J2T-DM-KGAD SAP v4

A Randomized, Double-Blind, Placebo-Controlled Trial to Evaluate the Efficacy and Safety of Lebrikizumab When Used in Combination With Topical Corticosteroid Treatment in Patients With Moderate-to-Severe Atopic Dermatitis

NCT04250337

Approval Date: 15-Oct-2021

1. Statistical Analysis Plan: J2T-DM-KGAD (DRM06-AD06): A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL TO EVALUATE THE EFFICACY AND SAFETY OF LEBRIKIZUMAB WHEN USED IN COMBINATION WITH TOPICAL CORTICOSTEROID TREATMENT IN PATIENTS WITH MODERATE-TO-SEVERE ATOPIC DERMATITIS

Lebrikizumab (LY3650150)

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol J2T-DM-KGAD
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly on 19 June 2020.
Statistical Analysis Plan Version 2 electronically signed and approved on by Lilly on 13 May 2021.
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly on 10 September 2021.
Statistical Analysis Plan Version 4 electronically signed and approved on by Lilly on the te provided below.

Approval Date: 15-Oct-2021 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table of Contents

Figure **Page**

Table of Contents

Appendix **Page**

3. Revision History

Statistical Analysis Plan (SAP) Version 1 was approved prior to any unblinding on 19 June 2020.

Statistical Analysis Plan Version 2 was approved prior to any unblinding on 13 May 2021.

Statistical Analysis Plan Version 3 was approved prior to any unblinding on 10 September 2021.

Following primary database lock, a site audit with a critical finding necessitated revision of the Statistical Analysis Plan (SAP). SAP Version 4 was prepared and approved by statisticians who were independent from the study team and blinded to study-level and patient-level data at the time of SAP amendment. The limited number of study team statisticians who were unblinded during the primary database lock did not participate in revision of the SAP. No other study team members were unblinded to the data during the primary database lock.

Changes in Version 2, Version 3, and Version 4 are documented in the following 3 tables. Minor corrections/additions may not be included.

Revisions in SAP Version 2

Section	Action
Section 4	Rearranged Table KGAD.4.1 to clearly list the major secondary endpoints in the table.
Section 4, 6.6, and 6.11	Added major secondary endpoints: Percentage of patients with a Sleep-loss score ≥2 points at Baseline who achieve a ≥2 points reduction from Baseline to Week 16.
Section 4 and 6.11	Added other secondary endpoints: mean weight of TCS use, percentage of patients with SQAAQ response, Percentage of patients rescued by visit.
Section 4	Removed Percentage of patients with Pruritus NRS change of ≥4 from Baseline by visit.
Section 4, 6,6 and 6.11	Changed "Percentage of patients who achieve a ≥4-point improvement from baseline to Week 16" to "Percentage of patients with a DLQI total score of ≥4-points at Baseline who achieve a ≥4-point improvement from baseline to Week 16".
Section 4, 6.6, and 6.11	Separated DLQI and CDLQI, DLQI will be included in the EMA major secondary endpoints.
Section 4 and 6.11	Removed endpoints related to patients with Pruritus NRS ≥5-points at Baseline: Percentage of patients who achieve a ≥4-point improvement from baseline to Week 16, Week 4, Week 2, Week 1.
Section 4, 6.6, and 6.11	Revised "Percent change from baseline" to "Change from Baseline" for sleep-loss score at Week 16.
Section 6.1.2	Added "For patients who are randomized but not dosed, the Treatment Period starts on the date of randomization." For Pruritus Numeric Rating Scale (NRS) and Sleep-Loss due to Pruritus collected via eDiary, the baseline period has been updated to the 7-day window on or prior to the first injection.

Section	Action
Section 6.2	This section has been amended to implement the definition of primary and supportive estimands following ICH E9(R1) addendum. Added the definition of supportive estimands for both categorical endpoints and continuous endpoints. Added the missing data imputation methods relative to each estimand.
Section 6.4	This section has been amended to align with the definition of estimands. Removed all missing values MCMC-MI from sensitivity analyses, keeping tipping point analyses as the only sensitivity analyses for the primary estimand. Updated tipping point analysis: all subjects who use rescue medication need to be imputed as nonresponders prior to varying the response and nonresponse rates for those with missing data.
Section 6.6	Updated graphical testing scheme for multiplicity control of primary and major secondary endpoints for US. Modified multiplicity strategy for EMA, replacing serial gatekeeping procedure with graphical testing scheme.
Section 6.8.1	Added following to baseline disease characteristics: Sleep loss due to pruritus: <2, ≥2; EQ-5D US Population-based index score; EQ-5D UK Population-based index score. Separated DLQI and CDLQI. Baseline ethnicity distribution is for US patients only.
Section 6.9	Updated instructions to calculate the total number of injections.
Section 6.10	*Prior medications* are those medications that start prior to the date of first dose and stop prior to or on the date of first dose of study treatment. Removed the description of summary of Atopic Dermatitis treatment of interest. Consolidated the summary of Atopic Dermatitis treatment of interest with the summary of rescue medications. Added definition of flare.

Section	Action
Section 6.11	Removed analyses for itch-free days and no sleep loss days. Updated the derivation of BSA Total. Updated the derivation of postbaseline weekly mean for Pruritus and Sleep loss to prorated weekly mean. Separated analysis for DLQI and CDLQI total scores. Updated the following analysis method to ANCOVA with LOCF: Change from baseline in SCORAD scores, ED-5D variables, PROMIS scores, and ACQ-5 score.
Section 6.11.2	This section has been updated to reflect the change in the sensitivity analyses for primary outcomes.
Section 6.14	This section has been updated to be in alignment with compound level safety standard. Added "Drug interruption time period due to the use of systemic rescue therapies will be removed from study drug exposure calculations as described in compound level safety standards." Added Section of Atopic Dermatitis Exacerbation and Section of Suicidal Ideation and Behavior. Removed listing of exposure.
Section 6.15	Clarified ethnicity subgroup analysis is for US patients only Removed subgroup analyses for: TE ADA subgroup and TE ADA Nab subgroup
Section 8	Added reference: Bretz 2009 and Ratitch 2013
Appendix 1	Updated: replaced "assessment date" with "Office visit date". If multiple assessments occurred on a single day, use the first assessment for analyses. Added a statement for PEOM data analysis. Clarified the weekly mean to be prorated mean.
Appendix 2	Added Appendix 2: Definition of Topical and Systemic Atopic Dermatitis Therapy
Appendix 3	Added Appendix 3: Details of Combining Estimates and Test Statistics for Categorial Endpoints with Multiple Imputation
Throughout the document	Minor editorial changes and/or clarifications were made.

Revisions in SAP Version 3:

Section	Description of Change	Rationale
Section 4	• Added back several endpoints as other secondary endpoints.	To be consistent with protocol and CT.gov
Section 4, Section 6.6	• Removed "Percentage of patients with a Pruritus NRS of ≥4-points at Baseline who achieve a ≥4-point reduction from Baseline to Week 1" from the list of multiplicity controlled major secondary endpoints for FDA and EMA. • Removed "Percentage of patients with a Pruritus NRS of ≥4-points at Baseline who achieve a ≥4-point reduction from Baseline to Week 2" from the list of multiplicity controlled major secondary endpoints for FDA and EMA • Removed "Percentage of patients with a Pruritus NRS of ≥4-points at Baseline who achieve a ≥4-point reduction from Baseline to Week 4" from the list of multiplicity controlled major secondary endpoints for FDA and EMA • Removed "Percentage of patients with an IGA score of 0 or 1 and a reduction ≥2 points at Week 2." from the list of multiplicity controlled major secondary endpoints for FDA. • Removed "Percentage of patients with an IGA score of 0 or 1 and a reduction ≥2 points at Week 4." from the list of multiplicity controlled major secondary endpoints for FDA • Removed "Percentage of patients achieving EASI-90 (≥90% reduction from Baseline in EASI score) at Week 4." from the list of multiplicity controlled major secondary endpoints for EMA • Removed "Percentage of patients with a Sleep-loss score ≥2 points at Baseline who achieve a ≥2 points reduction from Baseline to Week 16" from the list of multiplicity controlled major secondary endpoints for FDA and EMA • Moved "Percentage of patients with an IGA score of 0 or 1 and a reduction ≥2 points at Week 16 in adults" from the list of multiplicity controlled major secondary endpoints to the "Other secondary endpoint" category	Strategy change in multiplicity control
Section 4, Section 6.11	• Added following major secondary endpoint: Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both EASI-75 and a ≥4-point reduction in Pruritus NRS score from Baseline at Week 16.	To allow the analysis on composite endpoints in graphical testing
Section 4, Section 6.11	• Added following other secondary endpoints: • Percentage of patients with a Pruritus NRS	To allow the analysis on composite endpoints

Section	Description of Change	Rationale
	score of ≥4 points at Baseline who achieve both EASI-75 and a ≥4-point reduction in Pruritus NRS score from Baseline by visit • Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both an IGA score of 0 or 1 and a reduction of ≥2 points in IGA score from Baseline, and a ≥4-point reduction in Pruritus NRS score from Baseline by visit	
Section 4, Section 6.11	• Removed other secondary endpoint: mean weight of TCS use.	To be consistent with protocol objective
Section 6.6	• Updated the graphical testing scheme for multiplicity control of primary and major secondary endpoints for FDA. • Added the graphical testing scheme for multiplicity control of primary and major secondary endpoints for Induction Period for EMA.	To prespecify the graphical testing scheme for US and EU based on blinded study data
Section 6.8.1	• Updated the subcategories for Atopic Dermatitis treatment used in the past. • Added prior use of systemic treatment (yes, no).	Clarification
Section 6.11	• Updated analysis method of Proportion of patients achieving IGA [0] form MCMC-MI to NRI	To be consistent with analysis methods for other secondary objectives
Section 6.14.6.5	• Removed listing of patients with hypersensitivity.	Listing of patients with hypersensitivity will be provided in the context of evaluating immunogenicity
Section 6.14.6.9	• Updated the section heading for Suicide/Self-injury.	To reflect the search strategy using SMQ code
Section 6.15.1	• Added a subgroup “Prior use of systemic treatment (yes, no)” for efficacy subgroup analysis.	To prespecify the analysis for this subgroup
Section 6.15.2	• Updated text on safety subgroup analyses may be added.	To allow safety subgroup analysis to be added
Appendix 1	• Added “If an assessment could be mapped to different weeks, it will be mapped to the earlier week.”	Clarification
Appendix 2	• Revised the formula for the transformed CMH statistic.	Correction
Appendix 3	• Added “Route of topical treatments includes: Topical and Transdermal.”	Clarification

Revisions in SAP Version 4:

Section	Description of Change	Rationale
Section 6.1.1	• Added 2 analysis populations (mITT Population and Modified Safety Population). • Removed "Unless otherwise specified, efficacy and health outcomes analyses will be conducted on this population" from the ITT Population and added to the mITT Population. • Removed "Safety analyses will be conducted on this population" from the Safety Population and added to the Modified Safety Population. • Removed "ITT" and added "mITT" and "Modified Safety" to analysis population in Table KGAD.6.2 for treatment comparison.	In a directed site audit triggered by statistically implausible data at one study site, it was determined that some or all of the study participants at the site did not meet the eligibility criterion of having moderate-to-severe atopic dermatitis, and associated data wsd unreliable.
Section 6.7	• Added patient disposition summaries for the mITT Population.	See above.
Section 6.8	• Updated analysis population from ITT to mITT for a summary of - demographics and baseline characteristics, and - medical histories.	See above.
Section 6.9	• Added treatment compliance for the Modified Safety Population.	See above.
Section 6.10	• Updated analysis population from ITT to mITT for a summary of - prior medications, and - concomitant medications. • Clarified analyses of rescue medication for AD will be based on mITT. • Clarified a listing of patients who use rescue medication will be based on ITT.	See above.
Section 6.11	• Updated analysis population from ITT to mITT for all efficacy and health outcome analyses.	See above.
Section 6.11.1	• Updated analysis population from ITT to mITT for the primary analysis of the primary outcome (IGA of 0 or 1 with at least 2-point reduction from baseline at Week 16) and the additional EMA primary outcome (EASI-75 at Week 16).	See above.
Section 6.14	• Updated the Modified Safety Population as primary analysis population for safety evaluations (exposure, adverse events, clinical laboratory data, vital signs, immunogenicity, adverse events of special interest). • Updated safety evaluations based on the Safety Population as sensitivity analysis • Clarified listings of immunogenicity assessments will be based on the Safety	See above.

Section	Description of Change	Rationale
	Population.	
Section 6.15	• Updated analysis population from ITT to mITT for efficacy subgroup analyses.	See above.
Section 6.14.5	• Removed "The subgroup analysis of efficacy by TE-ADA status (positive, negative) is described in the Section 6.15.1."	• Assessment of relationship between immunogenicity and efficacy to be performed as part of the integrated analysis including other Phase 3 lebrikizumab AD trials.
Section 6.16	• Removed languages related to the per-protocol (PP) population. • Clarified a listing of IPDs will be provided for the ITT Population.	• Per-protocol set analyses not planned. • Clarification.

Abbreviations: AD = atopic dermatitis; EASI-75 = ≥75% reduction from Baseline in Eczema Area and Severity Index score; EMA = European Medicines Agency; IGA = Investigator's Global Assessment; ITT = intent-to-treat; mITT = modified intent-to-treat; TE-ADA = treatment-emergent antidrug antibody.

4. Study Objectives

Table KGAD.4.1 shows the objectives and endpoints of the study. In addition, the analysis of the endpoints is described in Section 6.11 to provide supportive evidence of efficacy.

Table KGAD.4.1. Objectives and Endpoints

Study Objective: To evaluate the safety and efficacy of lebrikizumab in combination with TCS compared with placebo in combination with TCS in patients with moderate-to-severe AD	
FDA Endpoints	**EMA Endpoints**
Primary Percentage of patients with an IGA score of 0 or 1 and a reduction ≥2 points from Baseline to Week 16.	Percentage of patients with an IGA score of 0 or 1 and a reduction ≥2 points from baseline to Week 16. Percentage of patients achieving EASI-75 (≥75% reduction from Baseline in EASI score) at Week 16.
Major Secondary • Percentage of patients achieving EASI-75 (≥75% reduction from Baseline in EASI score) at Week 16 • Percentage of patients achieving EASI-90 (≥90% reduction from Baseline in EASI score) at Week 16 • Percentage of patients with a Pruritus NRS of ≥4-points at Baseline who achieve a ≥4-point reduction from Baseline to Week 16 • Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both EASI-75 and a ≥4-point reduction in Pruritus NRS score from Baseline at Week 16	• Percentage of patients achieving EASI-90 (≥90% reduction from Baseline in EASI score) at Week 16 • Percentage change in Pruritus NRS score from Baseline to Week 16 • Percentage of patients with a Pruritus NRS of ≥4-points at Baseline who achieve a ≥4-point reduction from Baseline to Week 16 • Percentage change in EASI score from Baseline to Week 16 • Change from baseline in DLQI at Week 16 • Percentage of patients with a DLQI total score of ≥4-points at baseline who achieve ≥4-point improvement in DLQI from baseline to Week 16 • Change from Baseline in Sleep-loss score at Week 16 • Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both EASI-75 and a ≥4-point reduction in Pruritus NRS score from Baseline at Week 16

Objectives and Endpoints

Study Objective: To evaluate the safety and efficacy of lebrikizumab in combination with TCS compared with placebo in combination with TCS in patients with moderate-to-severe AD	
Evaluate the pharmacokinetics of lebrikizumab. Average serum lebrikizumab concentration	Defined in a separate PK analysis plan
Other Efficacy Endpoints	
• Proportion of patients with EASI-75, EASI-90 and EASI-50 by visit • Proportion of patients with IGA Score of 0 or 1 and a reduction ≥2 points from Baseline by visit • Percentage change from Baseline in EASI Score by visit • Percentage change from Baseline in Pruritus NRS by visit • Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve a ≥4-point reduction from Baseline by visit • Percentage of patients with a Pruritus NRS score of ≥5 points at Baseline who achieve a ≥4-point reduction from Baseline by visit • Percentage of patients with Pruritus NRS change of ≥4 from Baseline by visit • Change from Baseline in Sleep-Loss score by visit • Percent change from Baseline in Sleep-Loss score by visit • Change from Baseline in DLQI by visit • Change from Baseline in CDLQI by visit • Percentage of patients with a DLQI total score of ≥4-points at Baseline who achieve ≥4-point improvement in DLQI from Baseline by visit • Percentage of patients who achieve ≥4-point improvement in DLQI from baseline to Week 16 • Percentage of patients with a Sleep-loss score ≥2 points at Baseline who achieve a ≥2-point reduction from Baseline by visit • Change from Baseline in EQ-5D by visit • Change from Baseline in POEM by visit • Change from Baseline in PROMIS Anxiety measure by visit • Change from Baseline in PROMIS Depression measure by visit • Change in ACQ-5 score from Baseline to Week 16 in patients who have self-reported comorbid asthma • Percentage of patients with an IGA score of 0 or 1 and a reduction ≥2 points at by visit in adults • Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both an IGA score of 0 or 1 and a reduction ≥2 points from Baseline, and a ≥4-point reduction in Pruritus NRS score from Baseline by visit • Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both EASI-75 and a ≥4-point reduction in Pruritus NRS score from Baseline by visit	

Objectives and Endpoints

Other Efficacy Endpoints
• Percentage change from Baseline to Week 16 in SCORAD • Change from Baseline in BSA by visit • Percentage of patients who respond "Strongly Agree" or "Agree" for each item of the modified SQAAQ by data collection sequence • Proportion of TCS/TCI-free days from Baseline to Week 16 • Time (days) to TCS/TCI-free use from Baseline to Week 16 • Percentage of patients rescued by visit

Abbreviations: ACQ-5 = Asthma Control Questionnaire 5-item version; AD = atopic dermatitis; BSA = body surface area; CDLQI = Children's Dermatology Life Quality Index; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; EQ-5D = European Quality of Life–5 Dimensions–5 Levels; EMA = European Medicines Agency; FDA = Food and Drug Administration; IGA = Investigator Global Assessment; NRS = Numeric Rating Scale; PK = pharmacokinetic; POEM = Patient-Oriented Eczema Measure; PROMIS = Patient-Reported Outcomes Measurement Information System; SCORAD = SCORing Atopic Dermatitis; SQAAQ= subcutaneous administration assessment questionnaire; TCI = topical calcineurin inhibitors; TCS = topical corticosteroid.

For regulatory submissions, primary and major secondary endpoints will be adjusted for multiplicity. Details can be found in Section 6.2.

5. Study Design

5.1. Summary of Study Design

Study J2T-DM-KGAD (KGAD) [aka DRM06-AD06] is a randomized, double-blind, placebo-controlled, parallel-group study in adult and adolescent (≥12 to <18 years weighing ≥40 kg) patients with moderate-to-severe atopic dermatitis (AD). The study is designed to evaluate the safety and efficacy of lebrikizumab when used in combination with topical corticosteroid (TCS) treatment compared with placebo in combination with TCS treatment. Approximately 225 patients will be enrolled into the study. The study has 1 treatment periods (16-week). Patients completing this 16-week study will be offered continued treatment in a separate long-term extension study J2T-DM-KGAA (DRM06-AD07). Patients who early terminate or choose not to enter the long-term extension study will undergo a follow-up visit approximately 12 weeks after the last study drug injection for safety follow-up.

5.1.1. Screening Period

Screening Period: Patients will be evaluated for study eligibility before the baseline visit (Day 1). Electronic diary collection will begin at screening.

5.1.2. Baseline and Double-Blinded Treatment Period (Week 0 to Week 16)

At baseline visit (Day 1), patients who meet the study eligibility criteria will be 2:1 randomly assigned to their treatments with stratification based on geographic region (US versus EU versus rest of world), age (adolescent patients 12 to <18 versus adults ≥18 years) and disease severity (Investigator Global Assessment [IGA] 3 versus 4). The treatment groups in the Blinded Period are:

- Lebrikizumab 250 mg every 2 weeks (Q2W): 500 mg lebrikizumab administered at Baseline and Week 2 (loading dose; 2 pre-filled syringes with a pre-assembled needle safety device [PFS-NSD]) and 250 mg Q2W through Week 14.
- Placebo: 4 mL (2 PFS-NSD) administered at Baseline and Week 2 and 2 mL Q2W through Week 14.

5.1.3. Safety Follow-Up Visit

Patients who terminate early from the study or do not enroll in the long term extension study, J2T-DM-KGAA (DRM06-AD07), will undergo a follow up visit approximately 12 weeks after the last study drug injection.

Figure KGAD.5.1 illustrates the study design.

Figure KGAD.5.1. Illustration of study design for Clinical Protocol KGAD.

5.2. Determination of Sample Size

For FDA: Approximately 225 patients will be randomized at a 2:1 ratio to lebrikizumab or placebo (150 patients:75 patients). The assumed IGA score of 0 or 1 at Week 16 response rates are 38% for lebrikizumab 250 mg Q2W and 13% for placebo. The assumption for lebrikizumab is based on the DRM06-AD01 Phase 2b study, the proportion of patients who achieved an IGA score of 0 or 1 at Week 16 using the rescue medication non-response sensitivity analysis, adjusting for the allowed use of TCS/topical calcineurin inhibitors (TCI). The placebo response rates are based on the review of historical TCS clinical studies in AD. This study has power >95% for testing superiority of lebrikizumab to placebo based on a two-sided Fisher's exact test with alpha of 0.05.

For European Medicines Agency (EMA): Approximately 225 patients will be randomized at a 2:1 ratio to lebrikizumab or placebo (150 patients:75 patients). The assumed IGA score of 0 or 1 at Week 16 response rates are 38% for lebrikizumab 250 mg Q2W and 13% for placebo. The assumed Eczema Area and Severity Index (EASI) 75 response rate at Week 16 response rates are 58% for lebrikizumab 250 mg Q2W and 20% for placebo. The assumptions for lebrikizumab are based on the DRM06-AD01 Phase 2b study, the proportion of patients who achieved an IGA score of 0 or 1 and proportion of patients who achieved EASI75 response at Week 16 using the rescue medication non-response sensitivity analysis, adjusting for the allowed use of TCS/TCI. The placebo response rates are based on the review of historical TCS clinical studies in atopic dermatitis. This study has power >95% for testing superiority of lebrikizumab to placebo based on a two-sided Fisher's exact test with alpha of 0.05.

5.3. Method of Assignment to Treatment

All patients will be randomly allocated to receive the study treatment using an electronic data capture (EDC) system at the Baseline visit. The allocation to treatment will be prospectively stratified by geographic region (US versus EU versus rest of world), age group (adolescent patients 12 to <18 years versus adults ≥18 years) and disease severity (IGA 3 versus 4). At the Baseline visit (Day 1), once a patient is considered eligible to participate in the study, demographic and stratification information will be entered into the EDC system to receive a medication number assigning a kit to a patient.

6. A Priori Statistical Methods

6.1. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly). The latest version of the Medical Dictionary for Regulatory Activities (MedDRA) will be used.

Analyses and summaries from assessment of endpoints described in the protocol (for example, described in KGAD Protocol Table 1) are planned to be included in a clinical study report (CSR). Analyses and summaries for key safety data are also planned to be included in the CSR. Results from additional efficacy analysis and other safety analyses may also be provided in the CSR as deemed appropriate.

Any change to the data analysis methods described in the protocol will require a protocol amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol and the justification for making the change will be described in the CSR.

All statistical processing will be performed using SAS® unless otherwise stated. Some of the analyses described in this document will be incorporated into interactive display tools instead of or in addition to static displays. Except where noted, all statistical tests will be two-sided and will be performed at the 0.05 level of significance.

The Schedule of Visits and Procedures outlined in the protocol specifies the allowable windows for assessments. Assessments performed outside these windows will not be excluded from any analysis, unless specified otherwise.

6.1.1. Analysis Populations

Analysis populations are defined in Table KGAD.6.1 along with the analysis they will be used to conduct.

Table KGAD.6.2 describes the treatment groups and the comparisons for each study period and the analysis population.

Table KGAD.6.1. Analysis Populations

Population	Description
All Entered Patients	All patients who signed informed consent. Patient flow will be summarized.
ITT Population	All randomized patients, even if the patient does not take the assigned treatment, does not receive the correct treatment, or otherwise does not follow the protocol. Patients will be analyzed according to the treatment to which they were assigned.
Modified ITT (mITT) Population	ITT Population *excluding* all patients from PPD. Patients will be analyzed according to the treatment to which they were assigned. Unless otherwise specified, efficacy and health outcomes analyses will be conducted on this population.
Safety Population	All randomized patients who received at least 1 dose of study treatment.
Modified Safety Population	Safety Population *excluding* all patients from PPD. Unless otherwise specified, safety analyses will be conducted on this population.

Abbreviation: ITT = intent-to-treat.

Table KGAD.6.2. Treatment Groups and Comparisons and Analysis Population

Study Period	Analysis Population	Treatment Groups	Abbreviation	Inferential Comparisons When Applicable
Treatment Period	mITT; Modified Safety; Safety	Placebo; Lebrikizumab 250 mg Q2W	PBO; LEB250Q2W	LEB250Q2W vs PBO

Abbreviations: mITT = modified intent-to-treat; LEB = lebrikizumab; PBO = placebo; Q2W = every 2 weeks.

6.1.2. General Considerations for Analyses

Treatment period starts after the first injection of study treatment at Baseline Visit (Day 1) and ends after patient completed Week 16 visit or the early termination visit (ETV) (between Day 1 and Week 16). For patients who are randomized but not dosed, the Double-Blinded Treatment Period starts on the date of randomization.

Baseline will be defined as the last available value before the first injection for efficacy and health outcome analyses. In most cases, this will be the measure recorded at Baseline Visit (Day 1). If the patient does not take any injection, the last available value on or prior to randomization date will be used. Change from baseline will be calculated as the visit value of interest minus the baseline value.

For Pruritus Numeric Rating Scale (NRS) and Sleep-Loss due to Pruritus collected via eDiary, the baseline period is the 7-day window prior to the first injection. A patient must have responses on at least 4 of 7 days to calculate a baseline weekly mean. If a patient has responses for 3 or fewer days, the baseline mean value will be considered missing. eDiary data for Pruritus NRS and Sleep-loss due to Pruritus are mapped to study visit per Appendix 1.

For the safety analyses, the following baselines will be used. For safety analyses using a baseline period, the baseline period is defined as the time from Screening Visit to the date/time of the first injection in Treatment Period.

- Treatment-emergent adverse events (TEAEs): baseline will be all results recorded during the baseline period.
- Treatment-emergent abnormal laboratory and vital signs results: baseline will be all results recorded during the baseline period.
- Change from baseline to last observation or to each scheduled post baseline visit for laboratory and vital signs results: baseline will be the last scheduled non-missing assessment recorded during the baseline period.

The randomization to treatment groups is stratified by geographical region (US versus EU versus rest of world), age group (adolescent patients 12 to <18 years versus adults ≥18 years) and baseline disease severity (IGA 3 versus 4) as described in Section 5.3. The countries will be categorized into geographic regions for analysis (Section 6.3). Unless otherwise specified, the statistical analysis models for Treatment Period will adjust for geographic region, age group and baseline disease severity.

For assessments of the primary endpoints and other binary efficacy and health outcomes endpoints, the following will be provided:

- Crude proportions for each treatment group along with the 95% two-sided asymptotic (that is, not continuity corrected) confidence intervals (CIs).
- The estimated common risk difference along with 95% CIs. The common risk difference is the difference in proportions adjusted for the stratification factors as mentioned in Section 6.3. SAS PROC FREQ will be used for the estimates and CIs, where the CIs are calculated by using Mantel-Haenszel-Sato method (Sato 1989).
- Cochran-Mantel-Haenszel (CMH) test will be used to compare the treatment groups while adjusting for the stratification factors. The CMH p-value will be reported, and the CMH adjusted odds ratio along with the 95% two-sided asymptotic (that is, not continuity corrected) CIs.

Treatment comparisons of key continuous efficacy variables and health outcome variables at each postbaseline time point will be made using analysis of covariance (ANCOVA) with the following in the model: treatment group, baseline value, and stratification factors mentioned in Section 6.3. Type III tests for least squares (LS) means will be used for statistical comparison between treatment groups. The LS mean difference, standard error, p-value, and 95% CI, unless otherwise specified, will also be reported.

Treatment comparisons of other continuous efficacy variables and health outcome variables with multiple postbaseline measurements will be made using mixed-model for repeated measures (MMRM). When MMRM is used, the model includes treatment, baseline value, visit, the interaction of the baseline value-by-visit, the interaction of treatment-by-visit, and the stratification factors mentioned in Section 6.3 as fixed factors. The covariance structure to model the within-patient errors will be unstructured. If the unstructured covariance matrix results in a lack of convergence, the heterogeneous Toeplitz covariance structure, followed by the heterogeneous autoregressive covariance structure will be used. The restricted maximum likelihood (REML) will be used. The Kenward-Roger method will be used to estimate the denominator degrees of freedom. Type III tests for the LS means will be used for he statistical comparison; the 95% CI will also be reported.

For variables that are not collected at each postbaseline visit, data may exist at visits where the variable was not scheduled to be collected. In these situations, data from the early discontinuation visit that do not correspond to the planned collection schedule will be excluded from the MMRM analysis (Andersen and Millen 2013). Also for by-visit summaries/displays such as boxplots, the weeks when data was not scheduled to be collected may not be displayed. However, unscheduled assessments within any defined study period will still be used in the shift analyses, and for imputing values for the change from baseline to last observation carried forward (LOCF) endpoint analyses.

The Kaplan-Meier (KM) product limit method maybe used to estimate the survival for time to event analyses. The log-rank test stratified by the stratification factors mentioned in Section 6.3 will be reported. A Kaplan-Meier plot of the time to event by treatment group may be provided.

Unless specified otherwise, Fisher's exact test will be used for adverse events (AEs) and other categorical safety measures. Odds ratios will be created with lebrikizumab treatment as the numerator, and placebo as the denominator. Continuous vital sign and laboratory values will be analyzed by an ANCOVA with treatment and baseline value in the model.

6.2. Primary and Supportive Estimands

There will be 3 estimands of interest in analyzing primary and major secondary endpoints for the Double-Blinded Treatment Period. Two types of intercurrent events (ICEs) in terms of estimating the treatment effects for the treatment period will be considered: Initiation of rescue medication as defined in Appendix 2 and permanent treatment discontinuation.

6.2.1. Primary Estimand (Hybrid)

The primary estimand is a hybrid estimand representing the primary clinical question of interest: What is the difference between treatment conditions (that is, lebrikizumab vs placebo) in the target patient population in successful responses or means after 16 weeks achieved without use of rescue medication, and if all patients continued with treatment except those who discontinued due to lack of efficacy?

The primary estimand is described by the following attributes:

- Population: Defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval
- Endpoint: Apply to all primary and major secondary endpoints
- How to account for ICEs
    - Subjects who require any use of rescue medication or discontinued treatment due to lack of efficacy prior to Week 16 will be considered as treatment failures, that is, nonresponder, after the ICEs. Therefore, composite strategy is used for these types of ICEs.
    - For subjects who discontinue treatment due to reasons other than lack of efficacy prior to Week 16, a hypothetical strategy will be used to estimate what the treatment effect would have been if subjects continued with treatment. Therefore, hypothetical strategy is used for these types of ICEs.
- Population-level summary: Difference in response proportions or means between treatment conditions.

6.2.2. Supportive Estimand for Categorical Endpoints (Composite)

The supportive estimand for categorical endpoints is a composite estimand representing the supportive clinical question of interest: What is the difference between treatment conditions in the target patient population in successful responses after 16 weeks achieved without use of rescue medication or treatment discontinuation?

The supportive estimand is described by the following attributes:

- Population: Defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval
- Endpoint: Apply to categorical endpoints
- How to account for ICEs
    - Subjects who require any use of rescue medication or discontinued treatment prior to Week 16 will be considered as treatment failures, that is, nonresponders, after the ICEs. Therefore, composite strategy is used for these types of ICEs.
- Population-level summary: Difference in response proportions between treatment conditions.

6.2.3. Supportive Estimand for Continuous Endpoints (Hypothetical)

The supportive estimand for continuous endpoints is a hypothetical estimand representing the supportive clinical question of interest: What is the difference between treatment conditions in the target patient population in means after 16 weeks if rescue medication was not available and all patients adhered to the treatment?

The supportive estimand is described by the following attributes:

A. Population: Defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval

B. Endpoint: Apply to continuous endpoints

C. How to account for ICEs

    - For subjects who require any use of rescue medication or discontinued treatment prior to Week 16, a hypothetical strategy will be used to estimate what the treatment effect would have been if rescue medication was not available and all subjects adhered to the treatment. Therefore, hypothetical strategy is used for these types of ICEs.

- Population-level summary: Difference in means between treatment conditions.

Analytical details of the primary and secondary analyses are available in Sections 6.4 and 6.11.

Table KGAD.6.3 summarizes different analyses that will be conducted on the 3 estimands.

Table KGAD.6.3. Description of Primary and Supportive Estimands

Estimand	Analysis Strategy for Intercurrent Events			Missing Data Imputation Method
	Rescue Medication	Treatment Discontinuation		
		Due to Lack of Efficacy	Due to Any Other Reasons	
Primary Estimand (Hybrid)	Composite: Set to baseline	Composite: Set to baseline	Hypothetical: Set to missing	Primary analysis: MCMC-MI Sensitivity analysis: Tipping point analysis
Supportive Estimand for	Composite:	Composite:	Composite:	NRI

Categorical Endpoints (Composite)	Set to nonresponder	Set to nonresponder	Set to nonresponder	
Supportive Estimand for Continuous Endpoints (Hypothetical)	Hypothetical: Set to missing	Hypothetical: Set to missing	Hypothetical: Set to missing	MMRM, LOCF

Abbreviations: LOCF = last observation carried forward; MCMC-MI = Markov chain Monte Carlo multiple imputation; MMRM = mixed-model repeated measures; NRI = nonresponder imputation.

6.3. Adjustments for Covariates

Unless otherwise specified, the statistical analysis models for the Treatment Period efficacy and health outcome analysis will include the following stratification factors for Baseline randomization: geographic region (US versus EU versus rest of world), age group (adolescent patients 12 to <18 versus adults ≥18 years) and baseline disease severity (IGA 3 versus 4).

Below are the country allocations within each geographic region.

Table KGAD.6.4. Geographic Regions for Statistical Analysis

Geographic Region	Country or Countries
US	US
Europe	Germany, Poland
Rest of world	Canada

In general, when an MMRM is to be used for analyses, baseline value and baseline-by-visit interactions will be included as covariates; when an ANCOVA is to be used for analyses, baseline value will be included as a covariate.

6.4. Handling of Dropouts or Missing Data

Depending on the estimands being addressed, different methods will be used to handle missing data. Description of the estimands can be found in Section 6.2.

For efficacy analysis relative to the primary estimand, the primary method of handling missing data including those as a result of ICEs will be based on Markov Chain Monte Carlo multiple imputation (MCMC-MI). The description of the MCMC-MI method can be found in Section 6.4.1. Tipping point analysis as described in Section 6.4.2 will serve as the sensitivity analysis for the primary analysis.

For efficacy analysis relative to the supportive estimand for categorical endpoints, missing data including those as a result of ICEs will be imputed as nonresponder. The description of nonresponder imputation (NRI) can be found in Section 6.4.3.

For efficacy analysis relative to the supportive estimand for continuous endpoints collected only once postbaseline, missing data including those as a result of ICEs will be imputed using LOCF. The description of LOCF can be found in Section 6.4.4.

For efficacy analysis relative to the supportive estimand for continuous endpoints collected multiple times postbaseline, an MMRM will be performed without explicit imputation. The description of MMRM can be found in Section 6.4.5.

Table KGAD.6.5 describes the planned imputation methods for efficacy and health outcome endpoints for the Double-Blinded Treatment Period.

Table KGAD.6.5. Imputation Techniques for Various Variables During Double-Blinded Treatment Period

Type of Endpoints	Efficacy and Health Outcome Endpoints	Estimand (Analysis Strategy for Intercurrent Events)	Missing Data Imputation Method (Analysis Method)
Categorical	IGA, EASI, Pruritus NRS, DLQI, and Sleep-loss related categorical endpoints at prespecified timepoints	Primary Estimand (Hybrid)	MCMC-MI, Tipping point analysis (CMH)
		Supportive Estimand (Composite)	NRI (CMH)
	Remaining categorical endpoints	Supportive Estimand (Composite)	NRI (CMH)
Continuous	EASI percent change, Pruritus NRS percent change, Sleep-loss change from Baseline, and DLQI change from Baseline	Primary Estimand (Hybrid)	MCMC-MI (ANCOVA)
		Supportive Estimand (Hypothetical)	No imputation (MMRM)
	Remaining continuous endpoints collected at multiple postbaseline timepoints including BSA, POEM, and CDLQI	Supportive Estimand (Hypothetical)	No imputation (MMRM)
	Remaining continuous endpoints collected only once postbaseline	Supportive Estimand (Hypothetical)	LOCF (ANCOVA)

Abbreviations: ANCOVA = analysis of covariance; BSA = body surface area; CDLQI = Children's Dermatology Life Quality Index; CMH = Cochran-Mantel-Haenszel; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; IGA = Investigator's Global Assessment; LOCF = last observation carried forward; MCMC-MI = Markov chain Monte Carlo multiple imputation; MMRM = mixed-model repeated measures; NRI = nonresponder imputation; NRS = Numeric Rating Scale; POEM = Patient-Oriented Eczema Measure.

6.4.1. Primary Analysis - Markov Chain Monte Carlo Multiple Imputation (MCMC-MI)

The primary method of handling missing efficacy data relative to the primary estimand will be as follows for both binary and continuous endpoints:

For patients who receive rescue medication (high-potency TCS or systemic AD treatment, defined in Appendix 2) or discontinue treatment due to lack of efficacy, set to the patient's baseline value subsequent to this time through Week 16. The MCMC-MI will be used to handle the remaining missing data. Imputation will be conducted within each treatment group independently so the pattern of missing observations in one treatment group cannot influence missing value estimations in another. The SAS PROC MI with the MCMC option will be used to conduct the MCMC-MI. The imputation model will include the relevant baseline and postbaseline.

For each imputation process, 25 datasets with imputations will be calculated. The initial seed values are given in Table KGAD.6.6. Each complete data set will be analyzed with the specified analysis. The results from these analyses will be combined into a single inference using SAS PROC MIANALYZE.

Cochran-Mantel-Haenszel (CMH) test statistic will be transformed using the Wilson-Hilferty transformation and then standardized (Ratitch 2013) prior to combining them using SAS PROC MIANALYZE. Details of combining estimates and test statistics for categorial endpoints with multiple imputation can be found in Appendix 3.

For binary responses related to EASI and IGA, the binary response variables will be calculated based on the multiply imputed datasets that have been created. Because the MCMC algorithm is based on the multivariate normal model, imputed values for IGA will not generally be one of the discrete values used in IGA scoring (0, 1, 2, 3, or 4). Therefore, to derive the binary IGA response variable, standard rounding rules will be applied to the imputed values. For example, if a patient has an IGA score imputed as 1.4 (and assuming a Baseline IGA score of 3), the imputed value would be rounded down to 1, and the minimum change from Baseline of 2 would have been met. This patient would be considered a responder.

For derivation of an EASI-75 and EASI-90 response, no rounding will be performed. The imputed Week 16 EASI value will be compared directly to the observed Baseline EASI value to determine whether a reduction of at least 75% or 90% was achieved.

For derivation of the following Pruritus NRS responses, no rounding will be performed. The imputed Pruritus NRS value will be compared directly to the observed mean baseline Pruritus NRS value to determine whether a response was achieved:

- Percentage of patients with a Pruritus NRS of ≥4-points at Baseline who achieve a ≥4-point reduction from Baseline to Weeks 1, 2, 4, and 16.

Imputation of continuous data will parallel that of binary variables. The imputed values will be used for the following secondary endpoints:

- Percentage change in Pruritus NRS score from Baseline to Week 16.
- Percentage change in EASI score from Baseline to Week 16.

Table KGAD.6.6. Seed Values for MCMC-MI

Analysis	Seed Values Lebrikizumab 250 mg Q2W Placebo
Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement from Baseline to Week 16.	970309630 1477266806
Percent change and change from Baseline in EASI score at 16 weeks. EASI-75 and EASI-90 will leverage imputation from EASI and therefore use the same seed numbers.	353985587 1828572477
Percentage change and change in Pruritus NRS score from Baseline to Week 16. Proportion of patients achieving at least a 4-point improvement from Baseline to Weeks 1, 2, 4, and 16 will leverage imputation from Pruritus NRS, and therefore, use the same seed numbers.	1611917356 1087836192
Change and percent change in Sleep loss from Baseline to Week 16. Proportion of patients achieving at least a 2-point improvement from Baseline to Week 16 will leverage imputation from Sleep loss, and therefore, use the same seed numbers.	321568 765982
Change DLQI from Baseline to Week 16. Proportion of patients achieving at least a 4-point improvement from Baseline to Week 16 will leverage imputation from DLQI, and therefore, use the same seed numbers.	458734 525683
Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both an IGA score of 0 or 1 and a reduction ≥2 points from Baseline, and a ≥4-point reduction in Pruritus NRS score from Baseline by visit Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both EASI-75 and a ≥4-point reduction in Pruritus NRS score from Baseline by visit	(Joint imputation of IGA, EASI, Pruritus, Sleep-loss and DLQI) 234567 234568

Abbreviations: DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index score; IGA = Investigator's Global Assessment; MCMC-MI = Markov chain Monte Carlo multiple imputation; NRS = Numeric Rating Scale; Q2W = every 2 weeks.

6.4.2. Tipping Point Analysis

Tipping point analysis will be conducted as a sensitivity analysis for the primary endpoint of an IGA 0 or 1 and a ≥2-point improvement from Baseline to Week 16 and the following secondary endpoints: EASI-75 and EASI-90 at Week 16 and Pruritus NRS improvement ≥4-points, at Week 16. For each of these endpoints, the tipping point analysis will only be conducted if its primary or major secondary analyses results are statistically significant.

All subjects who use rescue medication or discontinue treatment due to lack of efficacy will be imputed as nonresponders. Assumptions on missing data as a result of treatment discontinuation due to reasons other than lack of efficacy or any other intermittent missing data will be varied to investigate if there will be any tipping points.

For all the categorical endpoints described above that will be assessed using tipping point analysis, the following process will be used the determine the tipping point:

- Missing responses in the lebrikizumab groups will be imputed with a range of response probabilities, including probabilities of 0, 0.2, 0.4, 0.6, 0.8, and 1.0.

- For missing responses in the placebo group, a range of responses probabilities (for example, probability = 0, 0.2 … 1) will be used to impute the missing values. Multiple imputed dataset will be generated for each response probability.
- Treatment differences between lebrikizumab and placebo are analyzed for each imputed dataset using CMH test (Section 6.1.2). Results across the imputed datasets are aggregated using SAS® Proc MIANALYZE in order to compute a p-value for the treatment comparisons for the given response probability. If the probability values do not allow for any variation between the multiple imputed datasets (for example, all missing responses in the placebo and lebrikizumab groups are imputed as responders and nonresponders, respectively, that is, extreme case), then the p-value from the single imputed dataset will be used.

For each imputed response probability of Lebrikizumab, the tipping point is identified as the response probability value within the placebo group that leads to a loss of statistical significance when evaluating lebrikizumab relative to placebo.

For tipping point analyses the number of imputed data sets will be m=25 and the seed values to start the pseudorandom number generator of SAS are given in Table KGAD.6.7.

Table KGAD.6.7. Seed Values for Tipping Point Analysis

Analysis	Seed Value
Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline to Week 16	123470
Proportion of patients achieving EASI-75 and EASI-90 at Week 16	123471
Proportion of patients achieving at least a 4-point improvement from baseline to Weeks 16	123475

Abbreviations: EASI = Eczema Area and Severity Index; IGA = Investigator Global Assessment.

6.4.3. Nonresponder Imputation

The nonresponder imputation (NRI) method will be used to handle missing data relative to the supportive estimand for categorical endpoints (composite). Patients who receive rescue medication (high potency TCS or systemic AD treatment, defined in Appendix 2), or discontinue treatment, will be set to non-response subsequent to this time through Week 16. Intermittent missing values will also be set to non-response.

The NRI method imputes missing values as non-responders and can be justified based on the composite strategy (ICH E9R1) for handling ICEs. In this strategy patients are defined as responders only if they meet the clinical requirements for response at the predefined time AND they remain on the assigned study treatment (that is, not using rescue medications and not having missing values due to other reasons). Failing either criteria by definition makes them nonresponders.

Randomized patients without at least 1 postbaseline observation will also be defined as nonresponders for all visits for the NRI analysis.

6.4.4. Last Observation Carried Forward

In this analysis, the values subsequent to rescue medication use (high-potency TCS or systemic AD treatment, defined in Appendix 2) or treatment discontinuation will be made missing. All missing values will be imputed using LOCF for the health outcome endpoints. Baseline value will be used for imputation if there is no postbaseline observation.

6.4.5. Mixed-effects Model for Repeated Measures (MMRM)

Mixed-model for repeated measures analyses will be performed on continuous endpoints to mitigate the impact of missing data. This approach assumes missing observations are missing-at-random (missingness is related to observed data) and borrows information from patients in the same treatment arm taking into account both the missingness of data through the correlation of the repeated measurements.

The values subsequent to rescue medication use (high-potency TCS or systemic AD treatment, defined in Appendix 2) or treatment discontinuation will be made missing before applying the MMRM. The MMRM is described in Section 6.1.2.

6.5. Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. Typically, a logistic regression with treatment, site, and treatment-by-site may be used to assess the consistence of treatment effect in sites. However, due to a large number of sites and countries and relatively small sample size in the study, this logistic regression model will not likely converge. Instead, the subgroup analysis on the region will be evaluated. The countries will be categorized into geographic regions as in Section 6.3. Subgroup analysis details are provided in Section 6.15.1.

For the analysis of the primary endpoint, the presence of a treatment-by-geographic region interaction will be tested at 10% significance level. Treatment group comparisons for the primary endpoint will be presented separately for each geographic region. When there is evidence of an interaction ($p<.10$), descriptive statistics may be used to assess whether the interaction is quantitative (that is, the treatment effect is consistent in direction but not size of effect) or qualitative (the treatment is beneficial for some but not other geographic regions or countries).

6.6. Multiple Comparisons/Multiplicity

6.6.1. Multiplicity Control for US Submission

A pre-specified graphical multiple testing approach (Bretz et al. 2009, 2011) will be implemented to control the overall Type I error rate at a 2-sided alpha of 0.05, for all primary and major secondary endpoints for US submission. A gatekeeping approach will be used for multiple comparisons to control the family-wise error rate.

The following is a list of primary and major secondary endpoints to be tested in sequential order for US submission.

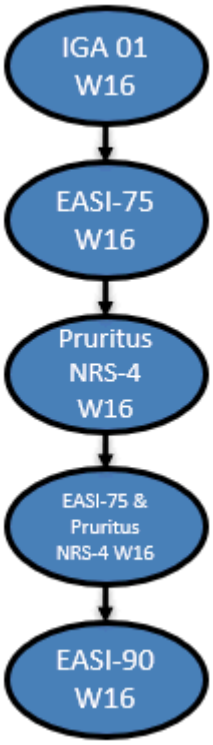
Primary Endpoint:

1. [IGA01 W16] Percentage of patients with an IGA score of 0 or 1 and a reduction of ≥2 points from Baseline to Week 16.

Major Secondary Endpoints:

2. [EASI-75 W16] Percentage of patients achieving EASI-75 (≥75% reduction from Baseline in EASI score) at Week 16.
3. [Pruritus NRS-4 W16] Percentage of patients with a Pruritus NRS of ≥4-points at Baseline who achieve a ≥4-point reduction from Baseline to Week 16.
4. [EASI-75 & Pruritus NRS-4 W16] Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both EASI-75 and a ≥4-point reduction in Pruritus NRS score from Baseline at Week 16
5. [EASI-90 W16] Percentage of patients achieving EASI-90 (≥90% reduction from Baseline in EASI score) at Week 16.

Figure KGAD.6.1 Describes the gatekeeping testing scheme for US submission.



Abbreviations: EASI = Eczema Area and Severity Index; IGA = Investigator Global Assessment; NRS = Numeric Rating Scale; W = week.

Figure KGAD.6.1. Gatekeeping Multiplicity Testing for US Submission.

6.6.2. Multiplicity Control for EU Submission

For all primary and major secondary endpoints, a pre-specified multiple testing approach (Bretz et al. 2009, 2011) will be implemented to control the overall Type I error rate at a 2-sided alpha of 0.05. A gatekeeping approach will be used for multiple comparisons to control the family-wise error rate.

The following is a list of primary and major secondary endpoints to be tested in sequential order for EMA.

Co-Primary Endpoints:

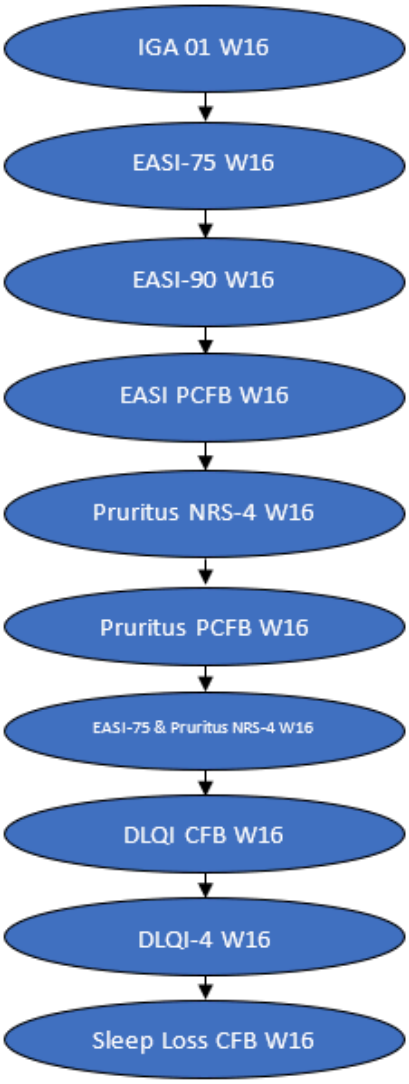
1. [IGA01 W16] Percentage of patients with an IGA 0 or 1 and a ≥2-point improvement from Baseline to Week 16.
2. [EASI-75 W16] Percentage of patients achieving EASI-75 (≥75% reduction from Baseline in EASI score) at Week 16.

Major Secondary Endpoints:

3. [EASI-90 W16] Percentage of patients achieving EASI-90 (≥90% reduction from Baseline in EASI score) at Week 16.
4. [EASI Percent Change from Baseline W16] Percentage change in EASI score from Baseline to Week 16.

5. [Pruritus NRS-4 W16] Percentage of patients with a Pruritus NRS of ≥4-points at Baseline who achieve a ≥4-point reduction from Baseline to Week 16.
6. [Pruritus Percent Change from Baseline W16] Percentage change in Pruritus NRS score from Baseline to Week 16.
7. [EASI-75 & Pruritus NRS-4 W16] Percentage of patients with a Pruritus NRS score of ≥4 points at Baseline who achieve both EASI-75 and a ≥4-point reduction in Pruritus NRS score from Baseline at Week 16
8. [DLQI CFB W16] Change from Baseline in DLQI at Week 16.
9. [DLQI W16] Percentage of patients with a Dermatology Life Quality Index (DLQI) total score of ≥4-points at Baseline who achieve a ≥4-point improvement from Baseline to Week 16.
10. [Sleep loss CFB W16] Change from Baseline in Sleep-loss score at Week 16

Figure KGAD.6.2 describes the gatekeeping testing scheme for EU submission.



Abbreviations: CFB = Change from baseline; DLQI = Dermatology Life Quality Index ; EASI = Eczema Area and Severity Index; IGA = Investigator Global Assessment; NRS = Numeric Rating Scale; PCFB = Percent change from baseline ; W = week.

Figure KGAD.6.2. Gatekeeping Multiplicity Testing for EU Submission.

6.7. Patient Disposition

The following patient disposition summaries will be provided (details of the analysis populations can be found in Section 6.1.1):

- Total number and percentage of patients entering each statistical analyses population defined in Section 6.1.1.

- The number and percentage of patients who entered the study, failed screening, were randomized at Baseline Visit (Day 1), completed Week 16, completed the safety Follow-Up Visit and entered long term extension study. Summary will be provided by the randomized treatment group (Analysis population: modified intent-to-treat [mITT];intent-to-treat [ITT]).
- The number and percentage of patients who completed the study, and the number and percentage of patients who discontinued the study at any time, by the randomized treatment group and primary reason for discontinuation (Analysis population: mITT; ITT).
- The number and percentage of patients who completed the treatment period and the number and percentage of patients who discontinued from the treatment period, by treatment group and primary reason for discontinuation, the number and percentage of patients who were rescued during treatment period (Analysis population: mITT; ITT).

All patients who were randomized (that is, in the ITT Population) and discontinued from study treatment during any period from the study will be listed together with the discontinuation reason, and the timing of discontinuation from the study will be reported.

Patient allocation by region, country, and center/site will be summarized with number of patients who entered the study, number of ITT patients for each treatment group, number of patients discontinued from study treatment, and number of patients discontinued from the study.

6.8. Patient Characteristics

6.8.1. Demographics and Baseline Characteristics

Patient demographic variables and baseline characteristics will be summarized by treatment group for the mITT Population. The continuous variables will be summarized using descriptive statistics and the categorical variables will be summarized using frequency counts and percentages. No formal statistical comparisons will be made between treatment groups unless otherwise specified. By-patient listings of basic demographic information for the ITT Population will also be provided.

The following demographic information will be included:

- Age
- Age group (Adolescents (12<18), Adults ≥18)
- Age group (Adolescents (12<18), Adults ≥18<65, ≥65 to < 75, ≥75)
- Sex (male, female)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple, Other, Not Reported)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not reported, Unknown) for US patients

- Region (as defined in Section 6.5)
- Country
- Weight (kg)
- Weight category (<60 kg, ≥60 to <100 kg, ≥100 kg)
- Height (cm)
- Body mass index (BMI) (kg/m^2)
- BMI category: Underweight (<18.5 kg/m^2), Normal (≥18.5 and <25 kg/m^2), Overweight (≥25 and <30 kg/m^2), Obese (≥30 and <40 kg/m^2), Extreme obese (≥40 kg/m^2)

The following baseline disease/clinical characteristics will be included:

- Age at onset (years): calculated as the difference between date of onset of AD and the date of birth collected on the case report form (CRF).
- Duration since AD onset (years): calculated as the difference between date of Informed Consent and the date of onset of AD collected on the CRF.
- Duration since AD onset category (0 to <2 years, 2 to <5 years, 5 to <10 years, 10 to <20 years, ≥20 years)
- Anatomical area affected by atopic dermatitis:
    - Head
    - Trunk (internal/medial axillae and groin)
    - Upper extremities (includes external axillae)
    - Lower extremities (includes buttocks and feet)
    - At least 2 areas
- Atopic dermatitis treatment used in the past
    - None
    - Topical corticosteroids
    - Topical calcineurin inhibitors
    - Immunosuppressive/immunomodulating drugs
        - Systemic corticosteroids
        - Cyclosporine
        - Mycophenolate-mofetil
        - IFN-γ
        - Janus kinase inhibitors

    - Azathioprine
    - Methotrexate
    - Phototherapy
    - Photochemotherapy (PUVA)
    - Dupilumab
    - Trilokinumab
    - Other biologics (eg, cell depleting biologics)
  - Other non-biologic medication/treatment
- Prior use of systemic treatment: Yes, No
- IGA score: 3 versus 4
- EASI score
- SCORing Atopic Dermatitis (SCORAD)
- Body surface area (BSA)
- Pruritus NRS
- Pruritus NRS <4, ≥4
- Pruritus NRS: <5, ≥5
- Sleep loss due to pruritus
- Sleep loss due to pruritus <2, ≥2
- Patient-Oriented Eczema Measure (POEM)
- Children Dermatology Life Quality Index (CDLQI)
- DLQI
- European Quality of Life–5 Dimensions (EQ-5D) Visual Analog Scale (VAS) score
- EQ-5D US Population-based index score
- EQ-5D UK Population-based index score
- Patient-Reported Outcomes Measurement Information System (PROMIS®) Anxiety and Depression scores
- Asthma Control Questionnaire (ACQ-5) (among patients who report comorbid asthma)

6.8.2. Medical History

Medical histories are defined as the conditions/events recorded on the *Medical History* eCRF with a start date prior to the first study drug injection.

The number and percentage of patients with medical histories will be summarized for the mITT Population by treatment group and by treatment and age groups using the MedDRA Preferred Term (PT) nested within System Organ Class (SOC).

The number and percentage of patients with specific medical history events of interest pre-specified on the *History Assessment* eCRF (hand dermatitis, facial dermatitis, conjunctivitis, herpes Zoster and others) will be summarized for the mITT Population by treatment group and by treatment and age groups.

6.9. Treatment Compliance

Treatment compliance with investigational product will be summarized for patients who have at least one dose in the treatment period for the Modified Safety Population and the Safety Population. Treatment compliance for each patient will be calculated as:

$$Treatment\ compliance\ (\%) = 100 \times \frac{Total\ number\ of\ injections\ administered}{Total\ number\ of\ injections\ expected}$$

- The number of injections expected can be derived from the study drug dispense related datasets.
- The total number of injections administered will be based *Study Drug Administration eCRF* page and the information from the *Dosing Diary* eCRF page.

The number of injections expected at each visit and total number of injections up to each visit during the treatment period are as follows:

Visit	Day1	W2	W4	W6	W8	W10	W12	W14[a]
# Injections at Each Visit	2	2	1	1	1	1	1	1
Total # Injections up to Each Visit	2	4	5	6	7	8	9	10

Abbreviation: W = week.

[a] Last injection during Treatment Period occurs on Week 14.

A patient will be considered compliant if he or she received ≥75% of the expected number of injections while enrolled in the study. Descriptive statistics for percent compliance will be summarized. Sub-intervals of interest, such as compliance between visits, may also be presented.

6.10. Prior and Concomitant Therapy

Medications will be classified into anatomical therapeutic chemical (ATC) drug classes using the latest version of the World Health Organization (WHO) drug dictionary. Medication start and stop dates will be compared to the date of first dose of treatment in each treatment period to allow medications to be classified as concomitant for each treatment period.

Prior medications are those medications that start prior to the date of first dose and stop prior to or on the date of first dose of study treatment. *Concomitant medications* are those medications that start before, on, or after the first day of study treatment of the defined treatment period and continue into the treatment period. Concomitant medications are assigned to the treatment period in which they are actually ongoing.

Prior medication will be summarized for mITT Population. Concomitant medication during the treatment period will be presented for the mITT Population.

6.10.1. Rescue Medication for Atopic Dermatitis(AD)

Rescue medication for AD for Treatment Period is defined as:

- any high or ultra-high-potency TCS defined as ATC code as "D07AC" or "D07AD."
- any systemic medication as defined in Appendix 2.

Patients in the mITT Population who use these rescue medications will be summarized. The summary will be provided for any rescue medication use, with high-potency TCS and systemic therapy summarized separately.

In addition, mild or moderate potency TCS/TCI and data collected in patient diary will also be summarized for the mITT Population.

Flare

Disease flares will be assessed based on rescue therapy usage; flare is defined as initiation or intensification of rescue therapy. A summary of the percentage of patients in the mITT Population rescued by week and a listing of patients in the ITT Population who use rescue medication will be provided.

6.11. Efficacy Analyses

Table KGAD.6.8 includes the description and derivation of the efficacy/health outcomes measures and endpoints. Table KGAD.6.9 provides the detailed analyses including analysis type, method and imputation, population, time point, and treatment comparisons for efficacy/health outcomes analyses.

Table KGAD.6.8. Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
Investigator's Global Assessment (IGA)	The IGA is a static assessment and rates the severity of the patient's AD. The IGA is comprised of a 5-point scale ranging from 0 (clear) to 4 (severe) and a score is selected using descriptors that best describe the overall appearance of the lesions at a given time point.	IGA score	Single item. Range: 0 to 4 0 represents "clear" 4 represents "severe"	Single item, missing if missing.
		Change from baseline in IGA score	Change from baseline: observed IGA score – baseline IGA score	Missing if baseline or observed value is missing.
		▪ IGA [0,1] with ≥2-point improvement ▪ IGA [0]	▪ Observed score of 0 or 1 and change from baseline ≤-2 ▪ Observed score of 0	▪ Missing if baseline or observed value is missing. ▪ Single item, missing if missing.
Eczema Area and Severity Index (EASI)	The EASI scoring system uses a defined process (Steps 1-5 below) to grade the severity of the signs of eczema and the extent affected. The extent of disease (percentage of skin affected: 0 = 0%; 1 = 1-9%; 2 = 10-29%; 3 = 30-49%; 4 = 50-69%; 5 = 70-89%; 6 = 90-100%) and the severity of 4 clinical signs (erythema, edema/papulation, excoriation, and lichenification) each on a scale of 0 to 3 (0 = none, absent; 1 = mild; 2 = moderate; 3 = severe) at 4 body sites (head and neck, trunk, upper limbs, and lower limbs). Half scores are allowed between severities 1, 2 and 3. Each body site will have a score that ranges from 0 to 72, and the final EASI score will be obtained by weight-averaging these 4 scores. Hence, the final EASI score will range from 0 to 72 for each time point.	EASI score	Derive EASI region score for each of head and neck, trunk, upper limbs, and lower limbs as follows: $EASI_{region}$ = (Erythema + edema/papulation + Excoriation + Lichenification) *(value from percentage involvement), where erythema, edema/papulation, excoriation, and lichenification are evaluated on a scale of 0 to 3 and value from percentage involvement is on a scale of 0 to 6. Then total EASI score is as follows: $EASI = 0.1*EASI_{head\ and\ neck} + 0.3*EASI_{trunk} + 0.2*EASI_{upper\ limbs} + 0.4*EASI_{lower\ limbs}$	N/A – partial assessments cannot be saved.
		▪ Change from baseline in EASI score ▪ Percent change from baseline EASI score	Change from baseline: observed EASI score – baseline EASI score % change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	Missing if baseline or observed value is missing.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		EASI-50	% Improvement in EASI score from baseline ≥ 50%: % change from baseline ≤-50	Missing if baseline or observed value is missing.
		EASI-75	% Improvement in EASI score from baseline ≥75%: % change from baseline ≤-75	Missing if baseline or observed value is missing.
		EASI-90	% Improvement in EASI score from baseline ≥90%: % change from baseline ≤-90	Missing if baseline or observed value is missing.
Body Surface Area (BSA)	The BSA assessment estimates the extent of disease or skin involvement with respect to AD and is expressed as a percentage of total body surface. BSA will be determined by the Investigator or designee using the patient palm = 1% rule	BSA score	BSA Total = $BSA_{head\ and\ neck}$ + BSA_{trunk} + $BSA_{upper\ limbs}$ + $BSA_{lower\ limbs}$	N/A – partial assessments cannot be saved.
		Change from baseline in BSA score	Change from baseline: observed BSA score – baseline BSA score	Missing if baseline or observed value is missing.
SCORing Atopic Dermatitis (SCORAD)	SCORAD is a validated clinical tool for assessing the extent and intensity of atopic dermatitis. There are 3 components to the assessment: • The extent of AD is assessed as a percentage of each defined body area and reported as the sum of all areas, with a maximum score of 100% (assigned as "A" in the overall SCORAD calculation).	SCORAD score	SCORAD = A/5 + 7B/2 + C, where A is extent of disease, range 0-100 B is disease severity, range 0-18 C is subjective symptoms, range 0-20	Missing if components A and B are missing or if component C is missing. Partial assessments performed by physician cannot be saved and partial assessments performed by subject cannot be saved.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
	• The severity of 6 specific symptoms of AD (redness, swelling, oozing/crusting, excoriation, skin thickening/lichenification, dryness) is assessed using the following scale: none (0), mild (1), moderate (2), or severe (3) (for a maximum of 18 total points, assigned as "B" in the overall SCORAD calculation). • Subjective assessment of itch and of sleeplessness is recorded for each symptom by the patient or relative on a VAS, where 0 is no itch (or sleeplessness) and 10 is the worst imaginable itch (or sleeplessness), with a maximum possible score of 20 (assigned as "C" in the overall SCORAD calculation.	▪ Change from baseline in SCORAD score ▪ Percent change from baseline in SCORAD score	Change from baseline: observed SCORAD score – baseline SCORAD score % change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	Missing if baseline or observed value is missing.
		SCORAD75	% Improvement in SCORAD from baseline ≥75%: % change from baseline ≤-75	Missing if baseline or observed value is missing.
		SCORAD90	% Improvement in SCORAD from baseline ≥90%: % change from baseline ≤-90	Missing if baseline or observed value is missing.
Pruritus Numeric Rating Scale (NRS)	The Pruritus NRS is a an 11-point scale used by patients to rate their worst itch severity over the past 24 hours with 0 indicating "No itch" and 10 indicating "Worst itch imaginable." Assessments will be recorded daily by the patient using an electronic diary (eDiary).	Pruritus NRS prorated weekly mean score	The prorated weekly mean is based on previous 7 days. If the patient has at least one daily score, the weekly mean is the prorated average of daily scores within the given week. Single item; range 0-10 eDiary data are mapped to study visit per Appendix 1.	Weekly mean score missing if the patient has no Pruritus-NRS responses within the week.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		▪ Change from baseline in Pruritus NRS prorated weekly mean score ▪ Percent change from baseline in Pruritus NRS prorated weekly mean score	Change from baseline: Observed Pruritus prorated weekly mean score – baseline Pruritus weekly mean score % change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	Missing if baseline or observed value is missing.
		4-point Pruritus improvement in Pruritus NRS prorated weekly mean score	Change from baseline ≤-4 in Pruritus NRS prorated weekly mean score	Missing if baseline is missing or observed value is missing.
Sleep-loss due to pruritus	Sleep-loss due to pruritus will be assessed by the patient. Patients rate their sleep based on a 5-point Likert scale [0 (not at all) to 4 (unable to sleep at all)]. Assessments will be recorded daily by the patient using an electronic diary.	Sleep-loss prorated weekly mean score	The prorated weekly mean is based on previous 7 days. If the patient has at least one daily score within the week, the weekly mean is the prorated average of daily scores within the given week. Single item; range 0 to 4.. eDiary data are mapped to study visit per Appendix 1.	Weekly mean score missing if the patient has no Sleep-loss score within the week.
		Change from baseline in Sleep-loss prorated weekly mean score Percent change from baseline in Sleep-loss prorated weekly mean score	Change from baseline: observed sleep loss prorated weekly mean score – baseline sleep loss weekly mean score % change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	Missing if baseline or observed value is missing.
		2-point improvement in Sleep-loss prorated weekly mean score	Change from baseline in Sleep-loss prorated weekly mean score ≤-2	Missing if baseline is missing or observed value is missing.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
Patient-Oriented Eczema Measure (POEM)	The POEM is a 7-item, validated, questionnaire used by the patient to assess disease symptoms over the last week. The patient is asked to respond to 7 questions on skin dryness, itching, flaking, cracking, sleep loss, bleeding and weeping. All 7 answers carry equal weight with a total possible score from 0 to 28 (answers scored as: No days=0; 1–2 days = 1; 3-4 days = 2; 5–6 days = 3; everyday = 4). A high score is indicative of a poor quality of life. POEM responses will be captured using an electronic diary and transferred into the clinical database.	POEM score	POEM total score: sum of questions 1 to 7, Range 0 to 28.	If a single question is left unanswered, then that question is scored as 0. If more than one question is unanswered, then the tool is not scored. If more than one response is selected, then the response with the highest score is used.
		Change from baseline in POEM score	Change from baseline: observed POEM score – baseline POEM score	Missing if baseline or observed value is missing.
		4-point improvement	Change from baseline ≤-4	Missing if baseline is missing or observed value is missing.
Dermatology Life Quality Index (DLQI)	DLQI is a validated, dermatology-specific, patient-reported measure that evaluates patient's health-related QoL. This questionnaire has 10 items that are grouped in 6 domains, including symptoms and feelings, daily activities, leisure, work and school, personal relationships, and treatment. The recall period of this scale is over the "last week". Response categories and corresponding scores are: Very much = 3 A lot = 2 A little = 1 Not at all = 0 Not relevant = 0	DLQI total score	A DLQI total score is calculated by summing all 10 question responses and has a range of 0-30 (less to more impairment) (Finlay and Khan 1994; Basra et al. 2008).	Score of 1 unanswered question = 0; If 2 or more questions are missing, the total score is missing. Note: #7B could be a valid missing while #7A is not "No." That is, #7 should be considered as 1 question.
		DLQI (0,1)	A DLQI (0,1) response is defined as a postbaseline DLQI total score of 0 or 1. A DLQI total score of 0 to 1 is considered as having no effect on a patient's HRQoL (Khilji et al. 2002; Hongbo et al. 2005).	Missing if DLQI total score is missing
		4-point improvement	Change from baseline ≤-4	Missing if baseline is missing or observed value is missing.
		DLQI total score and domain scores change from baseline	Calculated as: observed DLQI (total score or domain scores) – baseline DLQI (total score or domain scores)	Missing if baseline or observed value is missing

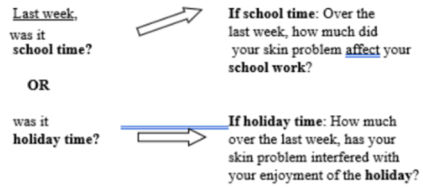
Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
	Scores range from 0-30 with higher scores indicating greater impairment of quality of life. A DLQI total score of 0 to 1 is considered as having no effect on a patient's health-related QoL (Hongbo et al. 2005), and a 4-point change from baseline is considered as the minimal clinically important difference threshold (Khilji et al. 2002; Basra et al. 2015)	DLQI symptoms and feelings domain	Sum of responses of questions #1 and #2: #1. How itchy, sore, painful or stinging has your skin been? #2. How embarrassed or self-conscious have you been because of your skin?	If 1 question in a domain is missing, that domain is missing.
		DLQI daily activities domain	Sum of responses of questions #3 and #4: #3. How much has your skin interfered with you going shopping or looking after your home or garden? #4. How much has your skin influenced the clothes you wear?	If 1 question in a domain is missing, that domain is missing.
		DLQI leisure domain	Sum of responses of questions #5 and #6: #5. How much has your skin affected any social or leisure activities? #6. How much has your skin make it difficult for you to do any sport?	If 1 question in a domain is missing, that domain is missing.
		DLQI work and school domain	Sum of responses of questions question #7A and #7B: #7A. Has your skin prevented you from working or studying? #7B. If No: how much has your skin been a problem at work or studying?	If the answer to question #7A is missing, this domain is missing. If #7A is No, and #7B is missing, this domain is missing.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		DLQI personal relationships domain	Sum of responses of questions #8 and #9: #8. How much has your skin created problems with your partner or any of your close friends or relatives? #9. How much has your skin caused any sexual difficulties?	If 1 question in a domain is missing, that domain is missing.
		DLQI treatment domain	Response of question #10: #10. How much of a problem has the treatment for your skin been, for example by making your home messy, or by taking up time?	If 1 question in a domain is missing, that domain is missing.
Children's Dermatology Life Quality Index (CDLQI)	The CDLQI is designed to measure the impact of any skin disease on the lives of children. Patients ≤16 years will complete the CDLQI and should continue to complete the CDLQI for the duration of the study. The scoring of each question is: • Very much = 3 • Quite a lot = 2 • Only a little = 1 • Not at all = 0 • Question unanswered = 0 • Question 7: 'Prevented school' (text-only questionnaire) = 3	CDLQI total score	A CDLQI total score is calculated by summing all 10 question responses and has a range of 0-30 (less to more impairment) (Waters et al. 2010).	Score of 1 unanswered question = 0; If 2 or more questions are missing, the total score is missing.
		CDLQI (0,1)	A CDLQI (0,1) response is defined as a postbaseline CDLQI total score of 0 or 1.	Missing if CDLQI total score is missing
		4-point improvement	Change from baseline ≤-4	Missing if baseline is missing or observed value is missing.
		CDLQI total score and domain scores change from baseline	Calculated as: observed CDLQI (total score or domain scores) – baseline CDLQI (total score or domain scores)	Missing if baseline or observed value is missing

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		CDLQI symptoms and feelings domain	Sum of responses of questions #1 and #2: #1. Over the last week, how itchy, "scratchy", sore, or painful has your skin been?	If 1 question in a domain is missing, that domain is missing.
			#2. Over the last week, how embarrassed or **self-conscious**, **upset,** or **sad** have you been because of your skin?	
		CDLQI sleep	Responses of questions 9 #9. Over the last week, how much has your **sleep** been affected by your skin problem?	Single item, missing if missing.
		CDLQI leisure domain	Sum of responses of questions #4, #5 and #6: #4. Over the last week, how much have you changed or worn **different** or **special clothes/shoes** because of your skin? #5. Over the last week, how much has your skin trouble affected **going out**, **playing**, or **doing hobbies**? #6. Over the last week, how much have you avoided **swimming** or **other sports** because of your skin trouble?	If 1 question in a domain is missing, that domain is missing.
		CDLQI school or holiday domain	Responses of questions 7: If select 'Prevented school,' score = 3 	Single item, missing if missing.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		CDLQI personal relationships domain	Sum of responses of questions #3 and #8: #3: Over the last week, how much has your skin affected your **friendships**? #8. Over the last week, how much trouble have you had because of your skin with other people **calling you names**, **teasing**, **bullying**, **asking questions** or **avoiding you**?	If 1 question in a domain is missing, that domain is missing.
		CDLQI treatment domain	Response of question #10: #10. How much of a problem has the treatment for your skin been?	Single item, missing if missing.
European Quality of Life–5 Dimensions–5 Levels (EQ-5D-5L)	EQ-5D comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The EQ VAS records the patient's self-rated health on a vertical visual analogue scale. The scores on these five dimensions can be presented as a health profile or can be converted to a single summary index number (utility) reflecting preferability compared to other health profiles	EQ-5D mobility EQ-5D self-care EQ-5D usual activities EQ-5D pain/ discomfort EQ-5D anxiety/ depression	Five health profile dimensions, each dimension has 5 levels: 1 = no problems 2 = slight problems 3 = moderate problems 4 = severe problems 5 = extreme problems. It should be noted that the numerals 1 to 5 have no arithmetic properties and should not be used as a primary score.	Each dimension is a single item, missing if missing.
		EQ-5D VAS	Single item. Range 0 to 100. 0 represents "worst health you can imagine" 100 represents "best health you can imagine"	Single item, missing if missing.
		Change from baseline in EQ-5D VAS	Change from baseline: observed EQ-5D VAS score – baseline EQ-5D VAS score	Missing if baseline or observed value is missing.
		EQ-5D-5L UK Population-based index score (health state index)	Derive EQ-5D-5L UK Population-based index score according to the link by using the UK algorithm to produce a patient-level index score between -0.59 and 1.0 (continuous variable).	N/A – partial assessments cannot be saved on the eCOA tablet.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		Change from baseline in EQ-5D-5L UK Population-based index score	Change from baseline: observed EQ-5D-5L UK score – baseline EQ-5D-5L UK score	Missing if baseline or observed value is missing.
		EQ-5D-5L US Population-based index score (health state index)	Derive EQ-5D-5L US Population-based index score according to the link by using the US algorithm to produce a patient-level index score between -0.11 and 1.0 (continuous variable).	N/A – partial assessments cannot be saved on the eCOA tablet.
		Change from baseline in EQ-5D-5L US Population-based index score	Change from baseline: observed EQ-5D-5L US score – baseline EQ-5D-5L US score	Missing if baseline or observed value is missing.
Patient-Reported Outcomes Measurement Information System (PROMIS®)	PROMIS® is a set of person-centered measures that evaluates and monitors physical, mental, and social health in adults and children. Pediatric and tools for anxiety and depression. Patients ≤17 years will complete pediatric versions for the duration of the study.	PROMIS anxiety total score PROMIS depression total score	A PROMIS anxiety has 8 questions on Emotion Distress-Anxiety (or Pediatric Anxiety) -Short Form 8a. Each ranges 1 to 5. Total raw scores are converted to T-Scores with higher scores representing greater anxiety. A PROMIS depression has 8 questions on Emotion Distress-Depression (or Pediatric Depressive Symptom) -Short Form 8a. Each ranges 1 to 5. Total raw scores are converted to T-score with higher scores representing greater depression.. Calculation is made by HealthMeasures Scoring Service, powered by Assessment Center℠	Total score can be derived even with partial response as instrument use item response theory method.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		Change from baseline in PROMIS anxiety total score Change from baseline in PROMIS depression total score	Change from baseline: observed score – baseline PROMIS anxiety total score Change from baseline: observed score – baseline PROMIS depression total score	Missing if baseline or observed value is missing.
Asthma Control Questionnaire (ACQ-5)	Patients who report comorbid asthma prior to enrollment will complete the Asthma Control Questionnaire in addition to other patient reported outcomes in this trial. The ACQ-5 has been shown to reliably measure asthma control and distinguish patients with well-controlled asthma (score ≤0.75 points) from those with uncontrolled asthma (score ≥1.5 points). It consists of 5 questions that are scored on a 7-point Likert scale with a recall period of 1 week. The total ACQ-5 score is the mean score of all questions; a lower score represents better asthma control.	ACQ-5 total score	An ACQ-5 total score is the mean score of all 5 questions.	If more than 1 question is missing, the ACQ-5 total score is missing.
		Change from baseline in ACQ-5 score	Change from baseline: observed ACQ-5 total score – baseline ACQ-5 total score	Missing if baseline or observed value is missing.
		MCID of 0.5	Change from baseline ≤-0.5	Missing if baseline is missing or observed value is missing.

Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
Topical corticosteroid (TCS) or topical calcineurin inhibitor (TCI) Use	A mid-potency TCS, triamcinolone acetonide 0.1% cream, and a low-potency TCS, hydrocortisone 1% cream (for use on sensitive skin areas) will be provided by the Sponsor or clinical site for use in this trial. Patients are to be instructed to return all used and unused TCS medication (tubes) to the study site for accountability purposes. All TCS and TCI use will be recorded daily by the patient using an electronic diary	Time (days) to TCS / TCI-free use from Baseline to Week 16	Days from first study drug injection to the day patient stop using all TCS/TCI (if a patient start and stop using low or mid-potency TCS/TCI multiple times, use the last stop date as the stop date for this patient)	If do not stop using the TCS/TCI, the patient will be censored at the date of their last visit
		Proportion of TCS/TCI free days from Baseline to Week 16	Number of the total TCS/TCI free days divided by total number of days during the treatment period	Missing data will be treated as TCS/TCI used days.
Modified Subcutaneous Administration Assessment Questionnaire (SQAAQ)	Adolescent patients from EU may complete the modified SQAAQ which uses 10 questions to assess the acceptability and tolerability with using a device to administer a subcutaneous injection. The person who administered the dose (adolescent patient or their parent/caregiver) should complete a 7-point Likert scale (from "Strongly Disagree" to "Strongly Agree") shortly after completing the injection.	Respond "Strongly Agree" or "Agree" for each self/caregiver administration of the study drug.	For each EU adolescent patient that have SQAAQ scale completed, the proportion of patients who answer "Strongly Agree" or "Agree" in each of the 10 questions.	Missing data will be treated as missing.

Abbreviations: AD = atopic dermatitis; MCID = minimally clinically important difference; QoL = quality of life; VAS = Visual Analog Scale.

Table KGAD.6.9. Description of Efficacy/Health Outcome Analyses

Measure	Variable	Analysis Method (Section 6.1)	Population (Section 6.1.1)	Comparison/Time Point	Analysis Type
Investigator's Global Assessment (IGA)	Proportion of patients achieving IGA [0,1] with a ≥2-point improvement	CMH analysis with MCMC-MI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Primary analysis: W16; Secondary analysis: other timepoints
		CMH analysis with NRI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Supplementary analysis
		CMH analysis with tipping point analysis	mITT	Leb 250 mg Q2W vs PBO; Week 16	Sensitivity analysis
	Proportion of patients achieving IGA [0]	CMH analysis with NRI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
	Proportion of patients achieving both IGA [0,1] with a ≥2-point improvement and a ≥4-point improvement in Pruritus Numeric Rating Scale (NRS)	CMH analysis with MCMC-MI	mITT with Baseline Pruritus NRS score at least 4	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
		CMH analysis with NRI	mITT with Baseline Pruritus NRS score at least 4	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
Eczema Area and Severity Index (EASI)	Change from baseline in EASI score	ANCOVA with MCMC-MI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis: percent change Week 16
	Percent change from baseline in EASI score	MMRM	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis

Description of Efficacy/Health Outcome Analyses

Measure	Variable	Analysis Method (Section 6.1)	Population (Section 6.1.1)	Comparison/Time Point	Analysis Type
	Proportion of patients achieving EASI-75 Proportion of patients achieving EASI-90	CMH analysis with MCMC-MI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Primary analysis: EASI-75, W16; Secondary analysis: EASI-90, W16; Secondary analysis: other timepoints
		CMH analysis with NRI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Supplementary analysis
		CMH analysis with tipping point analysis	mITT	Leb 250 mg Q2W vs PBO; Week 16	Sensitivity analysis
	Proportion of patients achieving EASI-50	CMH analysis with MCMC-MI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Induction Period	Secondary analysis
	Proportion of patients achieving both EASI-75 and a ≥4-point improvement in Pruritus NRS	CMH analysis with MCMC-MI	mITT with Baseline Pruritus NRS score at least 4	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
		CMH analysis with NRI	mITT with Baseline Pruritus NRS score at least 4	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
Body Surface Area (BSA) Affected by AD	Change from baseline in BSA score	MMRM	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis

Description of Efficacy/Health Outcome Analyses

Measure	Variable	Analysis Method (Section 6.1)	Population (Section 6.1.1)	Comparison/Time Point	Analysis Type
Pruritus Numeric Rating Scale (NRS)	Change from baseline in Pruritus NRS Percent Change from baseline in Pruritus NRS score	ANCOVA with MCMC-MI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis:
		MMRM	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
	Proportion of patients achieving at least 4-point improvement in pruritus NRS in patients who had baseline pruritus NRS ≥4	CMH analysis with MCMC-MI	mITT with Baseline Pruritus NRS score at least 4	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis: Weeks 1, 2, 4, and 16; Secondary analysis: other timepoints
		CMH analysis with NRI	mITT with Baseline Pruritus NRS score at least 4	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Supplementary analysis
		CMH analysis with tipping point analysis	mITT with Baseline Pruritus NRS score at least 4	Leb 250 mg Q2W vs PBO; Weeks 1, 2, 4, and 16	Sensitivity analysis
Sleep-loss Score	Percent Change from baseline in Sleep-loss Change from baseline in Sleep-loss	ANCOVA with MCMC-MI MMRM	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis: percent change, W16; Secondary analysis: other timepoints
	Proportion of patients achieving at least 2-point improvement Sleep-loss in patients who had baseline Sleep-loss ≥2	CMH analysis with MCMC-MI	mITT with Baseline Sleep-loss score at least 2	Leb 250mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis: Weeks 16; Secondary analysis: other timepoints
		CMH analysis with NRI	mITT with Baseline Sleep-loss score at least 2	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis: Weeks 1, 2, 4, and 16; Secondary analysis: other timepoints

Description of Efficacy/Health Outcome Analyses

Measure	Variable	Analysis Method (Section 6.1)	Population (Section 6.1.1)	Comparison/Time Point	Analysis Type
(Children) Dermatology Life Quality Index (DLQI/ CDLQI)	Change from baseline in DLQI total Score	ANCOVA with MCMC-MI MMRM	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis: W16 for DLQI change; Secondary analysis: other timepoints
	Proportion of patients achieving at least 4-point improvement in DLQI in patients who had baseline DLQI score ≥4	CMH analysis with MCMC-MI NRI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis: W16 for DLQI change; Secondary analysis: other timepoints
	Change from baseline in CDLQI Score	MMRM	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
SCORing Atopic Dermatitis (SCORAD)	Change from baseline in SCORAD score Percent change from baseline in SCORAD score	ANCOVA with LOCF	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
	Proportion of patients achieving SCORAD75 Proportion of patients achieving SCORAD90	CMH analysis with NRI	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
Patient-Oriented Eczema Measure (POEM)	Change from baseline in POEM score	MMRM	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits in Treatment Period	Secondary analysis
European Quality of Life–5 Dimensions–5 Levels (EQ-5D-5L)	Proportion of patients having no problem in each domain: • EQ-5D mobility • EQ-5D self-care • EQ-5D usual activities • EQ-5D pain/ discomfort EQ-5D anxiety/ depression	CMH analysis with NRI	mITT	Leb 250 mg Q2W vs PBO; Week 16	Secondary analysis

Description of Efficacy/Health Outcome Analyses

Measure	Variable	Analysis Method (Section 6.1)	Population (Section 6.1.1)	Comparison/Time Point	Analysis Type
	Change from baseline in • EQ-5D VAS • EQ-5D-5L UK Population-based index score • EQ-5D-5L US Population-based index score	ANCOVA with LOCF	mITT	Leb 250 mg Q2W vs PBO; Week 16	Secondary analysis
Patient-Reported Outcomes Measurement Information System (PROMIS®)	Change from baseline in PROMIS Anxiety score Change from baseline in PROMIS Depression score	ANCOVA with LOCF	mITT	Leb 250 mg Q2W vs PBO; Week 16	Secondary analysis
Asthma Control Questionnaire (ACQ-5)	Change from baseline in ACQ-5 score	ANCOVA with LOCF	mITT with self-reported comorbid asthma	Leb 250 mg Q2W vs PBO; Week 16	Secondary analysis
Topical corticosteroid (TCS) or topical calcineurin inhibitor (TCI) Use	Time (days) to TCS / TCI-free use from Baseline to Week 16	KM method with log-rank test	mITT	Leb 250 mg Q2W vs PBO;	Secondary analysis
	Proportion of TCS/TCI free days from Baseline to Week 16	MMRM	mITT	Leb 250 mg Q2W vs PBO; Week 16 and all scheduled visits	Secondary analysis

Description of Efficacy/Health Outcome Analyses

Modified Subcutaneous Administration Assessment Questionnaire (SQAAQ)	Proportion of patients who answer "Strongly Agree" or "Agree" in each of 10 questions in a visit	Observed case	Patients who complete SQAAQ at any visit	Leb 250 mg Q2W vs PBO; by sequence of each self/caregiver injection; (note, patient could start self/caregive injection at any visit, the visits will be aligned as: 1st self/caregiver injection, 2nd self/caregiver injection,…)	Secondary analysis

Abbreviations: AD =atopic dermatitis; ANCOVA = analysis of covariance; CMH = Cochran-Mantel-Haenszel; KM = Kaplan-Meier; Leb = lebrikizumab; LOCF = last observation carried forward; MCMC-MI = Markov chain Monte Carlo multiple imputation; mITT = modified intent-to-treat; MMRM = mixed-model repeated measures; NRI = nonresponder imputation; PBO = placebo; Q2W = every 2 weeks; VAS = Visual Analog Scale; W = week.

6.11.1. Primary Outcome and Methodology

The primary analysis of the study is to test the null hypotheses that lebrikizumab 250 mg Q2W is the same as placebo when evaluating the proportion of patients achieving IGA of 0 or 1 at Week 16 in the mITT Population. For EMA, an additional null hypothesis is that lebrikizumab 250 mg Q2W is the same as placebo when evaluating the proportion of patients achieving EASI-75 at Week 16 in the mITT Population.

The primary estimand for the primary analysis is described in Section 6.2.1. The missing values will be imputed using MCMC-MI based on missing at random assumption (Section 6.4.1).

A CMH test as described in Section 6.1.2 will be used for the comparisons. The odds ratio, the corresponding 95% CIs and p-value, as well as the treatment differences and the corresponding 95% CIs, will be reported.

Multiplicity controlled analyses will be performed on the primary and major secondary objectives to control the overall Type I error rate at a 2-sided alpha level of 0.05. A gatekeeping approach will be used to perform the multiplicity controlled analyses as described in Section 6.6.

Primary outcome IGA 0/1 and EASI-75 and their analysis are described in Table KGAD.6.8 and Table KGAD.6.9.

6.11.2. Sensitivity Analyses of Primary Outcome

Sensitivity analyses are included to demonstrate robustness of analyses methods using different missing data imputations, populations and analyses assumptions. Sensitivity analyses for primary outcomes include the tipping point analysis based on missing not at random assumption (Section 6.4.2).

Secondary analyses for both primary and secondary endpoints are described Table KGAD.6.8 and Table KGAD.6.9.

There will be no adjustment for multiple comparisons for additional analyses of the primary outcome.

6.11.3. Major Secondary Efficacy Analyses

Major secondary outcomes and their analyses are described in Table KGAD.6.8 and Table KGAD.6.9.

6.11.4. Other Secondary Efficacy Analyses

Other secondary outcomes and their analyses are described in Table KGAD.6.8 and Table KGAD.6.9.

6.12. Health Outcomes/Quality-of-Life Analyses

Analyses of POEM, DLQI, EQ-5D-5L, PROMIS, and ACQ-5 are described in Table KGAD.6.8 and Table KGAD.6.9.

6.13. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Details of pharmacokinetic (PK)/pharmacodynamic (PD) analyses can be found in a separate PK/PD analysis plan.

6.14. Safety Analyses

The planned analyses of safety data will be performed with an intent to maintain consistency with compound level standard safety analyses (PSAP). These standards are based on internal standards which were informed by Clinical Data Interchange Standards Consortium (CDISC) standards, regulatory guidance (for example, FDA Clinical Review Template), and cross-industry standardization efforts (for example, Pharmaceutical Users Software Exchange [PhUSE] white papers from the Standard Analyses and Code Sharing Working Group provided in the PhUSE Computational Science Deliverables Catalog).

Safety evaluations will be based on the Modified Safety Population (Treatment Period). A sensitivity analysis will also be conducted on the Safety Population (Treatment Period).

Analysis populations are fully defined in Table KGAD.6.1 while Table KGAD.6.2 describes the treatment groups and the comparisons for each analysis population.

For document writing purposes for safety, tests with two-sided p-values less than 0.05 will be referred to as having strong statistical evidence for a treatment difference, unless otherwise noted. However, p-values should not be over-interpreted for these safety analyses. Except for pre-specified hypotheses, they correspond to data-driven hypotheses and hence are only useful as a flagging mechanism.

Not all displays described in this section will necessarily be included in the CSRs. Any display described and not provided in the CSR would be available upon request. Not all displays will necessarily be created as a "static" display. Some may be incorporated into interactive display tools instead of or in addition to a static display. Any display created interactively will be included in the CSR if deemed relevant to the discussion.

6.14.1. Extent of Exposure

Duration of exposure to study treatment will be summarized by treatment group for the Modified Safety Population and the Safety Population.

The duration of exposure will be calculated as:

$$\begin{aligned} &Duration\ of\ exposure\ (days) \\ &= Date\ of\ last\ visit\ (scheduled\ or\ unscheduled)\ in\ the\ specified\ Treatment\ Period \\ &- Date\ of\ first\ dose\ in\ Treatment\ Period + 1 \end{aligned}$$

Drug interruption time period due to the use of systemic rescue therapies will be removed from study drug exposure calculations as described in compound level safety plan.

The number and percentage of patients in each of the following categories will be included in the summaries:

- >0, ≥7 days, ≥14 days, ≥30 days, ≥60 days, ≥90 days, ≥112 days, ≥120 days. Note that patients may be included in more than 1 category.
- >0 to <7 days, ≥7 to <14 days, ≥14 to <30 days, ≥30 to <60 days, ≥60 to <90 days, ≥90 to <112 days, ≥112 to <120 days , ≥120 days.

Additional exposure ranges may be considered if necessary. No p-values will be reported.

The summaries will also include the following information:

- Total exposure in patient years, calculated as:

$$Total\ exposure\ in\ patient\ years = \frac{Sum\ of\ exposures\ for\ all\ patients\ in\ treatment\ group}{365.25}$$

- Mean and median total dose. Total dose (in mg) is calculated by the number of active injections taken during the treatment period multiplied by dose. The total dose (in mg) taken during Treatment Period will be calculated as follows: *Total lebrikizumab dose=Total number of active injections (including loading doses, if any) received in* Treatment *Period ×250*
- The total number of injections administered will be based on the prompt *"Please select the number of syringes successfully injected"* in the *Dosing Diary* eCRF page.

6.14.2. Adverse Events

A TEAE is defined as an event that first occurred or worsened in severity after baseline. The MedDRA Lowest Level Term (LLT) will be used in the treatment-emergent computation. The maximum severity for each LLT during the baseline period will be used as baseline. The treatment period will be included as postbaseline for the analysis. For events with a missing severity during the baseline period, it will be treated as 'mild' in severity for determining treatment-emergence. Events with a missing severity during the postbaseline period will be treated as 'severe' and treatment-emergence will be determined by comparing to baseline severity. For events occurring on the day of first taking study medication, it will be assumed to be posttreatment.

The planned summaries for adverse events are provided in Table KGAD.6.10 are described more fully in compound level safety standards and in the adverse event-related PhUSE white paper [Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated Summary Document (PhUSE 2017)].

Summary tables as described in Table KGAD.6.10 will be presented for the following periods/analysis populations as indicated. Summary tables will include the number and percentage of patients reporting an event. For events that are gender-specific (as defined by MedDRA), the number of participants at risk will include only patients from the given gender.

Table KGAD.6.10. Summary Tables/Listings Related to Adverse Events

Analysis	Population (Section 6.1.1)
Overview of AEs	Modified Safety; Safety (sensitivity analysis)
Summary of TEAE by PTs	Modified Safety; Safety (sensitivity analysis)
Summary of TEAE by PTs occurring in ≥1% of patients	Modified Safety; Safety (sensitivity analysis)
Summary of TEAE possibly related to study drug by PTs within SOC	Modified Safety; Safety (sensitivity analysis)
Summary of TEAE by PTs within SOC	Modified Safety; Safety (sensitivity analysis)
Summary of TEAE PTs by maximum severity	Modified Safety; Safety (sensitivity analysis)
Summary of SAE by PT within SOC	Modified Safety; Safety (sensitivity analysis)
Summary of AEs leading to treatment discontinuation by PT with SOC	Modified Safety; Safety (sensitivity analysis)
Listing of SAEs (including Death)	ITT
Listing of primary AEs leading to study treatment discontinuation	ITT
Listing of TEAE (for Japan submission only)	Safety

Abbreviations: AE = adverse event; ITT = Intent-to-Treat; PT = Preferred Term; SAE = serious adverse event; SOC = System Organ Class; TEAE = treatment-emergent adverse event.

Statistical comparisons will be performed using Fisher's exact test. Odds ratio will be provided.

6.14.2.1. Common Adverse Events

The number and percentages of patients with TEAEs will be summarized by treatment using MedDRA PT for the common TEAEs (occurred in ≥1% before rounding in any column in the table).

6.14.2.2. Deaths, Other Serious Adverse Events and Other Notable Adverse Events

The number and percentage of patients reported with an SAE during the treatment period will be summarized by treatment using MedDRA PT. A listing of SAEs will be provided.

The number and percentage of patients who permanently discontinued from study treatment due to an AE (including AEs that led to death) during the treatment period will be summarized by treatment using MedDRA PT. Events will be ordered by decreasing frequency in the all treatment groups.

6.14.3. Clinical Laboratory Evaluation

As described more fully in compound level safety standards and in the laboratory-related PhUSE white papers (PhUSE 2013; PhUSE 2015), the clinical laboratory evaluations will be summarized as described in Table KGAD.6.11.

Table KGAD.6.11. Analysis for Clinical Laboratory Evaluations

Analysis	Population
Box plots of observed values by visit Box plots for change values by visit	Modified Safety; Safety (sensitivity analysis)
Change from baseline to last observations. ANCOVA model with treatment and baseline value in the model.	Modified Safety; Safety (sensitivity analysis)
Scatter plots of baseline-by-maximum values and baseline-by-minimum values	Modified Safety; Safety (sensitivity analysis)
Treatment-emergent abnormal high lab values (that is, patients shifting from a normal/low maximum baseline value to a high maximum postbaseline value) or abnormal low lab values (that is, patients shifting from normal/high minimum baseline value to a low minimum postbaseline value)	Modified Safety; Safety (sensitivity analysis)
Shift tables showing the number of patients who shift from each category of maximum (minimum) baseline observation to each category of maximum (minimum) postbaseline observation. Here categories may be low, normal, or high with cut-offs defined in the compound level safety standards.	Modified Safety; Safety (sensitivity analysis)
Listing of abnormal findings for laboratory analyte measurements, including qualitative measures	All Enrolled

Abbreviation: ANCOVA = analysis of covariance.

6.14.4. Vital Signs and Other Physical Findings

As described more fully in compound level safety standards and in the laboratory-related PhUSE white papers (PhUSE 2013; PhUSE 2015), vital signs will be summarized similarly to the clinical laboratory evaluation (Table KGAD 6.12). For vital signs, treatment emergent low and high are based on a combination of a specified value and a change or percentage change for adults and adolescents as defined in the compound level safety standards.

Table KGAD 6.12. Analysis Related to Vital Signs

Analysis	Population
Box plots for observed values by visit	Modified Safety; Safety (sensitivity analysis)
Box plots for change from baseline values by visit	Modified Safety; Safety (sensitivity analysis)
Scatterplots of baseline-by-maximum values and baseline-by-minimum values	Modified Safety; Safety (sensitivity analysis)
Tables with the number and percentage of subjects who shift from normal/high to low (that is, treatment-emergent low) and the number and percentage of subjects who shift from normal/low to high (that is, treatment-emergent high); the limits are defined in the compound level safety standards	Modified Safety; Safety (sensitivity analysis)

6.14.5. Immunogenicity

An individual sample is potentially examined multiple times in a hierarchical procedure to produce a sample anti-drug antibody (ADA) assay result and may yield a sample neutralizing ADA (NAb) assay result. Treatment-emergent ADA (TE-ADA) are defined as those with a titer

2-fold (1 dilution) greater than the minimum required dilution if no ADAs were detected at baseline (treatment-induced ADA) or those with a 4-fold (2 dilutions) increase in titer compared to baseline if ADAs were detected at baseline (treatment-boosted ADA). A patient is considered TE-ADA positive when at least 1 postbaseline ADA sample meets the definition of TE-ADA.

Compound level safety standards (described in PSAP) will be followed in the analyses of immunogenicity. Listings of immunogenicity assessments will be provided along with the summary of specified TEAEs by TE-ADA status for the Safety Population. The summary of TE-ADA and NAb status will be produced for the Modified Safety Population, where the postbaseline period for reporting is the same as described for AEs in Section 6.14.2. Additional assessments of the relationship between immunogenicity and efficacy will be performed as part of the integrated analysis including other Phase 3 lebrikizumab AD trials.

6.14.6. Special Safety Topics including Adverse Events of Special Interest

This section includes areas of interest whether due to observed safety findings, potential findings based on drug class, or safety topics anticipated to be requested by a regulatory agency for any reason. In general, potential adverse events of special interest (AESI) relevant to these special safety topics will be identified by one or more Standardized MedDRA Query(ies) (SMQs), by a Lilly-defined MedDRA PT listing based upon the review of the most current version of MedDRA, or by treatment-emergent relevant laboratory changes, as described below. Additional special safety topics may be added as warranted.

Unless otherwise specified, the special safety topics will be summarized for the Modified Safety Population and the Safety Population during the Treatment Period as described in Section 6.14.

Full details of the search terms and rules for deriving special safety topics in each of the sections below are described in the compound level safety standards along with information about the types of summaries and listings to be provided. In the event that the listing of terms or analysis changes for a special safety topic, it will be documented in the compound level safety standards which will supersede this document; it will not warrant an amendment to the individual study SAP.

6.14.6.1. Hepatic Safety

Hepatic labs include alanine aminotransferase (ALT), aspartate transaminase (AST), total bilirubin (TBL), and serum alkaline phosphatase (ALP).

Table KGAD.6.13. Summary Tables Related to Hepatic Safety

Analysis	Population
ALT and AST: The number and percentage of subjects with a measurement greater than or equal to 3 times (3X), 5 times (5X), and 10 times (10X) the performing lab upper limit of normal (ULN) during the treatment period for all subjects with a postbaseline value and for subsets based on various levels of baseline value TBL and ALP: The number and percentage of subjects with a measurement greater than or equal to 2 times (2X) the performing lab ULN during the treatment period will be summarized for all subjects with a postbaseline value and for subsets based on various levels of baseline value	Modified Safety; Safety (sensitivity analysis)
Plot of maximum postbaseline ALT vs maximum postbaseline total bilirubin	Modified Safety Population for All Periods: ever on lebri and never on lebri; Safety Population for All Periods: ever on lebri and never on lebri (sensitivity analysis)

Abbreviations: ALP = serum alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate transaminase; lebri = lebrikizumab; TBL = total bilirubin.

6.14.6.2. Eosinophilia and Eosinophil-Related Disorders

In addition to the standard laboratory analysis (Section 6.14.3), eosinophilia and eosinophil-related AE will be summarized. Details regarding eosinophil-related PTs are in Compound Level Safety Standard.

Table KGAD.6.14. Summary Tables Related to Eosinophil and Eosinophil-Related AE

Analysis	Population
Shift table summarizing the number and percentage of participants within each maximum baseline category versus each maximum postbaseline category by treatment	Modified Safety; Safety (sensitivity analysis)
Summary of eosinophil-related TEAE by PT	Modified Safety; Safety (sensitivity analysis)

Abbreviations: AE = adverse event; PT = Preferred Term; TEAE = treatment-emergent adverse event.

6.14.6.3. Infections, Including Herpes Infections and Relevant Parasitic Infections

Infections will be defined using the PTs from the MedDRA Infections and Infestations SOC. The MedDRA terms used to identify infections considered to be opportunistic infections (OI) in patients with immune mediated inflammatory conditions treated with immunomodulatory drugs are based on Winthrop et al. (2015) and are listed in the compound level safety standards. The

list contains narrow (more specific) and broad (less specific) PTs with respect to these prospectively defined OIs. Definitions of herpes infections, parasitic infections, and skin infections are listed in the compound level safety standards.

Table KGAD.6.15. Summary Tables/Listings Related to Infection Related AE

Analysis	Population
Summary of treatment-emergent infections by maximum severity	Modified Safety; Safety (sensitivity analysis)
Summary of serious infections by PT	Modified Safety; Safety (sensitivity analysis)
Summary of infection AEs resulting in permanent study drug discontinuation	Modified Safety; Safety (sensitivity analysis)
Treatment-emergent opportunistic and potential opportunistic infections	Modified Safety; Safety (sensitivity analysis)
Treatment-emergent adverse events – herpes and parasitic infections	Modified Safety; Safety (sensitivity analysis)
Treatment-emergent adverse events – skin infections	Modified Safety; Safety (sensitivity analysis)
Treatment-emergent adverse events – infections characterization from follow-up form	Modified Safety; Safety (sensitivity analysis)
A listing of patients with potential OI, serious infection, herpes and parasitic infections	Safety

Abbreviations: AE = adverse event; OI = opportunistic infections; PT = Preferred Term.

6.14.6.4. Conjunctivitis

Conjunctivitis are events of special interest and will be identified using PTs nested within the categories of conjunctivitis and Keratitis as described in the Compound Level Safety Standards:

Table KGAD.6.16. Summary Tables/Listings Related to Conjunctivitis

Analysis	Population
Summary of TEAE of conjunctivitis within categories	Modified Safety; Safety (sensitivity analysis)
TEAEs – conjunctivitis and eye inflammation characterization from follow-up form	Modified Safety; Safety (sensitivity analysis)
A listing of patients with conjunctivitis	Safety

Abbreviations: TEAE = treatment-emergent adverse event.

6.14.6.5. Hypersensitivity

Potential hypersensitivity reactions will be determined using the following SMQs: anaphylactic reaction, hypersensitivity, and angioedema. Potential hypersensitivity will be categorized as immediate (that is, occurring the same day as drug administration) and non-immediate (that is, occurring after the day of study drug administration but prior to subsequent drug administration). The planned summaries are provided in Table KGAD.6.17.

Table KGAD.6.17. Summary Tables/Listings Related to Hypersensitivity

Analysis	Population
Treatment-emergent adverse events – potential immediate hypersensitivity reactions (events occurring on day of study drug administration)	Modified Safety; Safety (sensitivity analysis)
Treatment-emergent adverse events – potential non-immediate hypersensitivity reactions (events occurring after the day of study drug administration)	Modified Safety; Safety (sensitivity analysis)

6.14.6.6. Injection Site Reactions (ISR)

Injection site reactions (ISRs) are AEs localized to the immediate site of the administration of a drug. The evaluation of study drug related ISRs will be through the unsolicited reporting of ISR TEAEs. Injection site reactions will be defined using the MedDRA High Level Term (HLT) of Injection Site Reaction, excluding certain PTs related to joints and Administration Site Reactions as described in the Compound Level Safety Standards.

Table KGAD.6.18. Summary Tables Related to Injection Site Reactions

Analysis	Population
Summary of TEAE of ISR overall, and by PT	Modified Safety; Safety (sensitivity analysis)

Abbreviations: ISR = injection site reaction; PT = Preferred Term; TEAE = treatment-emergent adverse event.

6.14.6.7. Malignancies

Malignancies will be defined using PTs from the Malignant tumors SMQ and summarized separately for the 2 categories: Non-melanoma skin cancer (NMSC) and Malignancies excluding NMSC.

Table KGAD.6.19. Summary Tables Related to Malignancies

Analysis	Population
Summary of TEAE of malignancies within categories of NMSC and malignancy excluding NMSC	Modified Safety; Safety (sensitivity analysis)

Abbreviations: NMSC = non-melanoma skim cancer; TEAE = treatment-emergent adverse event.

6.14.6.8. Atopic Dermatitis Exacerbation

Atopic dermatitis exacerbation will be defined using PTs specified in the Compound Level Safety Standards and summarized for Safety Population.

Table KGAD.6.20. Summary Tables Related to Atopic Dermatitis Exacerbation

Analysis	Population
Summary of TEAE of atopic dermatitis exacerbation	Modified Safety; Safety (sensitivity analysis)

Abbreviations: TEAE = treatment-emergent adverse event.

6.14.6.9. Suicide/Self-Injury

The PTs from the suicide/self-injury SMQ [20000037] will be summarized.

Table KGAD.6.21 Summary Tables Related to Suicidal Ideation and Behavior

Analysis	Population
Summary of TEAE of suicide/self-injury	Modified Safety; Safety (sensitivity analysis)

Abbreviations: TEAE = treatment-emergent adverse event.

6.15. Subgroup Analyses

6.15.1. Efficacy Subgroup Analyses

Subgroup analyses will be conducted for the primary endpoints IGA 0,1 and EASI-75 at Week 16 in the mITT Population using MCMC-MI approach as in primary analysis (Section 6.4.1). A logistic regression analysis with treatment, subgroup, and treatment-by-subgroup interaction as factors will be used. The treatment-by-subgroup interaction will be tested using the Firth correction (Firth 1993) at the 10% significance level. Treatment group differences will be evaluated within each subgroup using the chi-square test, regardless of whether the interaction is statistically significant. If any group within the subgroup (for example, yes, no) is <10% of the total population, only descriptive statistics will be provided for that subgroup (that is, no inferential testing).

Forest plots may be created to illustrate the treatment differences with 95% CIs between each of the lebrikizumab treatment groups and placebo group, by each subgroup category.

The following subgroups will be analyzed:

- Age group (12<18, ≥18)
- Age group (Adolescents (12<18), Adults ≥18 to < 65, ≥65 to < 75, ≥75)
- Sex (male, female)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple, Other, Not Reported)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not reported, Unknown) for US patients
- Region (as defined in Section 6.5)
- Weight category (<60 kg, ≥60 to <100 kg, ≥100 kg)

- BMI category Underweight (<18.5 kg/m^2), Normal (≥18.5 and <25 kg/m^2), Overweight (≥25 and <30 kg/m^2), Obese (≥30 and <40 kg/m^2), Extreme obese (≥40 kg/m^2)
- Duration since AD onset category (0 to <2 years, 2 to <5 years, 5 to <10 years, 10 to <20 years, ≥20 years)
- Baseline IGA 3 versus 4
- Baseline pruritus <4 versus ≥4
- Prior use of systemic treatment (yes, no)

Some additional subgroup analyses may be added to meet regulatory requirement. The analysis of additional subgroups will not require an amendment to the SAP.

6.15.2. Safety Subgroup Analyses

Subgroup analysis for safety related endpoints will be performed within the context of the integrated safety analysis. Subgroup analyses may be added to meet regulatory requirement. The analysis of additional subgroups will not require an amendment to the SAP.

6.16. Protocol Deviations

Protocol deviations will be identified throughout the study. Important protocol deviations (IPD) are defined as those deviations from the protocol likely to have a significant impact on the completeness, accuracy, and/or reliability of the study data or that may significantly affect a patient's rights, safety, or well-being.

Potential examples of IPDs include patients who violated the inclusion/exclusion criteria, used an interfering concomitant medication, significant non-compliance with study medication (<75% of expected injections). Refer to a separate document called "KGAD Trial Issues Management Plan" for the IPDs with categorizations.

The number and percentage of patients having IPD(s) will be summarized within category and subcategory of deviation by treatment group for Treatment Period using the ITT Population.

A by-patient listing of IPDs for the ITT Population will be provided.

6.16.1. Impact of COVID-19

Impact of pandemic (for example, COVID-19) on analyses may be addressed prior to study unblinding at Week 16 DBL, once the impact on study conducts are fully understood. In general, any missing assessments/visit window will be documented as protocol deviations. For patients who have missing assessments at Week 16 due to COVID-19, these patients may enter the escape arm. A summary or listing may be provided to summarize missing visits due to COVID-19.
Treatment discontinuation due to pandemic will be treated the same type of intercurrent event as treatment discontinuation due to reasons other than lack of efficacy. Strategies of how this type of intercurrent event will be handled are described in Section 6.2. Intermittent missing assessment due to pandemic will be treated the same as any other intermittent missing values. Details of how missing data will be handled are described in Section 6.4.

6.17. Interim Analyses and Data Monitoring

Data Monitoring Committee/Data Safety Monitoring Board (DMC/DSMB): The lebrikizumab Phase 3 AD programs' DSMB is an independent expert advisory group commissioned and charged with the responsibility of evaluating cumulative safety at regular intervals, as well as on an ad hoc basis, as needed. The DSMB will consist of members external to Lilly and follow the rules defined in the DSMB charter, focusing on potential and identified risks for this molecule. Data Monitoring Committee membership will include, at a minimum, a physician with expertise in dermatology and a statistician. No member of the DSMB may have contact with study sites. This committee will make recommendations as to a) continue the clinical studies without modification; or b) continue the clinical studies with modifications; or c) terminate one or more of the clinical studies. Details outlining the roles and responsibilities of the DMC are documented in the "Dermira DRM06 DSMB Program Charter" and the planned analyses are outlined in the DMC analysis plan prior to the first unblinded assessment.

Access to the unblinded safety data will be limited to the DSMB. The study team will not have access to the unblinded data. Only the DSMB is authorized to evaluate unblinded data. The purpose of the DSMB is to advise Lilly regarding patient safety; however, the DSMB may request key efficacy data to put safety observations into context and to confirm a reasonable benefit/risk profile for ongoing patients in the study. Hence, there will be no alpha adjustment for these interim assessments.

Week 16 Database lock (DBL): An unblinded interim analysis will be performed at the time (that is, a cut-off date) the last patient completes Week 16 or the ETV from the study. This database lock will include all data collected by the cut-off date. Treatment assignment will be unblinded at the time of this interim lock.

Final DBL: A final DBL will occur after all patients have completed the safety follow-up period of the study, discontinued current study, or enrolled into the long-term extension study DRM06-AD07.

6.18. Annual Report Analyses

Based on regulatory requirements for the Development Safety Update Report (DSUR), reports will be produced (if not already available from the study CSR) for the reporting period covered by the DSUR.

6.19. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

Summary of AEs, provided as a dataset, will be converted to an XML file. Both serious adverse events (SAEs) and 'Other' AEs are summarized by treatment group and by MedDRA PT.

- An AE is considered 'Serious' whether or not it is a TEAE.

- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event,
    - the number of participants who experienced each event term, and
    - the number of events experienced.
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

7. Unblinding Plan

Unblinding details are specified in a separated unblinding plan.

8. References

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med*. 2014;33(4):693–713. https://doi.org/10.1002/sim.5974

Andersen SW, Millen BA. On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharm Stat*. 2013;12(1):7–16. https://doi.org/10.1002/pst.1548

Basra MKA, Fenech R, Gatt RM, et al. The Dermatology Life Quality Index 1994-2007: a comprehensive review of validation data and clinical results .*Br J Dermatol*. 2008;159(5):997-1035. https://doi.org/10.1111/j.1365-2133.2008.08832.x

Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28:586–604.

Bretz F, Posch M, Glimm E, et al. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. Biom J. 2011;53(6):894-913.

Finlay AY, Khan GK. Dermatology Quality of Life Index (DLQI)--a simple practical measure for routine clinical use. *Clin Exp Dermatol*. 1994;19(3):210-216. https://doi.org/10.1111/j.1365-2230.1994.tb01167.x

Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38. https://doi.org/10.2307/2336755

Hongbo Y, Thomas CL, Harrison MA, et al. Translating the science of quality of life into practice: what do dermatology life quality index scores mean? *J Invest Dermatol*. 2005;125(4):569-664. https://10.1111/j.0022-202X.2005.23621.x

Khilji FA, Gonzalez M, Finlay AY. Clinical meaning of change in Dermatology Life Quality Index scores. *Br J Dermatol*. 2002;147(suppl 62):50.

PhUSE resources page. Analyses and Displays Associated with Measures of Central Tendency – Focus on Vital Sign, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Submission Documents. 2013. Available at: http://www.phusewiki.org/docs/CSS%20White%20Papers%202016/CSS_WhitePaper_CentralTendency_v1.0.pdf. Accessed June 09, 2020.

PhUSE resources page. Analyses and Displays Associated with Outliers or Shifts from Normal to Abnormal: Focus on Vital Signs, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Summary Documents. 2015. Available at http://www.phusewiki.org/docs/CSS%20White%20Papers%202016/CS_WhitePaper_OutliersShifts_v1.0.pdf. Accessed June 09, 2020.

[PhUSE]. Pharmaceutical Users Software Exchange. PhUSE Computational Science Standard Analyses and Code Sharing Working Group. Analysis and Display White Papers Project Team. Analysis and displays associated with adverse events: focus on adverse events in Phase 2-4 clinical trials and integrated summary documents. Version 1.0. Available at: https://www.phuse.eu/documents/working-groups/cs-whitepaper-adverseevents-v10-4442.pdf. Published 03 February 2017. Accessed June 09, 2020.

Ratitch B. Combining analysis results from multiple imputed categorical data. PharmaSUG. 2013. Paper SP03.

Sato T. On the variance estimator of the Mantel-Haenszel risk difference. *Biometrics*. 1989;45(4):1323–1324.

Waters A, Sandhu D, Beattie P, et al. Severity stratification of Children's Dermatology Life Quality Index (CDLQI) scores. *Br J Dermatol*. 2010;163 (suppl 1):121.

Winthrop KL, Novosad SA, Baddley JW, et al. Opportunistic infections and biologic therapies in immune-mediated inflammatory diseases: consensus recommendations for infection reporting during clinical trials and postmarketing surveillance. *Ann Rheum Dis*. 2015;74(12):2107-2116. https://doi.org/10.1136/annrheumdis-2015-207841

9. Appendices

Appendix 1. Study Visit Mapping for Pruritus NRS and Sleep-loss Diary and PEOM

Pruritus NRS and sleep loss are collected as a daily diary; entries will be mapped to study week by the following:

Week	Start Day	End Day
Baseline	Date of First Injection[a] - 7	Date of First Injection - 1
Week 1	Max(Date of First Injection, Week 2 Office Visit Date - 14)	Week 2 Office Visit Date - 8
Week 2	Week 2 Office Visit Date - 7	Week 2 Office Visit Date - 1
Week 4	Week 4 Office Visit Date - 7	Week 4 Office Visit Date - 1
Week 6	Week 6 Office Visit Date - 7	Week 6 Office Visit Date - 1
Week 8	Week 8 Office Visit Date - 7	Week 8 Office Visit Date - 1
Week 10	Week 10 Office Visit Date - 7	Week 10 Office Visit Date - 1
Week 12	Week 12 Office Visit Date - 7	Week 12 Office Visit Date - 1
Week 14	Week 14 Office Visit Date - 7	Week 14 Office Visit Date - 1
Week 16	Week 16 Office Visit Date - 7	Week 16 Office Visit Date - 1

[a] If date of first injection is missing, the randomization date will be used.

If multiple assessments on a single day are present, use the first assessment. If an assessment could be mapped to different weeks, it will be mapped to the earlier week. During the Double-Blinded treatment period, prorated weekly average score will be calculated if at least 1 of the 7 days score is not missing. If the range of 7 days are all missing daily assessments, then the weekly result is missing.

Patient-Oriented Eczema Measure (POEM) scores are collected every week via eDiary; the visit-week mapping will follow the following rule: the last collected POEM data before or on the visit date would be used, the evaluation window is injection date - 7 to injection date -1 for baseline and assessment date - 7 to assessment date -1 for postbaseline. For example, if a patient gets an injection/assessment on the 14th, the scale completed in between the 13th and the 7th would be used. It is preferred not to use a POEM score, even if it is completed on the 15th or 16th, as those scores are now considered biased by the new drug administration. It is also preferred to not use data from 14th, because it is impossible to determine if the entry on 14th is before or after injection/assessment. To avoid bias, the 14th is excluded from evaluation window.

Appendix 2. Definition of Topical and Systemic Atopic Dermatitis Therapy

The atopic dermatitis rescue therapy in this study is defined as: high-potency topical corticosteroid (TCS) and systemic atopic dermatitis therapy. The topical treatments and systemic treatments are defined as following:

1. **Topical Atopic Dermatitis Treatment (including topical corticosteroids, topical calcineurin inhibitors [TCIs], and crisaborole)**

Route of topical treatment includes: Topical or Transdermal

Topical Corticosteroids (TCS): Anatomical Therapeutic Chemical (ATC) code is D07

High-Potency TCS: ATC codes: D07AC or D07AD

Low or Moderate-Potency TCS: ATC code is D07, excluding D07AC or D07AD

TCI: Preferred Term includes: TACROLIMUS and PIMECROLIMUS

Crisaborole: Preferred Term includes: CRISABOROLE

2. **Systemic Atopic Dermatitis Treatment (including systemic corticosteroids, immunosuppressant, biologics and phototherapy/photochemotherapy)**

Route of systemic treatments administration includes: Oral, Intra-Arterial, Intramuscular, Intraperitoneal, Intravenous, Subcutaneous, and Transdermal. (This condition applies to the following categories except for phototherapies.)

Systemic Corticosteroids: ATC code is H02.

Immunosuppressant: Defined as: ATC2 is L04 or Preferred Terms (PTs) of Abrocitinib or Ruxolitinib.

Biologics: Defined as following PTs:

- Infliximab, Infliximabum, Etanercept, Etanerceptum, Adalimumab, Adalimumabum, Certolizumab, Certolizumabum, Certolizumab pegol,Golimumab, Golimumabum, Ozoralizumab, Afelimomab, Afelimomabum, Tumor Necrosis Factor Alpha (TNF-) Inhibitors, Tabalumab, Tregalizumab, Anakinra, Basiliximab, Basiliximabum, Daclizumab, Daclizumabum, Tocilizumab, Tocilizumabum, Mepolizumab, Mepolizumabum, Rilonacept, Rilonaceptum, Ustekinumab, Canakinumab, Briakinumab, Fezakinumab, Sirukumab, Sarilumab, Lebrikizumab, Secukinumab, Olokizumab, Gevokizumab, Brodalumab, Ladarixin, Ixekizumab, Dupilumab, Tildrakizumab, Tildrakizumabum, Reslizumab, Reslizumabum, Guselkumab, Guselkumabum, Olamkicept, Fletikumab, Bimekizumab, Mirikizumab, Risankizumab, Abatacept, Ligelizumab, Vedolizumab, Belimumab, Nemolizumab, Tralokinumab, Omalizumab

Phototherapy or Photochemotherapy:

If the programming search of the medication name (actual term or PT) contains ‘photo,’ then medical is to manually review to confirm whether the medication in question is indeed ‘Phototherapy’ or ‘Photochemotherapy.’

Appendix 3. Details of Combining Estimates and Test Statistics for Categorial Endpoints with Multiple Imputation

Following the implementation of Markov Chain Monte Carlo multiple imputation (MCMC-MI) imputation as specified in Section 6.4.1, the 25 data sets with imputations should be set together and sorted by imputation number. The following sections describe the processes for combining inferences for the individual imputed data sets into 1 inference for reporting. All calculations are performed in SAS software version 9.4.

Summarize Unadjusted Response Rate

The response rates, overall and by treatment arm, and their associated standard errors (SEs) are computed for each imputed data set using PROC FREQ with the *riskdiff* option specified for the appropriate column in the TABLES statement. The response rates and SEs from the resulting output are combined across the 25 imputed data sets using PROC MIANALYZE, separately for each arm and the overall group.

Note that the estimate and 95% confidence interval (CI) bounds output by PROC MIANALYZE are percents (that is, they are in terms of the response rate). To obtain the number of responders, the estimated percent is multiplied by the number of individuals in the analysis population and rounded to the nearest integer.

Compute Stratified Measures of Association

The common risk difference, common odds ratio (OR), and Cochran-Mantel-Haenszel (CMH) test statistic are computed for each imputed data set using PROC FREQ with the *riskdiff* option for the appropriate column (for risk difference) and the *cmh* option (for odds ratio and CMH test statistic) specified in the TABLES statement. Each of these analyses are stratified by geographic region, age group, and baseline disease severity via inclusion of these variables in the TABLES statement with the treatment and outcome variables.

Note that the PROC FREQ output corresponding to the Mantel-Haenszel method is used for the risk difference, and the output corresponding to the General Association statistic is used for the CMH statistic. PROC MIANALYZE is then called separately for each of these measures, with further details in the sections below.

Common Risk Difference

No transformation is necessary before using PROC MIANALYZE to combine the risk difference estimates and their associated SEs across the 25 imputed data sets. This procedure outputs an estimate of the common risk difference and the associated 95% CI bounds.

Common Odds Ratio

The OR from each imputed data set is first transformed using the natural logarithm. The SE for each log OR (SE_{lOR}) is derived from the OR 95% CI bounds (LB_{OR}, UB_{OR}) according to the following equation: $SE_{lOR} = (\ln(UB_{OR}) - \ln(LB_{OR}))/(2 * 1.96)$. The log OR and derived SE are then combined using PROC MIANALYZE, which outputs a combined estimate of the log OR and the associated 95% CI. Finally, these measures can be exponentiated to transform them back to the OR scale.

CMH Test

The CMH test statistic (CMH) from each imputed data set is transformed using the Wilson-Hilferty transformation and standardized so that it has approximately a standard normal distribution (Ratitch 2013). In particular, the transformed CMH statistic is computed as follows:

$$CMH_{WH} = \frac{(\frac{CMH}{df})^{\frac{1}{3}} - (1 - \frac{2}{9*df})}{\sqrt{\frac{2}{9*df}}},$$

where df is the degrees of freedom of the CMH statistic. Then the SE for each CMH_{WH} is 1, and PROC MIANALYZE is used to output a combined estimate of the transformed CMH statistic. Note that the 2-sided p-value output by PROC MIANALYZE is not used directly, but instead the 1-sided p-value is computed manually using both the t statistic and 2-sided p-value output by PROC MIANALYZE; if t statistic is >0, then the 1-sided p-value is computed as one half of the 2-sided p-value; otherwise, the 1-sided p-value is computed as one half of the 2-sided p-value. The resulting 1-sided p-value is reported as the pooled p-value for the CMH test.

Leo Document ID = e0255821-6fc0-46ae-83e9-32466b82a78f

PPD

Approval Date & Time: 15-Oct-2021 18:41:01 GMT

Signature meaning: Approved