Statistical Analysis Plan for Study M16-813

A Phase 2, Multicenter, Randomized, Placebo-Controlled, Double-Blind Study to Evaluate Risankizumab in Adult and Adolescent Subjects with Moderate to Severe Atopic Dermatitis

Date: 14 October 2020

Version 5.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for Risankizumab Study Protocol M16-813 A Phase 2, Multicenter, Randomized, Placebo-Controlled, Double-Blind Study to Evaluate Risankizumab in Adult and Adolescent Subjects with Moderate to Severe Atopic Dermatitis.

Study M16-813 examines the efficacy and safety of Risankizumab 150 mg and 300 mg in adult and adolescent subjects with moderate to severe AD with a confirmed diagnosis of moderate to severe AD and onset of symptoms at least 2 years prior to the Baseline Visit.

The analyses of pharmacokinetic endpoints and biomarker samples will not be covered in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) or later under the UNIX operating system.

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

The primary objective of this study is to assess the safety and efficacy of risankizumab for the treatment of moderate to severe atopic dermatitis (AD) in adult and adolescent subjects.

2.2 Study Design Overview

This is a Phase 2, global, randomized, double-blind, placebo-controlled multicenter study that will evaluate the efficacy and safety of two dose levels, 150 mg and 300 mg respectively, of risankizumab in adult (≥ 18 years of age) and adolescent (12 - 17 years of age) subjects with a confirmed diagnosis of moderate to severe AD and onset of symptoms at least 2 years prior to the Baseline Visit.

This study is comprised of an approximately 35-day screening period, a 16-week Double Blind (DB) Treatment Period A, a 36-week DB Treatment Period B, and a 20-week (140 days) Follow-up Visit.

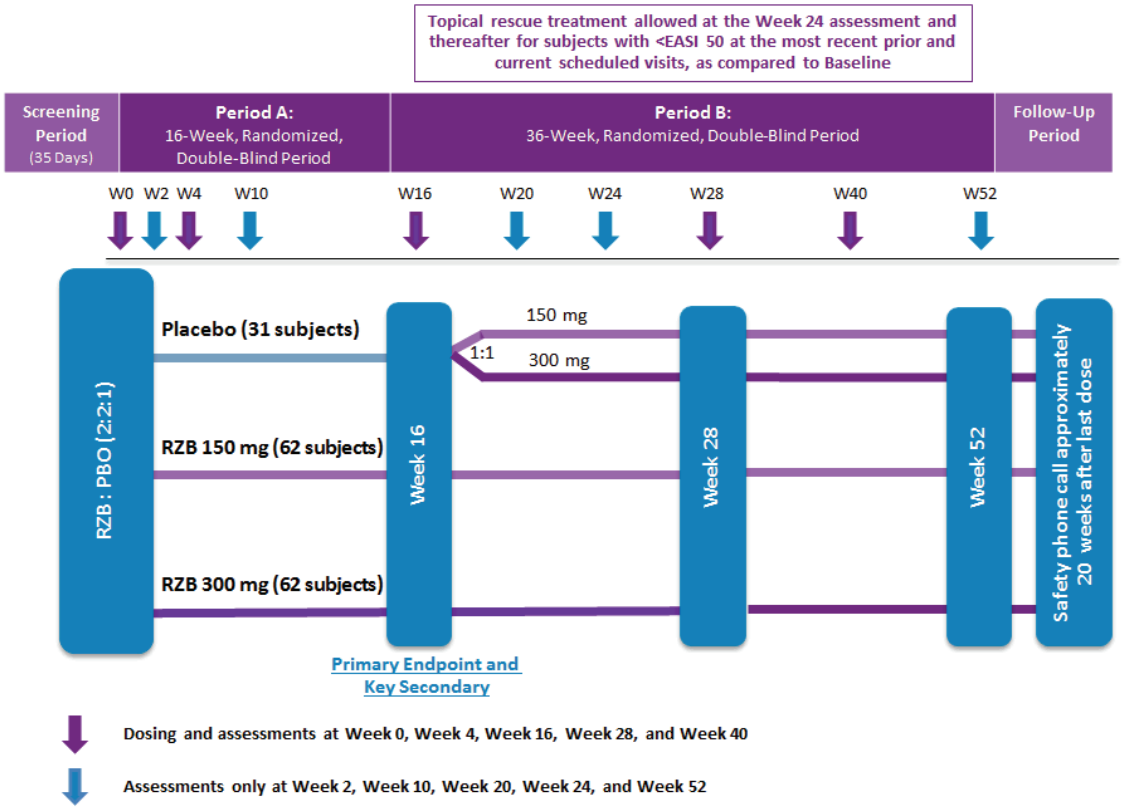
- Period A (Baseline/Week 0 – Week 16) is a 16-week double-blind, placebo-controlled treatment period. Subjects who meet the study's eligibility criteria (Protocol Section 5.1) will be randomized at the Baseline visit to receive risankizumab 150 mg, risankizumab 300 mg or matching placebo. Study drug administration for Period A will occur at Weeks 0 and 4. The last visit in Period A is planned at Week 16 when all scheduled efficacy and safety assessments will be performed.
- In Period B (Week 16 – Week 52), subjects receiving placebo in Period A will be re-randomized at Week 16 to receive risankizumab 150 mg or risankizumab 300 mg through Week 52. Subjects originally randomized to risankizumab 150 mg or 300 mg in Period A will stay on their previously assigned treatment through Week 52. Study drug administration for Period B will occur at Weeks 16, 28 and 40.
- A follow-up telephone call visit will occur 140 days after the last dose of study drug to determine the status of any ongoing AEs/SAEs or the occurrence of any new AEs/SAEs and use of any concomitant medications.

During Period B, subjects with < EASI 50 response as compared to Baseline at the most recent prior and current scheduled visits will be allowed to use approved concomitant rescue treatment as early as at Week 24 and at any visit thereafter. Rescue treatment is limited to up to twice-daily application of topical corticosteroids.

The primary analysis will be conducted after all ongoing subjects have completed Week 16. The Week 16 evaluation is the final analysis for the primary endpoint of EASI 75 and key secondary endpoints of vIGA-AD response and worst pruritus NRS 4 point reduction. The study sites and subjects will remain blinded to treatment assignments for the duration of the study.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



PBO = placebo; RZB = risankizumab; W = week

2.3 Treatment Assignment and Blinding

Each subject will be assigned a unique identification number by the IRT at the screening visit. Subjects who rescreen will use the same unique identification number from the initial screening visit. In addition, the IRT will assign a unique randomization number that will encode each subject's treatment group assignment according to the randomization schedule generated by the statistics department at AbbVie.

At the Baseline visit, subjects meeting eligibility criteria will be randomized in a 2:2:1 ratio to risankizumab 150 mg, risankizumab 300 mg, or matching placebo.

Randomization at Baseline will be stratified by baseline disease severity (moderate [vIGA-AD 3] vs. severe [vIGA-AD 4]) and geographic region (Japan vs. Rest of the World [RoW]).

At the Week 16 visit, subjects receiving placebo in Period A will be re-randomized in a 1:1 Ratio to receive either risankizumab 150 mg or risankizumab 300 mg through Week 52. Subjects originally randomized to risankizumab 150 mg or 300 mg in Period A will stay on their previously assigned treatment through Week 52. There is no stratification at the re-randomization at Week 16.

To maintain the blind, all risankizumab and placebo kits provided for the study will be identical in appearance. AbbVie personnel will remain blinded until the primary analysis at Week 16 is available. The study sites and subjects will remain blinded to treatment assignments for the entire duration of the study.

2.4 Sample Size Determination

Approximately 155 subjects will be randomized to risankizumab 150 mg, risankizumab 300 mg, or placebo in a ratio of 2:2:1 (62 subjects each for the risankizumab 150 mg and 300 mg groups and 31 subjects for the placebo group). Assuming an EASI 75 response rate of at most 15% in the placebo group, this sample size will provide more than 90% power to detect the treatment difference of at least 36% between each risankizumab group and placebo group using a two-sided test at a 0.025 significance level.

3.0 Endpoints

3.1 Primary Endpoint(s)

The primary endpoint is the proportion of subjects achieving at least a 75% reduction from Baseline in Eczema Area and Severity Index (EASI 75) at Week 16.

3.2 Secondary Endpoint(s)

The key secondary endpoints for this study are:

1. The proportion of subjects achieving validated Investigator Global Assessment Scale for Atopic Dermatitis (vIGA-AD) of "0" or "1" (on a 5-point scale) with a reduction from Baseline of ≥ 2 points at Week 16.
2. The proportion of subjects achieving a reduction of ≥ 4 points in worst pruritus numerical rating scale (NRS) from Baseline to Week 16.

Other secondary endpoints that are not included in the multiplicity testing procedure (see Section 12.0) are listed below and will be evaluated for each of the specified time points:

- Percent change in EASI from Baseline to Week 16, Week 28, and Week 52.
- Proportion of subjects achieving EASI 75 at Week 28 and Week 52.
- Proportion of subjects achieving EASI 50 at Week 16, Week 28, and Week 52.
- Proportion of subjects achieving EASI 90 at Week 16, Week 28, and Week 52.
- Proportion of subjects achieving vIGA-AD of "0" or "1" with a reduction from Baseline of ≥ 2 points at Week 28 and Week 52.
- Change in body surface area (BSA) from Baseline to Week 16, Week 28, and Week 52.
- Proportion of subjects achieving 50% improvement in SCORing Atopic Dermatitis (SCORAD 50) at Week 16, Week 28, and Week 52.
- Proportion of subjects achieving SCORAD 75 at Week 16, Week 28, and Week 52.
- Proportion of subjects achieving SCORAD 90 at Week 16, Week 28, and Week 52.
- Proportion of subjects achieving Dermatology Life Quality index (DLQI) of "0" or "1" at Week 16, Week 28, and Week 52.
- Proportion of subjects achieving Children's Dermatology Life Quality index (CDLQI) of "0" or "1" at Week 16, Week 28, and Week 52.
- Proportion of subjects achieving a DLQI improvement of ≥ 4 points at Week 16, Week 28, and Week 52 among subjects with a DLQI ≥ 4 at Baseline.
- Change in DLQI from Baseline to Week 16, Week 28, and Week 52.
- Change in CDLQI from Baseline to Week 16, Week 28, and Week 52.

- Change in worst pruritus NRS from Baseline to Week 16, Week 28, and Week 52.
- Proportion of subjects achieving a reduction of ≥ 4 points in worst pruritus NRS from Baseline to Week 28 and Week 52.

3.3 Other Efficacy Endpoint(s)

All variables listed as primary or secondary efficacy endpoints will be analyzed at all visits in addition to those specified above. In addition, the following additional endpoints will be evaluated at all applicable visits:

- Proportion of subjects achieving EASI 100.
- Proportion of subjects achieving vIGA-AD of "0" with a reduction from Baseline of ≥ 2 points
- Percent change in SCORAD from baseline.
- Change from Baseline in Patient Oriented Eczema Measure (POEM).
- Proportion of subjects achieving an improvement (reduction) in POEM of ≥ 4 from Baseline among subjects with a POEM ≥ 4 at Baseline.
- Change from Baseline in Patient Global Impression of Severity (PGIS).
- Proportion of subjects who report symptoms to be "Minimal" or "Absent" for PGIS.
- Proportion of subjects who have "Very much improved" or "Much improved" for Patient Global Impression of Change (PGIC).
- Change from Baseline in Atopic Dermatitis Symptom Scale (ADerm-SS) 7-item total symptom score (TSS-7) and 11-item total symptom score (TSS-11).
- Change from Baseline in ADerm-SS skin pain score.
- Change from Baseline in Atopic Dermatitis Impact Scale (ADerm-IS) domain scores (sleep, emotional state, daily activities).
- Proportion of subjects achieving an improvement (reduction) in ADerm-SS TSS-7 ≥ minimal clinically important difference (MCID) from Baseline for subjects with ADerm-SS TSS-7 ≥ MCID at Baseline.

- Proportion of subjects achieving an improvement (reduction) in ADerm-SS TSS-11 ≥ MCID from Baseline for subjects with ADerm-SS TSS-11 ≥ MCID at Baseline.
- Proportion of subjects achieving an improvement (reduction) in ADerm-SS skin pain score ≥ MCID from Baseline for subjects with ADerm-SS skin pain score ≥ MCID at Baseline.
- Proportion of subjects achieving an improvement (reduction) in ADerm-IS sleep domain score ≥ MCID from Baseline for subjects with ADerm-IS sleep domain score ≥ MCID at Baseline.
- Proportion of subjects achieving an improvement (reduction) in ADerm-IS emotional state domain score ≥ MCID from Baseline for subjects with ADerm-IS emotional state domain score ≥ MCID at Baseline.
- Proportion of subjects achieving an improvement (reduction) in ADerm-IS daily activities domain score ≥ MCID from Baseline for subjects with ADerm-IS daily activities domain score ≥ MCID at Baseline.
- Change from Baseline in EuroQoL-5D-5L (EQ-5D-5L).

Of note, MCIDs are planned to be established prior to the primary analysis database lock.

3.4 Safety Endpoint(s)

The following endpoints will be included in the safety analyses:

- Treatment emergent adverse events (TEAEs)
- Serious adverse events (SAEs)
- Areas of safety interest (ASIs)
- Adverse events (AEs) leading to study drug discontinuation
- Vital signs, laboratory tests and electrocardiogram (ECG) parameters.

3.5 Additional Endpoint(s)

The pharmacokinetic endpoints will be analyzed separately and not be covered in this SAP.

4.0 Analysis Populations

The following population sets will be used for the analyses.

The intent-to-treat (ITT) population, which includes all randomized subjects, will be used for all efficacy analyses. Subjects will be included in the analysis according to the treatment groups that they are randomized to. Subjects who are randomized to placebo in Treatment Period A and do not continue into Treatment Period B will be excluded from the analysis in Treatment Period B.

The following populations will be used for safety analyses:

- The safety population in Period A is defined as all subjects who are randomized at Baseline and received at least one dose of study drug in Treatment Period A. Subjects are assigned to a treatment group based on the treatment actually received.
- The safety population in Period B is defined as all subjects who received at least one dose of study drug in Treatment Period B. Subjects are assigned to a treatment group based on the treatment actually received.
- An All Risankizumab Treated Population (ALL_RZB) will consist of all subjects who received at least one dose of risankizumab in the study. This population will be used to provide a comprehensive summary of safety.

For the safety populations, subjects are assigned to a treatment group based on the "as treated" treatment group, regardless of the treatment randomized. The "as treated" is determined by the first dose subject will receive for Period A, and is determined by the first dose subject will receive for Period B.

5.0 Subject Disposition

The total number of subjects who were screened, randomized, and treated will be summarized. Reasons for exclusion, including screen failure, will be summarized.

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized for each treatment group:

- Subjects randomized in the study;
- Subjects who took at least one dose of study drug;
- Subjects who completed study;
- Subjects who prematurely discontinued study drug (all reasons and primary reason);
- Subjects in each analysis population, as applicable.

For end of study participation, the number and percentage of subjects who completed each treatment period (or did not with associated reasons) will be summarized overall and by treatment group.

6.0 Study Drug Duration and Compliance

For safety population, duration of treatment for each treatment period will be summarized for each treatment group and for all investigational study drug dose groups combined.

Study Drug Exposure (in Days) in Each Period:

In Period A, duration of treatment is defined as:

- For subjects who did not continue to Period B,
  - last injection date in Period A minus first injection date in Period A + 84 days.
- For subjects who continued to Period B, the minimum of
  - last injection date in Period B minus first injection date in Period A,
  - last injection date in Period A minus first injection date in Period A + 84 days.

In Period B, duration of treatment is defined as:

- last injection date in Period B minus first injection date in Period B + 84 days.

For ALL_RZB population, duration of treatment is defined as

- date of last injection of risankizumab minus date of first injection of risankizumab + 84 days.

Duration of treatment will be summarized using the number of subjects treated, mean, standard deviation, median, minimum and maximum.

Compliance

There will be a summary of the number of subjects receiving study drug at each study drug administration visit during each treatment period. This will be repeated on the cumulative number of doses.

When computing compliance at each study drug administration visit, the denominator will include all subjects in each analysis population who have not prematurely discontinued the study drug prior to the scheduled study drug injection. Subjects who have prematurely discontinued the study drug but have not prematurely discontinued the study are not used in the denominator.

7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications

Demographics, baseline or disease characteristics, medical history, and prior and concomitant medications will be summarized for the ITT population by treatment group and overall. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum).

7.1 Demographics and Baseline Characteristics

The following demographic and baseline characteristics will be summarized.

Subject Demographics

- Continuous demographic variables include age, weight, height, and body mass index (BMI).
- Categorical demographic variables include
  - Sex (male, female)
  - Ethnicity (Hispanic or Latino, Other)
  - Race (White, Black, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Multi Race)
  - Age categories (< 18 or ≥ 18 years; < 18, ≥ 18 - < 40, ≥ 40 – < 65, or ≥ 65 years),
  - BMI category (< 25, ≥ 25 – < 30, or ≥ 30 kg/m2)
  - Geographic region (Japan or rest of world)

Disease Characteristics

- Continuous variables including the baseline assessment of the following:
  - EASI overall score and body region scores
  - vIGA-AD score
  - Worst pruritus Numerical Rating Scale (NRS)
  - Body Surface Area (BSA) in percentage
  - Scoring Atopic Dermatitis (SCORAD)
  - Dermatology Life Quality Index (DLQI)
  - Children's Dermatology Life Quality Index (CDLQI)
  - Patient Oriented Eczema Measure (POEM)
  - Patient Global Impression of Severity (PGIS)
  - Patient Global Impression of Change (PGIC)
  - EuroQoL Dimensions 5 Levels (EQ-5D-5L)
  - Atopic Dermatitis Symptom Scale (ADerm-SS) 7-item total symptom score (TSS-7)
  - ADerm-SS 11-item total symptom score (TSS-11)

- ADerm-SS skin pain score
- Atopic Dermatitis Impact Scale (ADerm-IS) domain scores (sleep, emotional state, daily activities)
- AD disease duration since diagnosis (years),
- AD disease duration since symptoms started (years),
- duration between symptoms and diagnosis of AD disease (years)

- Categorical variables including:
  - Baseline vIGA-AD score (< 4, 4)
  - Baseline vIGA-AD score (0, 1, 2, 3, 4)
  - Baseline EASI (< median, ≥ median)
  - Baseline Electrocardiogram (ECG) categories, including normal, abnormal – not clinically significant, abnormal – clinically significant, unable to evaluate, and details of the abnormality when applicable.
  - Intrinsic/non-allergic subtype (IgE level < 200 kU/L, IgE level ≥ 200 kU/L)
  - Treatment-emergent ADA status during Week 0 to Week 16 (yes, no)
  - Treatment-emergent NAb status during Week 0 to Week 16 (yes, no)
  - Tobacco user (current, former, never, unknown)
  - Alcohol user (current, former, never, unknown)

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment group. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

Subjects with cardiovascular history and CV risk factors will also be summarized.

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name. A prior medication is defined as any medication taken prior to the date of the first dose of study drug. A concomitant medication is defined as any medication that started prior to the date of the first dose of study drug and continued to be taken after the first dose of study drug or any medication that started on or after the date of the first dose of study drug, but not after the date of the last dose of study drug. The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

8.0 Efficacy Analyses

8.1 General Considerations

All efficacy analyses will be conducted in the ITT Population.

The primary analysis will be performed after all ongoing subjects have completed Week 16 and the database has been locked. The Week 16 evaluation will be the only and final analysis for the primary endpoint of EASI 75 and key secondary endpoints of vIGA-AD response and worst pruritus NRS 4 point reduction.

Analysis of Categorical Variables

Categorical variables will be summarized by frequencies and percentage by each treatment group. Pairwise comparison of each risankizumab group versus placebo group will be performed using Cochran-Mantel-Haenszel (CMH) test, stratified by vIGA-AD categories (moderate [vIGA-AD 3] vs. severe [vIGA-AD 4]). Point estimates and 95% CIs for the difference in proportions between each risankizumab group and placebo will be provided. Construction of CIs for the common risk difference will be based on the Mantel-Haenszel estimate adjusting for the stratification factor.

Analysis of Continuous Variables

The model based mean and standard error will be provided for continuous variables. The Baseline and visit means will also be presented for each treatment group among subjects who have both Baseline and at least one post Baseline visit values. Treatment groups will be compared using MMRM model as described in Section 8.2 with treatment, vIGA-AD categories and the corresponding baseline value in the model. Point estimates and 95% CIs of mean change from baseline within treatment groups, and between each risankizumab treatment group and placebo will be provided.

Unless otherwise specified, any subject who is randomized based on a wrong stratum will be analyzed according to the actual stratum the subject belongs to.

"Baseline" refers to the last non-missing observation before the first administration of study drug or randomization if no study drug is given.

8.2 Handling of Missing Data

Missing data will be imputed using the following methods for the efficacy analyses:

- **N**on-**R**esponder **I**mputation incorporating Multiple Imputation (MI) to handle missing data due to **C**OVID-19 (NRI-C): the NRI-C analysis will categorize any subject who does not have evaluation during a pre-specified visit window (either due to missing assessment or due to early withdrawal from the study) as a non-responder for that visit. The exceptions are: 1) when the subject is a responder both before and after a specific visit window, in which case the subject will be categorized as a responder for the visit. (Only observations within the same analysis period will be used.) 2) missing data due to COVID-19 infection or logistical restriction will be handled by Multiple Imputation. In addition, all assessments after the start of rescue medications or prohibited medications will not be included in the analyses; as a result, subjects will be counted as non-responders thereafter and will not be imputed by MI. The NRI-C will be the primary approach in the analyses of categorical variables in Period A.

- A sensitivity analysis for categorical endpoints will use NRI with No special data handling for missing due to COVID-19 (NRI-NC). NRI-NC will be performed in the same way as NRI-C without the exception #2 above. That is, missing due to COVID-19 infection or logistical restriction will also be counted as non-responders. Subjects whose change/percent change from Baseline cannot be calculated because of a missing Baseline will be considered as a non-responder at all post-baseline visits in both NRI-C and NRI-NC approaches.
- Mixed-Effect Model Repeat Measurement (MMRM): The repeated measures analysis will be conducted using the mixed-effect model including observed measurements at all visits. The mixed-effect model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, main stratification factors at randomization (vIGA-AD categories [moderate vs severe]), and the continuous fixed covariates of baseline measurement. An unstructured variance covariance matrix will be used. Parameter estimation is based on the assumption of data being missing at random and using the method of restrictive maximum likelihood (REML). The fixed effects will be used to report visit means at corresponding visits. MMRM will be the primary approach in the analysis of continuous variables in Period A.
- Observed Cases (OC): The OC analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the OC analysis for that visit. OC will exclude values after a subject prematurely discontinue from study drug. OC analysis will be used for the summaries of long-term efficacy for all variables in Period B.
- Multiple Imputation (MI): The MI approach will be used as a sensitivity analysis for the primary efficacy endpoint and the two key secondary efficacy endpoints. MCMC will be first applied to augment data into monotonic missing pattern and PROC MI will be used to generate 30 'complete' datasets using the regression method including the following variables in the imputation model: baseline value, treatment group, stratum (vIGA-AD categories [moderate vs severe]), and measurements at each visit up to the end of the analysis period. The seed for PROC MI will be set as the numeric form of the date when the 1st subject in this study is randomized. The imputed post-

baseline measurements will be rounded to the same precision as the observed data before the determination of responder status. Using the CMH model adjusted by stratification factor (vIGA-AD categories [moderate vs severe]), the imputed endpoints will be analyzed using each of the 30 datasets. SAS PROC MIANALYZE will be used to generate the final inferences of the risk difference between treatment groups.

In both Period A and Period B, efficacy assessments after the first dose of rescue or prohibited medication will not be included in the analyses for that period. As a result, after the first dose of rescue or prohibited medication, the subjects will be considered as non-responders in the NRI approach and will be considered as missing in OC and MMRM analyses.

8.3 Primary Efficacy Endpoint(s) and Analyses

8.3.1 Primary Efficacy Endpoint(s)

The primary endpoint is the proportion of subjects achieving at least a 75% reduction from Baseline in Eczema Area and Severity Index (EASI 75) at Week 16.

The corresponding statistical null hypotheses are that there is no difference between each of the risankizumab dose group (150 mg and 300 mg, respectively) and placebo group in the proportion of subjects achieving EASI 75 at Week 16.

8.3.2 Handling of Missing Data for the Primary Efficacy Endpoint(s)

Missing data will be handled using NRI method as the primary approach. MMRM and MI method will be used as the sensitivity analysis.

8.3.3 Primary Efficacy Analysis

Pairwise comparison of the primary endpoint of EASI 75 will be conducted between each risankizumab group versus the placebo group using the CMH test adjusting for the stratification factor of vIGA-AD categories.

8.3.4 Additional Analyses of the Primary Efficacy Endpoint(s)

There are no additional analyses of the primary efficacy endpoint planned for this study.

8.4 Secondary Efficacy Analyses

8.4.1 Key Secondary Efficacy Analyses

The following ranked key secondary endpoints will be tested between each risankizumab dose group and the placebo group with a graphical multiple testing procedure as described in Section 12.0.

1. the proportion of subjects achieving validated Investigator Global Assessment Scale for Atopic Dermatitis (vIGA-AD) of "0" or "1" (on a 5-point scale) with a reduction from Baseline of ≥ 2 points at Week 16.
2. the proportion of subjects achieving a reduction of ≥ 4 points in worst pruritus NRS from Baseline to Week 16.

The corresponding null hypotheses for the ranked secondary endpoints are:

1. There is no difference between each of the risankizumab dose group (150 mg and 300 mg, respectively) and placebo group in the proportion of subjects achieving vIGA-AD of "0" or "1" (on a 5-point scale) with a reduction from Baseline of ≥ 2 points at Week 16.
2. There is no difference between each of the risankizumab dose group (150 mg and 300 mg, respectively) and placebo group in the proportion of subjects achieving a reduction of ≥ 4 points in worst pruritus NRS from Baseline to Week 16.

8.4.2 Other Secondary Efficacy Analyses

All other secondary efficacy endpoints as listed in Section 3.3 will be evaluated by treatment group at each specified time point.

For each of the key efficacy endpoints as well as each of the other binary secondary efficacy endpoints, pairwise comparison of each risankizumab group versus placebo group will be analyzed using NRI method. All other continuous secondary endpoints will be analyzed using MMRM method.

8.4.3 Supportive Secondary Efficacy Analyses

There are no supportive secondary efficacy analyses planned for this study.

8.5 Additional Efficacy Analyses

All variables listed in Section 3.5 are additional efficacy endpoints and will be evaluated at all applicable visits.

Categorical endpoints will be analyzed using NRI method. Continuous endpoints will be analyzed using MMRM method.

8.6 Efficacy Subgroup Analyses

To evaluate the consistency of the efficacy over demographic and other baseline characteristics, the primary efficacy endpoint will be analyzed in the following subgroups:

- Age Group 1 (< 18 years, ≥ 18 - < 40 years, ≥ 40 – < 65 years, ≥ 65 years)
- Age Group 2 (< 18 years, ≥ 18 years)
- Sex (male, female)
- BMI (normal: < 25, over weight: ≥ 25 – < 30, obese: ≥ 30)
- Race (White, Asian, Black, and Other)
- Weight (< median, ≥ median)
- Geographic regions (Japan, rest of world)
- Baseline vIGA-AD (< 4, 4)
- Baseline EASI (< median, ≥ median)
- Intrinsic/non-allergic subtype (IgE level < 200 kU/L vs. IgE level ≥ 200 kU/L)
- Tobacco use (current; former/never/unknown)

- Treatment-emergent ADA status during Week 0 to Week 16 (yes vs. no)
  Treatment-emergent NAb status during Week 0 to Week 16 (yes vs. no)

A subgroup will be combined with the adjacent subgroup if there are less than 10% subjects in the subgroup. Any Race subgroup that contains less than 10% subjects will be combined with the Other Race subgroup for analyses.

9.0 Safety Analyses

9.1 General Considerations

Safety analyses will include adverse events, laboratory, and vital sign measurements. Safety data will be summarized by treatment group using the safety populations as defined in Section 4.0. For the Week 16 analysis, safety analyses will be conducted among the Safety Population in Period A and among the ALL_RZB Population.

For the safety analysis, subjects are assigned to a treatment group based on the first dose actually received in each treatment period, regardless of the treatment randomized.

Missing safety data will not be imputed.

9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Treatment-Emergent Adverse Events

TEAEs are defined as follows for the safety population in each treatment period and the ALL_RZB population.

TEAEs for Period A:

For subjects who do not enter Period B, TEAEs are defined as any event with an onset or worsening date on or after the first dose of study drug and no more than 140 days after the last dose of study drug.

For subjects who enter Period B, TEAEs are defined as any event with an onset or worsening date on or after the first dose of study drug in Period A and before the first dose of study drug in Period B.

TEAEs for Period B:

TEAEs are defined as any event with an onset or worsening date on or after the first dose of study drug in Period B and no more than 140 days after the last dose of study drug in the study.

TEAEs for All_RZB:

TEAEs are defined as any event with an onset or worsening date on or after the first dose of risankizumab and no more than 140 days after the last dose of risankizumab in the study.

Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent. All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and Preferred Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

Pre-treatment AEs will be summarized separately.

9.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any TEAE
- Any TEAE related to study drug according to the investigator
- Any severe TEAE
- Any serious TEAE
- Any TEAE leading to discontinuation of study drug
- Any TEAE leading to death
- Treatment-emergent Areas of Safety Interest (ASIs) as specified in Appendix B
- All deaths

9.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

The above TEAE categories will be summarized by SOC and PT; by maximum relationship to study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum severity and SOC and PT; and by subject number and SOC and PT. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

In addition, treatment-emergent adverse events will be summarized by PT and sorted by decreasing frequency in the combined risankizumab group.

9.2.4 Treatment-Emergent Adverse Events per Patient-Years of Exposure

Exposure-adjusted AEs per 100 patient-years will be provided, where AEs per 100 patient-years of exposure are defined as the number of AEs divided by the total exposure in 100 patient-years.

Note that one event per preferred term per day per subject will be counted in the calculation of the number of AEs (i.e., a preferred term will not be counted twice on the same day for the same subject). See the calculation method below.

$$100 \times \frac{\text{Number of TEAEs}}{\text{Total Patient Years}},$$

where total patient years is defined as the sum of the study drug exposure (defined as date of last dose – date of first dose + 140 days (5 half-lives)) of all subjects normalized by 365.25 and rounded to one decimal place.

9.2.5 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

SAEs (including deaths) and AEs leading to study drug discontinuation will be summarized by SOC and PT and in listing format.

9.2.6 Area of Safety Interest

Area of safety interest will be summarized by SOC and PT. The list of ASIs with their respective search strategies for each ASI category are provided in Appendix B.

The final list will be based on the most updated final version of risankizumab Product Safety Statistical Analysis Plan, which is consistent to the most updated risankizumab Risk Management Plan.

Tabular listings of ASIs will be provided.

9.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for Baseline where SAE-related laboratory assessments on or before the first dose of study drug, or values observed more than 140 days after the last dose of study drug will be excluded. The clinical laboratory tests defined in the protocol operations manual (e.g., hematology and clinical chemistry) will be summarized.

Each laboratory variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from baseline to each applicable post-baseline visit will be summarized for selected laboratory variables, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline within each treatment group and difference between treatment groups (risankizumab vs. placebo).

Changes in laboratory parameters will be tabulated using shift tables either by NCI CTC criteria or categorized as low, normal, or high based on the normal ranges of the laboratory used for each sample. A shift table from baseline either to the worst value (based on NCI CTC criteria) during treatment or to minimum and maximum value (based on normal range), will be created. A similar shift table will be provided to summarize shifts from baseline to the final post-baseline value.

Laboratory abnormalities meeting CTC criteria grade 3 and 4 will be summarized.

Laboratory abnormalities will be evaluated based on potentially clinically significant (PCS) criteria (Appendix C). For each laboratory PCS criterion, the number and percentage of subjects who have a laboratory value meeting the criteria will be summarized. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCI criteria.

Additional summaries will be presented for liver function tests including ALT/SGPT, AST/SGOT, alkaline phosphatase, and total bilirubin, with laboratory values to be categorized as follows:

- < 1.5 × ULN
- ≥ 1.5 × ULN – < 3.0 × ULN
- ≥ 3.0 × ULN – < 5.0 × ULN
- ≥ 5.0 × ULN – < 10.0 × ULN
- ≥ 10.0 × ULN – < 20.0 × ULN
- ≥ 20.0 × ULN

A listing of potentially clinically significant liver function laboratory values will include all subjects who met any of the following four criteria:

- ALT ≥ 3 × ULN, or
- AST ≥ 3 × ULN, or
- Alkaline phosphatase ≥ 1.5 × ULN, or
- Total bilirubin ≥ 2 × ULN.

A listing of possible Hy's Law cases, defined as those who meet all of the following conditions will be provided.

- ALT of > 3 × ULN or AST of > 3 × ULN,
- Associated with an increase in bilirubin ≥ 2 × ULN,
- Alkaline phosphatase < 2 × ULN

9.4 Analysis of Vital Signs

Each vital sign variable will be summarized for all applicable time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from baseline to each applicable post-baseline visit will be summarized for each vital sign variable, with the number of

observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline within each treatment group and difference between treatment groups (risankizumab vs. placebo).

If applicable, summary statistics will also be provided for change from baseline height and weight at each scheduled visit for the adolescent group.

Vital sign variables will be evaluated based on potentially clinically significant (PCS) criteria (Appendix C). For each vital sign PCS criterion, the number and percentage of subjects who have a vital sign value meeting the criteria will be summarized. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCS criteria.

9.5 Safety Subgroup Analyses

There are no safety subgroup analyses planned for this study.

9.6 Other Safety Analyses

There are no other safety analyses planned for this study.

10.0 Other Analyses

There are no other analyses planned for this study.

11.0 Interim Analyses

There are no interim analyses planned for this study.

11.1 Data Monitoring Committee

An external data monitoring committee (DMC) composed of persons independent of AbbVie and with relevant expertise in their field will review unblinded safety data from

the ongoing study. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study.

A separate DMC charter describes the roles and responsibilities of the DMC members, frequency of data reviews, relevant data to be assessed, and general operations.

Since there are no efficacy analyses for early stopping, no alpha adjustment is needed.

12.0 Overall Type-I Error Control

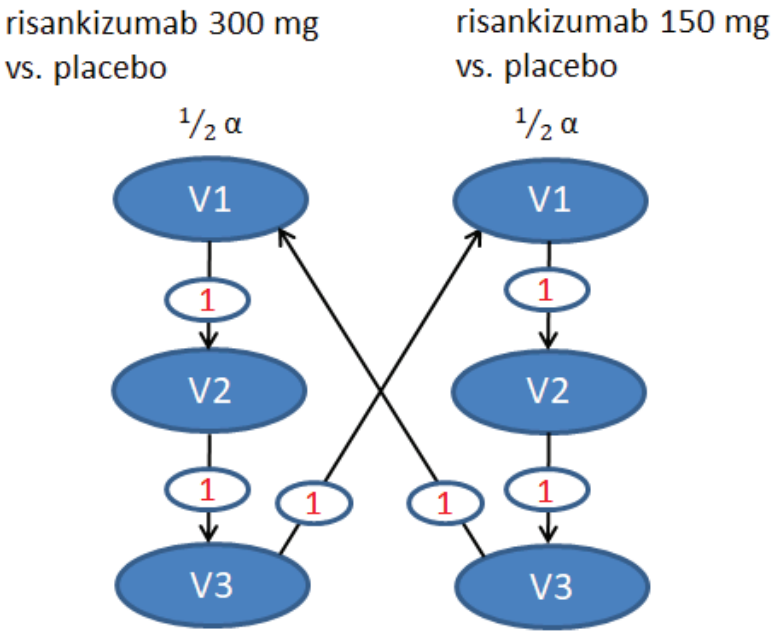
The primary and key secondary efficacy endpoints for risankizumab 150 mg and 300 mg will be tested with a graphical multiple testing procedure[1] to ensure strong control of family-wise type I error rate at two-sided significance level of 0.05. Specifically, the testing will utilize the endpoint sequence of primary endpoint followed by the ranked key secondary endpoints in the order as specified in Table 1, and will begin with testing the primary endpoint using two-sided ½ α of 0.025 for each comparison: risankizumab 150 mg vs. placebo and risankizumab 300 mg vs. placebo. Continued testing will follow a pre-specified α transfer path (Figure 2) which includes downstream transfer along the endpoints sequence within each dose as well as cross-dose transfer. Adjusted p-values for the primary and key secondary endpoints will be provided based on the testing procedure.

The graph for the testing procedure is provided in Figure 2. In the graph, the arrows specify the α transfer path. Once the null hypothesis for an endpoint is rejected (i.e., deemed significant) at its assigned significance level, its significance level will be transferred to the subsequent endpoint(s) following the arrow(s). The numbers on the arrows denote the weights for transferring and (possibly) splitting significance levels. Specifically, the weight 1 denotes 100% transfer of significance level.

Table 1. List of Primary and Key Secondary Endpoints

Abbreviation	Endpoint
V1	Proportion of subjects achieving EASI 75 from Baseline at Week 16.
V2	Proportion of subjects achieving vIGA-AD of 0 or 1 with a reduction from Baseline of ≥ 2 points at Week 16.
V3	Proportion of subjects achieving an improvement (reduction) of ≥ 4 points in worst pruritus NRS from Baseline to Week 16.

Figure 2. Graphical Approach for Multiplicity Adjustment



Since there are no efficacy analyses for early stopping planned for the DMC review, no alpha spending is needed due to the DMC review.

13.0 Version History

Table 2. SAP Version History Summary

Version	Date	Summary
1.0	01 May 2018	Original version
2.0	04 June 2018	Eligibility criterion 10, requesting a minimum weekly average of daily worst pruritus of at least 4 on an NRS, has been added to the Study Protocol Amendment 1. Therefore, the SAP language describing the NRS related endpoints have been updated accordingly to match the protocol requirement.
3.0	14 February 2019	Two additional efficacy endpoints are added to Study Protocol Amendment 2. Therefore, SAP is updated accordingly.
4.0	17 September 2019	Analysis details have been added to the SAP. In addition, actigraphy related endpoints have been removed due to removal of the actigraphy assessment from the protocol.
5.0	14 October 2020	Updated the method of NRI-C as the primary approach for analyzing categorical endpoints with handling missing data due to COVID-19. Also added details of the NRI-NC approach as sensitivity analysis with no special data handling for missing due to COVID-19. Clarified that OC analysis will be used for summary of all efficacy variables in Period B. Updated the number of imputed datasets from PROC MI to 30. Also updated the protocol deviation categories for summary and the area of safety interest categories to be consistent with the current product safety SAP.

14.0 References

1. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.

Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took excluded concomitant medication.
- Subject received medication not permitted during study phase.

Appendix B. Definition of Area of Safety Interest

Area of safety interest (ASI) will be identified using the following search criteria:

Area of Safety Interest	Search Criteria		Include in AE Overview (Y/N)
MACE	Adjudicated terms will be identified as described in PSSAP using CECAT and CETERM from the CE SDTM dataset.		Y
Extended MACE	Adjudicated terms will be identified as described in PSSAP (for MACE +) using CECAT and CETERM from the CE SDTM dataset.		N
Adjudicated Anaphylactic Reaction	Adjudicated terms will be identified using SDTM data (e.g., CE and PR domains).		Y
Serious Infections	Serious AEs in the Infections and Infestations SOC		Y
Tuberculosis	Active Tuberculosis CMQ (code 80000188)		Y
Opportunistic Infections excluding tuberculosis and herpes zoster	Opportunistic infections excluding tuberculosis and herpes zoster CMQ (code 80000189)		Y
Herpes Zoster	Herpes Zoster CMQ (code 80000175)		N
Malignant Tumours	Narrow	Malignant tumours (SMQ 20000194)	Y
Non-melanoma Skin Cancer (NMSC)	Broad	Skin malignant tumours (SMQ 20000204) excluding terms identified by the Melanoma CMQ (code 80000119)	N
Malignant Tumours excluding NMSC	'Malignant Tumours excluding NMSC' is identified by the 'Malignant Tumours' search excluding terms identified by the 'Non-melanoma skin cancer (NMSC)' search.		Y
Hypersensitivity	Narrow	Hypersensitivity (SMQ 20000214)	Y – serious events only
Hepatic Events	Broad	Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ 20000013)	N
	Broad	Hepatitis, non-infectious (SMQ 20000010)	N
	Broad	Cholestasis and jaundice of hepatic origin (SMQ 20000009)	N
	Broad	Liver related investigations, signs and symptoms (SMQ 20000008)	N
	Narrow	Liver-related coagulation and bleeding disturbances (SMQ 20000015)	N

Appendix C. Potentially Clinically Significant Criteria for Safety Endpoints

The criteria for potentially clinically significant (PCS) laboratory findings are described in Table C-1 and Table C-2, and the PCS criteria for vital sign findings are described in Table C-3.

Table C-1. Criteria for Potentially Clinically Significant Hematology Values

		Definition of Potentially Clinically Significant CTCAE (Version 4) Grade 3 or Greater
Hematology Variables	**Units**	**Very Low**
Hemoglobin	g/dL	< 8.0
Platelets count	10^9/L	< 50.0
WBC count	10^9/L	< 2.0
Neutrophils	10^9/L	< 1.0
Lymphocytes	10^9/L	< 0.5

Note: A post-baseline value must be more extreme than the baseline value to be considered a potentially clinically significant finding.

Table C-2. Criteria for Potentially Clinically Significant Chemistry Values

		Definition of Potentially Clinically Significant NCI CTCAE (Version 4) Grade 3 or Greater	
Chemistry Variables	**Units**	**Very Low**	**Very High**
TBL	mcmol/L		> 3.0 × ULN
ALP	U/L		> 5.0 × ULN
SGOT/AST	U/L		> 5.0 × ULN
SGPT/ALT	U/L		> 5.0 × ULN
Albumin	g/L	< 20	
Glucose	mmol/L	< 2.2	> 13.9
Triglycerides	mmol/L		> 5.7
Creatinine	mcmol/L		> 3.0 × ULN
Sodium	mmol/L	< 130	> 155
Potassium	mmol/L	< 3.0	> 6.0
Calcium	mmol/L	< 1.75	> 3.1
CPK	U/L		> 5.0 × ULN
Total Cholesterol	mmol/L		> 10.34
GGT			> 5.0 × ULN

Note: A post-baseline value must be more extreme than the baseline value to be considered a potentially clinically significant finding.

Table C-3. Criteria for Potentially Clinically Significant Vital Sign Values

Vital Signs Variables	**Criterion**	**Definition of Potentially Clinically Important Vital Signs**
Systolic Blood Pressure (mmHg)	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic Blood Pressure (mmHg)	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline