in an adjudication charter amendment, accounting for the event type numbering.

10.2.3.7 Hypersensitivity (including anaphylaxis)

- **Incidence of Anaphylactic Reaction TEAEs per 100 subject years by SOC, HLT and PT**

A separate table will be prepared based on the MedDRA anaphylaxis Algorithm (see Appendix 1) for acute anaphylactic events (reported on the same day as when an injection was administered or 1 day after). An AE glossary table will also be produced to summarize the MedDRA coding for these events. The glossary table will include the following fields: reported term, PT, LLT, HLT, and SOC.

A separate table will be prepared to summarize hypersensitivity events, identified using the SMQ "Hypersensitivity (SMQ)". All TEAEs which code to a PT included in the Scope=Narrow search will be included in this table. In addition, a separate table will be prepared to summarize serious hypersensitivity events, identified using the SMQ "Hypersensitivity (SMQ)". All serious TEAEs which code to a PT included in the Scope=Narrow search will be included in this table. An AE glossary table will also be produced to summarize the MedDRA coding for these events. The glossary table will include the following fields: reported term, PT, LLT, HLT, and SOC.

Furthermore, a separate table will be prepared to summarized injection site reactions, identified using the HLTs: "Administration site reactions NEC" and "Injection site reactions".

10.2.3.8 Hepatic events and PDILI

- **Incidence of hepatic events TEAEs per 100 subject years by SOC, HLT and PT**

A table for hepatic events will be created based on the SMQ of "Drug related hepatic disorders - comprehensive search (SMQ)". However, these 2 sub-SMQs are to be excluded: "Liver neoplasms, benign (incl cysts and polyps) (SMQ)" and "Liver neoplasms, malignant and unspecified (SMQ)". For each of the above SMQs, include all TEAEs which code to a PT included in the Scope=Broad and/or Scope=Narrow.

Note that all AEs meeting the above criteria are to be included. It will not be limited to events that the investigator determined to be related to study drug.

Cases of potential Hy's Law will be reported separately in a liver function test table.

10.3 Clinical laboratory evaluations

Laboratory values, including markedly abnormal laboratory values will be presented descriptively by treatment group for the SS, MS, and AMS.

The markedly abnormal tables and those based on common terminology criteria for AEs (CTCAE) grade will be produced only for selected laboratory variables.

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 (values observed more than 140 days after the last administration of study medication are not considered). 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 LLQ, 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 are required:

- 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 participants experiencing markedly abnormal values at any time while on treatment (assessment on or following the first dose of study treatment through the minimum of period of interest (Week 16) or date of last dose + 140 days) by laboratory variable and treatment group. Two separate tables will show results for the Initial Treatment Period (for the SS) and the Initial and Maintenance Treatment Period (for the AMS).
- A summary of the number and percentage of participants with a given CTCAE grade (0,1,2,3, or 4) based on minimum/maximum post-baseline value by laboratory variable and treatment group. Two separate tables will show results for the Initial Treatment Period (for the SS) and the Initial and Maintenance Treatment Period (for the AMS).
- A shift table of the number and percentage of participants experiencing CTCAE grade 0,1,2,3, or 4 values (as applicable) at Baseline to maximum post-Baseline CTCAE grade, by laboratory variable and treatment group. Two separate tables will show results for the Initial Treatment Period (for the SS) and the Initial and Maintenance Treatment Period (for the AMS).
- A by-participant listing of all laboratory data (including urinalysis) will be provided. This listing will be presented by treatment group and will include: center, participant identifier, age, sex, race, weight, visit, laboratory variable, result (with abnormal values flagged as "L" or "H" accordingly) and unit.

Markedly abnormal values are defined as those with a severity of Grade 3 and above based on the CTCAE criteria Version 4.03. Definitions of markedly abnormal values using the Grade 3 cut points are given in the tables below for age ranges of ≥17 years Table 10–4 for markedly abnormal liver function test values, Table 10–5 for markedly abnormal biochemistry values and Table 10–6 for markedly abnormal hematology values). Tables summarizing markedly abnormal values will include a summary (counts and percentages) of markedly abnormal labs observed at any time while on treatment (ie, treatment-emergent markedly abnormal [TEMA]). For this summary, Baseline values and values observed more than 140 days after the last administration of study medication are not considered. The laboratory results classified as Grade 3 or Grade 4 will be summarized and listed separately.

Table 10–4: Definitions of Markedly Abnormal Liver Function Values

Parameter name	Conventional		Standard		Abnormal Designation
	Unit	Criteria	Unit	Criteria	
ALP	U/L	>5.0 x ULN	U/L	>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 10–5: Definitions of Markedly Abnormal Biochemistry Values

Parameter name	Conventional		Standard		Abnormal Designation
	Unit	Criteria	Unit	Criteria	
Creatinine[1]	mg/dL	> 3.0 x Baseline or >3.0 x ULN	umol/L	> 3.0 x Baseline or >3.0 x ULN	AH
Glucose	mg/dL	<40 >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
Cholesterol	mg/dL	>400	mmol/L	>10.34	AH

1 The markedly abnormal definitions for creatinine are based on the logical or, if either criterion is met the creatinine value will be designated as abnormal high.

Table 10–6: Definitions of Markedly Abnormal Hematology Values

Parameter name	Conventional		Standard		Abnormal Designation
	Unit	Criteria	Unit	Criteria	
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; ALP = alkaline phosphatase; ALT= alanine aminotransferase; AST = aspartate aminotransferase; dL = deciliter; GGT: gamma-glutamyltransferase; L = liter; mg = milligram; mmol = millimoles; µg = microgram; ULN = upper limit of normal, WBC=white blood cells.

The table for markedly abnormal liver function tests (LFTs) will contain data beyond the CTCAE Grade 3 thresholds outlined in Table 10–4 above in order to allow for a more thorough review of elevated LFTs. There will be 1 table, which will list the count and percentage of participants meeting the below criteria at any time during the study:

- AST: >3xULN, >5xULN, >8xULN, >10xULN, >20xULN
- ALT: >3xULN, >5xULN, >8xULN, >10xULN, >20xULN
- AST or ALT: >3xULN, >5xULN, >8xULN, >10xULN, >20xULN
- Total Bilirubin: >1.5xULN, >2xULN
- ALP: >1.5xULN

For any participant with at least one markedly abnormal LFT (AST >3xULN, ALT >3xULN, bilirubin >3xULN, or ALP >1.5xULN) the New Ratio (nR) will be calculated as the ratio of either ALT or AST (whichever is higher) to ALP, all expressed as multiples of their ULN as follows:

- nR = [maximum(AST/ULN or ALT/ULN)]/(ALP/ULN)

Any pDILI will be summarized (all criteria must be met at the same assessment):

- (AST or ALT > 3xULN) and Total Bilirubin > 1.5xULN
- (AST or ALT > 3xULN) and Total Bilirubin > 2xULN

In addition, a table will be produced to summarize potential Hy's Law cases. The following definition will be used in that table:

- [AST ≥3xULN or ALT ≥3xULN] and Total Bilirubin ≥2xULN in the absence of ALP ≥2xULN

In order to meet the above potential Hy's Law criteria, a participant must experience the elevation in bilirubin and ALT or AST (and the absence of ALP elevation, if applicable) at the same assessment. For example, a participant who experiences a ≥2 x ULN elevation of bilirubin at one visit and a ≥3xULN elevation in ALT (or AST) at a subsequent visit has not fulfilled the Hy's Law criteria.

Potential hepatotoxicity (meeting one of the PDILI or Hy's Law laboratory criteria at least once) will be considered with and without symptoms potentially associated with hepatitis or hypersensitivity according to the investigator (reported on the Symptoms of Hepatitis and Hypersensitivity CRF page).

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 for the SS:

- 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 participants experiencing at least 1 markedly abnormal value for a vital sign variable as defined in Table 10–7, by treatment group and period (Initial Treatment Period [SS], and Initial and Maintenance Treatment Period [AMS]).

Unscheduled and repeat visits will be considered as long as they were collected in the period being summarized (values observed more than 140 days after the last administration of study medication are not considered).

Table 10–7: Definitions of Markedly Abnormal Blood Pressure Values

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

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

10.4.2 Electrocardiograms

Electrocardiogram data will be analyzed by treatment group and visit for the SS.

A summary of the number and percentage of participants with normal, abnormal ECG results, as determined by the central reader, will be presented for all applicable visits.

The following ECG variables will be summarized (absolute values and change from Baseline) by visit: QT corrected for heart rate using Friderica's formula (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 listing 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 participant who meet the ECG outlier criteria at any assessment post first dose will be summarized for each period.

Two separate by-participant listing of all 12-lead ECG data will be provided based on interpretation from central reader and from site, respectively.

10.4.3 Other safety endpoints

For by-visit summaries, unscheduled and repeat visits will not be summarized, but these data will be included in listings. By-visit tables should include the SFU visit. Summaries over a period of time (as in shift tables), unscheduled and repeat visits will be considered as long as they were collected in the period being summarized (values observed more than 140 days after the last administration of study medication are not considered).

10.4.3.1 Physical examination

Abnormal results of the physical examination together with details of abnormalities: abnormality clinically significant or not, will be listed by participant and visit for SS.

10.4.3.2 Columbia-Suicide Severity Rating Scale (C-SSRS)

The eC-SSRS questionnaire will be self-administered by the study participant and assessed by trained study personnel. This scale will be used to assess SIB that may occur during the study. Results of the eC-SSRS will be summarized using the number of participants and percentage with (i) suicidal ideation, (iii) suicidal behavior, (iii) suicidal ideation or behavior, and (iv) self-injurious behavior without suicidal intent.

Suicidal ideation is defined as an event in any of the following 5 categories:

[redacted]

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

[redacted]

Suicidal behavior or ideation is defined as an event in any of the above 9 categories.

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

The incidence of participants with suicidal ideation, suicidal behavior, suicidal ideation or behavior, and self-injurious behavior will be summarized for the Initial Treatment Period and the combined Initial and Maintenance Treatment Period by treatment group.

A by-participant listing of the eC-SSRS questionnaire data will be provided by treatment group.

10.4.3.3 Assessment and management of TB and TB risk factors

A summary of the number and percentage of participants with negative, positive, and indeterminate IGRA (Interferon-Gamma Release Assay) results at Screening and Week 44 will be presented.

A by-participant listing of the "Evaluation of signs and symptoms of tuberculosis" questionnaire data and IGRA results will be provided by treatment group.

By-participant listing of the result of chest x-ray for tuberculosis will be provided by treatment group.

10.4.3.4 Pregnancy testing

Pregnancy testing will consist of serum testing at the Screening. The pregnancy test will be urine at all other visits.

A by-participant listing of the pregnancy test data will be provided by treatment group.

10.4.3.5 Patient Health Questionnaire (PHQ)-9 scores

The PHQ-9 is a multipurpose instrument for screening, diagnosing, monitoring, and measuring the severity of depression. The PHQ-9 scores for depression range from 0 to 27 with higher scores indicating worse state. A score of 5 to 9 is considered to be minimal symptoms of depression. If any of the 9 questions are missing, then the score is treated as missing. A score of 10 to 14 is considered minor depression, dysthymia, or mild major depression. A score of 15 to 19 is considered to indicate moderately severe major depression, and a score ≥20 is considered to be severe major depression.

Change from Baseline in PHQ-9 is derived as post-Baseline score minus Baseline score.

A summary of the absolute and change from Baseline value will be presented by treatment group and visit.

The percentage of study participants with scores below 5, between 5 and 9, between 10 and 14, between 15 and 19, and greater than or equal to 20 in PHQ-9 will be summarized as a shift from Baseline by visit and treatment group based on observed values.

The percentage of study participants with scores ≥ 15 at any post-Baseline visit and the number and percentage of study participants with scores ≥ 20 at any post-Baseline visit will be summarized by treatment group based on observed values. This summary will also include the percentage of study participants with increase from baseline ≥ 5 at any post-Baseline visit.

The number and percentage of participants that complete the PHQ-9 will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS. The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit.

11 REFERENCES

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

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

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

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

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

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

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

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

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

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

Mallinckrodt CH, Lin Q, Lipkovich I, and Molenberghs G. A structured approach to choosing estimands and estimators in longitudinal clinical trials. Pharmaceutical Statistics 2012; 11:456-61.

O'Kelly, Michael, and Bohdana Ratitch. Clinical trials with missing data: A guide for practitioners. John Wiley & Sons, 2014

Ratitch, B., Lipkovich, I., O'Kelly, M. Combining Analysis Results from Mutiply Imputed Categorical Data, PharmaSUG, 2013, SP03

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

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

Sun H, Snyder E, Koch GG. Statistical planning in confirmatory clinical trials with multiple treatment groups, multiple visits, and multiple endpoints. J Biopharmaceut Stat. 2018;28(1):189-211.

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

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

12 APPENDICES

12.1 Appendix A: MedDRA algorithmic approach to anaphylaxis

The SMQ *Anaphylactic reaction* consists of three parts:

- A **narrow search** containing PTs that represent core anaphylactic reaction terms (Category A – core anaphylactic reaction terms)

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

- A **broad search** that contains additional terms that are added to those included in the narrow search. These additional terms are signs and symptoms possibly indicative of anaphylactic reaction and categorized in B, C or D

  - **Category B (Upper Airway/Respiratory Terms)**

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

Laryngeal oedema	Tracheal obstruction
Laryngospasm	Tracheal oedema
Laryngotracheal oedema	Upper airway obstruction
Mouth swelling	Wheezing

- **Category C (Angioedema/Urticaria/Pruritus/Flush terms)**

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

- **Category D (Cardiovascular/Hypotension terms)**

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

- An **algorithmic approach** which combines a number of anaphylactic reaction symptoms in order to increase specificity. A case must include one of the following where both occur on either the same day as when an injection was administered or one day after, and for scenarios where two events must have been reported, both events must have occurred within one day of each other (as anaphylaxis is an acute event, imputed dates should not be used in the algorithmic approach)**:**

- A narrow term or a term from Category A;
- A term from Category B - (Upper Airway/Respiratory) AND a term from Category C - (Angioedema/Urticaria/Pruritus/Flush);
- A term from Category D - (Cardiovascular/Hypotension) AND [a term from Category B - (Upper Airway/Respiratory) OR a term from Category C - (Angioedema/Urticaria/ Pruritus/Flush)]

12.2 Appendix B: Definition of CTCAE grades

Table 12–1: Definition of CTCAE grades by biochemistry parameters

Parameter	Definition	Unit	Grade 1	Grade 2	Grade 3	Grade 4
Creatinine[1]	High	umol/L	>1-1.5x Baseline or >ULN-1.5 x ULN	>1.5-3.0x Baseline or >1.5 – 3.0 x ULN	>3.0x Baseline or >3.0 – 6.0 x ULN	>6.0 x ULN
Sodium	Low	mmol/L	130-<LLN	N/A	120-<130	<120
Sodium	High	mmol/L	>ULN-150	>150-155	>155-160	>160
Potassium[2]	Low	mmol/L	3.0-<LLN	3.0-<LLN	2.5-<3.0	<2.5
Potassium	High	mmol/L	>ULN-5.5	>5.5-6.0	>6.0-7.0	>7.0
Calcium	Low	mmol/L	2.0-<LLN	1.75-<2.0	1.5-<1.75	<1.5
Calcium	High	mmol/L	>ULN-2.9	>2.9-3.1	>3.1-3.4	>3.4
Magnesium	Low	mmol/L	0.5-<LLN	0.4-<0.5	0.3-<0.4	<0.3
Magnesium	High	mmol/L	>ULN-1.23	N/A	>1.23-3.30	>3.30
Cholesterol	High	mmol/L	>ULN-7.75	>7.75-10.34	>10.34-12.82	>12.82

1 The CTCAE Grade definitions for creatinine are based on the logical or the highest applicable CTCAE grade should be assigned to a creatine value.
2 The decreased potassium criterion of 3.0-<LLN is specified for both CTCAE Grade 1 and Grade 2; values meeting this criterion will be counted as Grade 2.

Table 12–2: Definitions of CTCAE grades by hematology parameter

Parameter	Definition	Unit	Grade 1	Grade 2	Grade 3	Grade 4
Hemoglobin	Low	g/L	100-<LLN	80-<100	<80	N/A
Hemoglobin[1]	High	g/L	>0-20 above ULN or >0-20 above Baseline if Baseline is above ULN	>20-40 above ULN or >20-40 above Baseline if Baseline is above ULN	>40 above ULN or >40 above Baseline if Baseline is above ULN	N/A
Platelets	Low	10^9/L	75-<LLN	50-<75	25-<50	<25

Table 12–2: Definitions of CTCAE grades by hematology parameter

Parameter	Definition	Unit	Grade 1	Grade 2	Grade 3	Grade 4
WBC	Low	10^9/L	3-<LLN	2-<3	1-<2	<1
WBC	High	10^9/L	N/A	N/A	>100	N/A
Lymphocytes	Low	10^9/L	0.8-<LLN	0.5-<0.8	0.2-<0.5	<0.2
Lymphocytes	High	10^9/L	N/A	>4-20	>20	N/A
Neutrophils	Low	10^9/L	1.5-<LLN	1.0-<1.5	0.5-<1.0	<0.5

LLN=lower limit of normal; N/A=not applicable; ULN=upper limit of normal, WBC=white blood cells
1 The CTCAE Grade definitions to be applied are dependent on the Baseline hemoglobin value. If the baseline value is > ULN then the criteria relative to Baseline is applicable, otherwise the criteria relative to ULN is applicable.

Note that participants who meet the decreased potassium criterion of 3.0< LLN, which is specified as the decreased potassium lab criterion for both CTCAE Grade 1 and Grade 2, will be counted as Grade 2.

13 AMENDMENT TO THE STATISTICAL ANALYSIS PLAN

13.1 Amendment 1

Rationale for the amendment

The main purposes of this amendment were:

- General update to analyses to align with protocol amendment 3.
- Procedural clarifications from discussions and feedback provided at meetings
- Update to align with the bimekizumab program standards and safety topics of interest

Modifications and changes

Global Changes

Typos and formatting were updated throughout the document.

Global changes:

The following changes were made throughout the SAP:

- References to Section 3.9 were added through the efficacy sections to clarify definition and handling of intercurrent events
- PRO endpoint terminology was updated to match protocol (eg, Worst Pain score in HSSDD instead of Worst Pain in HSSDD)

Specific changes

In addition to the global changes, the following specific changes have been made (formats as missing spaces or redundant spaces are not listed, typos):

Change #1

The following abbreviations have been added:

CFS	COVID-19 Free Set
COVID-19	coronavirus disease 2019
NI	Negative Immunodepletion
nR	New Ratio
NS	Negative Screen
pDILI	potential drug induced liver injury
PI	Positive Immunodepletion
PS	Positive Screen

Change #2

Section 1 Introduction

The protocols were updated:

The SAP is based on the Protocol Amendment **3 ~~2~~, 3 February 2021** ~~16 December 2019~~.

Change #3

Section 2.2 Study endpoints

The following text was added:

The endpoints based on HS Symptom Daily Diary (HSSDD) and Hidradenitis Suppurativa Symptom Questionnaire (HSSQ) pain responses are based on the current definitions, which are continuous. It is anticipated that a responder (binary) endpoint will be defined for the HSSDD and HSSQ pain items as well as other symptom items prior to database lock and unblinding, based on separate ongoing, blinded, psychometric analyses aiming to determine threshold for within-patient clinically meaningful improvement.

The below HSSDD and HSSQ pain response endpoints and analyses will be adjusted accordingly in a future SAP amendment.

Change #4

Section 2.2.1.2 Secondary Efficacy endpoints

The secondary endpoints were updated:

The secondary efficacy endpoints are defined as:

- $HiSCR_{75}$ response (defined as at least a 75% reduction from Baseline in the total AN count with no increase from Baseline in abscess or draining tunnel count) at Week 16
- Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Baseline) by Week 16
- **Absolute change from Baseline in Dermatology Life Quality Index (DLQI) Total Score at Week 16**
- Absolute change from Baseline in Skin Pain score at Week 16, as assessed by the "worst pain" item (11-point numeric rating scale) in the HS Symptom Daily Diary (HSSDD)
- ~~Absolute change from Baseline in Dermatology Life Quality Index (DLQI) Total Score at Week 16~~

Change #5

Section 2.2.1.3 Other efficacy endpoints

The following text was added:

- Absolute change from Baseline in Hidradenitis Suppurativa Quality of Life (HiSQOL) domain scores (symptoms, psychosocial, activities and adaptations) **and Total score**

Change #6

Section 2.2.3.2 Other safety endpoints

The other safety topics of interest were updated:

- Other safety topics of interest: infections (serious, opportunistic, fungal, and TB), neutropenia, hypersensitivity **(including anaphylaxis),** suicidal ideation and behavior, ~~depression~~, major **adverse** cardiovascular events, **hepatic events and potential drug-induced** liver **injury (PDILI),**~~function test changes/enzyme elevations~~, malignancies, and inflammatory bowel disease.

Change #7

Section 2.3.1 Study description

The following sentence was deleted:

Enrollment of study participants currently using antibiotics will be capped at 30% of overall enrollment.

Change #8

Section 3.1 General presentation of summaries and analyses

The following text was added:

For PRO continuous variables, descriptive statistics will also include variable score, absolute and percentage changes from baseline, Q1 and Q3, 10th, and 90th percentiles.

If no participants have data at a given time point, then only n=0 will be presented. The other descriptive statistics will be left blank. If n < 3 then the n, minimum, and maximum only will be presented. The other descriptive statistics will be left blank. If n = 3 n, mean, median, minimum and maximum will be presented only. The other descriptive statistics will be left blank.

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

- For summaries of demographics and Baseline characteristics: percentages will be based on all study participants in the analysis set and a "Missing" category (corresponding to study participants with missing data for the variable being summarized) will be included as the last row in the list of categories being summarized.
- For summaries of efficacy and safety endpoints, unless otherwise specified: percentages will be based only on those study participants with observed data for the variable being summarized. As the denominator may be different from the number of study participants in the analysis set being considered, the denominator will be displayed in the table. The general format for displaying this will be "n/Nsub (%)."

Percentages will be presented to 1 decimal place. If the percentage is 100%, a decimal will not be presented. If the count is 0, the percentage will not be presented. Typically, the % sign will be presented in the column header, but not with each individual value.

For the purpose of the tabulations the lower and upper confidence limits for the percentages will be truncated at 0 and 100% respectively. **Confidence intervals (CIs) for the response rates in efficacy summaries based on nonresponder imputation (NRI) will be computed using the Wilson approximation.**

Change #9

Section 3.3 Definition of Baseline

The following text was added:

For randomized participants for whom no start date of treatment is available, the Baseline value will be considered as the last available value on or before the randomization date.

Change #10

Section 3.4 Protocol deviations

The following text was added:

Deviations related to the Coronavirus Disease 2019 (COVID-19) global pandemic are unavoidable deviations from the protocol due to confirmed COVID-19 infection, suspected COVID-19 infection, general circumstances around COVID-19 without infection or any other deviation from the protocol due to COVID-19. COVID-19 protocol deviations will also be reviewed separately as part of the ongoing data cleaning process.

Change #11

The following section was added:

COVID-19 Free Set (Section 3.5.9)

The COVID-19 Free Set (CFS) will consist of all study participants randomized into the study and who have no COVID-19 impact through Week 16. This analysis set will be used for sensitivity analysis of the primary efficacy endpoint.

Change #12

Section 3.8 Coding dictionaries

The following sentence was updated:

Medications will be coded according to the World Health Organization Drug Dictionary (WHODD) version **MAR2021 B3 or later**~~.SEP2015~~. Medical procedures will not be coded.

Change #13

The following section was added:

Definition of an intercurrent event (Section 3.9)

Handling of intercurrent events is one of the key elements for the analysis of efficacy endpoints.

An intercurrent event is defined as receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy (See Section 8.2.2).

Receipt of systemic antibiotic rescue medication is defined as initiating any systemic antibiotic on or after Baseline for any reason (including in response to an AE). The only exception to this rule is if a participant randomized to the antibiotic stratum on a tetracycline antibiotic interrupts their stable dose of tetracycline antibiotic during the study and subsequently restarts the same tetracycline antibiotic as confirmed using the coded preferred term. The restarted dose and frequency of the antibiotic must be the same or lower than the regimen prior to the interruption.

The dates of an intercurrent event are as follows:

- For receipt of systemic antibiotic rescue medication: start date of the antibiotic
- For discontinuation of study treatment due to an AE or lack of efficacy: Last study treatment date + 17 days. Note: study treatment discontinuation includes study discontinuation.

The choice of 17 days is intended to capture the interval between dosing and lesion assessments (14 days), as well as the visit window (3 days).

An additional sensitivity analysis will be conducted where missing data due to COVID-19 will be considered an intercurrent event and will be imputed as a nonresponse at that particular visit. This will be identified when there are missing data at a visit that has been impacted by COVID-19 according the COVID-19 impact CRF page. The date of this intercurrent event will be the date of the impacted visit.

Change #14

Section 3.10 Changes to protocol-defined analyses

The following text was added:

The HiSQOL endpoint was clarified to show that there are only 3 domains: symptoms, psychosocial, activities and adaptations **and to add total score**.

Also, the following endpoint was added to the list of protocol endpoints not included in the analysis:

- Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Week 16) by Week 48

Change #15

Section 4.1 Adjustments for covariates

The following text was added:

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

Change #16

Section 4.2.1.4 Missing Data Overview and Summary

The following text was added:

In summary, the approaches listed below will be used in this study for handling missing data for efficacy endpoints as appropriate:

- NRI: **Participants** who have missing data at the timepoint of interest are treated as though they did not respond to the treatment. **This approach is also referred to as Composite Estimand (NRI).**
- Multiple Imputation (MI) – MCMC / Monotone Regression: Using multiple imputation methodology, intermittent missing data are imputed based on the MCMC method, and monotone missing data are imputed using monotone regression.

- MI–MCMC / Reference-based imputation: Using multiple imputation methodology, intermittent missing data are imputed based on the MCMC method, and monotone missing data are imputed using an imputation model based on placebo (reference) data.
- LOCF: Post-Baseline missing data are imputed by carrying forward the last available observation (including Baseline).
- Tipping point analyses: Assumptions will be made about average outcomes among the subsets of **participants** who prematurely discontinued study treatment. Various "delta adjustments" will be made to the assumed responses among missing data in each treatment arm with varying degrees of plausibility in order to identify assumptions about the missing data under which the conclusions change (O'Kelly, 2014). Then, the plausibility of such assumptions is discussed.
- Observed case (OC): Missing data are not imputed. Only **participants** with available data who have not discontinued study treatment at the given timepoint are considered. Study participants with missing data or who have prematurely discontinued study treatment will be treated as missing. ~~For OC summaries, intercurrent events are not handled differently than other missing data.~~
- **Treatment policy strategy: All available data observed at the time point of interest will be considered, regardless of the occurrence of intercurrent events. This means the analysis includes on- and off-treatment values collected after study participants prematurely discontinued study treatment but agreed to remain on the study and continued to attend visits and provide assessments at those visits as well as values from study participants who received rescue antibiotic medication. Those observed values will be analyzed according to the study participant's randomized treatment. Study participants for whom efficacy data cannot be obtained at the week of interest, despite attempts to retain them in the study, will have their data imputed using MI – MCMC / monotone.**

Table 4-1 was updated:

Table 4–1: Missing data handling approach by endpoint priority and type

Endpoint Priority	Endpoint Type	Composite Estmi nad (NRI)	Modified Composite Estimand (MI)	MI (MCMC/ Reference-based)	Tipping Point	Treat-ment Policy	Hypo-thetical Estimand	OC
Primary	Responder	S[a]	P	S[a]	S	S[a]		S
Secondary **included in the statistical testing procedure**	Responder	S[a]	P					S
	Continuous						P	S
	Binary	X	X					X

Secondary not included in statistical testing procedure	**Continuous**						**X**	**X**
Other	**Responder**	**X[d]**	**X**					**X[d]**
	Continuous						**X**	**X[b]**
	Ordinal						**X[c]**	**X[c]**

~~B=Backup method, LOCF=Last observation carried forward, MCMC=Markov Chain Monte Carlo,~~ MI=multiple imputation, NRI=Nonresponder imputation, OC=Observed case, P=Primary method, S=Sensitivity method, **X=Method to be used (no priority designated).**

Note**: Composite estimand (NRI) refers** ~~Backup method is only applicable when the primary method is unable~~ to **the approach in which data preceded by the intercurrent event of study treatment discontinuation**~~converge~~ due to **AE or lack of efficacy or receipt of rescue antibiotic medication are imputed as nonresponse, and other missing data are also imputed as nonresponse.**

Note: Modified Composite Estimand (MI) refers to the approach in which data preceded by the intercurrent event of study treatment discontinuation due to AE or lack of efficacy or receipt of rescue antibiotic medication are imputed as nonresponse, and other missing data are imputed via a multiple imputation model.

Note: Hypothetical Estimand (MI) refers to the approach where outcomes for study participants without an intercurrent event of study treatment discontinuation are as observed, and outcomes for study participants~~challenges~~ **with the intercurrent event are imputed via a multiple** imputation model.

[a] Imputation method is applied on continuous data, and responder endpoint is derived from the continuous endpoint based on complete data set where applicable.

[b] Required only for by-visit summaries of variables whose value at Week 16 is part of the hierarchical testing procedure.

[c] **For variables with multiple categories, data will be summarized as observed with an additional missing row to capture missing data at a given visit.**

[e] ~~Participants with intercurrent events are imputed as nonresponders for all subsequent timepoints before the imputation method is applied for all other missing data.~~

[d] NRI/OC sensitivity analysis will be performed only for $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, $HiSCR_{100}$ summaries.

[e] ~~The treatment policy estimand will use the same MI-MCMC/Monotone Regression defined for the primary analysis, with the exception that participants with intercurrent events will not be treated as nonresponders for all subsequent timepoints before the imputation method is applied for missing data.~~

Change #17

Section 4.2.2 Missing data algorithms for efficacy analyses

The section has been moved from Section 4.2.1.5 and split into Section 4.2.2.1, Section 4.2.2.2, and Section 4.2.2.3.

Change #18

Section 4.2.2.1 MI – MCMC/Monotone Regression

This section was updated to the following text:

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. To investigate the efficacy results under the assumption of data being MAR, a multiple imputation method will be applied~~ as follows:~~

Binary endpoint

For a binary endpoint (eg, HiSCR$_{50}$), the procedure is as follows:

6. **Create a data set, one for each treatment group of participants with observed values and those needing estimation by multiple imputation. For the imputation step, a distinction is made between non-monotone missing values (ie, intermittent missing values between completed assessments) and monotone missing values (ie, where all participants data are missing after a given time point).**

c. **For the intermittent missing values, the missing values in each data set will be filled in using the MCMC method with multiple chain, monotone missing data imputing pattern, and non-informative prior for all parameters. Unless specified differently, the first 200 iterations will not be used (the "Burn-in" option). A total of 100 sets of imputations will be performed. The seed used for these imputations will be 762 and all other multiple imputation procedures described in this SAP will use this same seed as well. The resulting 100 imputed data sets will have a monotone missing pattern and will be imputed using a method for monotone missingness. Note** ~~Create a data set, one for each treatment group (~~note that a separate imputation procedure must be invoked in SAS for each treatment group as the seed cannot be set for by groups beyond the first when using a by statement~~), of participants with observed values and those needing estimation by multiple imputation. The intermittent missing lesion counts~~ **in PROC MI**.

d. ~~each data set (ie, missing values for a given subject that has available data before and after the missing timepoint) will be filled in using the MCMC method, with a total of 100 sets of imputations being performed. The seed used for these imputations will be 762 (note that all other multiple imputation procedures described in this SAP related to MCMC/Monotone regression analyses will use this same seed as well).~~ For monotone missing data ~~(ie, where all participant data is missing after a given timepoint),~~ monotone regression will ~~then~~ be used to impute missing data. **A separate regression model is estimated for each variable with missing values (ie, measurement at each time point). Based on the resulting model, a new regression model is then drawn and is used to impute the missing values for the variable. Since the data set has a monotone missing data pattern, the process is repeated sequentially for variables with missing values.** This **procedure** will be based on the100 **imputed datasets generated from** ~~sets of imputations already created using~~ the MCMC **procedure and** ~~method such that there~~ will be **performed by** imputation. **The SAS® PROC MI procedure will be used for the imputation.**

i. ~~data sets in total.~~ In both cases, Hurley Stage at Baseline, Baseline antibiotic use, and **value of the variable of interest** ~~lesion count~~ at Baseline and at each post-Baseline visit (**prior to the time point of interest** ~~in chronological order, see notes below about visits to include for different analysis sets~~) will be included in the imputation model. **The post-Baseline values will need to be specified in chronological order in the imputation model so that SAS® PROC MI imputes variables from left to right (eg, the Week 2 value will be first imputed using regression based on the Baseline value, and then Week 4 value will be imputed using regression based on Baseline and Week 2**

values, etc). Note that lesion count at earlier visits will also be used as predictors for the model of lesion count 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 **joint multivariate normal** ~~numeric~~ **variables**. Therefore, Hurley Stage at Baseline and Baseline antibiotic use will be re-coded as indicator variables. For Baseline antibiotic use, this will simply be 0 for Baseline antibiotic non-users and 1 for Baseline antibiotic users. For Hurley Stage at Baseline, this will be 0 for Hurley Stage II participants and 1 for Hurley Stage III participants. In order to achieve model convergence, Baseline antibiotic use may be dropped from the model. If convergence is still not obtained, then Hurley Stage at Baseline may also be dropped from the model. **Additionally, if a variable is dropped in order to allow convergence for one model in a study, that variable does not have to be dropped from other models in the study if the model converges without dropping the variable. In other words, model convergence should be evaluated for each efficacy variable independently.**

Note: The imputation of each lesion type (inflammatory nodule, abscess, draining tunnel, etc) will be performed separately. The 100 data sets obtained for each type will be merged by imputation number and subject number.

2. For each complete imputed data set, the dichotomous responder variable (**eg**, HiSCR 0 or 1) ~~based on the imputed %ΔAN and draining tunnel (fistula/sinus tract) count~~ will be computed. Each complete imputed data set will then be analyzed based on the logistic regression model.

Note: For derivation of HiSCR response, the AN, inflammatory nodule, abscess, and draining tunnel (fistula/sinus tract) counts at Week 16 in the imputed data sets will be compared directly to the observed Baseline counts to determine response. If values outside of the pre-defined range of values for lesion count (<0) are imputed, they will be cut off as appropriate after the multiple imputation procedure but before deriving the responder variable. For example, an imputed draining tunnel (fistula/sinus tract) count of **-1** would be changed to 0 before deriving the HiSCR responder variable. Additional ranges for values for secondary and other endpoints are defined in Table 4-2.

Note: Standard rounding rules will also be applied to the imputed values of endpoints that can only take integer values (eg, abscess count). For example, if a study participant has an abscess count imputed as 2.4, this imputed value would be rounded down to 2. This rounding step is performed after the multiple imputation but before deriving the responder variable.

Table 4–2: Imputation allowable ranges by variable

Variable	Minimum Value	Maximum Value	Integer Values Only
Lesion count[a]	0	--	Yes
DLQI total score	0	30	Yes
hs-CRP	**LLOQ/2**	**--**	**No**
HSSDD item score	0	10	Yes

Table 4–2: Imputation allowable ranges by variable

Variable	Minimum Value	Maximum Value	Integer Values Only
HSSQ item score	0	10	Yes
~~IHS4~~	~~0~~	~~–~~	~~No~~
~~HiSQOL total score~~	~~0~~	~~68~~	~~Yes~~
HiSQOL symptom status score	0	16	Yes
HiSQOL psychosocial impact score	0	20	Yes
HiSQOL impact on physical activities score	0	32	Yes
EQ-5D-3L VAS	0	100	Yes
WPAI dimension scores	0	100	No for variables: "Percent work time missed due to problem" and "Percent overall work impairment due to problem". Yes for variables: "Percent impairment while working due to problem" and "Percent activity impairment due to problem". These two variables can only take values that are multiples of 10.
~~PHQ-9~~	~~0~~	~~27~~	~~Yes~~

[a] Lesion counts will be imputed separately for each lesion type (abscesses, draining tunnels [fistulas/sinus tracts], inflammatory nodules, non-draining tunnels [fistulas/sinus tracts], non-inflammatory nodules, HS scars). **The imputed lesion counts will be used to derive the endpoints that are dependent on the lesion count data (eg, $HiSCR_{50}$).**

7. **Estimates of the adjusted responder rate for each treatment group and the associated SE are obtained** ~~The Week 16 results~~ **from the** ~~specified statistical analysis~~ (logistic regression ~~model per Section 8.2.2~~) of each of the 100 imputed data sets. **These estimates** will be combined for overall inference using Rubin's rules, which account for the uncertainty associated with the imputed values (Rubin, 1987)**, and the combined estimates and SEs will be used to construct 95% CIs using the logit scale. This will be done using SAS PROC MIANALYZE. The combined estimates and 95% CIs on the logit scale will be back-transformed using the inverse logit link function to obtain the adjusted responder rates (%) and associated 95% CIs)**. ~~This will be done using SAS PROC MIANALYZE.~~

   Note: The (unadjusted) proportion of responders will be calculated at each time point by treatment group from the imputed datasets using SAS PROC FREQ. These results will also be combined into an overall inference using SAS PROC MIANALYZE.

~~Note that this procedure indicates that the imputation model will be applied for each treatment group separately in PROC MI, which will be the default method. However, in the event that there are computational challenges with the imputation model (eg, due to a standard deviation of~~

~~0 for responses of a given imputation), it is acceptable to modify the imputation model to include treatment as a variable in the model rather than running a separate model for each treatment group. It should be noted that doing so assumes that treatment does not interact with any of the other variables in the imputation model.~~

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 models in Step 3 follow a log-normal distribution, a log transformation is needed to normalize these 100 odds ratio estimates. That is because the procedures for combining results from multiple imputed datasets assume that the statistics estimated from each imputed dataset are normally distributed. Therefore, the log of the odds ratio estimates from the logistic regression model are used when combining into a single inference **(Step 3). Additionally, the SE for the odds ratios are transformed as follows**:~~the use of PROC MIANALYZE in step 3). Appropriate transformations to the standard errors and p-values will also be made in order to get the correct confidence intervals. For the logistic regression using the p-value for the general association the Wilson-Hilferty transformation will be used (Ratitch, 2013).~~

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

Where UCL and LCL are the upper and lower confidence limit, respectively, for the CI of the odds ratio from the logistic regression model, and $Z_{\alpha/2}$ is the relevant critical value from the standard normal distribution (2.24 for a 97.5% CI). The estimates of the log odds ratio for Bimekizumab relative to placebo and the corresponding upper and lower CLs will be provided. The odds ratio will be then estimated by exponentiating the estimate of the log odds ratio. The odds ratio and the confidence limits of the odds ratio will be estimated as follows:

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

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

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

Where OR is the back-transformed estimate of the odds ratio just described, SE is the SE of the log odds ratio and $Z_{\alpha/2}$ is the relevant critical value from the standard normal distribution (2.24 for a 97.5% CI). These calculations will be done such that odds ratios and corresponding CIs are calculated for the odds ratio of bimekizumab vs. placebo.

Note: . If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose, with a corresponding $Z_{\alpha/2}$ of1.96**.**

In addition to calculating the odds ratio, associated CIs, and p-values for the pairwise comparisons of bimekizumab and placebo, the estimated proportion of responders (ie, estimated responder rate) and the difference in the proportion of responders between each bimekizumab treatment group and placebo will be estimated, and 2-sided 95% CIs will be

created for each difference. The creation of the estimates of the differences will be completed for each bimekizumab treatment group using the process detailed below:

8. **Use the logistic regression model to calculate:**

   Least squares mean estimates of the log odds of bimekizumab (G_B) and placebo (G_P), as well as their corresponding standard errors (S_B and S_P, respectively).

   Standard error of the least squares mean estimate of the log odds ratio (S_R)

9. **Compute estimates for predicted proportions using the following transformations:**

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

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

   The difference in proportions is then given by:

   $$D = P_B - P_P$$

10. **Estimate the standard error of D by:**

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

The MCMC method for multiple imputation, as previously outlined, ~~Missing data for continuous components of the primary endpoint and binary secondary efficacy endpoints will be imputed using MI as appropriate.~~

~~The above describes the procedure for binary endpoints. For continuous endpoints, the MI procedure will be similar to that described above with the following differences:~~

~~11. The absolute value of the given variable will be imputed. Once imputation has been performed across the 100 iterations specified, any values outside of the range of the given variable will be truncated accordingly.~~

~~12. The change from Baseline values will be computed based on the complete data sets.~~

~~13. The analysis model will be based on ANCOVA (Section 8.3.2 and Section 8.3.3) as opposed to logistic regression.~~

~~For other efficacy variables, MI~~ **will be used to account for missing values. The calculation steps** ~~impute missing data when possible and where specified. If the imputation model cannot converge, LOCF will be used. The MI procedure will also be similar to that~~ **described above will be based on the results provided from the logistic regression model of the multiple imputed datasets. An estimate for the difference in proportions, D, and corresponding standard error, SD, will be computed for each of these datasets. The results from these analyses will be combined into a single estimate of the difference in predicted proportion of response and a 2-sided 95% CI interval using SAS PROC MIANALYZE.**

Note that this procedure indicates that the imputation model will be applied for each treatment group separately in PROC MI, which will be the default method. , ~~for continuous and binary endpoints respectively.~~ **However, in the event there are computational challenges with the imputation model (eg, due to a standard deviation of 0 for responses of a given imputation), it is acceptable to modify the imputation model to include treatment as a variable in the model rather than running a separate model for each treatment group. It**

should be noted that doing so assumes that treatment does not interact with any of the other variables in the imputation model.

Continuous endpoint

For continuous endpoints (eg, Change from Baseline in DLQI total score at Week 16), the MI method will be applied as follows:

5. **The MCMC/monotone regression method described above in Step 1 for binary endpoints will be performed.**
6. **Based on the multiply imputed data sets obtained for the given variable, the change from Baseline will be derived for each of the 100 complete** ~~no inferential statistics will be calculated for the~~ **imputed data sets based on the observed Baseline value and the observed/imputed post-Baseline values. Note that if the value itself is being summarized, no additional derivation is needed.**
7. **If a statistical model is being used for the analysis of the variable, then that will be performed for each imputation in this step. If no statistical model is being used, then simple descriptive statistics will be calculated.**
8. **For data excluding hs-CRP, the following rules apply. The results of the 100 imputed data sets (based on the statistical model or descriptive statistics) are combined with means and standard errors** ~~Means and standard errors will be~~ calculated using Rubin's rules (via PROC MIANALYZE). **Note that for the** calculation of other descriptive statistics such as the median, min, and max, Rubin's rules do not apply. MI estimates will be computed by simply averaging the estimates from the multiple repetitions of the imputation algorithm. **Thus, for median, Q1, Q3, minimum and maximum, the following approach will apply**:that ~~will be used when summarizing continuous secondary efficacy variables by subgroup~~.
   - **The data will be summarized by treatment, visit, and imputation, and the summary statistics will be computed.**
   - **Results will be summarized by taking the mean value of each summary statistic at each visit across all imputations.**
   - **The number of decimal places will remain the same as the original for display purposes (ie, if the mean was presented to 1 decimal place, then the mean of the means will also be presented to 1 decimal place).**

For hs-CRP only, the following rules apply. The hs-CRP data will be presented using the geometric mean, 95% CI for the geometric mean, median, Q1, Q3, minimum and maximum. The change from Baseline will be expressed as the ratio to Baseline in the summaries. The following approach will be applied:

- **Following the MI procedure, the ratio to Baseline will be calculated for any of the imputed values**
- **The natural logarithm of the absolute values and of the ratios to Baseline will be calculated**

- **The logged values will be summarized (using PROC MEANS) by treatment, visit and imputation**
- **The datasets will be combined using PROC MIANALYZE in order to get the mean and 95% CI estimates from the absolute values and ratios to Baseline (based on logged data) across imputations**
- **The estimates of the mean and 95% CI will be back-transformed to obtain the geometric mean and 95% CI on the original scale**
- **For the median, Q1, Q3, minimum and maximum the procedure outlined above for the other endpoints will be followed**

If the imputation model cannot converge, LOCF will be used.

Change #19

Section 4.2.2.2 MI – MCMC/ Referenced-based imputation

The steps of the procedure were updated to:

The steps for the procedure are as follows:

1. For non-monotone (intermittent) missing data, MCMC will be used to impute lesion count data, with Baseline antibiotic use, Hurley Stage at Baseline, and lesion count at Baseline and at each post-Baseline visit (in chronological order) being included in the imputation model. This will be done only once for each participant in order to provide a dataset with monotone missing data.
2. Data will be processed sequentially by repeatedly calling SAS® PROC MI to impute missing outcome data at visits t=1, … T, where T is Week 16 for $HiSCR_{50}$.
    a. *Initialization*. Set t=1 (Baseline visit)
    b. *Iteration*. Set t=t+1. Create a data set combining records from bimekizumab- and placebo-treated participants with columns for covariates (Hurley Stage at Baseline and Baseline antibiotic use) and outcomes at visits 1 to t. Outcomes for all bimekizumab-treated participants are set to missing at visit t and set to observed or previously imputed values at visits 1 to t-1. Outcomes for placebo-treated participants are set to observed at visit t or observed or previously imputed values at visits 1 to t-1. **The outcomes should be sorted in chronological order in the model.**
    c. *Imputation*. ~~Run MCMC to impute~~ **Impute missing values** for visit t using previous outcomes for visits 1 to t-1, Baseline antibiotic use, and Hurley Stage at Baseline. Note that only placebo data will be used to estimate the imputation model since no outcome is available for bimekizumab-treated **participants at visit t. Consequently, the input dataset should include all study participants from placebo but only study participants from the bimekizumab arm that have values at timepoint t missing**.
    d. Repeat steps 2a-2c, 100 times with different seed values (seeds ranging from 853 to 952) to create 100 imputed complete data sets. **Study participants whose missing values were imputed in the last PROC MI call will be included in the input dataset for the next PROC MI call. Standard rounding rules will be applied to the imputed values. If the MI procedure yields a value outside of**

the pre-defined range for the given variable, the value will be updated to be within the predefined range of values for the endpoint of interest. For example, the imputed value for inflammatory nodules will be updated to 0 in the case of an imputed value less than 0.

e. *Analysis*. For each completed data set, the outcome (response or change from Baseline) will be calculated using the complete datasets (with no missing data).

3. Each complete imputed data set will then be analyzed based on the statistical model specified in this study (logistic regression). The Week 16 results from logistic regression 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.

Change #20

Section 4.2.2.3 Tipping Point Analysis

The steps for performing the tipping point analysis were updated:

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

The objective of the tipping point analyses is to **evaluate the sensitivity of results to departures from the missing at random assumption and to identify the point at which departures cause results to "tip" from statistically significant to statistically non-significant. As such, these** ~~identify assumptions about the missing data under which the conclusions from the main analysis change, ie, under which there is no longer evidence of a treatment effect. These~~ **tipping point analyses will only be performed if the primary efficacy analysis results in a statistically significant treatment effect** ~~(p<0.025). Note that each bimekizumab dose of 320mg Q2W and 320mg Q4W will be compared to placebo independently for these analyses. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the p-value in this analysis method will be 0.05 instead of 0.025 throughout for that dose.~~

For tipping point analyses, data for participants ~~**after As a first step**~~**, the intercurrent event date (See Section 3.9) will be changed to missing prior to imputation but will not be changed to non-response after imputation.**

The worst-case scenario will be evaluated first. All missing primary endpoint **values for** ~~(HiSCR$_{50}$ at Week 16). Specifically, all~~ study participants with a missing HiSCR$_{50}$ at Week 16 who have been randomized to bimekizumab (**where missing values include observations after the intercurrent event date and any other missing values) will be imputed as non-responders, while all missing values for placebo-randomized study participants** ~~with a missing HiSCR$_{50}$ at Week 16~~ **will be imputed as responders. While there is little justification for such an approach, it makes the most putative assumption possible against a bimekizumab treatment effect. After applying this imputation approach, a logistic regression model consistent with the one described for the primary analysis will be applied. If the p-value for the odds ratio of bimekizumab versus placebo remains significant** ~~is less than 0.025 for the particular bimekizumab dose regimen~~**, then no further tipping point analyses are needed.**

If this analysis based on the worst-case scenario results in a p-value **that is not significant (eg**, greater than 0.025), then additional tipping point analyses will be **performed to identify**. ~~Several assumptions will be made about average outcomes among~~ **the point at which results switch or "tip" from significant to non-significant. Note that** ~~subsets of study participants who prematurely discontinued study treatment and hence have a monotone missing data pattern (O'Kelly, 2014). In practice, it implies different delta adjustments will be made to the assumed responses on the monotone missing data in~~ **each bimekizumab dose of 320mg Q2W and 320mg Q4W will be compared to placebo independently for these analyses. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the** ~~with various degrees of plausibility with the goal to find for each treatment group the "tipping point" that will significantly reverse the primary result that yielded~~ a **p-value in this analysis method will be 0.05 instead of 0.025 throughout for that dose. In the tipping point analysis, a shift parameter or delta adjustment is applied to missing, and subsequently imported primary endpoint values (where missing values include observations after the intercurrent event and any other missing values).** These delta ~~adjustments will be done on the lesion count and will be~~ **implemented on the primary endpoint as follows**:

10. Data after intercurrent event date (See Section 3.9) will be set to missing.

11. **The same MCMC method described in Section 4.2.2.1 (Step 1a) will be implemented for non-monotone (intermittent) missing pattern values, using the same imputation model. This will be based on 100 sets of imputations**. ~~This will be done only once for each study participant to provide a dataset with monotone missing data.~~

12. **Based on the 100 datasets obtained in Step 2, a monotone regression model will be applied (using the same imputation model as in Step 2) as described in Section 4.2.2.1 (Step 1b).**1) while adjusting the imputed values by various delta adjustments. **This will be based on 1 imputation**.

13. Delta adjustments will be made to imputed lesion count values at Week 16, independently in each treatment group as described below.

14. Delta adjusted imputed values will be truncated so that they are within the range of allowable values for each component.

~~15.~~ **Following the delta adjustments for the** ~~individual components~~ **lesion counts**,~~of~~ the ~~composite endpoint~~ **HiSCR$_{50}$ will then be derived based on the delta-adjusted multiply imputed data sets obtained for each component** ~~endpoint of interest.~~

~~Several scenarios will be considered to define these shift parameters. Once defined, the same shift parameter value will be applied on the imputed endpoint value for all visits. Scenario 1 will assume that study participants randomized to bimekizumab and who have missing data have a lower probability of response compared to study participants randomized to placebo with missing data.~~

- ~~For endpoints for which high scores are associated with a more favorable outcome, it will mean that:~~
  - ~~A negative shift is applied to the imputed value for study participants randomized to bimekizumab to decrease the imputed value.~~

- ~~A positive shift is applied to the imputed value for study participants randomized to placebo to increase the imputed value.~~

~~For endpoints for which high scores are associated with a less favorable outcome, it will mean that:~~

- ~~A positive shift is applied to the imputed value for study participants randomized to bimekizumab to decrease the imputed value.~~
- ~~A negative shift is applied to the imputed value for study participants randomized to placebo to increase the imputed value.~~

~~For each continuous variable, a set of possible values will be first pre-defined for the shift parameter (example: 0, 1, 2, 3).~~

16. Each of the 100 imputed datasets will then be analyzed using a logistic regression model with factors of treatment group, , Baseline Hurley Stage, and Baseline antibiotic use.

17. The results obtained from the 100 logistic regression analyses in Step 7 will be combined for overall inference using Rubin's rules, and the results obtained for each shift parameter will be presented in a single table.

18. Steps 4 to 8 will be repeated so that, at each iteration, missing values are adjusted with a larger delta than at the previous iteration. ~~Depending on the results obtained, shift parameters with more granularity (eg, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9) may also be investigated.~~ The process will go on until the p-value for the odds ratio between bimekizumab and placebo is no longer statistically significant (eg, ≥ 0.025). **The odds ratio, 97.5% CI (or 95% depending on the significance level being used for testing), and p-values obtained for each value of delta will be combined in one single table.**

The delta adjustments result in study participants randomized to bimekizumab with missing data having a lower probability of response compared to study participants randomized to placebo with missing data. Since $HiSCR_{50}$ response is an endpoint for which high lesion counts are associated with a less favorable outcome:

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

To start, imputed values within each values within each lesion type, will be adjusted by the same value in each treatment arm. This adjustment will be 5% of the observed range within that lesion type. Depending on the results obtained, this adjustment will be multiplied for step 9 above (2 times, 3 times the initial adjustment) until the p-value is no longer statistically significant. ~~This can be an adjustment of preselected integer values (eg, 1, 2, and 3) or adjustments at intervals equal to a percentage of the allowable range of the component (eg, 5% of range of 10 to give 0.5, 1, 1.5 etc.). Depending on the results obtained, more granular adjustments (eg, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9) may also be implemented to better~~

~~identify the point at which results "tip". More robust primary analysis results will require larger adjustments to tip the results from significant to insignificant.~~

Additionally, study participants randomized to bimekizumab with an intercurrent event should be set to non-response, after applying the delta adjustment outlined in Step 6 above. This ensures study participants randomized to bimekizumab do not have a higher probability of response in the tipping point analyses compared to the primary analysis (ie, a study participant randomized to bimekizumab who is non-responder in the primary analysis cannot become a responder in the tipping point analyses).

Change #21

New section was added.

Rationale for estimand (Section 4.2.3):

Intercurrent events have been identified within the estimands for this study because their potential to impact efficacy assessments linked with the primary and secondary study objectives. In order to account for the effect of any observed post-randomization intercurrent events on the efficacy analyses, the following estimand strategies will be implemented when evaluating the primary and secondary efficacy endpoints:

- A composite estimand strategy will be used for the primary analysis of the binary secondary endpoints ($HiSCR_{50}$, $HiSCR_{75}$, flare),
- A hypothetical estimand will be used for the primary analysis of the continuous secondary endpoints (change from Baseline in DLQI total score and in “worst pain” item for the HSSDD).

Change #22

New section was added.

Composite estimand (Section 4.2.3.1):

A composite estimand strategy as defined in Section 8.2.2 allows incorporation of the two intercurrent events (eg, receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy) within the definition of the endpoint. These intercurrent events are considered meaningful to the efficacy outcome following receipt of study medication. For example, within the proposed composite estimand framework, a randomized study participant who discontinues from study treatment due to lack of efficacy prior to Week 16 will be considered a treatment failure at Week 16 regardless of the lesion count assessment performed at that visit.

The assumptions and robustness of the primary analysis (modified composite estimand as defined in Section 8.2.2) will be assessed through the sensitivity analyses defined in Section 8.2.3. The impact of intercurrent event handling and data imputation methods on endpoint derivation will also be assessed via the analyses of lesion counts and derived HiSCR variables as specified in Section 8.4.2.1 and Section 8.4.1.1, respectively.

Change #23

New section was added.

Hypothetical estimand (Section 4.2.3.2):

The hypothetical estimand is defined in Section 8.3 and involves a data-driven approach to account for the potential impact of intercurrent events (eg, receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy) on the analysis of continuous efficacy endpoints. Under this framework, outcomes for study participants without an intercurrent event are analyzed as observed. Conversely, outcomes for study participants with an intercurrent event are imputed via a multiple imputation model, ie any recorded data on or after the intercurrent event will be set to missing and imputed via multiple imputation following the strategy established in Section 4.2.2.1.

Change #24

Section 4.2.4 Dates and times

Partial stop and end date imputation rules were updated:

- Imputation of Partial Start Dates
  - If only the month and year are specified:
    - **If** the month and year of first dose **of study medication** is the same as the month and year of the **partial** start date, then **use the date of first dose of study medication,**
    - **Else, if the month and year of the partial start date are the same as the month and year of a study medication switch date, then use the date of study medication switch,**
    - **Otherwise, use the** 1st of the month **of the partial start date;**
  - If only the year is specified:
    - **If the year of first dose of study medication is the same as the year of the partial start date, then use the date of first dose of study medication,**
    - **Else, if the year of the partial date is the same as the year of a study medication switch date, then use the date of study medication switch,**
    - **Otherwise,** use the 1st of January of the year of the **partial** start date;
  - If the start date is completely unknown:
    - If the stop date is unknown or not prior to the date of first dose **of study medication**, then use the date of first dose **of study medication,**
    - If the stop date is prior to the date of first dose of **study medication**, then use the 1st of January of the year of the stop date.
- **Imputation of Partial Stop Dates**
  - **If only the month and year are specified, :**
    - **Use the last day of the month of the partial stop date;**
  - **If only the year is specified**
    - **use December 31st of the year of the partial stop date;**

- **If the stop date is completely unknown,**
  - **Do not impute the stop date.**

Note that if the stop date or the imputed stop date is prior to the imputed start date, then follow the procedure outlined below:

- **If only the year of the start date is specified:**
  - **If the year of start date is the same as the year of first dose of study medication and the imputed stop date is after the date of first dose of study medication, then set the start date to the date of first dose of study medication,**
  - **Otherwise, set the 1st January of the year of the start date;**
- **If only the month and year of start date are specified:**
  - **If the month and year of the start date is the same as the month and year of first dose of study medication and the imputed stop date is on or after the date of first dose of study medication then set the start date to the date of first dose of study medication,**
  - **If the month and year of the start date is the same as the month and year of first dose of study medication and the imputed stop date is before the date of first dose of study medication then set the start date to the 1st of the month of partial start date.**

Change #25

Section 4.6 Use of an efficacy subset of participants

The section was updated to:

A sensitivity analysis of the primary endpoint will be performed based on the FAS, the PPS, and **the CFS**.

Change #26

Section 4.8 Examination of subgroups

This section was updated to:

Subgroup analyses will be performed on the $HiSCR_{50}$, $HiSCR_{75}$, and flare endpoints by visit for the Initial Treatment Period and Maintenance Treatment Period separately**. Additional subgroup analyses will be performed on the change from Baseline in the worst pain score as measured by HSSDD and in the DLQI total score through Week 16 as described below.**

Along with the tables described, there will be tables for $HiSCR_{50}$, $HiSCR_{75}$, and flare which display the response difference and 95% CIs between each bimekizumab dose regimen versus placebo for each of the subgroups at Week 16. Corresponding forest plots will be prepared.

The following **subgroup variables** will be determined using Baseline data**, except for analgesic use, lesion intervention, and antibody positivity:**

- Age (<40 years, 40 to <65 years, ≥65 years)
- Gender (male, female)

- Disease duration (<median, ≥median)

The median disease duration will be calculated based on all participants in the analysis set used for analysis.

- Region (North America [Canada, USA], Western Europe [France, Germany, Ireland, Italy, Spain, United Kingdom], Central/Eastern Europe [Bulgaria, Czech Republic, Hungary, Poland], Asia/Australia [Australia, Israel, Japan])
- Weight (≤100 kg, >100 kg)
- BMI (<25 kg/m^2, 25 to <30 kg/m^2, ≥30 kg/m^2)
- Race (Black or African American, White, All Other Races **[American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other/Mixed])**
- Systemic antibiotic therapy at randomization (yes, no)
- Prior biologic therapy for any indication (yes, no)
- Prior biologic therapy for HS (yes, no)
- Hurley Stage at Baseline (II or III)
- Analgesic users (yes, no), separately for the Initial Treatment Period and the Maintenance Treatment Period (Section 6.4.2 specifies how participants are classified as analgesic users)
- Lesion intervention (yes, no), separately for the Initial Treatment Period and the Maintenance Treatment Period
- Antibody positivity (confirmatory assay: negative or positive)

~~Any analgesic rescue medication taken during the study, lesion, intervention (including new post-Baseline antibiotic use or dose adjustments) and antibody positivity are the only subgroups that are not determined by Baseline data. They will be presented in a separate table.~~

Subgroup analyses will also be performed by visit **The following subgroups for analysison the change from Baseline in the worst pain score as measured byHSSDD and in the DLQI total score** ~~through Week 16. The following subgroups for analysis~~ will be determined based on medication use during the Initial Treatment Period:

- Antihistamines users during the Initial Treatment Period (yes, no) (Section 6.4.2 specifies how participants are classified as antihistamine users)
- Analgesics users during the Initial Treatment Period (yes, no) (Section 6.4.2 specifies how participants are classified as analgesic users)
- Systemic antibiotic therapy start/increase after randomization during the Initial Treatment Period (yes, no)

All summaries will be based on imputed data as appropriate and will include descriptive statistics only.

Change #27

Section 5. 1 Study participant disposition

The following sentence was added:

Participants are defined as completing the Initial Treatment Period if they have a Week 16 visit, or if they fail to attend the Week 16 visit but attend at least one visit in the Maintenance Treatment Period.

The following summaries were also added:

To assess participant disposition (entry and periods in the study) during the COVID-19 pandemic, study participant disposition will also be assessed by period of the COVID-19 pandemic (pre – during – post), by comparing the dates of visits (or events) to the dates of the COVID-19 pandemic period. The dates to categorize the periods of the COVID-19 pandemic (pre/during/post) are defined below:

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

Change #28

The following new section was added:

Impact of COVID-19 (Section 5.2)

A listing of visits affected by COVID-19 will be presented based on the ES including the visit, date of visit, relationship to COVID-19, impact category and a narrative (short description) of the event. These data will be summarized for non-randomized participants and by treatment group and overall, for enrolled participants.

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

In addition, in order to assess the potential impact of COVID-19 on the collection and reporting of efficacy data, a separate summary on the RS will be presented to display missing data as well as data collected via an alternative modality (e.g.: phone, video call) for efficacy endpoints included in the hierarchy (Section 4.5). For these displays, missing data will be presented only for visits affected by COVID-19, as reported on the dedicated eCRF page. Missing data at other visits and for other reasons will not be included. Note that the remote contingencies for COVID-19 or other exceptional circumstances are not applicable to efficacy assessments and documentation (eg, lesion-based assessments, photography) that require direct face-to-face physician/participant interaction.

Change #29

Section 5.3 Protocol deviations

The following text has been added:

A separate summary of participants with protocol deviations related to COVID-19 will be provided.

A by-participant listing of protocol deviations will be provided. Protocol deviations (eg, missing assessments or visits) related to COVID-19 will be listed separately.

Change #30

Section 6.2 Other Baseline characteristics

The last 5 bullets were updated:

- Duration of disease (<median, ≥median)

The median disease duration will be calculated based on all participants in the analysis set used for the summary.

- Baseline antibiotic use (yes, no) **(According to the randomization strata)**
- **Baseline antibiotic use (yes, no) (Derived)**
- Hurley Stage at Baseline **(According to the randomization strata)**
- **Hurley Stage at Baseline (Derived)**

The following text was added or the additional summaries:

In addition, the following Baseline disease characteristics will be summarized by **the derived** Baseline Hurley Stage and by **the derived** Baseline antibiotic use and treatment group for the RS

Change #31

Section 6.4 Prior and concomitant medications

The following sentence was updated:

Prior medications include any medications that started **before** ~~prior to~~ the start date of study medication. Concomitant medications are **any medication that has a start date on or after the start date of study medication, or any medication that has a start date on or before the last dose of study medication + 28 days (whether placebo or bimekizumab).** ~~medications taken at least 1 day in common with dosing period~~.

The following sentence was added:

Additional summaries for the Initial Treatment Period and Maintenance Treatment Period will be presented for participants taking systemic antibiotic medications that qualify as intercurrent events as described in Section 3.9.

Change #32

Section 8 Efficacy Analyses

This section was updated:

All efficacy analyses of primary, ~~and~~ secondary, **and other** variables will be performed on the RS unless otherwise specified. ~~All efficacy analyses of other efficacy variables will be performed on the RS and MS unless otherwise specified.~~ All efficacy summary tables will be displayed by treatment **sequence** unless otherwise specified. The primary and secondary endpoints, and their components, will also be summarized by **the derived** Hurley Stage at

Baseline (grouping each stage and overall) and treatment **sequence** and by **the derived** Baseline antibiotic use (yes/no and overall) and treatment group.

Change #33

Section 8.2 Primary efficacy endpoint

The following rows in Table 8-2 were added or amended:

Table 8-1: Estimand Details and Attributes for Primary Endpoint

		Estimands for Primary Endpoint			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
Primary Objective: To evaluate the efficacy of bimekizumab in study participants with moderate to severe HS					
$HiSCR_{50}$	Sensitivity (Section 8.2.3.1)	$HiSCR_{50}$ response at Week 16	RS	**Composite strategy**, as for the primary analysis **where the occurrence of an intercurrent event will be handles by evaluating the corresponding participants as nonresponders**.	The odds ratio versus placebo based on a logistic regression. Missing values for any other reason will **also** be imputed as nonresponders.

Table 8-1: Estimand Details and Attributes for Primary Endpoint

		Estimands for Primary Endpoint			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
$HiSCR_{50}$	Sensitivity (Section 8.2.3.2)	$HiSCR_{50}$ response at Week 16	RS	**The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a hypothetical strategy, whereby all data at and after the intercurrent event will be treated as missing.** ~~**Composite strategy**, as for the primary analysis.~~**Composite strategy**~~, as for the primary analysis.~~	The odds ratio versus placebo based on a logistic regression. Missing values will be imputed using MI – Reference-Based Regression under a missing not at random assumption.
$HiSCR_{50}$	Sensitivity (Section 8.2.3.3)	$HiSCR_{50}$ response at Week 16	RS	Composite strategy[a], as for the primary analysis.	A tipping point analysis will be used where various delta adjustments will be made to the assumed responses on the monotone missing data in each treatment group independently with various degrees of plausibility. The odds ratio versus placebo is based on a logistic regression for each value of delta.

Table 8-1: Estimand Details and Attributes for Primary Endpoint

		Estimands for Primary Endpoint			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
$HiSCR_{50}$	**Sensitivity (Section 8.2.3.8)**	**$HiSCR_{50}$ response at Week 16**	**CFS**	**Composite strategy, as for the primary analysis.**	**The odds ratio versus placebo is based on a logistic regression, as for the primary analysis.**
$HiSCR_{50}$	**Sensitivity (Section 8.2.3.9)**	**$HiSCR_{50}$ response at Week 16**	**RS**	**The same two intercurrent events used for the primary analysis will be used. Any missing data due to COVID-19 will also be considered an intercurrent event. A composite strategy will be used, ie, the occurrence of an intercurrent event will be handled by evaluating the corresponding participants as treatment failures (nonresponders).**	**The odds ratio versus placebo is based on a logistic regression, as for the primary analysis.**
$HiSCR_{50}$	Sensitivity (Section 8.2.3.10)	$HiSCR_{50}$ response at Week 16	RS	Composite strategy, as for the primary analysis.	The odds ratio versus placebo based on a **stratified** Cochran-Mantel-Haenszel (CMH) test. Missing values **not preceded by an intercurrent event** will be imputed using MI – MCMC/Monotone Regression under a missing at random assumption.

CFS=Covid-19 Free Set; CMH=Cochran-Mantel-Haenszel; **COVID-19=coronavirus disease 2019**; FAS=Full Analysis Set; HiSCR=Hidradenitis Suppurativa Clinical Response; IES=intercurrent event(s) strategy; MCMC=Markov Chain Monte Carlo; MI= multiple imputation; PLS=Population-level summary; Pop=Population; PPS=Per-Protocol Set; RS=Randomized Set

[a] **The composite estimand strategy will be modified in the tipping point analysis such that participants with intercurrent events will be treated as nonresponders only in the bimekizumab treatment groups.**

Change #34

Section 8.2.2 Primary analysis o the primary efficacy endpoint

The following text was updated:

The primary analysis will be based on a logistic regression model including a fixed effect for treatment, Hurley stage at Baseline, and Baseline antibiotic use. The odds ratio versus placebo, p-value (from Wald test), and **97.5% CI will be calculated. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose with a corresponding $Z_{\alpha/2}$ of1.96** .

Change #35

New section was added:

Analysis on CFS (Section 8.2.3.8)

The primary efficacy analyses from Section 8.2.2 will be repeated based on the CFS.

Change #36

New section was added:

Analysis including COVID-19 as intercurrent event (Section 8.2.3.9)

An additional sensitivity analysis will include an additional intercurrent event. The composite estimand combines the clinically meaningful improvement from Baseline based on the $HiSCR_{50}$ response and completion of study treatment through Week 16 without receiving systemic antibiotic rescue medication or discontinuing IMP due to an AE or lack of efficacy.

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

5. Population=Study participants meeting the protocol-specified inclusion/exclusion criteria.
6. Study participant-level outcome=$HiSCR_{50}$ at Week 16.
7. Intercurrent event handling=An intercurrent event is defined as receipt of systemic antibiotic rescue medication, discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16, or missing data due to COVID-19. More information is provided in Section 3.9. A composite strategy will be implemented in which a positive clinical outcome is defined as achieving $HiSCR_{50}$ at Week 16 and not receiving systemic antibiotic rescue medication, not discontinuing study treatment due to an AE or lack of efficacy through Week 16, and not having missing data due to COVID-19. Any missing data at Week 16 that are not preceded by an intercurrent event will be imputed using multiple imputation as defined in Section 4.2.1.
8. Population-level summary measure=conditional odds ratio comparing bimekizumab to placebo adjusted for stratification factors used in the randomization.

The same imputation techniques and analysis model as in the primary efficacy analyses will then be used.

Change #37

Section 8.3 Secondary efficacy endpoints

The following text has been added:

Sensitivity analyses of the secondary endpoints will be performed on the CFS.

Change #38

Section 8.3.2 Flare by Week 16

The following text was deleted:

Disease flare status will be summarized by treatment group and visit using frequencies and percentages. The denominator for the percentage calculations will be the number of participants with non-missing data in each treatment group. A bar chart of percentage of subjects with flare, by visit and treatment, will be presented.

In addition, for each participant, the number of flares during the Initial Treatment Period will be calculated and summarized by treatment group. A corresponding histogram summarizing the number of flares during the Initial Treatment Period will be presented.

Change #39

Section 8.3.3 DLQI Total Score at Week 16

The following paragraph was updated to:

Change from Baseline in DLQI total score will be presented by treatment group. The analysis model will be based on an ANCOVA with fixed effects of treatment, Hurley Stage at Baseline, Baseline antibiotic use and Baseline value as a covariate. The least square mean (LSM), standard error (SE), **and 95% CI for the LSM** will be presented by treatment group. For the comparison between placebo and bimekizumab: the difference between the LSM, the associated **97.5**~~95~~% CI for the contrasts, and the corresponding p-value **will be presented. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose with a corresponding $Z_{\alpha/2}$ of1.96.**

Change #40

Section 8.3.4 Skin Pain score at Week 16, as assessed by the "worst pain" item in the HSSDD

This section was updated:

The items on the HSSDD assess patients' perception of the core symptoms of HS experienced in the past 24 hours: pain, smell or odor, drainage or oozing from HS lesions, and itch on an 11-point numeric rating scale (NRS). Two items assess skin pain: worst skin pain and average skin pain.

Weekly averages will be derived for each of the items of the HSSDD **for weeks matching the post-Baseline dosing weeks** up to Week 16. A weekly average is the sum of the scored item over the course of the study week divided by the number of days on which the item was completed. The weekly averages will be relative to the respective visit date except for Baseline, which will be anchored to the first dose of study drug. A weekly average will only be calculated if at least 4 non-missing values (not necessarily consecutive) are available. Otherwise, the HSSDD weekly average for the given question will be set to missing.

Baseline will be computed as the average from the first 7 consecutive day period in which there are at least 4 non-missing entries. That is, first consider the first 7 consecutive days prior to the Baseline visit, but not including the Baseline visit day itself. If there are at least 4 non-missing values (not necessarily consecutive), then the Baseline average will be calculated. If there are less than 4 values, the 7 consecutive day period will move one day earlier. If there are at least 4 non-missing values (not necessarily consecutive) in that period, then the Baseline average will be calculated. This will continue until there are at least 4 non-missing values in a 7 consecutive day period in the 14 days prior to Baseline. If there is no period in which there are at least 4 non-missing entries, then the Baseline value will be set to missing. ~~Baseline will be computed as the average from the 2 weeks prior to Baseline, up to and including the data from the Baseline visit. If less than 7 non-missing values are available for a given question, the Baseline for the given question will be set to missing~~.

Change from Baseline in worst skin pain score is defined as the average Week 16 worst skin pain score minus the Baseline worst skin pain score. Missing data imputation described in Section 4.2.1.2 will be applied to the weekly averages and not to the individual daily PRO data.

Change from Baseline in worst skin pain score will be presented by treatment group. The analysis model will be based on an ANCOVA with fixed effects of treatment, Hurley Stage at Baseline, Baseline antibiotic use, analgesic use (Section 6.4.2) and Baseline value as a covariate. A treatment-by-analgesic-use interaction term will also be added to the model and removed if not significant.

The LSM, SE, **and 95% CI for the LSM** will be presented by treatment group. For the comparison between placebo and bimekizumab, the difference between the LSM, the associated **97.5** ~~95~~% CI for the contrasts, and the corresponding p-value will be presented. **If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose** with a corresponding $Z_{\alpha/2}$ of1.96.

Change #41

Section 8.4.1.2 Time to response of HiSCR$_{25}$, HiSCR$_{50}$, HiSCR$_{75}$, HiSCR$_{90}$, and HiSCR$_{100}$

The following text was updated:

Initial Treatment Period

Time to HiSCR$_{25}$, HiSCR$_{50}$, HiSCR$_{75}$, HiSCR$_{90}$, and HiSCR$_{100}$ (in days) during the Initial Treatment Period will each be calculated as:

> Min (Date of first HiSCRxx response, Date of Week 16 visit) – Date of **first dose of study medication** ~~Baseline visit~~ + 1, here xx represents 25, 50, 75, 90, 100 respectively. All visits including unscheduled visits are considered.

Participants who discontinue study treatment without achieving a given HiSCR response prior to Week 16 visit will be censored at the date of **last lesion count assessment**.~~discontinuation~~. Participants who reach the Week 16 Visit without achieving the given response will be censored at the date of the Week 16 Visit. Participants who experience an intercurrent event **prior to achieving a HiSCR response** will be censored at the date of the intercurrent event. Participants will be censored at Baseline if there is no Baseline lesion count assessment or no Post-Baseline lesion count assessment.

The following text was deleted:

Combined Initial and Maintenance Treatment Period

An additional time to $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ (in days) during the combined Initial and Maintenance Treatment Period will be calculated as above, where the Week 48 visit is considered instead of Week 16.

Time to $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response during the combined Initial and Maintenance Treatment Period will each be estimated and presented using the Kaplan-Meier product-limit method for each treatment group. ~~These summaries will be limited to participants randomized to bimekizumab.~~

Kaplan-Meier plots of time to HiSCR responses will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 (since there are no responders at Week 0) and will increase over time, representing time to achieving the response.

The median time to response, including the 2-sided 95% confidence interval, will be calculated for each treatment.

~~Maintenance Treatment Period~~

~~For participants randomized to placebo, an additional time to $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ (in days) during the Maintenance Treatment Period will be calculated as:~~

> ~~Min (Date of first HiSCRxx response, Date of Week 48 visit) – Date of Week 16 visit + 1, here xx represents 25, 50, 75, 90, 100 respectively. All visits including unscheduled visits are considered.~~

~~Participants who discontinue study treatment without achieving a given HiSCR response prior to Week 48 visit will be censored at the date of discontinuation. Participants who reach the Week 48 Visit without achieving the given response will be censored at the date of the Week 48 Visit. Participants who experience an intercurrent event will be censored at the date of the intercurrent event. Participants will be censored at Week 16 if there is no Post-Week 16 lesion count assessment.~~

~~Time to $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response during the Maintenance Treatment Period will each be estimated and presented using the Kaplan-Meier product-limit method for the placebo/bimekizumab 320mg Q2W treatment group.~~

~~The median time to response, including the 2-sided 95% confidence interval, will be calculated for each treatment.~~

Change #42

New section was added:

$HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ among Week 16 Responders (Section 8.4.1.4)

See Section 8.4.1.1 for the derivation of HiSCR response.

Summaries of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ responders at each visit from Week 16 through Week 48 will be summarized based on a subset of participants in the MS who achieve response at Week 16. The summaries will be as follows:

- $HiSCR_{50}$ responder rate based on participants who achieved $HiSCR_{50}$ response at Week 16
- $HiSCR_{75}$ responder rate based on participants who achieved $HiSCR_{75}$ response at Week 16
- $HiSCR_{90}$ responder rate based on participants who achieved $HiSCR_{90}$ response at Week 16
- $HiSCR_{100}$ responder rate based on participants who achieved $HiSCR_{100}$ response at Week 16

Line plots of the above HiSCR responder rate categories over time (from Week 16 to Week 48), by treatment group, will be produced.

Change #43

Section 8.4.1.5 Time to loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ in Week 16 responders

The following text has been updated:

See Section 8.4.1.1 for the derivation of HiSCR response.

Time to loss of response will be based on the MS and include only participants who had the corresponding HiSCR response at Week 16 (considering intercurrent event handling from the composite estimand described in Section 8.2.2).

Time to loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ (in days) is defined as: Date of loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ - Date of Week 16 treatment **administration** + 1.

Time to loss of response will be estimated and presented using the Kaplan-Meier product-limit method for each treatment. Participants who experience an intercurrent event **prior to loss of response** will be considered as having lost response on the date of intercurrent event. Participants who reach the Week 48 Visit without loss of response will be censored at the date of the Week 48 Visit. Participants who discontinue treatment or study, for reasons other than those already defined for an intercurrent event, and who have not yet displayed loss of response by the time of withdrawal, will be censored at the date of **the last lesion count assessment** ~~withdrawal~~.

~~The summary for each HiSCR response will include only participants who had the corresponding HiSCR response at Week 16.~~

Change #44

Section 8.4.1.6 Partial Response

The following paragraph was updated:

The number and percentage of participants who are partial responders at Week 16 and become $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ responders in the Maintenance Treatment Period will be summarized by treatment group and visit. These analyses will be based on the subset of participants in the **MS** ~~RS~~ that are partial responders but not $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ responders, respectively, at Week 16. **These summaries will be based on observed case data and will not consider the occurrence of intercurrent events.**

Change #45

New section was added:

Flare relative to Baseline (Section 8.4.3)

See Section 8.3.2 for the derivation of flare.

Disease flare status will be summarized by treatment group and visit using frequencies and percentages. The denominator for the percentage calculations will be the number of participants with non-missing data in each treatment group. This summary will also include the number of participants with any flare in the Initial Period, Maintenance Period, and the combined Initial and Maintenance Period. A bar chart of percentage of participants with flare, by visit and treatment, will be presented.

In addition, for each participant, the number of flares during the Initial Treatment Period will be calculated and summarized by treatment group. A corresponding histogram summarizing the number of flares during the Initial Treatment Period will be presented.

Change #46

The following section was deleted:

Flare by Week 48 (Section 8.4.3)

See Section 8.2.1 for the derivation of AN count.

Disease flare by Week 48 is defined when at least a 25% increase in AN count with an absolute increase of ≥2 AN relative to Week 16 is observed by Week 48. A participant's disease flare status (yes/no) will be determined at each visit in the Maintenance Treatment Period using these criteria and will be listed with the other lesion count assessment data in the data listings.

The number of participants who experience at least 1 disease flare by Week 48 will be summarized by treatment group.

Disease flare status during the Maintenance Period will also be summarized by treatment group and visit using frequencies and percentages. The denominator for the percentage calculations will be the number of participants with non-missing data in the Maintenance Treatment Period in each treatment group. A bar chart of percentage of subjects with flare, by visit and treatment, will be presented.

In addition, for each participant, the number of flares during the Maintenance Treatment Period will be calculated and summarized by treatment group. A corresponding histogram summarizing the number of flares during the Maintenance Treatment Period will be presented.

Change #47

Section 8.4.4 Time to flare by Week 16

This section was updated as follows:

See Section 8.3.2 for the derivation of flare ~~by Week~~.

Time to flare (in days) during the Initial Treatment Period will each be calculated as:

> Min (Date of first flare, Date of Week 16 visit) – Date of **first dose of study medication** ~~Baseline visit~~ + 1. All visits in the Initial Treatment Period including unscheduled visits are considered.

Participants who discontinue study treatment without experiencing a flare prior to Week 16 Visit will be censored at the date of **last lesion count assessment**. ~~discontinuation~~. Participants who reach the Week 16 Visit without experiencing a flare will be censored at the date of the ~~the~~ Week 16 Visit. Participants who experience an intercurrent event **prior to experiencing a flare** will be treated as experiencing a flare at the date of the intercurrent event. Participants will be censored at Baseline if there is no Baseline lesion count assessment or no Post-Baseline lesion count assessment.

Time to flare will be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.

Kaplan-Meier plots of time to flare will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 and will increase over time, representing time to achieving the response.

The median time to **flare** ~~response~~, including the 2-sided 95% confidence interval, will be calculated for each treatment. Comparisons of bimekizumab vs placebo will be analyzed using a log-rank test stratified by Hurley Stage at Baseline and Baseline antibiotic use.

Change #48

Section 8.4.5 Time to flare by Week 48

This section was updated as follows:

See Section 8.3.2 for the derivation of flare **relative to Baseline** ~~by Week 48~~.

~~Maintenance Treatment Period~~

~~Time to flare (in days) during the Maintenance Treatment Period will each be calculated as:~~

~~Min (Date of first flare, Date of Week 48 visit) - Date of Week 16 visit + 1. All visits in the Maintenance Treatment Period including unscheduled visits are considered.~~

~~Participants who discontinue study treatment without experiencing a flare prior to Week 48 visit will be censored at the date of discontinuation. Participants who reach the Week 48 Visit without experiencing a flare will be censored at the date of the Week 48 Visit. Participants who experience an intercurrent event will be treated as experiencing a flare at the date of the intercurrent event. Participants will be censored at Week 16 if there is no Baseline lesion count assessment or no Post-Week 16 lesion count assessment~~.

~~Time to flare will each be estimated and presented using the Kaplan-Meier product-limit method for each treatment group.~~

~~Kaplan-Meier plots of time to flare will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 and will increase over time, representing time to achieving the response.~~

~~The median time to response, including the 2-sided 95% confidence interval, will be calculated for each treatment.~~

~~Combined Initial and Maintenance Treatment Period~~

Time to flare (in days) during the combined Initial and Maintenance Treatment Period will be calculated as:

Min (Date of first flare, Date of Week 48 visit) – Date of **first dose of study medication** ~~Baseline visit~~ + 1. All visits in the up to Week 48 including unscheduled visits are considered.

Flare will be defined relative to the Baseline visit. Participants who discontinue study treatment without experiencing a flare prior to Week 48 visit will be censored at the date of **last lesion count assessment.** ~~discontinuation.~~ Participants who reach the Week 48 Visit without experiencing a flare will be censored at the date of the Week 48 Visit. Participants who experience an intercurrent event **prior to experiencing a flare** will be treated as experiencing a flare at the date of the intercurrent event. Participants will be censored at Baseline if there is no Baseline lesion count assessment or no Post-Baseline lesion count assessment.

Time to flare will each be estimated and presented using the Kaplan-Meier product-limit method for each treatment group. ~~This summary will be limited to participants randomized to bimekizumab.~~

~~Kaplan-Meier plots of time to flare will be presented by treatment group. In these Kaplan-Meier plots, the line will start at 0 and will increase over time, representing time to achieving the response.~~

The median time to **flare** ~~response~~, including the 2-sided 95% confidence interval, will be calculated for each treatment.

Change #49

Section 8.4.6 International Hidradenitis Suppurativa Severity Score System (IHS4)

The following sentence was added:

The observed IHS4 score, change and percentage change from Baseline will be summarized by treatment group and visit. **Missing IHS4 scores will be imputed using the multiple imputation procedure specified in Section 4.2.2.1, where IHS4 scores will be derived based on the imputed lesion counts.**

Change #50

Section 8.4.10 Time to initiation of systemic rescue therapy in the Initial Treatment Period

The following text was updated:

See Section 3.9 for the definition of a systemic antibiotic rescue therapy.

Time to initiation of systemic rescue therapy (in days) during the Initial Treatment Period will be calculated as:

Min (Date of initiation of rescue therapy, Date of change in the dose/type of current antibiotic, Date of Week 16 visit) – Date of **first dose of study medication** ~~Baseline visit~~ + 1.

Participants who discontinue **the** study ~~treatment~~ without initiating systemic rescue therapy prior to Week 16 visit will be censored at the date of discontinuation. Participants who reach the Week 16 Visit without initiating systemic rescue therapy will be censored at the date of the Week 16 Visit. Participants will be censored at Baseline if there is no ~~Baseline lesion count assessment or no~~ Post-Baseline **visit** ~~lesion count assessment~~.

Change #51

Section 8.4.11 Time to an intercurrent event in the Initial Treatment Period

The following text was updated:

See Section 3.9 for the definition of an intercurrent event.

Time to an intercurrent event (in days) during the Initial Treatment Period will be calculated as:

> Min (Date of **intercurrent event** ~~initiation of rescue therapy, Date of change in the dose/type of current antibiotic, Date of withdrawal due to AE or lack of efficacy~~, Date of Week 16 visit) – Date of **first dose of study medication** ~~Baseline visit~~ + 1.

Participants who discontinue **the** study ~~treatment~~ without experiencing an intercurrent event prior to Week 16 visit will be censored at the date of discontinuation. **That includes participants who discontinue from the study for reasons other than Adverse Event and Lack of Efficacy.** Participants who reach the Week 16 Visit without experiencing an intercurrent event will be censored at the date of the Week 16 Visit. Participants will be censored at Baseline if there is no ~~Baseline lesion count assessment or no~~ Post-Baseline **visit** ~~lesion count assessment~~.

Change #51

Section 8.4.12 Hidradenitis Suppurativa Symptom Daily Diary (HSSDD)

The following sentence was updated:

Change from Baseline in Worst Pain **score** and Worst Itch **score** ~~at Week 16~~ will additionally be summarized **by visit and** by analgesic and antihistamine use status (Section 6.4.2), respectively.

Change #53

Section 8.4.13 Hidradenitis Suppurativa Symptom Questionnaire (HSSQ)

The following text was updated:

HSSQ response for pain **item** ~~score~~ is defined as at least a 30% reduction and at least a 1-unit reduction from Baseline in HSSQ among study participants with a score of ≥3 at Baseline.

The number and percentage of responders for pain **item** ~~score~~ will be summarized by treatment group and visit **based on the MS**.

The number and percentage of participants who were responders at any timepoint in the Maintenance Treatment Period will be summarized by treatment group for the skin pain score **based on the MS.**

Change from Baseline in pain score and itch score will additionally be summarized by visit and by analgesic and antihistamine use status (Section 6.4.2), respectively.

Change #54

Section 8.4.15 Hidradenitis Suppurativa Quality of Life (HiSQOL)

The following text was updated:

Summary statistics of the actual values and change from Baseline values will be used to summarize HiSQOL **domain and total scores** for each visit by treatment group. The table will

display descriptive statistics for the Baseline, followed by descriptive statistics for the change from Baseline for all visits.

Missing data for the continuous change from Baseline will be handled by using MI via the MCMC and monotone regression method specified in Section 4.2.1.3. **The imputed HiSQOL total score will be derived based on the imputed subscales.**

The number and percentage of participants that complete the HiSQOL will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS (or MS, as appropriate). The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit.

A by-participant listing of the HiSQOL questionnaire, HiSQOL responses, **domain and total scores** and change from Baseline will be provided.

Change #55

Section 9.1 Pharmacokinetics

The following sentence was updated:

All ~~However, all~~ PK concentrations **collected will be listed irrespective of the dosing or sampling occurring out of window.**

Change #56

Section 8.2.3.1 Derivation of palmoplantar IGA response

Anti-bimekizumab antibodies (ADAb) will be measured using a **3**-tiered assay approach: screening assay, confirmatory assay, and titration assay. Samples confirmed as positive within the confirmatory assay will be further evaluated **for the presence of** in a **neutralizing anti**-assay ~~to evaluate the potential of the ADAb to neutralize the activity of~~ **bimekizumab antibodies specific to IL-17AA, IL-17FF**(~~IL17A or IL17F~~, **or both**) ~~in vitro~~. Samples will be taken at Baseline, then at study Weeks 4, 8, 12, 16, 20, 24, 36 and 48, and at PEOT and SFU timepoints.

The screening ~~Screening, confirmatory, and titer~~ cut **point will be used to determine** ~~points of~~ the **status of anti-bimekizumab antibodies in** ~~respective assays will be determined by~~ **the test sample as Positive Screen (PS) or Negative Screen (NS). For** ~~bioanalytical laboratory based either on commercially available drug naïve~~ samples **presenting anti-bimekizumab antibody levels that are PS, a further confirmatory assay will be performed, and the result of which will be reported as either Positive Immunodepletion (PI) or Negative Immunodepletion (NI).**

ADAb status for each sample will be derived as follows:

- **Sample values that are either NS, or PS and NI and where the bimekizumab concentration is less than the validated ADAb assay drug tolerance limit will be defined as anti-bimekizumab antibody negative.**
- **Sample values that are either NS, or PS and NI and where the bimekizumab concentration exceeds the validated ADAb assay drug tolerance limit will be defined as inconclusive.**

- **Sample values that are PS and PI will be defined as ADAb positive (regardless of availability of a titer value)**
- **Missing or non-evaluable samples will be defined as missing**

Positive immunodepletion samples will be titrated, and the ADAb titer (reciprocal dilution factor including minimum required dilution) will be reported. Subsequently, PI samples will also be subject to a neutralizing assay to evaluate the potential of ADAb to neutralize the target binding of bimekizumab (IL-17AA or IL-17FFIL17F or both) in **vitro.**

~~The following definitions will be applied regarding ADAb status of each test samples:~~

- ~~An ADAb status will be confirmed as positive for any sample with an ADAb level that is positive screen and positive immunodepletion.~~
- ~~An ADAb status of negative will be concluded for any sample with an ADAb level that is either negative screen or (positive screen and negative immunodepletion).~~

~~If the titer for an ADAb level that is positive screen and positive immunodepletion is missing, then a conservative approach will be used and ADAb status will be consider as positive. No imputation rules apply for the missing titer. If the ADAb level is positive screen but no confirmatory result could be determined, then a conservative approach will be used and ADAb status will be consider as positive.~~

~~Anomalous values will be not included in summaries/analysis and will be reviewed and flagged. by the Clinical Pharmacologist~~.

There are different levels of classification for ADAb status, the definitions are as follows:

~~For each participant an overall ADAb status will be derived:~~

- ~~Overall Positive is defined as having at least one value that is confirmed positive during the treatment period.~~
- ~~Overall Negative is defined as having no values that are confirmed positive at any time in the treatment period.~~

~~The treatment period does not include Baseline/pre-treatment samples or SFU.~~

~~Furthermore, the following subcategories for each subject will be derived:~~

- **Pre ADAb negative – treatment-emergent ADAb negative (Category 1): Includes study participants who are anti-bimekizumab antibody negative at Baseline and anti-bimekizumab antibody negative at all sampling points during the period of interest (one post-Baseline missing/inconclusive sample is allowed for subjects with pre- anti-bimekizumab antibody negative sample). This group also includes study participants who have a missing or inconclusive sample (either missing or inconclusive or insufficient volume) at Baseline (ie, pre-treatment) with all post-Baseline samples as ADAb negative.**
- **Pre ADAb negative – treatment-emergent ADAb positive (Category 2): Includes study participants who are ADAb negative at Baseline and ADAb positive at any sampling points post-Baseline during the period of interest. This group also includes study**

participants who have a missing sample (either missing or insufficient volume) at Baseline (ie, pre-treatment) with 1 or more post-Baseline samples as ADAb positive.

- **Pre ADAb positive – treatment-emergent reduced ADAb (Category 3): Includes study participants who are ADAb positive at Baseline, and ADAb negative at all sampling points post-Baseline during the period of interest.**
- **Pre ADAb positive – treatment-emergent unaffected ADAb positive (Category 4): Includes study participants who are ADAb positive at Baseline and are ADAb positive at any sampling point post-Baseline during the period of interest with titer values of the same magnitude as Baseline (ie, less than a predefined fold difference from the Baseline titer).**
- **For this analysis, this is set at an increase of less than the validated Minimum Significant Ratio (MSR) of 2.07-fold from Baseline.**
- **Pre ADAb positive – treatment-emergent ADAb boosted positive (Category 5): Includes study participants who ADAb positive at Baseline and are ADAb positive at any sampling point post-Baseline during the period of interest with increased titer values compared to Baseline (equal to or greater than a predefined fold difference increase from Baseline titer which will be defined within the validation of the assay).**
  - **For this analysis, this is set at an increase equal to or greater than the validated MSR of 2.07-fold from Baseline.**
  - **Note: for any study participant who is ADAb positive at Baseline and ADAb positive at a post-Baseline time point during the period of interest, but for whom titers are not available to determine treatment unaffected or treatment boosted status, the study participant will be considered as treatment boosted, assuming no other samples are available.**
- **ADAb Inconclusive (Category 6): Includes study participants who have an ADAb positive Baseline (pre-treatment) sample and some post-Baseline samples during the period of interest are missing or inconclusive, while other post-Baseline samples are ADAb negative.**
- **Total treatment-emergent ADAb positivity (Category 7 [Categories 2 and 5 combined]): Includes study participants who are pre ADAb negative – treatment-emergent ADAb positive (Category 2) and pre ADAb positive – treatment boosted ADAb positive (Category 5).**
- **Total prevalence of pre- ADAb positivity (Category 8 [Categories 3, 4, 5 and 6 combined]): Study participants that are tested ADAb positive at Baseline.**
- **Missing: Includes study participants who are ADAb negative, missing, or inconclusive at Baseline with some post-Baseline samples as missing or inconclusive, and other samples as ADAb negative.**

~~Derivation for above classification will be different for the interim analysis and the final analysis. SFU data will be considered only for the final data analysis. That is, each instance of "excluding SFU" in the categories above, should be changed to "including SFU."~~

In the case that a sample is collected 1 or more days following the scheduled visit date in which the drug was administered, the **ADAb** ~~anti-bimekizumab antibody~~ results for that sample will be associated with the scheduled visit and summarized accordingly. Such samples will also be considered when **ADAb** ~~anti-bimekizumab antibody~~ results are summarized over a given study period.

Analysis

Immunogenicity will be assessed through summary tables and figures, and listing of individual results by participants. All analyses will be run on the **AMS** ~~Active Medication Set~~, unless specified otherwise.

- Summary of **ADAb** ~~anti-bimekizumab antibody~~ status overall and by each visit separated by treatment group
- Summary of the time-point of the first occurrence of **ADAb** ~~anti-bimekizumab antibody~~ positivity during the treatment period by treatment group. A plot of the titer by time to first **ADAb** ~~anti-bimekizumab antibody~~ positivity will be prepared.
- All individual participant-level **ADAb** ~~anti-bimekizumab antibody~~ results will be listed.
- The number and percentage of participants in each of the **8 ADAb** ~~anti-bimekizumab antibody~~ categories during the treatment period by treatment group~~., with an additional category combining participants in categories 2 and 5, summarized as total treatment emergent. In addition, the count and percentage of participants who are pre anti-bimekizumab positive will be calculated (this is the sum of categories 3, 4, and 5).~~
- The prevalence of immunogenicity, separated by treatment group, and defined subcategory, will be reported by visit, defined as (cumulative) proportion of participants having confirmed positive **ADAb** ~~anti-bimekizumab antibody~~ samples at any visit up to and including that visit. Missing samples will not be included in the denominator.
- The time to achieving treatment-emergent **ADAb** ~~anti-bimekizumab antibody~~ positivity, separated by treatment group and **defined subcategory** ~~sub-categories 2 & 5 above~~, will be analyzed based on Kaplan-Meier methods. Participants will be considered to have an event at the time point at which treatment emergent **ADAb** ~~anti-bimekizumab antibody~~ positive is first achieved (taking the MSR into consideration for sub-category 5). Participants classified as treatment-emergent **ADAb** ~~anti-bimekizumab antibody~~ negative will be censored at the time of the last available **ADAb** ~~anti-bimekizumab antibody~~ result.
- A summary of $HiSCR_{50}$ responders at Week 16, separated by treatment group, as a function of ADAb titer will be presented graphically. This will be repeated for $HiSCR_{75}$ responders.
- Individual plots of plasma bimekizumab concentrations/ **ADAb** ~~anti-bimekizumab antibody~~ titer both plotted on the Y-axes by visit (x-axis) for the full treatment period (excluding SFU for interim analyses and including SFU for final analyses) will be presented for participants with and without $HiSCR_{50}$ response at Week 16.
- Spaghetti plots of ADAb titer (y-axis) by visit (x-axis), separated by treatment group for all **ADAb** ~~anti-bimekizumab antibody~~ positive participants, including Baseline positive participants.

- Box plots of ADAb titer (logscale) by time to first ADAb positivity by treatment group.

For purposes of efficacy subgroup analyses based on anti-bimekizumab antibody status, 2 categories will be used:

- **ADAb** ~~Anti-bimekizumab antibody~~ positive – This is defined as participants who have **ADAb** ~~anti-bimekizumab antibody~~ levels above the specified cut point on at least 2 time points while on treatment (ie, excluding Baseline, excluding SFU).
- **ADAb** ~~Anti-bimekizumab antibody~~ negative – Participants who are not defined as anti-bimekizumab positive (as described above) will be defined as **ADAb** anti-bimekizumab antibody negative.

The groups for defining **ADAb** anti-bimekizumab antibody status for safety subgroup analyses are as follows:

- AEs starting before first **ADAb** ~~anti-bimekizumab antibody~~ positive result
- AEs starting on or after first **ADAb** ~~anti-bimekizumab antibody~~ positive result
- AEs for participants who were always **ADAb** ~~anti-bimekizumab antibody~~ negative

Change #57

Section 10.1.1 Exposure during the Initial Treatment Period

This section was split into 2 subsection 10.1.1.1 and 10.1.1.2 for exposure duration (days) and time at risk (days), respectively.

Change #58

Section 10.1.1.1 Study medication duration (days)

The following text was updated:

Definitions for study medication duration (days) are provided as follows:

- Date of last dose in the Initial Treatment Period – Date of first dose in the Initial Period + 14 days.

Note: The use of 14 days assumes a Q2W dosing interval (bimekizumab 320mg Q2W and placebo). For participants randomized to bimekizumab 320mg Q4W, this will be adjusted based on the dosing interval (eg, use Date of last bimekizumab dose in the Initial Treatment Period – Date of first dose in the Initial Period + 28 days).

Note: If date of last dose in the Initial Treatment Period + 14 days (or ~~date of last bimekizumab dose in the Initial Treatment Period~~ + 28 days in the case of Q4W dosing) extends to a date beyond the date of first dose in the Maintenance Treatment Period, then this calculation reverts to:

  - Date of first dose in the Maintenance Treatment Period – Date of first dose in the Initial **Treatmen**t Period + 1.

Change #59

Section 10.1.1.2 Time at risk (days)

The section was updated:

Definitions for time at risk (days) are provided as follows:

~~*For participants s who permanently discontinue study treatment:*~~

- ~~Date of last dose - date of first dose + 14 days~~

~~The use of 14 days assumes a Q2W dosing interval. For participants randomized to bimekizumab 320mg Q4W, this will be adjusted based on the dosing interval (eg, use Date of last bimekizumab dose in the Initial Treatment Period – date of first dose in the Initial Period + 28 days).~~

~~Note: If date of last dose + 14 days (or date of last dose of bimekizumab + 28 days for Q4W dosing) extends to a date beyond the final visit date (including PEOT, but not including SFU), then this calculation reverts to:~~

- ~~Final visit date (including PEOT, but not including SFU) - date of first dose + 1.~~

~~Time at risk (days)~~

- For participants who complete the Week 16 visit and continue to the Maintenance Treatment Period:
  - Date of first dose in the Maintenance Treatment Period – Date of first dose in the Initial Period + 1.
- For participants who discontinue on or prior to the final visit of the Initial Period, use the minimum of the following:
  - **Date of last dose in the Initial Treatment Period – Date of first dose in the Initial Treatment Period + 141**
  - The total number of days in the Initial Treatment Period (112 days). For AEs that emerged after 112 days but still within the 140 days window, those AEs would be classified as TEAE, but will be excluded from the output based on the Initial Treatment Period. However, these AEs will be included in the AE summaries for Maintenance Treatment Period.
  - Date of last clinical contact – Date of first dose in the Initial Treatment Period + 1.
- For participants who die prior to the final visit of the Initial Treatment Period: Date of death – date of first dose in the Initial Period +1.

Change #60

Section 10.1.2 Exposure during the Maintenance Treatment Period

This section was split into 2 subsection 10.1.2.1 and 10.1.2.2 for exposure duration (days) and time at risk (days), respectively.

Change #61

Section 10.1.2.1 Study medication duration (days)

The section was updated:

Definitions for study medication duration (days) are provided as follows:

- Date of last dose in the Maintenance Treatment Period – date of first dose in the Maintenance Treatment Period + 14 days.

The use of 14 days assumes a Q2W dosing interval. For participants randomized to bimekizumab 320mg Q4W in the Maintenance Treatment Period, this will be adjusted based on the dosing interval (eg, use Date of last bimekizumab dose in the Maintenance Treatment Period – date of first dose in the Maintenance Treatment Period + 28 days).

Note: If date of last dose in the Maintenance Treatment Period + 14 days (or ~~date of last bimekizumab dose in the Maintenance Treatment Period~~ + 28 days in the case of Q4W dosing) extends to a date beyond the final visit date of the Maintenance Treatment Period (not including SFU), then this calculation reverts to:

  - Final visit date of the Maintenance Treatment Period (not including SFU) – date of first dose in the Maintenance Treatment Period + 1.

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

  - Date of death – Date of first dose in the Maintenance Treatment Period + 1.

~~*For participants who permanently discontinue study treatment:*~~

~~- Date of last dose - date of first dose + 14 days~~

~~The use of 14 days assumes a Q2W dosing interval. For participants randomized to bimekizumab 320mg Q4W in the Maintenance Treatment Period, this will be adjusted based on the dosing interval (eg, use Date of last bimekizumab dose in the Maintenance Treatment Period - date of first dose in the Maintenance Treatment Period + 28 days).~~

~~Note: If date of last dose + 14 days (or date of last dose of bimekizumab + 28 days for Q4W dosing) days extends to a date beyond the final visit date (including PEOT, but not including SFU), then this calculation reverts to:~~

~~- Final visit date (including PEOT, but not including SFU) - date of first dose + 1.~~

Change #62

Section 10.1.2.2 Time at risk (days)

The text was updated:

Definitions for time at risk (days) are provided as follows:

- For participants who complete the Maintenance Treatment Period as planned and continue into an extension study (and, therefore, do not have the SFU visit in the feeder study): ~~Final visit date of the Maintenance Treatment Period - date of first dose in the Maintenance Treatment Period + 1.~~

  - **Date of last visit of the Maintenance Treatment Period – Date of first dose in the Maintenance Treatment Period + 1.**

- For participants who die prior to the final visit of the Maintenance Treatment Period: ~~Date of death - date of first dose in the Maintenance Period + 1.~~

  - **Date of death – Date of first dose in the Maintenance Period + 1.**

- For all other participants, use the minimum of the following:
    - Date of last dose in the Maintenance Treatment Period – Date of first dose in the Maintenance Treatment Period + **141**~~140~~ days.

Change #63

Section 10.1.3 Exposure during the Initial and Maintenance Treatment Period

This section was split into 2 subsection 10.1.3.1 and 10.1.3.2 for exposure duration (days) and time at risk (days), respectively.

Change #64

Section 10.1.3.1 Study medication duration (days)

The section was updated:

Definitions for study medication duration (days) are provided as follows:

- **Sum of study medication durations from the individual Initial and Maintenance Treatment Period.**

Note: The algorithms for calculating these durations are specified in Section 10.1.1.1 and Section 10.1.2.1.

Note: If date of last dose in the Initial Treatment Period + 14 days (or + 28 days in the case of Q4W dosing) extends to a date beyond the date of first dose in the Maintenance Treatment Period, then this calculation reverts to:

- **Sum of study medication durations from the individual Initial and Maintenance Treatment Periods - 1.**

~~*For participants who do not switch study treatments:*~~

- ~~Date of last dose – Date of first dose + 14 days.~~

~~The use of 14 days assumes a Q2W dosing interval. For participants randomized to bimekizumab 320mg Q4W, this will be adjusted based on the dosing interval (eg, use Date of last bimekizumab dose – date of first dose + 28 days).~~

~~Note: If date of last dose + 14 days (or date of last bimekizumab dose in the Maintenance Treatment Period + 28 days in the case of Q4W dosing) extends to a date beyond the final visit date (including PEOT, not including SFU), then this calculation reverts to:~~

- ~~Final visit date (including PEOT, not including SFU) – Date of first dose + 1.~~

- ~~For participants who die, if date of last dose + 14 days (or + 28 days in the case of Q4W dosing) extends to a date beyond the date of death, then this calculation reverts to:~~
    - ~~Date of death – Date of first dose + 1.~~

~~*For participants who switch study treatments (between Initial and Maintenance Treatment Periods):*~~

- ~~Initial Treatment Period (attributed to initially randomized treatment):~~
    - ~~Date of last dose in the Initial Period – Date of first dose in the Initial Period + 14 days.~~

~~Note: Participants who switch study treatments are on a Q2W dosing schedule for the Initial Treatment Period.~~

~~Note: If date of last dose in the Initial Treatment Period + 14 days extends to a date beyond the date of first dose in the Maintenance Treatment Period, then this calculation reverts to:~~

- ~~Date of first dose in the Maintenance Treatment Period – Date of first dose in the Initial Period + 1.~~

- ~~Maintenance Treatment Period (attributed to the treatment initiated in the Maintenance Treatment Period):~~
  - ~~Use the study medication duration algorithm specified for the Maintenance Treatment Period in Section 10.1.2.1.~~

Change #65

Section 10.1.3.2 Time at risk (days)

This section was updated:

~~*For participants who do not switch study treatments:*~~ Definitions for time at risk (days) are provided as follows:

- For participants who complete the Maintenance Treatment Period as planned and continue into an extension study (and, therefore, do not have the SFU visit in the feeder study):
  - Final visit date – Date of first dose + 1.
- For participants who die prior to the final visit:
  - Date of death – Date of first dose in the + 1.
- For all other participants, use the minimum of the following:
  - Date of last dose – Date of first dose + 141 days.
  - Date of last clinical contact – Date of first dose + 1.

  Note: This group could include participants who discontinue early, participants who complete the Maintenance Treatment Period as scheduled but choose not to continue into an extension study, or participants who are ongoing in the SFU period at the time of the data snapshot (in the case of the interim analysis).

~~*For participants who switch study treatments (between Initial and Maintenance Treatment Periods):*~~

- ~~Initial Treatment Period (attributed to initially randomized treatment):~~
  - ~~Date of first dose in the Maintenance Treatment Period – Date of first dose in the Initial Period + 1.~~

    ~~(Note: This assumes that anyone in this category has completed the Initial Treatment Period and doses [with a new study treatment] in the Maintenance Treatment Period.)~~

- ~~Maintenance Treatment Period (attributed to the treatment initiated in the Maintenance Treatment Period):~~

~~– Use the time at risk algorithm specified for the Maintenance Treatment Period in Section 10.1.2.2.~~

Change #66

Section 10.2.1 Data considerations

The following sentence was added:

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. **Note that if the seriousness of an adverse event is unknown, every attempt should be made to resolve this prior to a snapshot for an interim analysis or database lock; in the exceptional case that the seriousness of an adverse event is still missing then no imputation should be applied for this characteristic.**

Change #67

Section 10.2.2 AE summaries

The following text was deleted:

The following summaries will be provided by treatment group for the Initial Treatment Period, ~~Maintenance Treatment Period,~~ and the Initial and Maintenance Treatment Period combined based on the SS, ~~MS,~~ and AMS respectively.

The following summaries were added:

- Incidence of TEAEs – Suspected and Confirmed COVID-19 cases by SOC, HLT and PT

Suspected and confirmed COVID-19 cases will be identified with the preferred terms “Corona virus infection” or “Corona virus test positive”.

The following subset of tables will also be presented for the Maintenance Treatment Period using the MS:

- Incidence of TEAEs – Overview
- Incidence of TEAEs per 100 subject years by SOC, HLT, and PT
- Incidence of Serious TEAEs per 100 subject years by SOC, HLT, and PT
- Incidence of TEAEs Leading to Discontinuation per 100 subject years by SOC, HLT, and PT
- Incidence of TEAEs Leading to Death by SOC, HLT, and PT

Change #68

Section 10.2.3.1 Infections (serious, opportunistic, fungal and TB)

The following text was updated:

- **Incidence of Fungal Infection TEAEs per 100 subject years by SOC, HLT and PT**

  Fungal infections will be summarized in a stand-alone table. The table will include all TEAEs (serious and non-serious) that code into the **High Level Group Term (HLGT)** ~~HLT~~ “Fungal infectious disorders”

- **Incidence of Opportunistic Infection TEAEs per 100 subject years by SOC, HLT and PT**

Opportunistic infections (including tuberculosis) will be summarized in a stand-alone table. The table will include all TEAEs identified using UCB-defined search criteria.~~(refer to Excel spreadsheet on "OI - MedDRA v19.0.xlsx" in "Bimekizumab-Safety-Topics-of-Interest.docx").~~

Change #69

Section 10.2.3.3 Major adverse cardiac event

The following sentence was added:

A separate table and listing will present the adjudicated cardiovascular events by type. For each cardiovascular event type (24 total), the individual PTs that fall within each event type will be summarized. **The other 10 MACE events not listed in the table are described in the adjudication committee charter.**

Change #70

Section 10.2.3.5 Suicidal ideation and behaviors

The following paragraph was added:

A separate table will present the adjudicated neuropsychiatric events by type. For each neuropsychiatric event type (6 total), the individual PTs which fall within each event type will be summarized. It will include events adjudicated as SIB and events adjudicated as non-suicidal. Note that the event type Suicidal ideation can be classified as either SIB or non-suicidal.

Change #71

Section 10.2.3.6 Inflammatory bowel disease

This section was updated:

An external inflammatory bowel disease (IBD) adjudication committee will evaluate potential IBD events and will classify each one as follows:

- Event Type Code 1: Possible IBD – Crohn's Disease
- Event Type Code 2: Probable IBD – Crohn's Disease
- Event Type Code 3: Definite IBD – Crohn's Disease
- Event Type Code 4: Possible IBD – Ulcerative Colitis
- Event Type Code 5: Probable IBD – Ulcerative Colitis
- Event Type Code 6: Definite IBD – Ulcerative Colitis
- Event Type Code 7: Possible IBD – Unclassified
- Event Type Code 8: Probable IBD – Unclassified
- Event Type Code 9: Definite IBD – Unclassified
- Event Type Code 10: Symptoms not consistent with IBD
- **Event Type Code 11: Possible Inflammatory Bowel Disease – Microscopic Colitis**
- **Event Type Code 12: Probable Inflammatory Bowel Disease – Microscopic Colitis**

- **Event Type Code 13: Definite Inflammatory Bowel Disease – Microscopic Colitis**
- **Event Type Code 14: Possible Inflammatory Bowel Disease – no further differentiation possible**
- **Event Type Code 15: Probable Inflammatory Bowel Disease – no further differentiation possible**
- **Event Type Code 16: Definite Inflammatory Bowel Disease – no further differentiation possible**
- Event Type Code 99: Not enough information to adjudicate

A table for adjudicated ~~definite~~ IBD events (event type codes **1, 2,** 3, **4, 5,** 6, **7, 8, 9, 11, 12, 13, 14, 15** and **16**~~9~~) as determined by the adjudication committee will be produced. **It will summarize events determined by the adjudication committee as definite IBD (event type codes 3, 6, 9, 13, and 16), probable IBD (event type codes 2, 5, 8, 12, and 15) and possible IBD (event type codes 1, 4, 7, 11, and 14). Definite and probable IBD will also be aggregated and summarized.** This table will be ~~produced overall, as well as~~ stratified by subjects with or without a previous medical history of IBD. Previous medical history of IBD will be determined using the information recorded on the **History of IBD** ~~Extra-Articular Assessment at Screening~~ CRF page ("Does subject have a history of IBD?").

A **separate** table **will present the** ~~for~~ adjudicated gastrointestinal ~~probable IBD~~ **events by type. For each gastrointestinal** event type **(17 total), the individual PTs which fall within each event type will be summarized. It will include events** ~~codes 2, 5, and 8) as~~ **determined by the adjudication committee as definite IBD probable IBD and** ~~will be produced. This table will be produced overall, as well as stratified by participants with or without a previous medical history of IBD.~~

~~A table for adjudicated~~ **possible IBD. It** ~~events (event type codes 1, 4, and 7) as determined by the adjudication committee will be produced. This table~~ **will also include events determined as Symptoms not consistent with** ~~be produced overall, as well as stratified by participants with or without a previous medical history of~~ **IBD (event type code 10) and Not enough information to adjudicated (event type code 99).**

A listing of all events identified for potential review by the IBD adjudication committee will be produced. This listing will indicate whether each event was escalated to the committee for formal review/adjudication.

A separate ~~table and~~ **listing** will present the adjudicated IBD events by type. For each IBD event type (event type codes 1 through **16**~~10~~ and 99; **17**~~11~~ total), the individual PTs which fall within each event type will be **listed**~~summarized~~.

A third listing will present the individual diagnostic criteria met for each adjudicated IBD event.

Change #72

Section 10.2.3.7 Hypersensitivity (including anaphylaxis)

The following text was updated:

A separate table will be prepared to summarize hypersensitivity events, identified using the SMQ "Hypersensitivity (SMQ)". All TEAEs which code to a PT included in the Scope=Narrow search will be included in this table. **In addition, a separate table will be prepared to summarize serious hypersensitivity events, identified using the SMQ "Hypersensitivity (SMQ)". All serious TEAEs which code to a PT included in the Scope=Narrow search will be included in this table**. An AE glossary table will also be produced to summarize the MedDRA coding for these events. The glossary table will include the following fields: reported term, PT, LLT, HLT, and SOC.

Furthermore, **a separate table will be prepared to summarized injection site reactions, identified using the** HLTs: "Administration site reactions NEC" and "Injection site reactions".

Change #73

Section 10.3 Clinical laboratory evaluations

The following text was added:

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 **(values observed more than 140 days after the last administration of study medication are not considered).** All summaries will be presented in SI units and will be based on observed case values.

CTCAE grading was updated:

Markedly abnormal values are defined as those with a severity of Grade 3 and above based on the CTCAE criteria **Version 4.03**.~~(U.S. Department of Health and Human Services 2017).~~

And Table 10-2 was updated:

Table 10–2: Definitions of Markedly Abnormal Biochemistry Values

Parameter name	Conventional		Standard		Abnormal Designation
	Unit	Criteria	Unit	Criteria	
Creatinine[1]~~Creatinine~~	mg/dL	> 3.0 x **Baseline or >3.0 x** ULN	**umol**~~mmo l~~/L	> 3.0 x **Baseline or >3.0 x** ULN	AH
Glucose	mg/dL	<40 >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

Table 10–2: Definitions of Markedly Abnormal Biochemistry Values

Parameter name	Conventional		Standard		Abnormal Designation
	Unit	Criteria	Unit	Criteria	
Sodium	mmol/L	>155 <130	mmol/L	>155 <130	AH AL
Cholesterol	**mg/dL**	**>400**	**mmol/L**	**>10.34**	**AH**

1 The markedly abnormal definitions for creatinine are based on the logical or, if either criterion is met the creatinine value will be designated as abnormal high.

And the following text was added:

- **Total** Bilirubin: >1.5xULN, >2xULN
- **ALP: >1.5xULN**

For any participant with at least one markedly abnormal LFT (AST >3xULN, ALT >3xULN, bilirubin >3xULN, or ALP >1.5xULN) the New Ratio (nR) will be calculated as the ratio of either ALT or AST (whichever is higher) to ALP, all expressed as multiples of their ULN as follows:

- **nR = [maximum(AST/ULN or ALT/ULN)]/(ALP/ULN)**

Any pDILI will be summarized (all criteria must be met at the same assessment):

- **(AST or ALT > 3xULN) and Total Bilirubin > 1.5xULN**
- **(AST or ALT > 3xULN) and Total Bilirubin > 2xULN**

In addition, a table will be produced to summarize potential Hy's Law cases. The following definition will be used in that table:

- [AST ≥3xULN or ALT ≥3xULN] and Total Bilirubin ≥2xULN in the absence of ALP ≥2xULN

In order to meet the above **potential Hy's Law** criteria, a participant must experience the elevation in bilirubin and ALT or AST (and the absence of ALP elevation, if applicable) at the same assessment. For example, a participant who experiences a ≥2 x ULN elevation of bilirubin at one visit and a ≥3xULN elevation in ALT (or AST) at a subsequent visit has not fulfilled the Hy's Law criteria.

Potential hepatotoxicity (meeting one of the PDILI or Hy's Law laboratory criteria at least once) will be considered with and without symptoms potentially associated with hepatitis or hypersensitivity according to the investigator (reported on the Symptoms of Hepatitis and Hypersensitivity CRF page).

Change #74

Section 10.4.1 Vital signs

The following text was added:

Unscheduled and repeat visits will be considered as long as they were collected in the period being summarized (values observed more than 140 days after the last administration of study medication are not considered).

Change #75

Section 10.4.3 Other safety endpoints

The following text was added:

For by-visit summaries, unscheduled and repeat visits will not be summarized, but these data will be included in listings. By-visit tables should include the SFU visit. Summaries over a period of time (as in shift tables), unscheduled and repeat visits will be considered as long as they were collected in the period being summarized (values observed more than 140 days after the last administration of study medication are not considered).

Change #76

Section 10.4.3.2 Columbia-Suicide Severity Rate Scale (C-SSRS)

The following text was updated:

The incidence of participants with suicidal ideation, suicidal behavior, suicidal ideation or behavior, and self-injurious behavior will be summarized **for the Initial Treatment Period and the combined Initial and Maintenance Treatment Period** by treatment group ~~for each treatment~~.

Change #77

Section 10.3.4.5 Patient Health Questionnaire (PHQ)-9 scores

The following text was updated:

A ~~In addition, a categorical~~ **summary of the absolute and change from Baseline value** ~~scores~~ **will be presented by treatment group and visit**.

The percentage of study participants with scores below ~~a corresponding shift table. The following categories will be presented: 0-4; 5~~, **between 5 and** 9, **between** 10 **and** 14, **between** 15 **and** 19, **and greater than or equal to 20 in PHQ-9 will be summarized as a shift from Baseline by visit and treatment group based on observed values**~~; ≥20~~.

The percentage of study participants with scores ≥ 15 at any post-Baseline visit and the number and percentage of study participants with scores ≥ 20 at any post-Baseline visit will be summarized by treatment group based on observed values. This summary will also include the percentage of study participants with increase from baseline ≥ 5 at any post-Baseline visit.

~~Different to other safety variables, PHQ-9 will be summarized using the MCMC/monotone regression approach described for continuous variables.~~

The number and percentage of participants that complete the PHQ-9 will be calculated for each visit by treatment group. The percentage will be based on the number of participants in the RS. ~~(or MS, as appropriate).~~ The summary will be repeated where the percentage is based on the subset of participants who complete each particular visit.

Change #78

Section 11 References

The following reference was updated:

Common Terminology Criteria for Adverse Events (CTCAE); Version **4** ~~5~~.0 June **2010** ~~2017~~.U.S. Department of Health and Human Services

Change #79

Section 12.1 Appendix A: MedDRA algorithmic approach to anaphylaxis

The algorithm for identifying anaphylaxis was updated:

The SMQ *Anaphylactic reaction* consists of three parts:

- A **narrow search** containing PTs that represent core anaphylactic reaction terms (Category A – core anaphylactic reaction terms)

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

- A **broad search** that contains additional terms that are added to those included in the narrow search. These additional terms are signs and symptoms possibly indicative of anaphylactic reaction and categorized in B, C or D

  - **Category B (Upper Airway/Respiratory Terms)**

Acute respiratory failure	Nasal obstruction
Asthma	Oedema mouth
Bronchial oedema	Oropharyngeal spasm
Bronchospasm	Oropharyngeal swelling
Cardio-respiratory distress	Respiratory arrest
Chest discomfort	Respiratory distress
Choking	Respiratory failure
Choking sensation	Reversible airways obstruction

Circumoral oedema	Sensation of foreign body
Cough	Sneezing
Cyanosis	Stridor
Dyspnoea	Swollen tongue
Hyperventilation	Tachypnoea
Irregular breathing	Throat tightness
Laryngeal dyspnoea	Tongue oedema
Laryngeal oedema	Tracheal obstruction
Laryngospasm	Tracheal oedema
Laryngotracheal oedema	Upper airway obstruction
Mouth swelling	Wheezing

- **Category C (Angioedema/Urticaria/Pruritus/Flush terms)**

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

- **Category D (Cardiovascular/Hypotension terms)**

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

Diastolic hypotension
Hypotension

- An **algorithmic approach** which combines a number of anaphylactic reaction symptoms in order to increase specificity. A case must include one of the following where both occur on either the same day as when an injection was administered or one day after, and for scenarios where two events must have been reported, both events must have occurred within one day of each other**:**
  - A narrow term or a term from Category A;
  - A term from Category B - (Upper Airway/Respiratory) AND a term from Category C - (Angioedema/Urticaria/Pruritus/Flush);
  - A term from Category D - (Cardiovascular/Hypotension) AND [a term from Category B - (Upper Airway/Respiratory) OR a term from Category C - (Angioedema/Urticaria/ Pruritus/Flush)]

Change #80

Section 912.2 Appendix B: Definition of CTCAE grades

Table 12-1 was updated:

Table 12–1: Definitions of CTCAE grades by biochemistry parameter

Parameter	Definition	Unit	Grade 1	Grade 2	Grade 3	Grade 4
Creatinine[1]	High	**umol/L** ~~mmol/L~~	**>1-1.5x Baseline or** >ULN-1.5 x ULN	**>1.5-3.0x Baseline or** >1.5 – 3.0 x ULN	**>3.0x Baseline or** >3.0 – 6.0 x ULN	>6.0 x ULN
Sodium	Low	mmol/L	130-<LLN	N/A	120-<130	<120
Sodium	High	mmol/L	>ULN-150	>150-155	>155-160	>160
Potassium[2]	Low	mmol/L	3.0-<LLN	3.0-<LLN	2.5-<3.0	<2.5
Potassium	High	mmol/L	>ULN-5.5	>5.5-6.0	>6.0-7.0	>7.0
Calcium	Low	mmol/L	2.0-<LLN	1.75-<2.0	1.5-<1.75	<1.5
Calcium	High	mmol/L	>ULN-2.9	>2.9-3.1	>3.1-3.4	>3.4
Magnesium	Low	mmol/L	0.5-<LLN	0.4-<0.5	0.3-<0.4	<0.3
Magnesium	High	mmol/L	>ULN-1.23	N/A	>1.23-3.30	>3.30
Cholesterol	**High**	**mmol/L**	**>ULN-7.75**	**>7.75-10.34**	**>10.34-12.82**	**>12.82**

1 The CTCAE Grade definitions for creatinine are based on the logical or and the highest applicable CTCAE grade should be assigned to a creatine value.
2 The decreased potassium criterion of 3.0-<LLN is specified for both CTCAE Grade 1 and Grade 2; values meeting this criterion will be counted as Grade 2.

And Table 12-2 was updated:

Table 12–2: Definitions of CTCAE grades by hematology parameter

Parameter	Definition	Unit	Grade 1	Grade 2	Grade 3	Grade 4
Hemoglobin	Low	g/L	100-<LLN	80-<100	<80	N/A
Hemoglobin[1]	High	g/L	>0-20 above ULN **or >0-20 above Baseline if Baseline is above ULN**	>20-40 above ULN **or >20-40 above Baseline if Baseline is above ULN**	>40 above ULN or **>40 above Baseline if Baseline is above ULN**	N/A
Platelets	Low	10^9/L	75-<LLN	50-<75	25-<50	<25
WBC	Low	10^9/L	3-<LLN	2-<3	1-<2	<1
WBC	High	10^9/L	N/A	N/A	>100	N/A
Lymphocytes	Low	10^9/L	0.8-<LLN	0.5-<0.8	0.2-<0.5	<0.2
Lymphocytes	High	10^9/L	N/A	>4-20	>20	N/A
Neutrophils	Low	10^9/L	1.5-<LLN	1.0-<1.5	0.5-<1.0	<0.5

LLN=lower limit of normal; N/A=not applicable; ULN=upper limit of normal, WBC=white blood cells
1 The CTCAE Grade definitions to be applied are dependent on the Baseline hemoglobin value. If the baseline value is > ULN then the criteria relative to Baseline is applicable, otherwise the criteria relative to ULN is applicable.

13.2 Amendment 2

Rationale for the amendment

The main purposes of this amendment were:

- General update to analyses to align with protocol amendment 4.
- Procedural clarifications from discussions and feedback provided at meetings
- Update to align with the bimekizumab program standards and safety topics of interest

Modifications and changes

Global Changes

The following changes were made throughout the SAP:

- Typos and formatting were updated throughout the document
- HSSDD worst pain and average pain were updated to worst skin pain and average skin pain, respectively

Specific changes

In addition to the global changes, the following specific changes have been made (typos such as missing spaces or redundant spaces are not listed):

Change #1

List of Abbreviations

The following abbreviations have been added:

CFB	change from Baseline
CV-CAC	Cardiovascular Event Adjudication Committee
eCDF	empirical cumulative distribution function
IBD-CAC	Inflammatory Bowel Disease Adjudication Committee

Change #2

Section 1 Introduction

The protocols were updated:

The SAP is based on the Protocol Amendment **4** ~~3~~**, 9 May 2022** ~~9 February 2021~~.

Change #3

Section 2.2 Study endpoints

The following text was deleted:

The endpoints based on HS Symptom Daily Diary (HSSDD) and Hidradenitis Suppurativa Symptom Questionnaire (HSSQ) pain responses are based on the current definitions, which are continuous. It is anticipated that a responder (binary) endpoint will be defined for the HSSDD and HSSQ pain items as well as other symptom items prior to database lock and unblinding, based on separate ongoing, blinded, psychometric analyses aiming to determine threshold for within-patient clinically meaningful improvement.

The below HSSDD and HSSQ pain response endpoints and analyses will be adjusted accordingly in a future SAP amendment.

Change #4

Section 2.2.1.2 Secondary efficacy endpoints

The following bullet was added:

- Skin pain response based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) at Week 16 among study participants with a score of ≥3 at Baseline

Change #5

Section 2.2.1.3 Other efficacy endpoints

The following bullets were updated:

- ~~Skin pPain response statusain response, as assessed by the "worst pain" item in the HSSDD, defined as an improvement from baseline in the weekly worst skin pain score of at least 3 units~~ **Skin pain response based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) among study participants with a score of ≥3 at Baseline**

- **Skin pain response** ~~Response in HS Skin Pain (11-point numeric rating scale) assessed by the HSSDD at Week 16Response~~ **(at least a 30% reduction and at least a 1-point reduction from Baseline in HSSDD weekly worst skin pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline**
- **Skin pain response (at least a 30% reduction and at least a 1-point reduction from Baseline in HSSDD weekly average skin pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline**
- ~~HS Symptom Questionnaire (HSSQ) response (at least a 3-unit reduction from Baseline in worst HS Skin Pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline assessed HSSQ by the HSSDD in the Initial Treatment Period, and assessed by the HSSQ in the Maintenance Treatment Period~~**Skin pain response (at least a 30% reduction and at least a 1-point reduction from Baseline in HS Skin Pain score [11-point numeric rating scale]) assessed by the HSSQ among study participants with a score of ≥3 at Baseline**
- Absolute change from Baseline in DLQI Total Score
- DLQI Total Score of 0 or 1
- Minimum clinically important difference (MCID) (improvement from Baseline of 4 or more) in the DLQI Total Score among study participants with a Baseline score of at least 4
- Absolute change from Baseline in Hidradenitis Suppurativa Quality of Life (HiSQOL) domain scores (symptoms, psychosocial, activities and adaptations) and Total score
- Patient Global Impression of HS Severity (PGI-S-HS)
- Patient Global Impression of Change of HS Severity (PGI-C-HS)
- Patient Global Impression of Severity of Skin Pain (PGI-S-SP)
- Patient Global Impression of Change of Skin Pain (PGI-C-SP)
- Absolute change from Baseline in each of the other HS Symptoms - itch, drainage or oozing of HS lesions, and smell or odor.
- ~~Responders~~ **Response** on other HS Symptoms (11-point numeric rating scale) - itch, drainage or oozing of HS lesions, and smell or odor

Change #6

Section 2.4 Determination of sample size

The power to detect a statistically significant difference for each of the endpoints are shown in Table 2–1. Notably, with a 2-sided significance level of 0.025, the sample size of 140:70 provides 73% power for detecting at least a difference of 1.5 (bimekizumab Q4W vs placebo) for the Worst **Skin** Pain **change from Baseline (CFB)** endpoint.

Given the high level of power for each of the primary and key secondary endpoints at the 0.025 significance level for the Q2W comparison (power≥0.89), and per the alpha spending strategy, there is a high likelihood that the Q4W comparison of Worst **Skin** Pain **CFB** vs placebo ~~for Worst Pain change from Baseline~~ will be allowed to be tested against the 0.05 level of

significance. The power for this latter test is 81%. The sample size is thus ultimately driven by the Worst Skin Pain CFB endpoint. Furthermore, the randomization ratio of 2:2:2:1 has been chosen to provide study participants with a high probability (6/7 ~ 86%) of being randomized to active study drug.

After randomization for this study was complete, an additional endpoint to assess Worst Skin Pain response was included in the sequential testing procedure. This additional endpoint is based on the threshold for clinically meaningful change and is defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) at Week 16 among study participants with a score of ≥3 at Baseline. Note that the power calculations reported in Table 2–1 for this endpoint are based on the sample size that was initially driven by the Worst Pain CFB endpoint as described above. With a 2-sided significance level of 0.025, the sample size of 104:52 in the subset of participants reporting Baseline HSSDD worst skin pain score at or beyond the threshold for clinically meaningful change (ie, Baseline HSSDD ≥3) provides 53% power for detecting a statistically significant difference between bimekizumab Q4W and placebo in the proportion of Worst Pain responders.

Note that the power at the 0.025 level of significance associated with this endpoint for the comparison between bimekizumab Q2W and placebo is 95%. The Q4W comparison of Worst Skin Pain response vs placebo against the 0.05 level of significance is therefore likely, and the power at this significance level is 65%. Given the strength of the power for the bimekizumab Q2W arm vs placebo across endpoints, it is considered acceptable to have a relatively low power for this final endpoint in the testing sequence for the bimekizumab Q4W treatment arm vs placebo.

Change #7

Section 2.4 Determination of sample size

Table 2-1 was updated:

Table 2–1: Power calculation assumptions and methods

Endpoint	Power α = 0.025, 2-sided		Assumptions		
	Q2W	Q4W	Week 16 Bimekizumab Q2W N=280[a]	Week 16 Bimekizumab Q4W N=140	Week 16 Placebo N=70
$HiSCR_{50}$	0.99	0.90	Proportion responders=0.60	Proportion responders=0.50	Proportion responders=0.25
$HiSCR_{75}$	0.99	0.98	Proportion responders=0.45	Proportion responders=0.35	Proportion responders=0.10
Flare	0.99	0.99	Proportion of participants with flare by Week 16=0.09	Proportion of participants with flare by Week 16=0.19	Proportion of participants with flare by Week 16=0.52
DLQI	0.99	0.96	Mean CFB=-5.4; SD=6.8	Mean CFB=-4.8; SD=6.8	Mean CFB=-0.8; SD=6.6

Table 2–1: Power calculation assumptions and methods

Worst Skin Pain **CFB**[b]	0.89	0.73	Mean CFB=-2.2; SD=3.2	Mean CFB=-2.0; SD=3.2	Mean CFB=-0.5; SD=3.7
Worst Pain Response[c]	**0.95**	**0.53**	**Proportion responders=0.53**	**Proportion responders=0.43**	**Proportion responders=0.23**

CFB=change from Baseline; Q2W=every 2 weeks; Q4W=every 4 weeks; SD=standard deviation
Note: Estimates for HS0004 are based on Week 12 data from the HS0001 study.
[a] Pooled Q2W at Week 16 from Q2W/Q2W and Q2W/Q4W arms
[b] Within-participant average of Worst **Skin** Pain according to 24-hour recall.
[c] **Assumes N=208, 104, 52 in Q2W, Q4W, and placebo, respectively, to account for Worst Skin Pain score at or above 3 (ie, the threshold for clinically meaningful change from Baseline).**

Change #8

Section 3.1 General presentation of summaries and analyses

The following text was added:

Per protocol, visit windows are ±3 days from the date of first dose. The 20-week SFU Visit window is ±7 days from the date of the final dose. All by-visit summaries will contain nominal (ie, scheduled) visits only. Unscheduled visits will not be mapped to scheduled visits except for assessments that occur within a 3-day time window of a scheduled visit. In that case, the assessment will be mapped to the corresponding scheduled visit and will be used for the analysis. This will only occur for selected vendor data. **The only exception to this rule is for unscheduled assessments that occur up to 3 days after the Baseline visit. These unscheduled visits will remain as unscheduled as the Baseline assessment cannot be after the first dose of study drug administration. See Section 3.3 for more details on the definition of Baseline values.**

Change #9

Section 3.5.8 Pharmacokinetics Per-Protocol Set

The following text was deleted:

The Pharmacokinetics Per-Protocol Set (PK-PPS) will consist of study participants who received at least 1 full dose of bimekizumab and provided at least 1 quantifiable plasma concentration post-dose without important protocol deviations that would affect the PK. ~~The Pharmacokinetics Per-Protocol Set is defined separately for each of the treatment periods (ie, separately for the Initial Treatment Period and the Maintenance Treatment Period).~~

Change #10

Section 3.10 Changes to protocol-defined analyses

The following text was updated:

The following other efficacy endpoints are included in the protocol but will not be included as part of the analysis:

- ~~Responders~~ **Response** on other HS Symptoms (11-point numeric rating scale) - itch, drainage or oozing of HS lesions, and smell or odor
- Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Week 16) by Week 48

The calculation of nominal p-values has been added for selected efficacy endpoints. **These nominal p-values are not controlled for multiplicity and should not be used to declare statistical significance.**

The protocol defines the PK-PPS separately by period, but there will only be one PK-PPS for the overall study.

Change #11

Section 4.1 Adjustments for covariates

The following sentence was updated:

The Worst Skin Pain s**econdary endpoints (change from Baseline continuous** ~~secondary~~**endpoint and pain response binary endpoint)** will also include analgesic use as a covariate.

Change #12

Section 4.2.1.2 Handling of missing data for the secondary efficacy endpoints

The following paragraph was updated:

For secondary continuous efficacy endpoints, MI-MCMC/monotone regression is the primary method for imputing missing data, regardless of whether the missing data are preceded by an intercurrent event. That is, if an intercurrent event occurs on or before a visit, the result for that visit will be treated as missing and **then** imputed~~with the missing data~~. If the imputation model cannot converge, last observation carried forward (LOCF) will be used. The OC method will be performed as a sensitivity analysis.

Change #13

Section 42.2.1 MI – MCMC/Monotone Regression

The imputation rule for HSSDD was updated in Table 4-2.

Table 13–2: Imputation allowable ranges by variable

Variable	Minimum Value	Maximum Value	Integer Values Only
Lesion count[a]	0	--	Yes
DLQI total score	0	30	Yes
hs-CRP	LLOQ/2	--	No
HSSDD item score	0	10	**No**~~Yes~~
HSSQ item score	0	10	Yes
HiSQOL symptom status score	0	16	Yes

Table 13–2: Imputation allowable ranges by variable

Variable	Minimum Value	Maximum Value	Integer Values Only
HiSQOL psychosocial impact score	0	20	Yes
HiSQOL impact on physical activities score	0	32	Yes
EQ-5D-3L VAS	0	100	Yes
WPAI dimension scores	0	100	No for variables: "Percent work time missed due to problem" and "Percent overall work impairment due to problem". Yes for variables: "Percent impairment while working due to problem" and "Percent activity impairment due to problem". These two variables can only take values that are multiples of 10.

[a] Lesion counts will be imputed separately for each lesion type (abscesses, draining tunnels [fistulas/sinus tracts], inflammatory nodules, non-draining tunnels [fistulas/sinus tracts], non-inflammatory nodules, HS scars). The imputed lesion counts will be used to derive the endpoints that are dependent on the lesion count data (eg, $HiSCR_{50}$).

Change #14

Section 4.2.3 Rationale for estimand

The following text was added to the bullet:

- A composite estimand strategy will be used for the primary analysis of the primary and binary secondary endpoints ($HiSCR_{50}$, $HiSCR_{75}$, flare, **HS worst skin pain response**),

Change #15

Section 4.5 Multiple comparisons/multiplicity

Under this framework, each bimekizumab dose of 320mg Q2W and 320mg Q4W will be compared to placebo in the first instance at a familywise error rate of 0.025 (α/2). Simultaneously within each dose, closed testing for the primary and secondary efficacy endpoints will be performed as follows:

1. Step 1: Test $HiSCR_{50}$ at significance level 0.025.
2. Steps 2 to ~~6~~**5** – If Step 1 is significant at 0.025 then test each secondary endpoint sequentially in the order shown in Figure 4-1, moving to the next step only if significance achieved at 0.025.
3. In the event that Step ~~6~~**5** is significant at 0.025 for a given dose, then Steps 1 to 6 will be repeated for the other dose using a significance level of 0.05.

The secondary efficacy variables supporting the primary efficacy variable are listed below, and will be included in the multiplicity adjustment using the analysis methods specified in Section 8.3:

1. Proportion of study participants who achieve $HiSCR_{75}$ at Week 16.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

2. Proportion of study participants who experience at least 1 flare by Week 16, with flare defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Baseline.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

3. CFB in DLQI Total Score at Week 16.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

4. Absolute change from Baseline in Skin Pain Score at Week 16, as assessed by the "worst skin pain" item (11-point numeric rating scale) in the HSSDD.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

5. Skin pain response at Week 16, based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) at Week 16 among study participants with a score of ≥3 at Baseline.

a. bimekizumab 320mg Q2W vs placebo

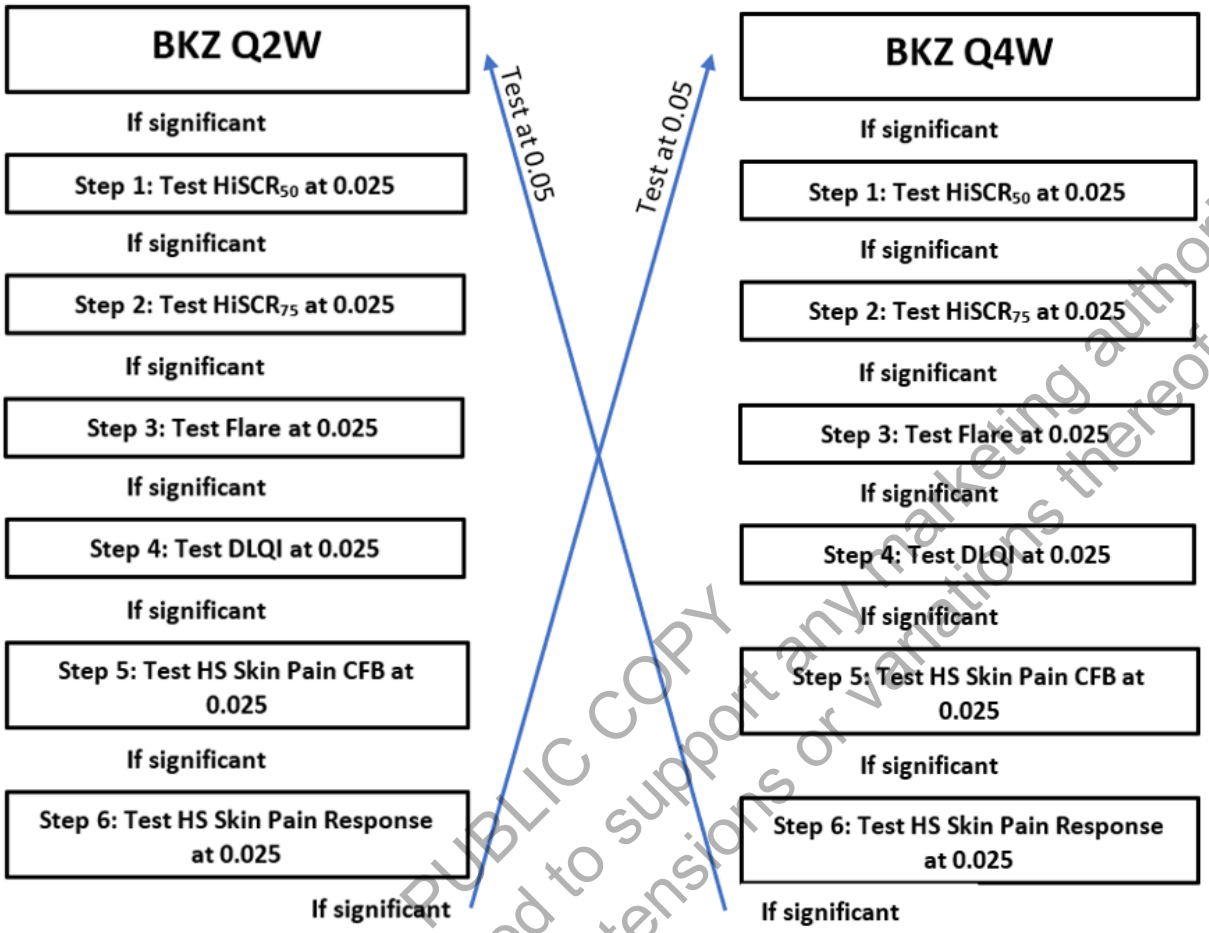
b. bimekizumab 320mg Q4W vs placebo

Change #16

Section 4.5 Multiple comparisons/multiplicity

Figure 4-1 was updated to add the new secondary endpoint:

Figure 4-1: Sequence of testing



AN=abscess and inflammatory nodule; DLQI=Dermatology Life Quality Index; $HiSCR_{50}$=a 50% reduction in the total AN count with no increase from Baseline in abscess or draining tunnel count; $HiSCR_{75}$=a 75% reduction in the total AN count with no increase from Baseline in abscess or draining tunnel count; HS=hidradenitis Suppurativa; Q2W=every 2 weeks; Q4W=every 4 weeks

HS skin pain response is tested among study participants with a score of ≥3 at Baseline.

Change #17

Section 4.8 Examination of subgroups

The following text was updated:

Subgroup analyses will be performed on the $HiSCR_{50}$, $HiSCR_{75}$, ~~and~~flare, **and worst skin pain response** endpoints by visit for the Initial Treatment Period and Maintenance Treatment Period. Additional subgroup analyses will be performed on the CFB in the worst skin pain score as measured by HSSDD and in the DLQI total score through Week 16 as described below.

Along with the tables described, there will be tables for $HiSCR_{50}$, $HiSCR_{75}$, ~~and~~ flare, **and skin pain response endpoints** which display the response difference and 95% CIs between each bimekizumab dose regimen versus placebo for each of the subgroups at Week 16. Corresponding forest plots will be prepared.

Additionally, the following bullets were clarified:

- Antibody positivity (confirmatory assay: negative or positive. **See Section 9.3.2**)
- **Antihistamines users during the Initial Treatment Period (yes, no) (Section 6.4.2 specifies how participants are classified as antihistamine users) (applicable only to the skin pain response endpoint)**

Change #18

Section 6.4 Prior and concomitant medications

The following text was added:

The number and percentage of study participants with concomitant vaccines for COVID-19 will be summarized by treatment group, overall and by World Health Organization Drug Dictionary Standardized Drug Grouping (SDG), presenting SDG subgroup, and preferred term. The SDG subgroup Vaccines for COVID-19 will be used to identify vaccines for COVID-19 using the narrow scope; this subgroup is divided further into separate subgroups which is the level that will be presented. The number of individual occurrences of the vaccine for COVID-19 will also be summarized.

A listing of concomitant vaccines for COVID-19 will be provided.

Change #19

Section 6.4.2 Classification of participants as analgesic, antihistamine users

The section was updated as follows:

If a participant has taken a new analgesic/increased regimen of analgesic, or taken an antihistamine, on 1 or more days (need not be consecutive) ~~for a given week~~ **in a study period (Initial Treatment Period or Maintenance Treatment Period)**, then for that ~~week~~ **period** the participant will be classified as an analgesic or antihistamine user, respectively. The ~~week~~ **period** under consideration is to match the **period** as defined for the HSSDD ~~week~~ for the Initial Treatment Period or HSSQ ~~week~~ for the Maintenance Treatment Period, based on dates/times of the medications taken. ~~If there is a visit date but no HSSDD available at the visit, then the analgesic/antihistamine user status for that week will be derived based on the visit date. If there is no visit available, then the weekly analgesic/antihistamine user status will default to the analgesic/antihistamine status for the overall study period.~~

New analgesic/increased regimen of analgesic use, regardless of indication, is defined as an analgesic medication with start date on or after the first dose of study medication. Stable analgesics (ie, analgesics which were taken already before randomization) will not be included in this category of analgesic user. This classification will be used ~~to adjust the formal analysis of the Worst Pain secondary endpoints and~~ for selected subgroup analyses.

Antihistamine use is identified by considering the ATC classification. This classification is used for analyzing the Worst Itch endpoint and for selected subgroup analyses, by visit, for the Initial Treatment Period and Maintenance Period as applicable.

Additionally, if a participant has taken a new analgesic/increased regimen of analgesic on 1 or more days (need not be consecutive) prior to the Week 16 visit, then for that week the participant will be classified as an analgesic user. This classification will be used to adjust the formal analysis of the Worst Skin Pain secondary endpoints. If there is a visit date but

no HSSDD available at the visit, then the analgesic/antihistamine user status for that week will be derived based on the visit date. If there is no visit available, then the weekly analgesic/antihistamine user status will default to the analgesic/antihistamine status for the overall study period.

Change #20

Section 8.2 Primary efficacy endpoint

The intercurrent event strategy for the sensitivity analysis was updated in Table 8-1:

Table 8–1: Estimand Details and Attributes for Primary Endpoint

	Estimands for Primary Endpoint			
Statistical Category (Section)	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
Primary Objective: To evaluate the efficacy of bimekizumab in study participants with moderate to severe HS				
Sensitivity (Section 8.2.3.2)	$HiSCR_{50}$ response at Week 16	RS	**Composite strategy, as for the primary analysis.** ~~The intercurrent events will be handled using a~~ **~~hypothetical strategy~~**~~, whereby all data at and after the intercurrent event will be treated as missing.~~	The odds ratio versus placebo based on a logistic regression. Missing values will be imputed using MI – Reference-Based Regression under a missing not at random assumption.

Table 8–1: Estimand Details and Attributes for Primary Endpoint

	Estimands for Primary Endpoint			
Statistical Category (Section)	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
Sensitivity (Section 8.2.3.5)	$HiSCR_{50}$ response at Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, ~~whereby only participants still on randomized treatment at Week 16 are included.~~ **whereby only participants with a lesion count assessment at Week 16 who have not had an intercurrent event on or before Week 16 are included.**	The odds ratio versus placebo is based on a logistic regression. Missing values will not be imputed.

Change #21

Section 8.2.3.5 Analysis on observed cases

The following text was updated:

An additional supportive analysis will be based on observed data only for study ~~participants who are still on the randomized treatment at Week 16. Study participants with missing data or who have prematurely discontinued study treatment will be treated as missing (see Section 4.2.2).~~ **participants with a lesion count assessment at Week 16 who have not had an intercurrent event on or before Week 16. Study participants with missing data at Week 16 will be treated as missing (see Section 4.2.2).**

Change #22

Section 8.2.3.11 Center-by-Treatment Interaction

The following sentence was added: In order to achieve model convergence, other explanatory variables eg, Baseline Hurley Stage and Baseline antibiotic use may be dropped from the model.

If model convergence is still not achieved, region and a region-by-treatment interaction term will be added to the model instead. Regions are defined in Section 3.7.

Change #23

Section 8.3 Secondary efficacy endpoints

The following sensitivity analyses as well as the new skin pain response secondary analysis were added to Table 8-2:

Table 8–2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
Secondary Objective: Evaluate the efficacy of bimekizumab on other measures of disease activity in study participants with moderate to severe HS					
DLQI	Secondary - Sensitivity (Section 8.3.2.2)	Change from Baseline in DLQI total score to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a DLQI total score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the DLQI total score for participants receiving bimekizumab versus placebo. Missing values will not be imputed.

Table 8–2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
HSSDD	Secondary (Section 8.3.3.2)	Change from Baseline in worst skin pain score, as assessed by "worst skin pain" item in HSSDD to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a HSSDD worst skin pain score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the worst skin pain score for participants receiving bimekizumab versus placebo. Missing values will not be imputed.
HSSDD	Secondary (Section 8.3.4.1)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improvement in the weekly worst skin pain score of at least 3 points	RS[a]	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo based on a logistic regression, as for the primary analysis.

Table 8–2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
HSSDD	Secondary Sensitivity (Section 8.3.4. 2.1)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improve-ment in the weekly worst skin pain score of at least 3 points	RS[a]	**Composite strategy**, as for the primary analysis where the occurrence of an intercurrent event will be handled by evaluating the corresponding participants as nonresponders.	The odds ratio versus placebo based on a logistic regression. Missing values for any other reason will also be imputed as nonresponders.
HSSDD	Secondary Sensitivity (Section 8.3.4. 2.2)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improve-ment in the weekly worst skin pain score of at least 3 points	RS[a]	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a HSSDD worst skin pain score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	The odds ratio versus placebo is based on a logistic regression. Missing values will not be imputed.

AE=adverse event; ANCOVA=analysis of covariance; DLQI=Dermatology Life Quality Index; HiSCR=Hidradenitis Suppurativa Clinical Response; HSSDD=Hidradenitis Suppurativa Symptom Daily Diary; IES=intercurrent event(s) strategy; LSMD=Least Squares Mean Difference; MCMC=Markov Chain Monte Carlo; MI=multiple imputation; PLS=Population-level summary; Pop=Population; RS=Randomized Set

[a] **Analysis includes all study participants in the RS with a Baseline HSSDD Worst Skin Pain score of 3 or higher.**

Change #24

Section 8.3.3.1 Primary analysis of change from Baseline in DLQI Total Score

Section 8.3.3.1 was separated from Section 8.3.3 due to the added sensitivity analysis for DLQI Total Score. The following text was added:

Change from Baseline in DLQI total score will be presented by treatment group. The analysis model will be based on an ANCOVA with fixed effects of treatment, Hurley Stage at Baseline, Baseline antibiotic use and Baseline value as a covariate. The least square mean (LSM), standard error (SE), and 95% CI for the LSM will be presented by treatment group. For the comparison between placebo and bimekizumab: the difference between the LSM, the associated 97.5% CI for the contrasts, and the corresponding p-value will be presented. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose with a corresponding $Z_{\alpha/2}$ of 1.96. **Estimand and intercurrent event details are specified in Table 8–2.**

Change #25

The following section was added:

Sensitivity analysis of change from Baseline in DLQI Total score at Week 16 (Section 8.3.3.2)

A sensitivity analysis using the same analysis model as in Section 8.3.2.1 will be used, based on observed data. Estimand and intercurrent event details are specified in Table 8–2.

Change #26

Section 8.3.4.1 Primary analysis of change from Baseline in skin pain score at Week 16

Section 8.3.4.1 was separated from Section 8.4.3 due to the added sensitivity analysis for HSSDD worst skin pain score and removed 'A treatment by analgesic use interaction term will also be added to the model and removed if not significant' from section.

Change #27

The following section was added:

Sensitivity analysis of change from Baseline in skin pain score at Week 16 (Section 8.3.4.2)

A sensitivity analysis using the same analysis model as in Section 8.3.3.1 will be used, based on observed data. Estimand and intercurrent event details are specified in Table 8–2.

Change #28

The following sections were added:

HSSDD skin pain response at Week 16 (Section 8.3.5)

The analysis set for the analyses of the skin pain response will be restricted to those study participants in the RS with a Baseline worst skin pain score of 3 or higher. The weekly scores and Baseline score are derived as specified in Section 8.3.3.

Primary analysis of skin pain response at Week 16 (Section 8.3.5.1)

Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, is defined as an improvement in the weekly worst skin pain score of at least 3 points versus Baseline.

The primary analysis will be based on a logistic regression model including a fixed effect for treatment, Hurley stage at Baseline, Baseline antibiotic use, and analgesic use (Section 6.4.2).

The odds ratio versus placebo, p-value (from Wald test), and 97.5% CI will be calculated. If one dose regimen is tested at the 0.05 significance level as determined in Section 4.5, then the confidence interval will be 95% instead of 97.5% for that dose. Missing data will be handled as specified in Section 4.2.1.2. Estimand and intercurrent event details are specified in Table 8–2.

The number and percentage of participants who are pain responders at Week 16 will be summarized by treatment group.

By-participant listings of pain response status will be provided.

Sensitivity analyses of Skin Pain Response at Week 16 (Section 8.3.5.2)

Nonresponse imputation (Section 8.3.5.2.1)

As a sensitivity analysis, any missing data at Week 16 that are not preceded by an intercurrent event (Table 8–2) will be imputed as nonresponse. That is, participants who experience an intercurrent event will be imputed as nonresponder at the timepoint of the event and all subsequent timepoints (including any recorded data after the event), and all missing data will also be imputed as nonresponse.

The same analysis model as Section 8.3.4.1 will then be used on the imputed data set.

Analysis on observed case (Section 8.3.5.2.2)

An additional supportive analysis will be based on observed data only for study participants with a worst skin pain score at Week 16 who have not had an intercurrent even on or before Week 16. Study participants with missing data or who have prematurely discontinued study treatment will be treated as missing (see Section 4.2.2).

The same analysis model as in Section 8.3.4.1 will then be used on the imputed data set.

Change #29

Section 8.4.1.3 HiSCR$_{25}$, HiSCR$_{50}$, HiSCR$_{75}$, HiSCR$_{90}$, and HiSCR$_{100}$ response at both Weeks 16 and 48

The following text was added:

The number and percentage of HiSCR$_{25}$, HiSCR$_{50}$, HiSCR$_{75}$, HiSCR$_{90}$, and HiSCR$_{100}$ responders at both Weeks 16 and 48 will be summarized based on the RS **and MS**.

Change #30

Section 8.4.12 Hidradenitis Suppurativa Symptom Daily Diary (HSSDD)

The following text was updated:

See Section 8.3.3 for details on HSSDD Baseline **and weekly average** definitions and derivations.

Percent change from Baseline in HSSDD responses for worst and average skin pain score is defined as

$$\text{Percent change from Baseline } = 100 \times \frac{\text{Post Baseline HSSDD score} - \text{Baseline HSSDD score}}{\text{Baseline HSSDD score}}$$

Change from Baseline in each HSSDD item (worst skin pain, average skin pain, smell or odor, itch at its worst, and amount of drainage or oozing) score will be summarized using descriptive statistics by treatment group and visit, based on weekly averages. The table will display descriptive statistics for the Baseline, followed by descriptive statistics for the change from Baseline for all visits. Percentage change will be summarized for the worst and average skin pain items.

Additionally, change from Baseline in each HSSDD item will be evaluated by treatment group at Week 16 via continuous empirical cumulative distribution function (eCDF) plots showing the absolute change from Baseline on the horizontal axis and the cumulative percent of participants experiencing that change on the vertical axis.

Change from Baseline in Worst **Skin** Pain score and Worst Itch score will additionally be summarized by visit and by analgesic and antihistamine use status (Section 6.4.2), respectively.

HSSDD response **based on clinically meaningful change** for the worst skin pain ~~and average skin pain~~ items~~s~~ is defined as at least **a 3-point** ~~30% reduction and at least a 1 point~~ reduction from Baseline in HSSDD among study participants with a score of ≥3 at Baseline, based on weekly averages.

The number and percentage of responders based on clinically meaningful change for ~~each item~~ **the worst skin pain item** will be summarized by treatment group and visit.

The number and percentage of participants who were responders **based on clinically meaningful change** at any timepoint in the Initial Treatment Period will be summarized by treatment group for the worst ~~HS~~skin pain ~~score~~ **item**.

HSSDD response for the worst skin pain and average skin pain items is defined as at least a 30% reduction and at least a 1-point reduction from Baseline among study participants with a score of ≥3 at Baseline. The number and percentage of responders for each item will be summarized by treatment group and visit.

The number and percentage of participants who were responders based on clinically meaningful change at any timepoint in the Initial Treatment Period will be summarized by treatment group for the worst skin pain and average skin pain items.

The number and percentage of participants that complete the HSSDD will be calculated for each visit by treatment group. A participant will be counted as completing the HSSDD at a visit if the minimum number of daily entries is present to calculate the weekly average (see Section 8.3.3). The percentage will be based on the number of participants in the RS. A participant will be considered a completer at a visit if the weekly average can be calculated for that visit.

Change #31

Section 8.4.13 Hidradenitis Suppurativa Symptom Questionnaire

The following text was updated:

Additionally, change from Baseline in each HSSQ item will be evaluated by treatment group at Week 16 and at Week 48 via continuous eCDF plots showing the absolute change from Baseline on the horizontal axis and the cumulative percent of participants experiencing that change on the vertical axis.

Missing data for the continuous change from Baseline will be handled by using MI via the MCMC and monotone regression method specified in Section 4.2.1.3.

HSSQ response for skin pain item is defined as at least a 30% reduction and at least a 1-point reduction from Baseline in HS Skin Pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline.

Change #32

Section 8.4.15 Hidradenitis Suppurativa Quality of Life

The following text was added:

Additionally, change from Baseline in each HiSQOL subscale will be evaluated by treatment group at Week 16 and at Week 48 via continuous eCDF plots showing the absolute change from Baseline on the horizontal axis and the cumulative percent of participants experiencing that change on the vertical axis.

Change #33

Section 8.5 Additional statistical analyses of other efficacy endpoints

The following analysis was added

- Skin Pain response per HSSDD at Week 12

Change #34

Section 9.3.2 Anti-bimekizumab antibodies

The section was updated as follows:

Anti-bimekizumab antibodies (ADAb) will be measured using a 3-tiered assay approach: screening assay, confirmatory assay, and titration assay. Samples confirmed as positive within the confirmatory assay will be further evaluated for the presence of neutralizing anti-bimekizumab antibodies specific to IL-17AA, IL-17FF or both. Samples will be taken at Baseline, then at study Weeks 4, 8, 12, 16, 20, 24, 36 and 48, and at PEOT and SFU timepoints.

ADAb samples are not analyzed when study participants are on a treatment other than bimekizumab. For study participants who switch from placebo to bimekizumab, samples are analyzed starting at the visit when the switch to bimekizumab occurs (Week 16). The sample at Week 16 will act as the Baseline for that treatment group.

The screening cut point will be used to determine the status of anti-bimekizumab antibodies in the test sample as Positive Screen (PS) or Negative Screen (NS). For samples presenting anti-bimekizumab antibody levels that are PS, a further confirmatory assay will be performed, and the result of which will be reported as either Positive Immunodepletion (PI) or Negative Immunodepletion (NI).

ADAb status for each sample will be derived as follows:

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

Positive immunodepletion samples will be titrated, and the ADAb titer (reciprocal dilution factor including minimum required dilution) will be reported. Subsequently, PI samples will also be subject to a neutralizing assay to evaluate the potential of ADAb to neutralize the target binding of bimekizumab (IL-17AA or IL-17FF or both) in vitro.

Cumulative ~~There are different levels of classification for~~ **ADAb status will be derived as follows:**

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

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

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

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

Overall ADAb status will be derived as follows:

A study participant will be classified as:

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

ADAb categories will be derived ~~definitions are~~ as follows:

- **Pre ADAb negative – treatment-emergent ADAb negative (Category 1)**: Includes study participants who are anti-bimekizumab antibody negative at Baseline and anti-bimekizumab antibody negative at all sampling points during the period of interest (one post-Baseline

missing/inconclusive sample is allowed for subjects with pre- anti-bimekizumab antibody negative sample). This group also includes study participants who have a missing or inconclusive sample (either missing or inconclusive or insufficient volume) at Baseline (ie, pre-treatment) with all post-Baseline samples as ADAb negative.

- **Pre ADAb negative – treatment-emergent ADAb positive (Category 2)**: Includes study participants who are ADAb negative at Baseline and ADAb positive at any sampling points post-Baseline during the period of interest. This group also includes study participants who have a missing sample (either missing or insufficient volume) at Baseline (ie, pre-treatment) with 1 or more post-Baseline samples as ADAb positive.
- **Pre ADAb positive – treatment-emergent reduced ADAb (Category 3)**: Includes study participants who are ADAb positive at Baseline, and ADAb negative at all sampling points post-Baseline during the period of interest.
- **Pre ADAb positive – treatment-emergent unaffected ADAb positive (Category 4)**: Includes study participants who are ADAb positive at Baseline and are ADAb positive at any sampling point post-Baseline during the period of interest with titer values of the same magnitude as Baseline (ie, less than a predefined fold difference from the Baseline titer).
  - For this analysis, this is set at an increase of less than the validated Minimum Significant Ratio (MSR) of 2.07-fold from Baseline.
- **Pre ADAb positive – treatment-emergent ADAb boosted positive (Category 5)**: Includes study participants who ADAb positive at Baseline and are ADAb positive at any sampling point post-Baseline during the period of interest with increased titer values compared to Baseline (equal to or greater than a predefined fold difference increase from Baseline titer which will be defined within the validation of the assay).
  - For this analysis, this is set at an increase equal to or greater than the validated MSR of 2.07-fold from Baseline.
  - Note: for any study participant who is ADAb positive at Baseline and ADAb positive at a post-Baseline time point during the period of interest, but for whom titers are not available to determine treatment unaffected or treatment boosted status, the study participant will be considered as treatment boosted, assuming no other samples are available.
- **ADAb Inconclusive (Category 6)**: Includes study participants who have an ADAb positive Baseline (pre-treatment) sample and some post-Baseline samples during the period of interest are missing or inconclusive, while other post-Baseline samples are ADAb negative.
- **Total treatment-emergent ADAb positivity** (**Category 7** [Categories 2 and 5 combined]): Includes study participants who are pre ADAb negative – treatment-emergent ADAb positive (Category 2) and pre ADAb positive – treatment boosted ADAb positive (Category 5).
- **Total prevalence of pre- ADAb positivity (Category 8** [Categories 3, 4, 5 and 6 combined]): Study participants that are tested ADAb positive at Baseline.
- **Missing:** Includes study participants who are ADAb negative, missing, or inconclusive at Baseline with some post-Baseline samples as missing or inconclusive, and other samples as ADAb negative.

For purposes of efficacy subgroup analyses based on anti-bimekizumab antibody status, the following categories can also be used:

- **ADAb positive – This is defined as study participants who are anti-bimekizumab antibody positive on at least 2 time points while on treatment (ie, excluding Baseline, excluding SFU).**
- **ADAb negative – Study participants for who either:**
  - **All samples (including Baseline) are ADAb negative and there are no missing or inconclusive samples**
  - **Only 1 sample is ADAb positive and all other samples (including Baseline) are ADAb negative or missing or inconclusive**
  - **Only 1 sample is missing or inconclusive and the remaining ADAb samples are negative.**
- **ADAb missing - Defined as study participants who do not fulfil the criteria for one of the 2 groups listed above.**

The rationale for requiring at least 2 time points in which ADAb levels are above the specified cut point is to exclude those study participants who have only one occurrence of ADAb levels during the course of treatment. Including such study participants would increase the number of ADAb positive study participants with potentially no impact on efficacy.

In the case that a sample is collected 1 or more days following the scheduled visit date in which the drug was administered, the ADAb results for that sample will be associated with the scheduled visit and summarized accordingly. Such samples will also be considered when ADAb results are summarized over a given study period.

Analysis

Immunogenicity will be assessed through summary tables and figures, and listing of individual results by participants. All analyses will be run on the AMS, unless specified otherwise.

- Summary of ADAb status overall and by each visit separated by treatment group
- Summary of the time-point of the first occurrence of ADAb positivity during the treatment period by treatment group. A plot of the titer by time to first ADAb positivity will be prepared.
- All individual participant-level ADAb results will be listed.
- The number and percentage of participants in each of the 8 ADAb categories during the treatment period by treatment group.
- The prevalence of immunogenicity, separated by treatment group, and defined subcategory, will be reported by visit, defined as (cumulative) proportion of participants having confirmed positive ADAb samples at any visit up to and including that visit. Missing samples will not be included in the denominator.

- The time to achieving treatment-emergent ADAb positivity, separated by treatment group and defined subcategory, will be analyzed based on Kaplan-Meier methods. **This will be shown only for Categories 2 and 8 above.** Participants will be considered to have an event at the time point at which treatment emergent ADAb positive is first achieved (taking the MSR into consideration for sub-category 5). Participants classified as treatment-emergent ADAb negative will be censored at the time of the last available ADAb result.
- A summary of $HiSCR_{50}$ responders at Week 16, separated by treatment group, as a function of ADAb titer will be presented graphically.~~This will be repeated for $HiSCR_{75}$ responders.~~
- Individual plots of plasma bimekizumab concentrations/ ADAb titer both plotted on the Y-axes by visit (x-axis) for the full treatment period (excluding SFU for interim analyses and including SFU for final analyses) will be presented for participants with and without $HiSCR_{50}$ response at Week 16.
- Spaghetti plots of ADAb titer (y-axis) by visit (x-axis), separated by treatment group for all ADAb positive participants, including Baseline positive participants.
- Box plots of ADAb titer (logscale) by time to first ADAb positivity by treatment group.

~~For purposes of efficacy subgroup analyses based on anti-bimekizumab antibody status, 2 categories will be used:~~

- ~~ADAb positive – This is defined as participants who have ADAb levels above the specified cut point on at least 2 time points while on treatment (ie, excluding Baseline, excluding SFU).~~
- ~~ADAb negative – Participants who are not defined as anti-bimekizumab positive (as described above) will be defined as ADAb negative.~~

The groups for defining ADAb status for safety subgroup analyses are as follows:

- AEs starting before first ADAb positive result
- AEs starting on or after first ADAb positive result
- AEs for participants who were always ADAb negative

This is further explained in Section 10.2.2.

Change #35

The following section was added:

COVID-19 related considerations (Section 10.2.1.1)

To assess the impact of COVID-19 mass vaccination on TEAEs, a sensitivity analysis will present all TEAEs excluding TEAEs assessed as exclusively related to COVID-19 vaccine by the investigator. TEAEs recorded as related to both study medication and COVID-19 vaccination should not be excluded. A complementary table and listing of TEAEs related to COVID-19 vaccine will be presented.

Another sensitivity analysis will present all TEAEs excluding TEAEs with start date on or up to 5 days after date of COVID-19 vaccine. Note that study participants may receive more than one

administration of COVID-19 vaccine. A complementary table and listing of TEAEs with start date on or up to 5 days after date of COVID-19 vaccine will also be presented.

Change #36

Section 10.2.2 AE summaries

The following AE summaries were added:

- Incidence of TEAEs Excluding TEAEs Exclusively Related to COVID-19 Vaccine by SOC, HLT, and PT
- Incidence of TEAEs Exclusively Related to COVID-19 Vaccine by SOC, HLT, and PT
- Incidence of COVID-19 Vaccine Interval Censored TEAEs by SOC, HLT, and PT
- Incidence of COVID-19 Vaccine Interval TEAEs by SOC, HLT, and PT

Change #37

Section 10.2.2 AE summaries

The following text was added:

The following table will be presented for the combined Initial and Maintenance Treatment Period. **This summary will include only AEs that occur while a participant is on bimekizumab. Any AEs in the Initial Treatment Period that begin while a participant is on placebo will be excluded.**

- Incidence of TEAEs per 100 subject years by SOC, HLT, and PT and by Time of Onset Relative to Anti-bimekizumab Antibody Status. **This will include columns for the following:**
  - **TEAEs starting before the first ADAb positive result (includes ADAb categories 2 and 5) where TEAEs have occurred before the following events: a) the first positive ADAb result for subjects in category 2 and b) the first post-Baseline boosted ADAb titer result for subjects with titer results and the first post-Baseline positive ADAb result for subjects with positive ADAb at Baseline with no other samples with titer available for subjects in category 5**
  - **TEAEs starting on the same date or after the first ADAb positive result (includes ADAb Categories 2, 3, 4, 5 and 6) where TEAEs have occurred on or after the following events: a) the first positive ADAb results for subjects in categories 2, 3, 4 and 6, and b) the first post-Baseline boosted ADAb titer result for subjects with titer results and the first post-Baseline positive ADAb result for subjects with positive ADAb at Baseline with no other samples with titer available for subjects in category 5**
  - **TEAEs for subjects who are ADAb negative at all timepoints (includes ADAb Category 1)**

The tables with risk differences will also be accompanied by figures (dot plots) which show the incidence of the adverse events and corresponding 95% risk difference confidence intervals. These will be ordered by descending order of risk difference (bimekizumab vs placebo).

Change #38

Section 10.2.3.3 Major adverse cardiac event

The entire section was updated:

Potential cardiovascular events are adjudicated by the independent Cardiovascular Event Adjudication Committee (CV-CAC) according to the CV-CAC Charter (version 6.0). Adjudicated events are classified by the CV-CAC to one of the event types as defined in Table 10–1. The classification of an event as a Major Adverse Cardiac Event (MACE) is also determined by the CV-CAC. Events which are classified by the CV-CAC as any of the event types identified in the third column of Table 10–1 will be considered an extended MACE. Note that extended MACE is determined programmatically and includes a broader scope definition of MACE.

MACE as determined by the CV-CAC will be presented in a table and listing. Extended MACE will be presented separately in a table and listing.

Another table and listing will present the adjudicated cardiovascular events by type. For each cardiovascular event type, the individual PTs which fall within each event type will be summarized. This listing will indicate whether each event was determined to be a MACE and/or an extended MACE.

Additionally, a listing of all events identified for potential review by the CV-CAC will be produced. This listing will indicate whether each event was identified by the CV-CAC Chair for full committee review.

Table 13–1: Cardiovascular event classifications

Event Type Code	Event Type	Extended MACE
1	Non-Fatal Myocardial Infarction (MI)	Yes
2	Non-Fatal Stroke: hemorrhagic	Yes
3	Non-Fatal Stroke: ischemic	Yes
4	Non-Fatal Stroke: embolic	Yes
5	Non-Fatal Stroke: undeterminable	Yes
6	Hospitalization or ER for Unstable Angina with urgent revascularization	Yes
7	Hospitalization or ER for Unstable Angina without urgent revascularization	No
8	Hospitalization for Heart Failure	Yes
9	Transient Ischemic Attack (TIA)	No
10	Coronary Revascularization Procedures (e.g. percutaneous coronary intervention, coronary artery bypass grafting)	Yes

Table 13–1: Cardiovascular event classifications

Event Type Code	Event Type	Extended MACE
11	Urgent Revascularization Procedures (i.e. due to symptoms of brain ischemia or pending infarction)	Yes
12	Arrhythmia (not associated with ischemia)	No
13	Peripheral Arterial Event	No
14	Venous Thromboembolic Event: pulmonary embolism (PE)	No
15	Venous Thromboembolic Event: deep vein thrombosis (DVT)	No
16	Venous Thromboembolic Event: PE and DVT	No
17	Other CV Event	No
18	Death due to Myocardial Infarction (MI)	Yes
19	Death due to Stroke	Yes
20	Sudden Cardiac Death	Yes
21	Other CV Death (e.g. heart failure, pulmonary embolism, cardiovascular procedure-related)	Yes
22	Cardiovascular: Undetermined Cause of Death (i.e. cause of death unknown)	Yes
23	Non-Cardiovascular Death	No
24	Non-Cardiovascular Event	No
99	Inadequate information to adjudicate	No

CV=Cardiovascular; DVT=Deep Vein Thrombosis; ER=Emergency Room; MACE= Major Adverse Cardiac Event; MI=Myocardial Infarction; PE=Pulmonary Embolism; TIA=Transient Ischemic Attack.
MACE is determined by the adjudication committee and is not identified programmatically based on event type.

Change #39

Section 10.2.3.5 Suicidal Ideation and Behavior

The entire section was updated:

Potential neuropsychiatric events are adjudicated by the independent Neuropsychiatric Adjudication Committee according to the Neuropsychiatric Adjudication Committee (version 8.0). Adjudicated events are classified by the Committee as Suicidal or Non-suicidal. Adjudicated events are also further classified by the Committee to one of the event types as

defined in Table 10–2. Suicidal Ideation and Behavior (SIB) is defined as events classified by the Committee as Suicidal.

A table and listing will present SIB events.

Another table and listing will present the adjudicated neuropsychiatric events by type. For each neuropsychiatric event type, the individual PTs which fall within each event type will be summarized. This listing will indicate whether each event was determined to be Suicidal or Non-Suicidal. For event type suicidal ideation, the listing will also indicate if intent was present and if the suicidal ideation was clinically significant.

Additionally, a listing of all events identified for potential review by the Committee will be produced. This listing will indicate whether each event was identified by the Neuropsychiatric Event Adjudication Committee Chair for full committee review.

Table 13–2: Neuropsychiatric event classifications

Event Type Code	Event Classification	Event Type
1	Suicidal	Suicidal events/completed suicide
2	Suicidal	Suicide attempt
3	Suicidal	Preparatory acts toward imminent suicidal behavior
4	Suicidal/Non-suicidal[a]	Suicidal ideation
7	Non-suicidal	Nonsuicidal Self-injurious behavior
8	Non-suicidal	Nonsuicidal Other
99	Not applicable	Inadequate information to adjudicate

[a] Suicidal ideation event types can be classified by the Neuropsychiatric Adjudication Committee as Suicidal or Non-suicidal depending on whether intent to die was present.

Change #40

Section 10.2.3.6 Inflammatory bowel disease

The entire section was updated:

Selected gastrointestinal events are adjudicated by the independent Inflammatory Bowel Disease (IBD) Adjudication Committee (IBD-CAC) according to the IBD-CAC Charter (version 3.0). Adjudicated events are classified by the IBD-CAC into one of the diagnostic types as defined in Table 10–3. The events will further be classified as definite, probable or possible IBD.

An overview of adjudicated IBD events will be stratified by subjects with or without a previous medical history of IBD. Previous medical history of IBD will be determined using the information recorded on the History of IBD CRF page ("Does subject have a history of IBD?"). This overview table will present events adjudicated by the IBD-CAC as either possible, probable or definite IBD. Definite and probable IBD will also be aggregated and summarized in this table.

In addition, this table will summarize each IBD event classification (possible, probable or definite) separately.

Another table and listing will present the adjudicated IBD events by type. For each IBD event type, the individual PTs which fall within each event type will be summarized.

Additionally, a listing of all events identified for potential review by the IBD-CAC will be produced. This listing will indicate whether each event was identified by the IBD-CAC Chair for full committee review.

A further supportive listing will present the individual diagnostic criteria met for each adjudicated IBD event.

Table 13–3: IBD event classifications

Event Type Code	Event Type (Classification and diagnosis)	Classification
1	Possible Inflammatory Bowel Disease – Crohn's Disease	Possible
2	Probable Inflammatory Bowel Disease – Crohn's Disease	Probable
3	Definite Inflammatory Bowel Disease – Crohn's Disease	Definite
4	Possible Inflammatory Bowel Disease – Ulcerative Colitis	Possible
5	Probable Inflammatory Bowel Disease – Ulcerative Colitis	Probable
6	Definite Inflammatory Bowel Disease – Ulcerative Colitis	Definite
7	Possible Inflammatory Bowel Disease – type unclassified	Possible
8	Probable Inflammatory Bowel Disease – type unclassified	Probable
9	Definite Inflammatory Bowel Disease – type unclassified	Definite
10	Symptoms not consistent with Inflammatory Bowel Disease	Not applicable
11	Possible Inflammatory Bowel Disease – Microscopic Colitis	Possible
12	Probable Inflammatory Bowel Disease – Microscopic Colitis	Probable
13	Definite Inflammatory Bowel Disease – Microscopic Colitis	Definite
14	Possible Inflammatory Bowel Disease – no further differentiation possible	Possible
15	Probable Inflammatory Bowel Disease – no further differentiation possible	Probable
16	Definite Inflammatory Bowel Disease – no further differentiation possible	Definite
99	Not enough information to adjudicate	Not applicable

IBD=inflammatory bowel disease.
Note: IBD diagnoses of "microscopic colitis" and "no further differentiation possible" were added in an adjudication charter amendment, accounting for the event type numbering.

Change #41

Section 10.2.3.8 Hepatic events and PDILI

The following word was added:

Cases of **potential** Hy's Law will be reported separately in a liver function test table.

Change #42

Section 12.1 Appendix A: MedDRA algorithmic approach to aanphylaxis

The following text was added:

- An **algorithmic approach** which combines a number of anaphylactic reaction symptoms in order to increase specificity. A case must include one of the following where both occur on either the same day as when an injection was administered or one day after, and for scenarios where two events must have been reported, both events must have occurred within one day of each other **(as anaphylaxis is an acute event, imputed dates should not be used in the algorithmic approach):**

13.3 Amendment 3

Rationale for the amendment

The main purpose of this amendment was:

- Change the Flare by Week 16 endpoint from a secondary endpoint included in the statistical hierarchy to an other efficacy endpoint, in alignment with changes described in protocol amendment 5
- Clarify the tipping point analysis procedure

Modifications and changes

Global Changes

The following changes were made throughout the SAP:

- Moving the Flare by Week 16 endpoint from any description of secondary efficacy endpoints, and re-inserting this endpoint as an other efficacy endpoint

Specific Changes

In addition to the global changes, the following specific changes have been made (typos such as missing spaces or redundant spaces are not listed):

Change #1

Section 1 Introduction

The protocols were updated:

The SAP is based on the Protocol Amendment 5, 27 September 2022.

Change #2

Flare by Week 16 was removed as a secondary efficacy endpoint:

2.2.1.2 Secondary efficacy endpoints

The secondary efficacy endpoints are defined as:

- $HiSCR_{75}$ response (defined as at least a 75% reduction from Baseline in the total AN count with no increase from Baseline in abscess or draining tunnel count) at Week 16
- ~~Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Baseline) by Week 16~~
- Absolute change from Baseline in Dermatology Life Quality Index (DLQI) Total Score at Week 16
- Absolute change from Baseline (CFB) in Skin Pain score at Week 16, as assessed by the "worst skin pain" item (11-point numeric rating scale) in the HS Symptom Daily Diary (HSSDD)
- Skin pain response based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) at Week 16 among study participants with a score of ≥3 at Baseline

Change #3

Flare by Week 16 was added as an other efficacy endpoint:

2.2.1.3 Other efficacy endpoints

The other efficacy endpoints are defined as:

- Time to response of $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$
- $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$
- Absolute change from Baseline in International Hidradenitis Suppurativa Severity Score System (IHS4)
- Change from Baseline in the HS-Physician's Global Assessment 6-point scale
- Absolute and percentage change from Baseline in high-sensitivity C-reactive protein (hs-CRP)
- Initiation of systemic antibiotic rescue therapy
- $HiSCR_{25}$, $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ at both Weeks 16 and 48
- Time to loss of response of $HiSCR_{50}$, $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ in Week 16 responders
- Partial responders (defined as a ≥25% reduction in AN count from Baseline [Week 0]) at Week 16 who progress to $HiSCR_{50}$ during the Maintenance Treatment Period
- Partial responders (defined as a ≥25% reduction in AN count from Baseline [Week 0]) at Week 16 who progress to $HiSCR_{75}$, $HiSCR_{90}$, and $HiSCR_{100}$ response during the Maintenance Treatment Period
- Change and percentage change from Baseline in lesion counts (abscess count, inflammatory nodule count, AN count, and draining tunnel count)
- AN count of 0, 1, or 2

- AN_{25}, AN_{50}, AN_{75}, AN_{90}, AN_{100} (defined as a 25%, 50%, 75%, 90%, 100% reduction in the total AN count relative to Baseline)
- **Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Baseline) by Week 16**
- Flare (defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Week 16) by Week 48
- Time to flare from Weeks 0 to 16
- Time to flare from Week 16 to 48
- Absolute and percentage change (worst and average skin pain) from Baseline in HS Skin Pain score (11-point numeric rating scale)
- Skin pain response based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) among study participants with a score of ≥3 at Baseline
- Skin pain response (at least a 30% reduction and at least a 1-point reduction from Baseline in HSSDD weekly worst skin pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline
- Skin pain response (at least a 30% reduction and at least a 1-point reduction from Baseline in HSSDD weekly average skin pain score [11-point numeric rating scale]) among study participants with a score of ≥3 at Baseline
- Skin pain response (at least a 30% reduction and at least a 1-point reduction from Baseline in HS Skin Pain score [11-point numeric rating scale]) assessed by the HSSQ among study participants with a score of ≥3 at Baseline
- Absolute change from Baseline in DLQI Total Score
- DLQI Total Score of 0 or 1
- Minimum clinically important difference (MCID) (improvement from Baseline of 4 or more) in the DLQI Total Score among study participants with a Baseline score of at least 4
- Absolute change from Baseline in Hidradenitis Suppurativa Quality of Life (HiSQOL) domain scores (symptoms, psychosocial, activities and adaptations) and Total score
- Patient Global Impression of HS Severity (PGI-S-HS)
- Patient Global Impression of Change of HS Severity (PGI-C-HS)
- Patient Global Impression of Severity of Skin Pain (PGI-S-SP)
- Patient Global Impression of Change of Skin Pain (PGI-C-SP)
- Absolute change from Baseline in each of the other HS Symptoms - itch, drainage or oozing of HS lesions, and smell or odor
- Response on other HS Symptoms (11-point numeric rating scale) - itch, drainage or oozing of HS lesions, and smell or odor

- Responses to the European Quality of Life-5 Dimensions-3 Level questionnaire (EQ-5D-3L), absolute and changes from Baseline in EQ-5D-3L visual analog scale (VAS) scores
- Absolute change from Baseline in Work Productivity and Activity Impairment Questionnaire-Specific Health Problem (WPAI-SHP) v2.0 adapted to HS scores
- Domain Scores (effectiveness, convenience and global satisfaction) on the Treatment Satisfaction Questionnaire – Medication-9 (TSQM-9)

Change #4

Flare was removed from the table of power calculations:

Table 2–1: Power calculation assumptions and methods

Endpoint	Power α = 0.025, 2-sided		Assumptions		
	Q2W	**Q4W**	**Week 16 Bimekizumab Q2W N=280[a]**	**Week 16 Bimekizumab Q4W N=140**	**Week 16 Placebo N=70**
$HiSCR_{50}$	0.99	0.90	Proportion responders=0.60	Proportion responders=0.50	Proportion responders=0.25
$HiSCR_{75}$	0.99	0.98	Proportion responders=0.45	Proportion responders=0.35	Proportion responders=0.10
DLQI	0.99	0.96	Mean CFB=-5.4; SD=6.8	Mean CFB=-4.8; SD=6.8	Mean CFB=-0.8; SD=6.6
~~Flare[b]~~	~~0.99~~	~~0.99~~	~~Proportion of participants with flare by Week 16=0.09~~	~~Proportion of participants with flare by Week 16=0.19~~	~~Proportion of participants with flare by Week 16=0.52~~
Worst Skin Pain CFB[b]	0.89	0.73	Mean CFB=-2.2; SD=3.2	Mean CFB=-2.0; SD=3.2	Mean CFB=-0.5; SD=3.7
Worst Skin Pain Response[c]	0.95	0.53	Proportion responders=0.53	Proportion responders=0.43	Proportion responders=0.23

CFB=change from Baseline; Q2W=every 2 weeks; Q4W=every 4 weeks; SD=standard deviation
Note: Estimates for HS0003 are based on Week 12 data from the HS0001 study.
[a] Pooled Q2W at Week 16 from Q2W/Q2W and Q2W/Q4W arms
[b] Within-participant average of Worst Skin Pain according to 24-hour recall.
[c] Assumes N=208, 104, 52 in Q2W, Q4W, and placebo, respectively, to account for Worst Skin Pain score at or above 3 (ie, the threshold for clinically meaningful change from Baseline).

Change #5

The language in **Section 4.2.2.3 Tipping Point Analysis** was updated:

For tipping point analyses, data for participants after the intercurrent event date (See Section 3.9) will be changed to missing prior to imputation ~~but~~ **and, for the bimekizumab treated participants**, will ~~not~~ be changed to non-response after imputation.

Change #6

The language in **Section 4.2.3 Rationale for estimand** was updated:

- A composite estimand strategy will be used for the primary analysis of the primary and binary secondary endpoints ($HiSCR_{50}$, $HiSCR_{75}$, ~~flare,~~ HS worst skin pain response).

Change #7

Flare was removed from the testing procedure to control for multiplicity:

4.5 Multiple comparisons/multiplicity

To control the overall type I error rate at 0.05 for the multiple comparisons in the primary and secondary efficacy endpoints, a closed testing procedure under a parallel gatekeeping framework will be applied (Sun, 2018).

Under this framework, each bimekizumab dose of 320mg Q2W and 320mg Q4W will be compared to placebo in the first instance at a familywise error rate of 0.025 (α/2). Simultaneously within each dose, closed testing for the primary and secondary efficacy endpoints will be performed as follows:

1. Step 1: Test $HiSCR_{50}$ at significance level 0.025.
2. Steps 2 to 5~~6~~ – If Step 1 is significant at 0.025 then test each secondary endpoint sequentially in the order shown in Figure 4-1, moving to the next step only if significance achieved at 0.025.
3. In the event that Step 5~~6~~ is significant at 0.025 for a given dose, then Steps 1 to 5~~6~~ will be repeated for the other dose using a significance level of 0.05.

The secondary efficacy variables supporting the primary efficacy variable are listed below, and will be included in the multiplicity adjustment using the analysis methods specified in Section 8.3:

1. Proportion of study participants who achieve $HiSCR_{75}$ at Week 16.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

~~2 Proportion of study participants who experience at least 1 flare by Week 16, with flare defined as a ≥25% increase in AN count with an absolute increase in AN count of ≥2 relative to Baseline.~~

~~a. bimekizumab 320mg Q2W vs placebo~~

~~b. bimekizumab 320mg Q4W vs placebo~~

2. Absolute CFB in DLQI Total Score at Week 16.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

3. Absolute CFB in Skin Pain Score at Week 16, as assessed by the "worst skin pain" item (11-point numeric rating scale) in the HSSDD.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

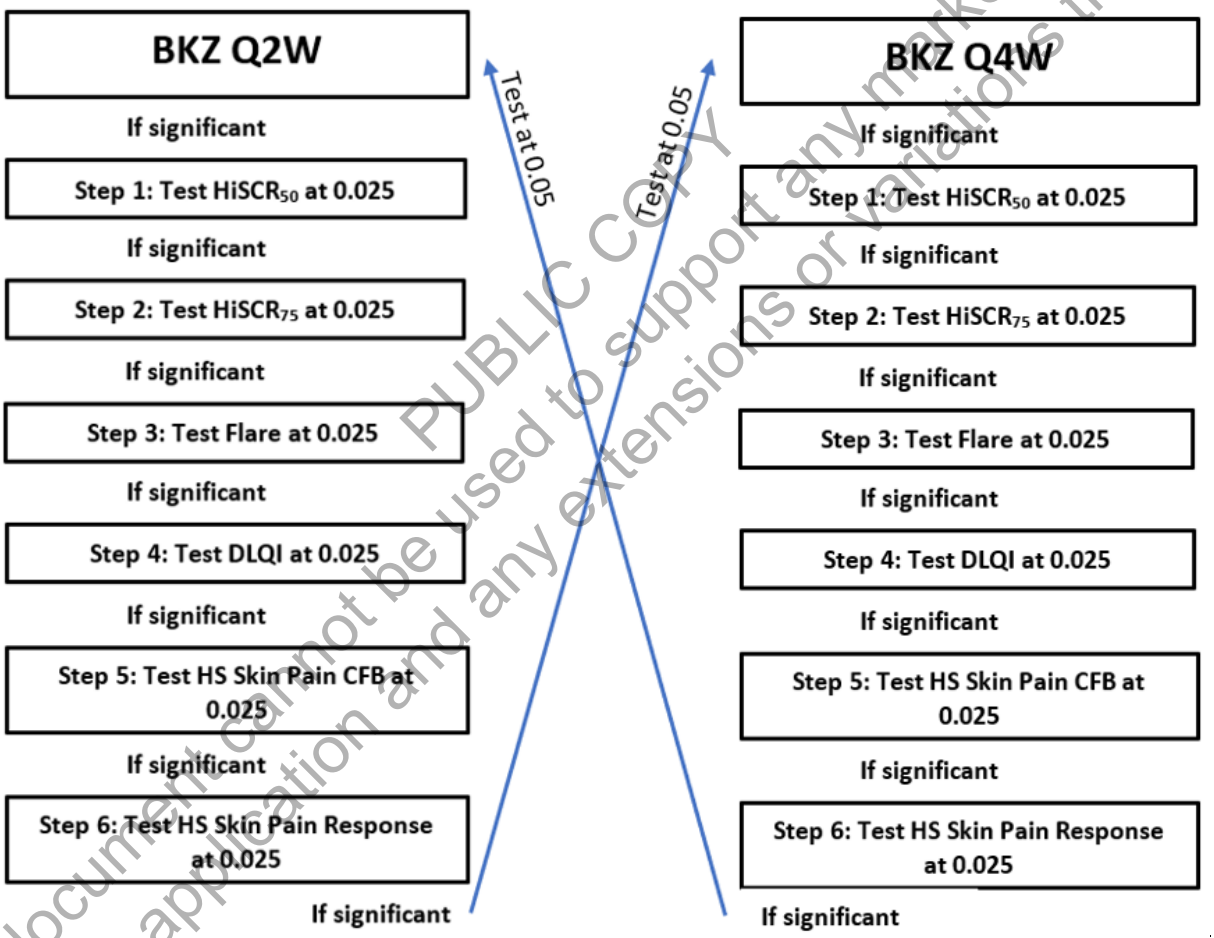
4. Skin pain response at Week 16, based on the threshold for clinically meaningful change (defined as at least a 3 point decrease from Baseline in HSSDD weekly worst skin pain score) at Week 16 among study participants with a score of ≥3 at Baseline.

a. bimekizumab 320mg Q2W vs placebo

b. bimekizumab 320mg Q4W vs placebo

Change #8

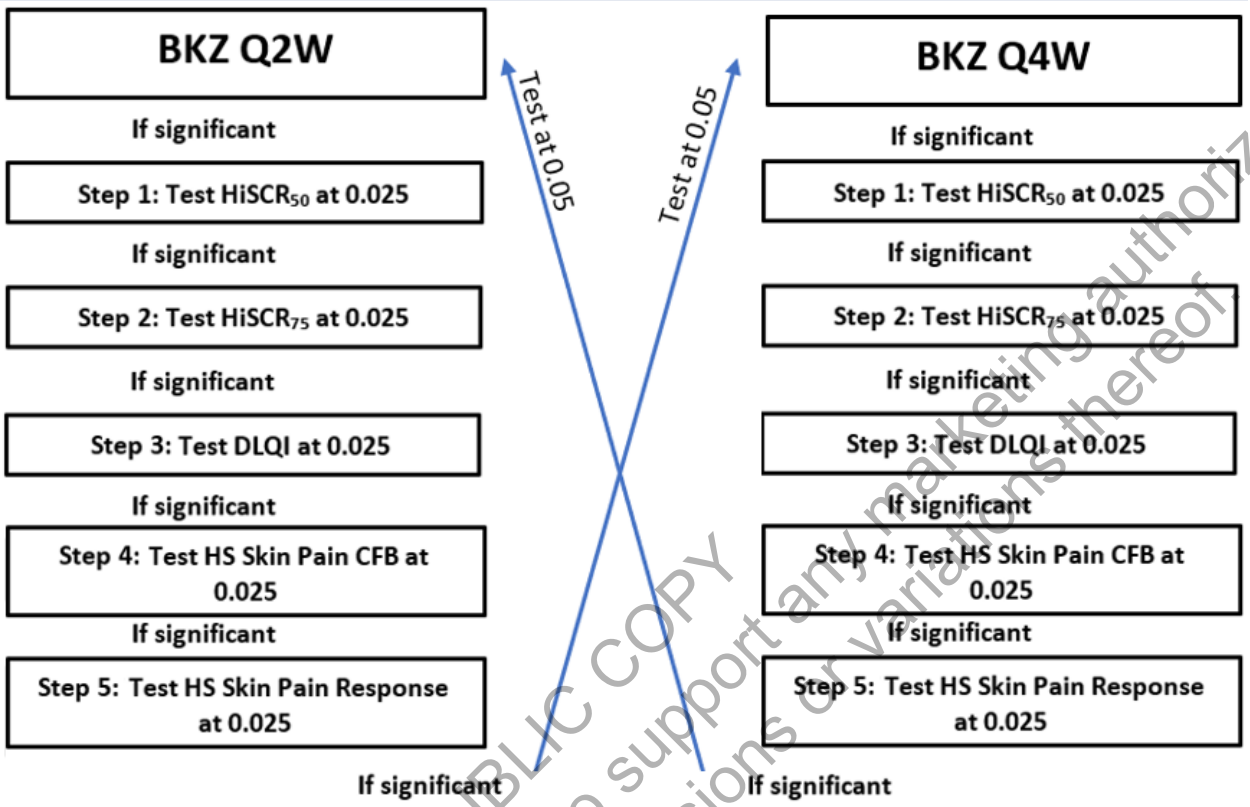
The sequence of testing was changed from the 6-Step schematic below:



AN=abscess and inflammatory nodule; DLQI=Dermatology Life Quality Index; $HiSCR_{50}$=a 50% reduction in the total AN count with no increase from Baseline in abscess or draining tunnel count; $HiSCR_{75}$=a 75% reduction in the total AN count with no increase from Baseline in abscess or draining tunnel count; HS=hidradenitis Suppurativa; Q2W=every 2 weeks; Q4W=every 4 weeks
HS skin pain response is tested among study participants with a score of ≥3 at Baseline.

To the 5-Step schematic below, which reflects the change to the statistical hierarchy:

Figure 4-1: Sequence of testing



AN=abscess and inflammatory nodule; DLQI=Dermatology Life Quality Index; $HiSCR_{50}$=a 50% reduction in the total AN count with no increase from Baseline in abscess or draining tunnel count; $HiSCR_{75}$=a 75% reduction in the total AN count with no increase from Baseline in abscess or draining tunnel count; HS=hidradenitis Suppurativa; Q2W=every 2 weeks; Q4W=every 4 weeks
HS skin pain response is tested among study participants with a score of ≥3 at Baseline.

Change #9

The language in **Section 4.8 Examination of subgroups** was updated as follows:

Subgroup analyses will be performed on the $HiSCR_{50}$, $HiSCR_{75}$, ~~flare,~~ and worst skin pain response endpoints by visit for the Initial Treatment Period and Maintenance Treatment Period. Additional subgroup analyses will be performed on the CFB in the worst skin pain score as measured by HSSDD and in the DLQI total score through Week 16 as described below.

Along with the tables described, there will be tables for $HiSCR_{50}$, $HiSCR_{75}$, ~~flare,~~ and skin pain response endpoints which display the response difference and 95% CIs between each bimekizumab dose regimen versus placebo for each of the subgroups at Week 16. Corresponding forest plots will be prepared.

Change #10

Flare was removed from **Table 8-2: Estimand Details and Attributes for Secondary Analyses** as follows:

Table 8-2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
Secondary Objective: Evaluate the efficacy of bimekizumab on other measures of disease activity in study participants with moderate to severe HS					
$HiSCR_{75}$	Secondary (Section 8.3.1)	$HiSCR_{75}$ response at Week 16	RS	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo based on a logistic regression, as for the primary analysis.
~~Flare~~	~~Secondary (Section 8.3.2)~~	~~Flare by Week 16~~	~~RS~~	~~The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. A **composite strategy** will be used, ie, the occurrence of an intercurrent event will be handled by evaluating the corresponding participants as treatment failures (flare).~~	~~The odds ratio versus placebo based on a logistic regression, as for the primary analysis.~~

Table 8-2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
DLQI	Secondary (Section 8.3.2.1)	Change from Baseline in DLQI total score to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **hypothetical strategy**, whereby all data at and after the intercurrent event will be treated as missing.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the DLQI total score for participants receiving bimekizumab versus placebo. Missing values will be imputed using MI – MCMC/Monotone Regression under a missing at random assumption.
DLQI	Secondary - Sensitivity (Section 8.3.2.2)	Change from Baseline in DLQI total score to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a DLQI total score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the DLQI total score for participants receiving bimekizumab versus placebo. Missing values will not be imputed.

Table 8-2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
HSSDD	Secondary (Section 8.3.3.1)	Change from Baseline in worst skin pain score, as assessed by "worst skin pain" item in HSSDD to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **hypothetical strategy**, whereby all data at and after the intercurrent event will be treated as missing.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the worst skin pain score for participants receiving bimekizumab versus placebo. Missing values will be imputed using MI – MCMC/Monotone Regression under a missing at random assumption.
HSSDD	Secondary (Section 8.3.3.2)	Change from Baseline in worst skin pain score, as assessed by "worst skin pain" item in HSSDD to Week 16	RS	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a HSSDD worst skin pain score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	Difference in mean change from Baseline (LSMD from ANCOVA) to Week 16 in the worst skin pain score for participants receiving bimekizumab versus placebo. Missing values will not be imputed.

Table 8-2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
HSSDD	Secondary (Section 8.3.4.1)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improve-ment in the weekly worst skin pain score of at least 3 points	RS[a]	**Composite strategy**, as for the primary analysis.	The odds ratio versus placebo based on a logistic regression, as for the primary analysis.
HSSDD	Secondary Sensitivity (Section 8.3.4.2.1)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improve-ment in the weekly worst skin pain score of at least 3 points	RS[a]	**Composite strategy**, as for the primary analysis where the occurrence of an intercurrent event will be handled by evaluating the corresponding participants as nonresponders.	The odds ratio versus placebo based on a logistic regression. Missing values for any other reason will also be imputed as nonresponders.

Table 8-2: Estimand Details and Attributes for Secondary Analyses

		Estimands for Secondary Endpoints			
Objective Clinical Category	**Statistical Category (Section)**	**Variable/ Endpoint**	**Pop**	**IES**	**PLS (Analysis)**
HSSDD	Secondary Sensitivity (Section 8.3.4.2.2)	Skin pain response at Week 16, as assessed by the "worst skin pain" item in the HSSDD, defined as an improvement in the weekly worst skin pain score of at least 3 points	RS[a]	The main intercurrent events are receipt of systemic antibiotic rescue medication or discontinuation of study treatment due to an AE or lack of efficacy prior to Week 16. The intercurrent events will be handled using a **while on treatment strategy**, whereby only participants with a HSSDD worst skin pain score at Week 16 who have not had an intercurrent event on or before Week 16 are included.	The odds ratio versus placebo is based on a logistic regression. Missing values will not be imputed.

AE=adverse event; ANCOVA=analysis of covariance; DLQI=Dermatology Life Quality Index; HiSCR=Hidradenitis Suppurativa Clinical Response; HSSDD=Hidradenitis Suppurativa Symptom Daily Diary; IES=intercurrent event(s) strategy; LSMD=Least Squares Mean Difference; MCMC=Markov Chain Monte Carlo; MI=multiple imputation; PLS=Population-level summary; Pop=Population; RS=Randomized Set

[a] Analysis includes all study participants in the RS with a Baseline HSSDD Worst Skin Pain score of 3 or higher.

Change #11

Section 8.3.2 Flare by Week 16 was removed, and re-inserted **as Section 8.4.3 Flare by Week 16**. The number of surrounding sections was subsequently updated.

Change #12

Section 8.5 Flare by Week 16 was added to the list of nominal tests.

- $HiSCR_{50}$ at Week 12
- $HiSCR_{75}$ at Week 12
- $HiSCR_{90}$
- $HiSCR_{100}$
- Flare by Week 12
- Flare by Week 16

- Time to flare by Week 12 (based on time-to-event analysis per Section 8.4.5 and adjusted appropriately)
- IHS4 change from Baseline
- IHS4 percentage change from Baseline
- HS Physician's Global Assessment: rate of participants who are Clear or Mild
- DLQI total score change from Baseline at Week 12
- Worst Skin Pain per HSSDD change from Baseline at Week 12
- Skin Pain response per HSSDD at Week 12

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

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

Approval Signatures

Name: hs0003-sap-amend-3

Version: 1.0

Document Number: CLIN-000204829

Title: hs0003-sap-amend-3

Approved Date: 07 Oct 2022

Document Approvals	
Approval Verdict: Approved	Name: [redacted] Capacity: Clinical Date of Signature: 07-Oct-2022 14:23:43 GMT+0000
Approval Verdict: Approved	Name: [redacted] Capacity: Clinical Date of Signature: 07-Oct-2022 16:26:25 GMT+0000