JAHV stat analysis plan

A Randomized, Double-Blind, Placebo-Controlled, Withdrawal, Safety and Efficacy Study of Oral Baricitinib in Patients From 2 Years to Less Than 18 Years Old With Juvenile Idiopathic Arthritis (JIA)

NCT03773978

Approval Date: 14-Mar-2022

1. Statistical Analysis Plan: I4V-MC-JAHV: A Randomized, Double-Blind, Placebo-Controlled, Withdrawal, Safety and Efficacy Study of Oral Baricitinib in Patients from 2 Years to Less Than 18 Years Old with Juvenile Idiopathic Arthritis (JIA)

Baricitinib (LY3009104) Juvenile Idiopathic Arthritis

Study I4V-MC-JAHV (JAHV) is a multicenter, randomized, double-blind, placebo-controlled, medication-withdrawal study with a safety/pharmacokinetic (PK) assessment period, an open-label lead-in (OLLI) period, and a double-blind withdrawal (DBW) period in patients with JIA who have had an inadequate response or intolerance to treatment with at least 1 other conventional or biologic DMARD (bDMARD).

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I4V-MC-JAHV
Phase 3

Statistical Analysis Plan electronically signed and approved by Lilly on date provided below.

Approval Date: 14-Mar-2022 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

List of Figures

Figure **Page**

3. Revision History

Statistical Analysis Plan (SAP) Version 1 was approved prior to the first patient visit.

Statistical Analysis Plan (SAP) Version 2 was approved prior to the primary outcome database lock and includes the following changes:

- Updated the secondary and exporatory endpoints based on protocol I4V-MC-JAHV(d).
- Updated the study design figure footnote on enrollment by age group for clarity based on the current protocol I4V-MC-JAHV(d).
- Updated the analysis population for clarity regarding analyses in the open-label periods. Added OLLI population 2 that includes patients from the Safety/PK assessment period for efficacy/health outcomes analyses in the open-label periods. Added clarification on all participents in the open-label period including Safety/PK and OLLI population.
- Added clarification of baseline and postbaseline measures in Section 6.1.2.
- Clarified variables adjusted for Cox proportional hazards regression.
- Clarified time to disease flare calculation in Section 6.10.1.
- Updated subgroup analysis section by demographic and clinical characteristics subgroups.
- Safety analysis update:
    - Updated analysis strategy and corresponding period and population is defined in Section 6.1.1 and Table JAHV.6.10.
    - Updated details on SMQ in corresponding safety topic
    - Adding COVID-19 Trial Impact in Section 6.19
- Other minor typographical corrections and clarifications not affecting content.

4. Study Objectives

4.1. Primary Objective

Study objectives are listed in Table JAHV.4.1.

Estimands (International Conference on Harmonisation [ICH] E9 R1) of the study are defined based on the following 4 attributes:

- The population of interest is patients with juvenile idiopathic arthritis (JIA) who have had an inadequate response or intolerance to treatment with at least 1 other conventional or biologic disease-modifying antirheumatic drug (bDMARD). Analysis populations are defined in Section 6.1.1. The analysis population corresponding to each of the efficacy and health outcome endpoints is specified in Table JAHV.6.4, and Table JAHV.6.8.
- Primary, major secondary, and exploratory endpoints/variables are listed in Table JAHV.4.1, Table JAHV.4.2, and Table JAHV.4.3. A full list of efficacy and health outcome endpoints/variables are given in Table JAHV.6.4, Table JAHV.6.7, and Table JAHV.6.8.
- Population Level Summary: categorical variables will be summarized by proportion and continuous variables will be summarized by average. Details are given in Section 6.7.
- Intercurrent Event(s) Strategy: intercurrent events will be handled by the while-on-treatment strategy and the composite strategy approaches as primary strategies. The hypothetical strategy will also be used in supplementary analysis. Specific statistical methods to be used for handling intercurrent events under different strategies are described in Section 6.3. Statistical methods corresponding to each of the efficacy and health outcome variables are summarized in Table JAHV.6.4 and Table JAHV.6.8.

Table JAHV.4.1. Primary Objective and Endpoint

Objectives	Endpoint
Primary To evaluate the efficacy of baricitinib compared to placebo in children with JIA	Time to disease flare (flare defined as worsening of ≥30% in at least 3 of the 6 PedACR core criteria for JIA and an improvement of ≥30% in no more than 1 of the criteria) from the beginning of the double-blind withdrawal (DBW) period to the end of the DBW period

Abbreviations: JIA = juvenile idiopathic arthritis; PedACR = Pediatric American College of Rheumatology.

4.2. Secondary Objectives

Table JAHV.4.2. Secondary Objectives and Endpoints

Objectives	Endpoints
Secondary • To evaluate the efficacy of baricitinib in children with JIA	During the open-label lead-in (OLLI) period: • PedACR30/50/70/90/100 response rates • Changes in each of the 6 individual core set components variables of the PedACR Core Set as follows: o Number of active joints o Number of joints with limited range of motion o Physician's Global Assessment of Disease Activity o Parent's Global Assessment of Patient's Overall Well-Being o Physical function as measured by the Childhood Health Assessment Questionnaire (CHAQ) o Acute-phase reactant (high-sensitivity c-reactive protein [hsCRP]) and erythrocyte sedimentation rate (ESR) • Proportion of patients with inactive disease (as defined by Wallace et al. 2011) • Proportion of patients with minimal disease activity (as defined by Consolaro et al. 2012) • Change from baseline in the Physical Summary Score (PhS) and Psychosocial Summary Score (PsS) of the Child Health Questionnaire-Parent Form 50 (CHQ-PF50) • Changes from baseline in caregiver burden as measured by the Parental Impact-Time and Parental Impact-Emotion scales of the CHQ-PF50 • Change from baseline in Juvenile Arthritis Disease Activity Score-27(JADAS27) • Changes from baseline in arthritis-related pain as measured by the CHAQ pain severity Visual Analogue Scale (VAS) item

Secondary Objectives and Endpoints

Objectives	Endpoints
• To evaluate the efficacy of baricitinib compared to placebo in children with JIA	During the DBW period: • Proportion of patients with disease flare • PedACR30/50/70/90/100 response rates • Changes from baseline in each of the 6 individual components of the PedACR Core Set variables (due to disease flare or completion) as follows: o Number of active joints o Number of joints with limited range of motion o Physician's Global Assessment of Disease Activity o Parent's Global Assessment of Patient's Overall Well-Being o Physical function as measured by the CHAQ o Acute-phase reactant (hsCRP) and ESR • Proportion of patients with inactive disease (as defined by Wallace et al. 2011) • Proportion of patients with minimal disease activity (as defined by Consolaro et al. 2012) • Proportion of patients in remission (as defined by Wallace et al. 2011) • Change from baseline in JADAS-27 • Change from baseline in the PhS and PsS of the CHQ-PF50 • Change from baseline in caregiver burden as measured by the Parental Impact-Time and Parental Impact-Emotion scales of the CHQ-PF50Changes from baseline in arthritis-related pain as measured by the CHAQ pain severity VAS item
• To assess the efficacy of baricitinib in children with JPsA	• Change in Psoriatic Area and Severity Index (PASI) score during the OLLI period
• To assess the efficacy of baricitinib compared to placebo in children with JPsA	• Change from baseline in PASI score during the DBW period
• To assess the efficacy of baricitinib in children with ERA or JPsA	• Change in Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis index during the OLLI period • Change in Juvenile Spondyloarthritis Disease Activity Index (JSpADA) during the OLLI period
• To assess the efficacy of baricitinib compared to placebo in children with ERA or JPsA	• Change from baseline in SPARCC enthesitis index during the DBW period • Change from baseline in JSpADA during the DBW period

Secondary Objectives and Endpoints

Objectives	Endpoints
• To evaluate the potential effects of baricitinib on the cellular and humoral immune system	• Change in immunoglobulin levels and peripheral blood immunophenotyping (including T and B cells, T cell subsets, and NK cells) from baseline and at Week 4, Week 12, and Week 44 • Change of IgG titers from pre-vaccination to 4 weeks and 12 weeks post vaccination in patients eligible for vaccination with tetanus, diphtheria, and pertussis (TDaP) and/or pneumococcal conjugate vaccine according to local guidelines
• To characterize baricitinib PK in the JIA population and explore relationships between baricitinib exposure and study endpoints	• Population PK of baricitinib in patients with JIA • Proportions of patients achieving PedACR30/50/70/90/100 response rates by PK exposure • Time to disease flare in patients with JIA by PK exposure • Change in JADAS-27 by PK exposure
• To assess the patient acceptability and palatability of baricitinib tablets and oral suspension	• Assessment of tablet or oral suspension product acceptability and palatability during the OLLI period
• To assess the safety of baricitinib compared to placebo in patients with JIA	• Adverse events including serious adverse events • Permanent discontinuation of investigational product • Temporary interruption of investigation product

Abbreviations: CHQ-PF50 = Child Health Questionnaire-Parent Form 50; DBW = double-blind withdrawal; ERA = enthesitis-related juvenile idiopathic arthritis; hsCRP = high-sensitivity C-reactive protein; HRQoL = Health-related Quality of Life; IgA = immunoglobulin A; IgG = immunoglobulin G; JIA = juvenile idiopathic arthritis; JADAS = Juvenile Arthritis Disease Activity Score; JPsA = juvenile psoriatic arthritis; NK = natural killer; OLLI = open-label lead-in; PedACR = Pediatric American College of Rheumatology; PK = pharmacokinetic(s); VAS = Visual Analog Scale.

4.3. Exploratory Objectives

Table JAHV.4.3. Exploratory Objectives and Endpoints

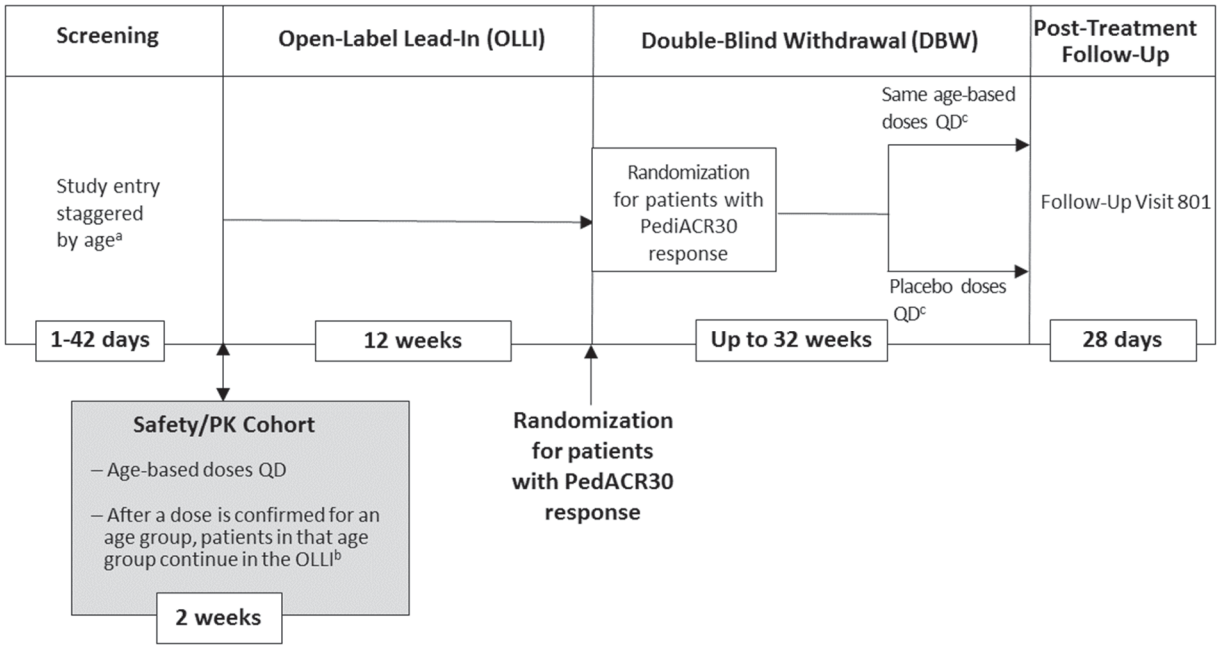
Objectives	Endpoints
Exploratory	
• To evaluate the quality of life (QOL) in children with JIA treated with baricitinib	• Change in European Quality of Life-5 Dimensions–Youth version (EQ-5D-Y) scores during the OLLI period
• To evaluate the QOL in children with JIA treated with baricitinib compared to placebo	• Change from baseline in EQ-5D-Y scores during the DBW period • Change from baseline in the individual scales (Global Health; Physical Functioning; Role/Social Limitations-Physical; Role/Social Limitations-Emotional/Behavioral; Bodily Pain/Discomfort; Behavior; Global Behavior Item; Mental Health; Self-Esteem; General Health Perception; Change in Health; Parental-Impact-Time; Parental Impact-Emotion; Family-Activities; Family-Cohesion) as measured by the CHQ-PF50
• To evaluate the efficacy of baricitinib compared to placebo in structural joint damage	• Change from the baseline (Week 0) to the end of the double-blind withdrawal (DBW) period of radiographic images of the hands/wrists using modified Total Sharp Score (mTSS [van der Heijde method]; van der Heijde 2000; Ravelli et al. 2007)
• Describe the skeletal age in pediatric patients	• Change from baseline (Week 0) to the end of the DBW in skeletal age based on hand/wrist radiograph and chronological age.

Abbreviations: DBW = double-blind withdrawal; JIA = juvenile idiopathic arthritis; OLLI = open-label lead-in.

5. Study Design

5.1. Summary of Study Design

Study I4V-MC-JAHV (JAHV) is a multicenter, randomized, double-blind, placebo-controlled, medication-withdrawal study with a safety/pharmacokinetic (PK) assessment period, an open-label lead-in (OLLI) period, and a double-blind withdrawal (DBW) period in patients with JIA who have had an inadequate response or intolerance to treatment with at least 1 other conventional or bDMARD. This includes patients with polyarticular JIA (rheumatoid factor positive or rheumatoid factor negative), extended oligoarticular course JIA, enthesitis-related JIA (ERA), and juvenile psoriatic arthritis (JPsA) as defined by the International League of Associations for Rheumatology (ILAR) criteria. The safety and tolerability data from this study are intended to establish an understanding of the benefit/risk relationship for baricitinib in patients with nonsystemic JIA.



Abbreviations: OLE = open-label extension; PedACR30 = Pediatric American College of Rheumatology 30 criteria; PK = pharmacokinetic(s); QD = once daily.

[a] Staggered approach to enrollment by age group (12 to <18 years, 9 to <12 years, 6 to <9 years, 2 to <6 years) will be implemented with older groups completing the safety/PK assessment period before younger groups are enrolled.

[b] Once the PK and safety profiles for an age group are confirmed, subsequent patients in that age group may enroll directly into the OLLI period. If the comparability assessment in the safety/PK period for an age group is inconsistent with baricitinib 4-mg exposures in adults with RA such that baricitinib dosage for the age group needs to be adjusted, the patients on the inconsistent dosage will discontinue the study and may enter the separate OLE study (JAHX).

[c] Patients who experience a disease flare during the DBW period will discontinue the study and may proceed directly to the separate OLE study (JAHX).

Figure JAHV.1. Study design for Clinical Protocol I4V-MC-JAHV.

5.2. Determination of Sample Size

A total of 128 patients will be randomized in a 1:1 ratio to baricitinib or placebo (64 per treatment arm) in the DBW period. This sample size will provide approximately 80% power to detect the difference in time to disease flare between the 2 treatment groups using a 2-sided test with a significance level of 0.05, assuming that the expected percentages of patients experiencing disease flare in the DBW period are 35% for baricitinib and 60% for placebo and that the dropout rate is no greater than 10% in this period.

It is estimated that 197 patients are required to enter the OLLI period to allow 128 patients to be randomized into the DBW period (assuming that 65% of the patients meet the Pediatric American College of Rheumatology (PedACR) 30 criteria at the end of the OLLI period). The nonresponder and dropout rate will be monitored during the OLLI period to adjust the overall sample size to ensure that a minimum of 128 patients will be randomized in the DBW. If the

PedACR30 response rate during the OLLI period is higher than the assumed rate of 65%, fewer than 197 patients may be required.

The above sample size and power estimates are based on nQuery®+nTerim 4.0.

5.3. Method of Assignment to Treatment

Patients who enter the PK lead-in period will be assigned a dose based on age. Doses of subsequent patients who enter the PK lead-in period may be adjusted until a dose is found that produces exposures within the range produced by baricitinib 4-mg in adults with RA. All patients who enter the OLLI period will receive a fixed age-based dose of baricitinib as determined in the PK lead-in period.

Patients who meet the PedACR30 response criteria at the end of the OLLI period will be randomized in a 1:1 ratio (baricitinib age-based dose or matching placebo) to double-blind treatment at Week 12. The randomization will be stratified by:

- history of prior bDMARD use (Yes versus No)
- the JIA categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined])
- Predose exposure erythrocyte sedimentation rate (ESR) categories (elevated [>20 mm/hour] and not elevated) in the polyarticular JIA patients

Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS). The IWRS will be used to assign bottles containing double-blind investigational product to each patient. All patients <6 years of age will receive oral suspension. Patients ≥ 6 to <12 years old have the option of receiving the oral suspension. Patients >12 years will be supplied tablets only. Site personnel will enter the confirmation number found on the bottles into the IWRS, which confirms that they have located the correct bottles.

6. A Priori Statistical Methods

6.1. General Considerations

This plan describes *a priori* statistical analyses for efficacy, health outcomes, and safety that will be performed.

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly). The statistical analyses will be performed using SAS® Version 9.4 or higher.

Not all displays described in this SAP will necessarily be included in the clinical study report (CSR). 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 described in this SAP and not included in the CSR would be available upon request.

Statistical tests of treatment effects and confidence intervals (CIs) will be performed at a 2-sided significance level of 0.05, unless otherwise stated.

Data collected at early termination visits will be mapped to the next scheduled visit number for that patient. For by-visit summaries, only visits in which a measure was scheduled to be collected will be summarized. Any unscheduled visit data will be included in the patient-level listings. However, the data might still be used in other analyses, including change from baseline to endpoint using last observation carried forward (LOCF) analyses and other categorical analyses.

6.1.1. Analysis Populations

For purposes of analysis, the following populations are defined based on the different treatment period as shown in Table JAHV.6.1.

Table JAHV.6.1. Analysis Populations

	Population	Description
	Entered population	All participants who sign informed consent.
Efficacy Analysis population	Safety/PK population	All patients who received at least 1 dose of investigational product in the Safety/PK assessment period.
	OLLI population	All participants who take at least 1 age-based final dose, as confirmed by PK assessments of investigational product, in the OLLI period, other than the Safety/PK population.
	OLLI population 2	All patients who received at least 1 dose of investigational product in the OLLI period.
	Safety/PK and OLLI population	All enrolled patients who were initially assigned to the open-label investigational product in Safety/PK assessment period and OLLI period, following intent-to-treat (ITT) principles.
	DBW population	All randomized patients in the DBW period following intent-to-treat (ITT) principles.

	Population	Description
Safety Analysis population	DBW safety population	All randomized patients in the DBW period who receive at least 1 dose of investigational product.
	General safety population	All patients who received at least 1 dose of investigational product, which is baricitinib.

Abbreviations: DBW = doubleblind withdrawal; OLLI = open-label lead-in; PK = pharmacokinetic.

Efficacy analyses will be conducted by treatment periods with the corresponding analysis population, where the primary efficacy analysis will focus on the DBW period with the DBW population. Safety analyses will be performed by treatment periods with the corresponding safety population.

The patients who participate in the Safety/PK period and participate in the OLLI period with the age-based final dose will be eligible to participate in the DBW period if the patient is a PedACR30 responder at the end of the OLLI period. Thus, such patients from the Safety/PK assessment period will be included in the OLLI population 2 and DBW population. Also, Safety/PK assessment period and OLLI period may be combined for efficacy/safety analysis with the corresponding population that includes such patients.

In the rare situation where a patient is lost-to-follow-up at the first postbaseline visit but some safety data exist (e.g., unscheduled laboratory assessments) after first dose of study drug, a listing of the data or a patient profile will be provided.

6.1.2. Definition of Baseline and Postbaseline Measures

The baseline value for the efficacy (except flare), health outcomes is defined as the last non-missing measurement on or prior to the date of first study baricitinib administration. For patients who participate in the Safety/PK period, this value will be recorded on or before Visit 2 (Day 1); for the other patients who started with the OLLI period, this value will be recorded on or before Visit 5 (Week 0).

The baseline value for disease flare analyses is defined as the last non-missing measurement on or prior to the date of randomization (beginning of the DBW period). This value will be recorded on or before Visit 9 (Week 12).

Baseline for the safety analyses is defined in Section 6.18. Postbaseline measurements are collected after study drug administration through Visit 17 (Week 44) or early discontinuation visit. For data collected in the electronic Clinical Outcomes Assessment (eCOA) tablet (including Patient-Reported Outcomes [PRO] and Clinician-Reported Outcomes [ClinRO]) and related to efficacy assessments, unscheduled postbaseline visits that fall within the visit windows defined by Lilly will be summarized in the by-visit analyses if there is no scheduled visit available. Refer to clinical protocol I4V-MC-JAHV(d) for detail of the visit windows. If there is more than 1 unscheduled visit within the defined visit window and no scheduled visit is available, the unscheduled visit closest to the scheduled visit date will be used. If 2 unscheduled visits of equal distance are available, then the latter of the 2 will be used.

Postbaseline measures for the safety analyses are defined as the nonmissing scheduled (planned) measurements after the date of first study drug administration for continuous measures by-visit analyses and all nonmissing measurements after the date of first study drug administration for all other analyses.

6.1.3. Definition of OLLI and DBW Analysis Periods

The OLLI period starts at:

- The completion of Safety/PK assessment period for Safety/PK population patients
- The time of first age-based dose of baricitinib for OLLI population patients

The OLLI period ends for the following study events:

- Week 12 (Visit 9) for patients who do not meet the randomization criteria, or
- Visit of early discontinuation of study treatment in the OLLI period.

The DBW period starts at the time of randomization and ends at the following study events:

- Patient experience disease flare (defined in Table JAHV.6.4)
- Week 44 (Visit 17) for patients who do not flare, or
- Visit of early discontinuation of study treatment in the randomized double-blind withdrawal period.

6.1.4. Analysis Methods

The primary endpoint will be the time-to-flare during the DBW period for randomized patients. Patients who discontinue or complete the DBW period without experiencing a flare will have their data censored at the time of their discontinuation date or completion date. The censored patients will contribute to the at-risk set in the stratified logrank test. Survival curves will be estimated using the Kaplan–Meier method in the DBW period. Treatment comparisons will be performed using a stratified log-rank test across JIA categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined]) for the DBW population in the DBW period. Treatment comparisons may also be analyzed using a Cox proportional hazards model adjusted for stratification variables. The hazard ratio with 95% CIs will be reported. And the similarity of treatment effect across JIA categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined]) will be evaluated using Cox proportional hazards regression adjusted for stratification variables. If any of these variables are redundant for a particular model, they will be dropped.

Secondary efficacy and health outcomes analyses (Table JAHV.6.4 and Table JAHV.6.5) will be based on treatment period and respective study populations.

Efficacy, health outcomes, and safety data collected in OLLI period will be summarized without inferential statistics.

In the DBW period, the main analysis of categorical efficacy variables and health outcomes variables will use a logistic regression analysis with treatment, JIA categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined]), predose exposure ESR categories (elevated [>20 mm/hour] and not elevated), history of prior bDMARD use (Yes or No) as covariates in the model. The p-value and 95% CI for the odds ratio from the logistic regression model are used for statistical inference.

In the DBW period, the main analysis for all continuous efficacy and health outcomes variables will use analysis of covariance (ANCOVA) model with treatment, JIA patient categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined]), predose exposure ESR categories(elevated [>20 mm/hour] and not elevated), history of prior bDMARD use (Yes or No), and baseline score as covariates in the model. For ESR and high-sensitivity C-reactive protein (hsCRP) analysis, predose ESR level categories will not be included as a covariate. Type III sums of squares for the least-squares mean (LSM) will be used for the statistical comparison of treatment groups, and the LSM difference, standard error, p-value and 95% CI will also be reported. The LOCF approach will be used to impute missing data.

The mixed model repeated measures (MMRM) analysis will also be considered as a supplemental analysis method. The MMRM will use a restricted maximum likelihood (REML) estimation. The model will include treatment, JIA patient categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined]), predose exposure ESR categories (elevated [>20 mm/hour] and not elevated), history of prior bDMARD use (Yes or No), visit, treatment-by-visit-interaction, baseline score and baseline-by-visit-interaction as fixed effects. If any of these are redundant for a particular model, they will be dropped. An unstructured (co)variance structure will be used to model the between- and within-patient errors. If this analysis fails to converge, the heterogeneous autoregressive [ARH(1)], followed by the heterogeneous compound symmetry (CSH), followed by the heterogeneous Toeplitz (TOEPH) will be used. The Kenward-Roger method will be used to estimate the degrees of freedom. Type III tests for the LSM will be used for the statistical comparison. The LSM difference, standard error, p-value, and 95% CI will also be reported.

The primary safety analysis will be conducted during the DBW period for the DBW safety population. Fisher's exact test will be used for the adverse events (AEs), discontinuation, and other categorical safety data for between-treatment group comparisons. Continuous vital signs, body weight, and other continuous safety variables, including laboratory variables will be analyzed by an ANCOVA with treatment group and baseline value in the model. The significance of within-treatment group changes from baseline will be evaluated by testing whether or not the treatment group LSM changes from baseline are different from zero; the standard error and within-group p-value for the LSM change will also be displayed. In addition, the LSM difference between baricitinib and placebo groups, the corresponding p-value, and 95% CI will be provided. Treatment-emergent high/low for categorical laboratory safety analyses will also be produced.

6.2. Adjustments for Covariates

The randomization to treatment groups at Week 12 (Visit 9) is stratified by:

- history of prior bDMARD use (Yes versus No)
- the JIA categories
    - polyarticular and extended oligoarticular [combined] or
    - enthesitis-related JIA and JPsA [combined]
- predose exposure ESR categories (elevated [>20 mm/hour] and not elevated) in the polyarticular JIA patients

Unless otherwise specified, the statistical analysis models will adjust for JIA patient category (polyarticular and extended oligoarticular versus ERA and JPsA), history of prior bDMARD use (Yes or No), and predose exposure ESR categories (above 20 mm/L or below 20 mm/L). If any of these are redundant for a particular model, they will be dropped. The covariates used in the ANCOVA model for continuous data generally will include the parameter value at baseline. When an MMRM analysis is performed, baseline value and baseline-by-visit interactions will be included as covariates.

6.3. Handling of Dropouts or Missing Data

Intercurrent events (ICH E9R1) are events which occur after the treatment initiation and make it impossible to measure a variable or influence how it should be interpreted. Examples of such events include treatment discontinuation due to death or AEs, and loss to follow-up. In this study, intercurrent events will be handled by the while-on-treatment strategy and the composite strategy approaches as primary strategies. The hypothetical strategy will also be used in supplementary analysis.

The *censoring rule*: The efficacy and health outcome data collected after permanent study drug discontinuation will be excluded from the analyses. This censoring rule will be applied to all continuous and categorical efficacy and health outcome endpoints.

For binary response variables that are derived from component scores, the following steps will be implemented unless otherwise specified: If all component scores are missing at a visit, the response status will be set to nonresponse; if at least 1 component score is nonmissing at a visit, the LOCF approach will be applied to impute the missing components within a treatment period. If the binary response status still cannot be determined after LOCF, the binary response status will be set to nonresponse. The detail of the missing data definition and imputation rule is documented in Table JAHV.6.4 and Table JAHV.6.5.

6.3.1. Last Observation Carried Forward (LOCF)

According to ICH E9 R1, the while-on-treatment strategy could be applied based on the last postbaseline value at or before the visit of interest while the patient was still on study drug. The primary endpoint for efficacy evaluation, time to disease flare, is an endpoint that reflects loss of

treatment effect. Therefore, for patients who completed DBW via flare, the LOCF values appropriately estimate their disease status at this time and beyond.

All continuous endpoints will be imputed using the LOCF methodology after applying the censoring rule to those who discontinued study drug. For patients who permanently discontinue study treatment or discontinue from the study for any reason at any time, the last nonmissing postbaseline observation on or prior to discontinuation will be carried forward to subsequent time points for evaluation. LOCF will not be applied across OLLI and DBW analysis periods.

6.3.2. Non-Responder Imputation (NRI)

A non-responder imputation (NRI) imputation method can be justified based on the composite strategy (ICH E9R1) for handling intercurrent events. In this strategy a patient is defined as a responder only if (i) they meet the clinical requirements for response at the predefined time and (ii) they remain on the assigned study treatment. Failing either criteria by definition makes them a non-responder.

Binary efficacy and health outcome variables will be imputed using NRI after applying the censoring rule to those who discontinued study drug. Patients will be considered a non-responder for the NRI analysis if they do not meet the clinical response criteria or are entirely missing the visit at the analysis time point. Randomized patients without at least 1 postbaseline observation will also be defined as non-responders for the NRI analysis.

6.3.3. Mixed Model for Repeated Measures (MMRM)

An MMRM method can be justified based on the hypothetical strategy (ICH E9 R1) for handling intercurrent events. In this strategy the effect of study treatment is assessed in a hypothetical trial where all patients have complete data and continue to take study treatment without dropping out of the study. The MMRM method assumes that missing data can bias results, but the bias can be attenuated by modeling random effects using the within-subject error correlation structure. These correlations between the repeated measurements provide the platform used to account for the bias from subject dropout by assuming the dropout follows missing at random (MAR) assumption.

Selected efficacy and health outcomes variables will be assessed using MMRM as a supplementary analysis. The details are described in Table JAHV.6.5.

6.4. Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. The countries will be categorized into geographic regions, as described in Table JAHV.6.2.

6.5. Multiple Comparisons/Multiplicity

No multiplicity control measures will be used.

6.6. Patient Disposition

An overview of patient disposition will be summarized by treatment periods with the corresponding safety population. Frequency counts and percentages of patients excluded by

primary reason for exclusion will be provided for patients who failed to meet study entry requirements during screening.

Frequency counts and percentages of patients who complete the study treatment or discontinue early from the study, along with their discontinuation reasons and whether they completed follow-up or enrolled into the extension, will be summarized. A by-treatment summary will provide for the DBW period with the DBW safety population.

A listing of patient disposition will be provided for all enrolled patients, with the extent of their participation in the study and the reason for discontinuation. A listing of all randomized patients with their treatment assignment will also be provided.

6.7. Patient Characteristics

Patient characteristics including demographics, baseline characteristics, historical illnesses, and preexisting conditions will be summarized descriptively at the study baseline by treatment periods with the corresponding analysis population. They will also be summarized descriptively by treatment group for the DBW period with the DBW population. No formal statistical comparisons will be made among treatment groups unless otherwise stated.

Table JAHV.6.2 describes the specific variables and how they may be summarized.

Table JAHV.6.2. Patient Characteristics

Variable	Continuous measure Summary	Categorical Summary
Age [a]	Yes	12 to <18 years, 9 to <12 years, 6 to <9 years, and 2 to <6 years
Sex	No	Male, Female
Race	No	American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian or other Pacific Islander, White, or Multiple
Geographic region	No	By Country
	No	Asia, South America, Europe and rest of world
Height (cm)	Yes	None
Weight (kg)	Yes	None
BMI[b]	Yes	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)
Tobacco use	No	Never, Current, Former
Prior biologic JIA therapy[c]	No	Never used, Ever used
Number of prior biologics JIA[c]	No	0, 1, 2, >2
Prior non-biologic JIA therapy[d]	No	Never used, Ever used
Number of non-biologic JIA therapies[d]	No	0, 1, 2, >2
MTX usage at baseline	No	Yes, No
Duration of JIA diagnosis (years)[e]	Yes	0 to <2 years, 2 to <5 years, 5 to <10 years, ≥10 years
Age at JIA onset (years)[e]	Yes	12 to <18 years, 6 to <12 years, and 2 to <6 years
Baseline number of active joints	Yes	None
Baseline number of joints with limited range of motion	Yes	None
Baseline Physician's Global Assessment of Disease Activity	Yes	None
Baseline Parent's Global Assessment of Patient's Overall Well-Being	Yes	None
Baseline Physical function measured by CHAQ	Yes	None
Baseline ESR	Yes	Elevated (above 20 mm/L) or Not Elevated (not above 20 mm/L)
Baseline hsCRP	Yes	above 3 mg/L or not above 3 mg/L
Baseline Juvenile Arthritis Disease Activity Score (JADAS)-27	Yes	None
Baseline Physical Summary Score (PhS) and Psychosocial Summary Score (PsS) of the Child Health Questionnaire-Parent Form 50 (CHQ-PF50)	Yes	None
Baseline caregiver burden as measured by Parental Impact-Time and Parental Impact-	Yes	None

Variable	Continuous measure Summary	Categorical Summary
Emotion scales of the CHQ-PF50		
Baseline Arthritis-related pain measured by CHAQ pain severity	Yes	None
Variable	**Continuous measure Summary**	**Categorical Summary**
Baseline SPARCC [f, g]	Yes	None
Baseline JSpADA [f, g]	Yes	None
Baseline PASI score [f]	Yes	None
Latest screening period renal function status	No	impaired (eGFR <60 mL/min/1.73 m^2) or not impaired (eGFR ≥60 mL/min/1.73 m^2)
Immunoglobulin level (Including T and B cells, T cell subsets and NK cells)	Yes	None
Peripheral blood immunophenotyping (Including T and B cells, T cell subsets and NK cells)	Yes	None
pre-vaccination IgG titer[h]	Yes	None

Abbreviations: BMI = body mass index; CHAQ = Childhood Health Assessment Questionnaire; CHQ-PF50 = Child Health Questionnaire-Parent Form 50; ESR = erythrocyte sedimentation rate; IgG = immunoglobulin G; JIA = juvenile idiopathic arthritis; JSpADA = Juvenile Spondyloarthritis Disease Activity Index; NK = natural killer; PASI = Psoriasis Area and Severity Index; SPARCC = Spondyloarthritis Research Consortium of Canada.

a Age in years will be calculated as length of the time interval from the imputed date of birth (July 1st in the year of birth collected in the electronic case report form(eCRF)) to the informed consent date.

b Body Mass Index (BMI) will be calculated as: $BMI\ (kg\ /\ m^2) = Weight\ (kg)\ /(Height\ (m))^2$.

c Biologic JIA therapies for example: anakinra, canankinumab, tocilizumab, sarilumab, rituximab, etanercept, adalimumab, abatacept, tocilizumab, and golimumab

d Non-biologics JIA therapies for example: methotrexate,celecoxib, diclofenac, ibuprofen and corticosteroids.

e Age at diagnosis in years will be calculated as the time interval from the imputed date of birth (July 1st in the year of birth collected in the eCRF) to the date of JIA diagnosis.

f For JPsA patients who have this score available.

g For ERA patients who have this score available.

h In patients eligible for vaccination with tetanus, diphtheria, and pertussis (TDaP) and/or pneumococcal conjugate vaccine according to local guidelines.

6.7.1. Historical Illness and Pre-existing Conditions

Historical illnesses are defined as those conditions recorded in the Preexisting Conditions and Medical History electronic case report form (eCRF) with an end date prior to the informed consent date. The number and percentage of patients with selected historical diagnoses will be summarized for the OLLI population, also they will be summarized by treatment group using the DBW population. Historical diagnoses will be categorized using the Medical Dictionary for

Regulatory Activities (MedDRA®, most current available version) algorithmic Standardized MedDRA Queries (SMQs) or similar predefined lists of Preferred Terms (PTs) of interest.

Preexisting conditions are defined as those conditions recorded in the Preexisting Conditions and Medical History eCRF or the Adverse Events eCRF with a start date prior to the date of informed consent and an end date after informed consent or missing. For events occurring on the day of the first dose of study treatment, the date and time of the onset of the event will both be used to determine if the event was preexisting. Conditions with a partial or missing start date (or time if needed) will be assumed to be 'not preexisting' unless there is evidence, through comparison of partial dates, to suggest otherwise. Preexisting conditions will be categorized using the MedDRA SMQs or similar predefined lists of PTs of interest. Frequency counts and percentages of patients with selected preexisting conditions will be summarized for the OLLI population, and also will be summarized by treatment group using the DBW population.

6.8. Treatment Compliance

Treatment compliance with investigational product will be summarized by treatment periods with corresponding population. Patients will be considered compliant for each treatment period if they miss <20% of the expected doses. Proportions of patients compliant will be summarized.

Similarly, patients will be considered noncompliant if they are judged by the investigator to have intentionally or repeatedly taken more than the prescribed amount of study medication.

Compliance in the period of interest up to Visit *x* will be calculated as follows:

$$\text{Compliance} = \frac{\text{total number of tablets (or weight of suspension) dispensed} - \text{total number of tablets (or weight of suspension)returned}}{\text{expected number of total tablets (or weight of suspension)}}$$

where

- Total number of tablets (or weight of suspension) dispensed: sum of tablets (or weight of suspension) dispensed in the period of interest prior to Visit *x*;
- Total number of tablets (or weight of suspension) returned: sum of the tablets (or weight of suspension) returned in the period of interest prior to and including Visit *x*;
- Expected number of tablets (or weight of suspension): number of days in the period of interest*number of tablets (or weight of suspension) taken per day = [(date of visit – date of first dose + 1) – number of days of temporary drug interruption]*number of tablets (or weight of suspension) taken per day

Descriptive statistics for percent compliance and noncompliance rates will be summarized for the OLLI population for Week 0 through Week 12 and will also be summarized for the DBW population by treatment group for Week 12 through Week 44. Subintervals of interest, such as compliance between visits, may also be presented. The number of expected tablets or suspension dispensed, tablets or suspension returned, and percent compliance may be listed by patient for Week 0 through Week 44.

6.9. Concomitant Therapy

Summaries of previous and concomitant medications will be provided by analysis population. .

At screening, previous and current JIA treatments are recorded for each patient. A summary of previous medications used for JIA, including vaccination with tetanus, diphtheria, and pertussis (TDaP) and/or pneumococcal conjugate and medications that are discontinued after screening and before the first dose of study drug, will be prepared using frequency counts and percentages by preferred medication name, with preferred medication names sorted by frequency. Concomitant therapy will be recorded at each visit and will be classified similarly. An additional summary for previous medications used for JIA will be created containing the reason of discontinuation.

Concomitant therapy for the treatment period is defined as therapy that starts before or during the treatment period and ends during the treatment period or is ongoing (has no end date or ends after the treatment period). Should there be insufficient data to make this comparison (for example, the concomitant therapy stop year is the same as the treatment start year, but the concomitant therapy stop month and day are missing), the medication will be considered as concomitant for the treatment period.

Summaries of previous medications will be provided for the following categories:

- previous JIA therapies
- previous JIA therapies including reason for discontinuation

Summaries of concomitant medications will be provided for the following categories:

- concomitant medications for JIA
- concomitant medications for non-JIA

Table JAHV.6.3. Concomitant JIA Therapies

Drug Class	As Needed	Chronic Use	Conditions for Use
MTX[a]	No	Yes	If on MTX, must be on a stable average dose of ≤20 mg/m^2/week for the 8 weeks preceding screening and must continue at that dose throughout the study
cDMARDs other than MTX[a]	No	Yes	If receiving cDMARDs (other than MTX), must be on a stable dose for at least 4 weeks prior to the screening and must continue at that dose throughout the study.
Oral corticosteroids	No	Yes	If receiving oral corticosteroids, daily doses of ≤10 mg/day or 0.2 mg/kg/day prednisone equivalent, whichever is less. Must be on stable dose for at least 2 weeks prior to screening and 6 weeks prior to baseline; the dose must be continued throughout the study.
NSAIDs[b] • including cyclooxygenase-2 inhibitors, e.g., celecoxib	No	Yes	• Must be on stable dose at least 1 week prior to baseline. • Changes in dose, discontinuation and/or introduction of new NSAIDs are only allowed for treatment of an AE.
Analgesics • including local anaesthetics, e.g., lidocaine, and • topical anaesthetics, e.g., EMLA cream.	No	Yes	• Must be on stable dose at least 1 week prior to baseline. • Changes in dose, discontinuation and/or introduction of new analgesics are only allowed for treatment of an AE.

Abbreviations: cDMARD = conventional disease-modifying antirheumatic drug; NSAID = nonsteroidal anti-inflammatory drug; MTX = methotrexate.

[a] Concomitant use of >2 of any cDMARDs (including MTX) is not allowed.

[b] For use as an anti-inflammatory agent.

6.10. Efficacy Analyses

6.10.1. Primary Outcome and Methodology

Time to disease flare (time is measured by week) is used as the primary outcome, which will be calculated as follows:

$$(Date\ of\ Flare - Date\ of\ Randomization\ in\ the\ DBW + 1)/7$$

The flare-ESR is calculated based on the flare definition and PedACR core criteria defined in Section 4.1, Table JAHV.4.1, Table JAHV.4.2, and Table JAHV.6.6, and uses ESR as acute-phase reactant. The detail of disease flare calculation is documented in separate disease flare and PedACR response calculation algorithm document. Patients completing the DBW period without flare will be censored at the date of completion (ie, the date of the last visit in the period).

Patients who discontinued treatment early in the DBW period without flare will be censored at the treatment discontinuation date.

The primary analysis of the study is to test for the hypotheses that baricitinib is superior to placebo in prolonging the time to disease flare in PedACR 30 responders from Week 12 to Week 44 using the DBW population.

A stratified logrank test will be used as the primary analysis method, in which the log-rank test will be stratified on the JIA categories:

- polyarticular and extended oligoarticular[combined]
- enthesitis-related JIA and JPsA[combined]

The p-value and median time to flare (if applicable) by treatment group will be reported. Treatment comparisons may also be analyzed using a Cox proportional hazards model adjusted for stratification variables. The hazard ratio with 95% CIs will be reported. The Kaplan-Meier product limit method will be used to estimate the survival curves for time-to flare. Time-to flare will also be summarized graphically by treatment group using Kaplan-Meier techniques.

6.10.2. Secondary Efficacy Analyses

Secondary efficacy and health outcomes analyses (Table JAHV.6.4) will be based on treatment periods with the respective population at each visit. The analysis method is described in Table JAHV.6.5. There will be no adjustment for multiple comparisons.

Table JAHV.6.4. Description and Derivation of Primary, Secondary Efficacy Endpoints and Health Outcomes

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
Disease flare	Disease flare is defined as a worsening of ≥30% in at least 3 of the 6 PedACR core criteria for JIA and an improvement of ≥30% in no more than 1 of the criteria from the patient's condition at randomization baseline. In disease flare-ESR, ESR measure is used as acute-phase reactant in the core criteria. • If either the number of joints with active arthritis or the number of joints with limitation of motion are used in the calculation of flare for a study visit, then a minimum worsening of at least 2 active joints or 2 joints with limitation of motion must be present. • An active joint is defined as a joint with swelling or, in the absence of swelling, limitation of motion accompanied by pain on motion and/or tenderness. • If either the Physician's Global Assessment of Disease Activity or the Parent's Global Assessment of Well-Being are used in the calculation of flare for a study visit, then a minimum worsening of at least 20 mm in 0-100 mm visual analogue scale (VAS) must be present. • If ESR is used in the definition of "flare" and counts towards worsening, then the second value for ESR used in the calculation must be above the upper limit of normal for ESR (>20 mm/hour).	Time to disease flare	The details of disease flare calculation are documented in separate disease flare and PedACR response calculation algorithm document. Calculation: Date of disease flare-ESR minus date of randomization+1	Component is missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values. If more than 3 components are missing, there will not be sufficient information; thus, the patient will be considered as a non-flare patient. The details of handling of missing data are documented in separate disease flare and PedACR response calculation algorithm document.
		Proportion of patients with disease flare	The detail of disease flare calculation is documented in separate disease flare and PedACR response calculation algorithm document.	

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
PedACR response	Pediatric American College of Rheumatology (PedACR x%) responder is defined as patients who demonstrate disease response improvement of ≥x% in at least 3 of 6 Pediatric American College of Rheumatology (PedACR) core response variables and >30% worsening in not more than 1 of the remaining variables. (PedACR30; Giannini et al. 1997). When PedACR response is analyzed as secondary endpoint, ESR measure is used as acute-phase reactant in the core criteria.	PedACR 30/50/70/90/100 response rates (compare to patient's condition prior to first dose of investigational product)	The detail of disease flare calculation is documented in separate disease flare and PedACR response calculation algorithm document.	Component is missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values. If more than 3 components are missing, there will not be sufficient information; thus, the patient will be considered as a non-responder. The detail of handling of missing data is documented in separate disease flare and PedACR response calculation algorithm document.

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
Active joint count	Number of active joints (defined as a joint that is swollen or in the absence of swelling has loss of passive motion accompanied by either pain on motion or joint tenderness) in 73 joints.	Change from baseline of active joint count	Calculation: Change from baseline: observed active joint count at visit minus baseline active joint count	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.
Limited range of motion joint count	Number of joints with limited range of motion in 69 joints	Change from baseline of limited of motion joint count	Calculation: Change from baseline: observed limited range of motion joint count at visit minus baseline limited range of motion joint count	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.
Physician's global assessment of disease activity (VAS)	The Physician's Global Assessment of Disease Activity is used to assess the patient's current disease activity, as it relates to their signs and symptoms. The instrument uses a 21 circle VAS ranging from 0 to 10 (using 0.5 increments) where 0 = "no activity" and 10 = "maximum activity" (Filocamo et al. 2010).	Change from baseline of physician's global assessment of disease activity	Calculation: Change from baseline: observed score at visit minus baseline score	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.
Parent's global assessment of overall well-being measured by CHAQ (10cm VAS)	The parent is generally asked to make a global assessment of the child's overall well-being on a 10 cm VAS, with anchors of '0 = very good' and '10 = very poor', which is located at the bottom of the Childhood Health Assessment Questionnaire (CHAQ) questionnaire	Change from baseline of parent's global assessment of overall well-being	Calculation: Change from baseline: observed score at visit minus baseline score	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
Physical function as measured by CHAQ (0-3)	The Childhood Health Assessment Questionnaire (CHAQ) assesses health status and physical function over the past week in children with juvenile arthritis over the past week, which the parent or legal guardian completes, regardless of the age of the patient. Physical function is measured by CHAQ from disability index. The Disability Index contains 30 items grouped into the following 8 domains (not including assistive devices/aids questions): physical function, dressing, and grooming (4 items), arising (2 items), eating (3 items), walking (2 items), hygiene (5 items), reach (4 items), grip (5 items), and activities (5 items). Each item is scored from 0 to 3 (0 = no difficulty; 1 = some difficulty; 2 = much difficulty; 3 = unable to do or not applicable) (Singh et al. 1994). The total score is the average of domain score, which is range from 0 to 3. At least 6 domains are required in calculation.	Change from baseline of physical function	The details of physical function calculations are documented in separate CHAQ instruction document. Calculation: Change from baseline: observed score at visit minus baseline score	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.
hsCRP (mg/L)	The hsCRP test is a highly sensitive quantification of C-Reactive protein (CRP), an acute-phase protein that increases during inflammation.	Change from baseline of hsCRP	Calculation: Change from baseline: observed lab value at visit minus baseline lab value	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
ESR (mm/hr)	An erythrocyte sedimentation rate (ESR) is a type of blood test that measures how quickly erythrocytes (red blood cells) settle at the bottom of a test tube that contains a blood sample. Normally, red blood cells settle relatively slowly. A faster-than-normal rate may indicate inflammation in the body.	Change from baseline of ESR	Calculation: Change from baseline: observed lab value at visit minus baseline lab value	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.
Physical Summary Score (PhS) and Psychosocial Summary Score (PsS)	Physical Summary Score (PhS) and Psychosocial Summary Score (PsS) from Child Health Questionnaire-Parent Form 50 (CHQ-PF50) will be used to address physical and psychosocial health status. The CHQ-PF50 is a generic QOL instrument designed to capture the physical, emotional, and social components of health status of children as well as caregiver burden over the past 4 weeks (HealthActCHQ 2013). The CHQ-PF50 is completed by the parent and has been validated for use in JIA patients (Ruperto et al. 2001).	Change from baseline in the Physical Summary Score (PhS) and Psychosocial Summary Score (PsS) of the Child Health Questionnaire-Parent Form 50 (CHQ-PF50)	The Physical Summary Score (PhS) and Psychosocial Summary Score (PsS) will be scored according to the manual, “CHQ Scoring and Interpretation manual”. (HealthActCHQ 2013). Calculation: Change from baseline: observed score at visit minus baseline score	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.
Caregiver burden	Caregiver burden is measured by Parental Impact-Time and Parental Impact-Emotion scales from CHQ-PF50 will be used to address burden to the parent/legal guardian.	Change from baseline of caregiver burden score measured by the Parental Impact-Time and Parental Impact-Emotion scales of the CHQ-PF50	The caregiver burden will be scored according to the manual, “CHQ Scoring and Interpretation manual”. (HealthActCHQ 2013). Calculation: Change from baseline: observed score at visit minus baseline score	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values.

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
Inactive disease	Inactive disease is indicated by the presence of all of the following (Wallace et al. 2011): • No joints with active arthritis based on JADAS-27 • No fever, rash, serositis, splenomegaly, hepatomegaly, or generalized lymphadenopathy attributable to JIA as assessed by the investigator • No active uveitis as assessed by the investigator • Normal erythrocyte (ESR) or hsCRP (ie, within normal limits in the local laboratory or, if elevated, not attributable to JIA) • Physician's Global Assessment of Disease Activity indicating no active disease (best possible score on scale [0]) • Duration of morning stiffness less than 15 minutes • In inactive disease analysis as secondary endpoint, ESR measure is used as acute-phase reactant measure.	Proportion of patients who achieved inactive disease	• Step 1: If joint with active arthritis isn't 0, stop; else • Step 2: If there is fever, rash, serositis, splenomegaly, hepatomegaly, or generalized lymphadenopathy attributable, stop; else • Step 3: If uveitis, stop; else • Step 4: If ESR>20, stop; else • Step 5: If physician's global assessment of disease activity isn't 0, stop; else • Step 6: If stiffness >15min, stop; else • Step 7: Conclude the patient achieved inactive disease status.	Missing if any component is missing. NRI will be applied to missing data.

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
Minimal disease activity	Minimal disease activity is calculated based on the scores from the Physician's Global Assessment of Disease Activity, Parent's Global Assessment of Patient's overall Well-Being, and the number of swollen joints as described in Consolaro et al. 2012.	Proportion of patients who achieved minimal disease activity	• if physician's global assessment of disease activity ≤3.5, and • if parent's global rating of patient's overall well-being ≤2.5, and • if swollen joint count ≤1, then conclude the patient reaches minimal disease activity; all else conclude the minimal disease activity is not reached.	Component is missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing values. If all components are missing after LOCF, NRI will be applied to missing data.

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
Disease remission	Remission is defined as inactive disease for at least 24 consecutive weeks (Wallace et al. 2011). In inactive disease as secondary endpoint, ESR measure is used as acute-phase reactant in the core criteria. Thus, the disease remission as secondary endpoint is based on ESR.	Proportion of patients who achieved minimal disease activity	• At Visit 13, check if there are inactive disease status in all Visit 7, Visit 8, Visit 9, Visit 10, Visit 11, Visit 12, if yes, conclude the patient has achieved disease remission status. • At Visit 14, check if there are inactive disease status at Visit 8, Visit 9, Visit 10, Visit 11, Visit 12, Visit 13, if yes, conclude the patient has achieved disease remission status. • At Visit 15, check if there are inactive disease status at Visit 9, Visit 10, Visit 11, Visit 12, Visit 13, Visit14, if yes, conclude the patient has achieved disease remission status. • At Visit 16, check if there are inactive disease status at Visit 10, Visit 11, Visit 12, Visit 13, Visit 14, Visit 15, if yes, conclude the patient has achieved disease remission status. • At Visit 17, check if there are inactive disease status at Visit 11, Visit 12, Visit 13, Visit 14, Visit 15, Visit 16, if yes, conclude the patient has achieved disease remission status.	Missing if item is missing. NRI will be applied to missing data.

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
JADAS-27	The JADAS-27 score is a validated composite disease activity measure for JIA (Consolaro et al. 2012). Recently, the scoring system was adapted to use the 27-joint count (Bazso et al. 2009) and hsCRP or ESR for the inflammatory marker component (Nordal et al. 2012). The JADAS 27 includes the following joints: cervical spine, elbows, wrists, metacarpophalangeal joints (from first to third), proximal interphalangeal joints, hips, knees, and ankles. In JADAS-27 analysis as secondary endpoint, ESR measure is used as acute-phase reactant measure,	Change from baseline of JADAS-27 score	JADAS-27 score (0-57) will be determined based on 4 components: • PHY: Physician's Global Assessment of Disease Activity (0-10 VAS) • PGA: Parent's Global Assessment of Patient's overall Well-Being (0-100 mm VAS) • Joint27: Number of joints with active disease (27-joint assessment) • ESR (mm/hr) Calculation: NormESR: Normalized ESR (0-10) NormESR=(ESR-20)/10 Scaled PGA= scaled Parent's Global Assessment of Patient's overall Well-Being (0-10) Scaled PGA=PGA/10 JADAS(ESR)27=PHY+ScaledPGA+joint27+NormESR Change from baseline: observed JADAS(ESR)-27 at visit minus baseline JADAS-27	Missing if baseline is missing. For observed value, missing if all the components are missing. LOCF will be applied for post baseline missing value. If at least one component is non-missing, LOCF will be applied on the missing components prior to LOCF being applied to the composite score

Secondary Efficacy Endpoint Analyses

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
Pain (VAS)	The pain assessment item will be collected from the CHAQ.	Change from baseline of pain vas score	Calculation: Change from baseline: observed score at visit– baseline score	Missing if baseline or postbaseline value is missing. LOCF will be applied for post baseline missing value.

Abbreviations: CHQ-PF50 = Child Health Questionnaire-Parent Form 50; ESR = erythrocyte sedimentation rate; hsCRP = high-sensitivity C-reactive protein; JADAS = Juvenile Arthritis Disease Activity Score; JIA = juvenile idiopathic arthritis; LOCF = last observation carried forward; NRI = nonresponder imputation; PedACR = Pediatric American College of Rheumatology; VAS = Visual Analog Scale.

Table JAHV.6.5. Description of Analysis Period and Analysis Method of Secondary Endpoint and Health Outcome

Measure	Variable	Population	Analysis Timepoint	Treatment Comparisons	Analysis Method	Analysis Type
Disease flare	Proportion of patients with disease flare	DBW	Week 16 to Week 44	Baricitinib versus placebo	Logistic regression	Secondary analysis
PedACR response	PedACR30/50/70/90/100 response rates (compare to patient's condition prior to first dose of investigational product)	Safety/PK, OLLI, OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	Logistic regression	Secondary analysis
Active joint count	Change from baseline of active joint count	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
					MMRM	Supplementary analysis
Limited of range motion joint count	Change from baseline of limited range of motion joint count	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
					MMRM	Supplementary analysis
Physician's global assessment of disease activity	Change from baseline of physician's global assessment of disease activity	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
					MMRM	Supplementary analysis
Physical function as measured by CHAQ	Change from baseline of physical function	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
					MMRM	Supplementary analysis

Description of Analysis Period and Analysis Method of Secondary Endpoint and Health Outcome

Measure	Variable	Population	Analysis Timepoint	Treatment Comparisons	Analysis Method	Analysis Type
Parent's global assessment of overall well-being as measured by CHAQ	Change from baseline of parent's global assessment of overall well-being	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
					MMRM	Supplementary analysis

Description of Analysis Period and Analysis Method of Secondary Endpoint and Health Outcome

Measure	Variable	Population	Analysis Timepoint	Treatment Comparisons	Analysis Method	Analysis Type
hsCRP (mg/L)	Change from baseline of hsCRP	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
					MMRM	Supplementary analysis
ESR (mm/hr)	Change from baseline of ESR	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
					MMRM	Supplementary analysis
Physical Summary Score (PhS) and Psychosocial Summary Score (PsS)	Change from baseline in the Physical Summary Score (PhS) and Psychosocial Summary Score (PsS) of the Child Health Questionnaire-Parent Form 50 (CHQ-PF50)	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
Caregiver burden	Change from baseline of caregiver burden score, as measured by the Parental Impact-Time and Parental Impact-Emotion scales of the CHQ-PF50	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
Inactive disease	Proportion of patients who achieved inactive disease	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	Logistic regression	Secondary analysis
Minimal disease activity	Proportion of patients who achieved minimal disease activity	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	Logistic regression	Secondary analysis
Disease remission	Proportion of patients who achieved minimal disease activity	DBW	Week 16 to Week 44	Baricitinib versus placebo	Logistic regression	Secondary analysis

JADAS-27	Change from baseline of JADAS-27 score	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis
					MMRM	Supplementary analysis

Description of Analysis Period and Analysis Method of Secondary Endpoint and Health Outcome

Measure	Variable	Population	Analysis Timepoint	Treatment Comparisons	Analysis Method	Analysis Type
Pain (VAS)	Change from baseline of pain VAS item score	OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
		DBW	Week 16 to Week 44	Baricitinib versus placebo	ANCOVA	Secondary analysis

Abbreviations: ANCOVA = analysis of covariance; CHAQ = Childhood Health Assessment Questionnaire; CHQ-PF50 = Child Health Questionnaire-Parent Form 50; DBW = DBW = double-blind withdrawal; hsCRP = high-sensitivity C-reactive protein; JADAS = Juvenile Arthritis Disease Activity Score; MMRM = mixed model repeated measures; OLLI = open-label lead-in; PedACR = Pediatric American College of Rheumatology; VAS = Visual Analog Scale.

6.10.3. More Supplementary Analysis

For the composite measures/endpoints which have acute-phase reactant ESR or CRP as component, in general, the ESR version of derivation will be used as the main version of derivation for the first and secondary analysis. The hsCRP version of derivation will be used as a supplementary version for the supplementary analysis and will be provided for selected visits only, as defined in Table JAHV.6.6 and Table JAHV.6.7.

Table JAHV.6.6. Description of the Use of ESR and hsCRP in Endpoints Derivation

Endpoint	Version	Calculation difference between ESR version and hsCRP version	Measure	Type of the endpoint	Analysis
flare	ESR version	Derivation refer to Table JAHV.6.4	time to flare	primary endpoint	primary analysis
	ESR version	Derivation refer to Table JAHV.6.4	proportion of patients who have disease flare	secondary endpoint	secondary analysis
PedACR response	ESR version	Derivation refer to Table JAHV.6.4	proportion of patients who achieve PedACR response	secondary endpoint	secondary analysis
	PedACR (hsCRP) response	Derivation logic is same as ESR version	proportion of patients who achieve PedACR-hsCRP response	supplementary	supplementary analysis
JADAS-27	ESR version	Derivation refer to Table JAHV.6.4	change from baseline of JADAS-27	secondary endpoint	secondary analysis
	hsCRP version	Derivation logic is same as ESR version, use Norm hsCRP=(hsCRP-10)/10 instead of NormESR	change from baseline of JADAS-hsCRP-27	supplementary	supplementary analysis
inactive disease	ESR version	Derivation refer to Table JAHV.6.4	proportion of patients who achieve inactive disease status	secondary endpoint	secondary analysis
	hsCRP version	Derivation logic is same as ESR version, use hsCRP>3 instead of ESR>20	proportion of patients who achieve inactive disease status (hsCRP)	supplementary	supplementary analysis
disease remission	ESR version	Derivation refer to Table JAHV.6.4	proportion of patients who achieve disease remission status	secondary endpoint	secondary analysis
	hsCRP version	Derivation logic is same as ESR version, use hsCRP>3 instead of ESR>20	proportion of patients who achieve disease remission status (hsCRP)	supplementary	supplementary analysis

Abbreviations: ESR = erythrocyte sedimentation rate; hsCRP = high-sensitivity C-reactive protein; JADAS = Juvenile Arthritis Disease Activity Score; PedACR = Pediatric American College of Rheumatology.

Table JAHV.6.7. Description hsCRP Related Supplementary Analysis

Measure	Variable	Population	Analysis Time Point	Treatment Comparisons	Analysis Method	Analysis Type
PedACR-hsCRP response	PedACR-hsCRP 30/50/70/90/100 response rates (compare to patient's condition prior to first dose of investigational product)	OLLI2	Week 12	No comparison	Summary statistics	Supplementary analysis
		DBW	Week 16, week 32, Week 44	Baricitinib versus placebo	Logistic regression	Supplementary analysis
Inactive disease status (hsCRP)	Proportion of patients who achieved inactive disease status (hsCRP)	OLLI2	Week 12	No comparison	Summary statistics	Supplementary analysis
		DBW	Week 16, Week 32, Week 44	Baricitinib versus placebo	Logistic regression	Supplementary analysis
Disease remission status (hsCRP)	Proportion of patients who achieved disease remission status (hsCRP)	DBW	Week 16, Week 32, Week 44	Baricitinib versus placebo	Logistic regression	Supplementary analysis
JADAS-hsCRP-27	Change from baseline of JADAS-hsCRP-27 score	OLLI2	Week 12	No comparison	Summary statistics	Supplementary analysis
		DBW	Week 16, Week 32, Week 44	Baricitinib versus placebo	ANCOVA	Supplementary analysis
					MMRM	Supplementary analysis

Abbreviations: DBW = double-blind withdrawal ; hsCRP = high-sensitivity C-reactive protein; JADAS = Juvenile Arthritis Disease Activity Score; OLLI = open-label lead-in; PedACR = Pediatric American College of Rheumatology.

6.11. Planned Exploratory Analyses

In consideration of the limited sample size in JIA categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined]), a non-stratified logrank test may be applied to analysis of time to flare as an exploratory analysis. The p-value and median time to flare (if applicable) by treatment group will be reported. Treatment comparisons may also be analyzed using a Cox proportional hazards model. The hazard ratio with 95% CIs will be reported. The Kaplan-Meier product limit method will be used to estimate the survival curves for time-to flare. Time-to flare will also be summarized graphically by treatment group using Kaplan-Meier techniques.

6.11.1. X-ray and Structure Data

Semiannual wrist, hand, finger, and knee radiographs to monitor bone age and long bone growth. Imaging will be required until skeletal maturity is attained, and this should be determined by a qualified physician at the site. For patients already enrolled at the time of protocol amendment (d), the x-ray procedures will be optional. For patients that consent to the x-ray procedures, x-rays must be completed within 30 days from time of consent/assent and every 6 months ± 30 days thereafter.

Descriptive statistics will be used to summarize the modified total Sharp score (mTSS) (when data are available), skeletal age, and chronological age at baseline for all the patients who have a baseline value available.

For patients in the DBW period, an analysis of covariance model with baseline mTSS, presence of erosion at baseline (yes or no), and treatment group in the model will test the treatment difference between baricitinib and placebo in change from baseline to Week 44 in mTSS when data are available. Linear extrapolation will be applied for patients who are missing the radiograph at Week 44 but who have baseline and postbaseline radiographs available.

The skeletal age can deviate from the chronological age calculated from the date of birth. Results will include a description of the skeletal age, the chronological age, and the difference between skeletal age and chronological age at baseline and post baseline when data are available.

6.12. Acceptability and Palatability Analysis

Acceptability and palatability data will be collected and analyzed to address secondary and exploratory objectives of this study. The detail will be described by Lilly in separate acceptability and palatability analysis plans.

6.13. Health Outcomes/Quality-of-Life Analyses

Besides the health outcomes/ QOL measures discussed in Section 6.10.2 and Table JAHV.6.3. The change from baseline in European Quality of Life-5 Dimensions–Youth version (EQ-5D-Y) scores by treatment periods with the corresponding analysis population will be analyzed as exploratory analyses using methods described for continuous or categorical data as described for efficacy measures. The EQ-5D-Y will be scored according to the EQ-5D-Y user guide (Van Reenen et al. 2014).

Change from baseline in the individual scales (Global Health; Physical Functioning; Role/Social Limitations-Physical; Role/Social Limitations-Emotional/Behavioral; Bodily Pain/Discomfort; Behavior; Global Behavior Item; Mental Health; Self-Esteem; General Health Perception; Change in Health; Parental-Impact-Time; Parental Impact-Emotion; Family-Activities; Family-Cohesion) as measured by the CHQ-PF50 by treatment group during the DBW period will also be included as exploratory analyses. The individual scales of the CHQ-PF50 will be scored according to the manual, "CHQ Scoring and Interpretation manual". (HealthActCHQ 2013).

6.14. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Pharmacokinetic, pharmacodynamic and biomarker analyses to address secondary and exploratory objectives of this study will be described by Lilly in separate PK/PD and Biomarker analysis plans.

6.15. Evaluation of Immunological Measures

Change from baseline in immunoglobulin levels and peripheral blood immunophenotyping (including T and B cells, T cell subsets, and natural killer [NK] cells) at Weeks 4, 12, and at the end of the DBW period will be evaluated and summarized using descriptive statistics. Patients who are immunized with TDaP or pneumococcal conjugate vaccines will have their immunoglobulin G (IgG) antibody titers to the antigens evaluated preimmunization and at 4 and 12 weeks postimmunization. A primary immune response will be assessed in patients who have never received TDaP or pneumococcal conjugate vaccines previously, and secondary/booster responses will be assessed if the patients have previously received the vaccines.

6.16. Subgroup Analysis

Cox proportional hazards regression will be used to evaluate the similarity of treatment effect across the JIA categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined]). The model will have treatment and stratification variables as factors in the model. If any of these variables are redundant for a particular model, they will be dropped.

For the primary endpoint, selected secondary endpoints, and selected growth measurements, following subgroup analysis may be conducted:

- Gender: (Male; Female)
- Age group (12 to <18 years, 9 to <12 years, 6 to <9 years, 2 to <6 years)
- Geographic region (Asia, South America, Europe and rest of world)
- JIA categories (polyarticular and extended oligoarticular [combined] versus ERA and JPsA [combined])
- JIA categories (polyarticular, extended oligoarticular, ERA, JPsA)
- Predose exposure ESR categories (elevated [>20 mm/hour] and not elevated)
- history of prior bDMARD use (Yes or No)
- MTX usage at baseline (Yes or No)

6.17. Analysis on Subsets of Patients

Additional analysis on subpopulation of patients may be conducted based on country or geographic region.

The summary statistical analysis will be conducted for the following subsets of patients at each visit.

In the subset of patients with JPsA:

- change from baseline in Psoriasis Area and Severity Index (PASI) score during the OLLI period and during the DBW period.

In the subset of patients with JPsA or ERA:

- change from baseline in Spondyloarthritis Research Consortium of Canada (SPARCC) index of entheseal assessment during the OLLI period and during the DBW period
- change from baseline in Juvenile Spondyloarthritis Disease Activity Index (JSpADA) during the OLLI period and during the DBW period

The detail of the calculation is documented in Table JAHV.6.8.

Table JAHV.6.8. Description and Derivation of Endpoints for Subsets of Patient

Measure	Description	Variable	Derivation/Comment	Imputation approach if with missing components
Psoriasis Area and Severity Index (PASI)	Psoriasis Area and Severity Index (PASI): combines assessments of the extent of body-surface involvement in 4 anatomical regions (head and neck, trunk, arms, and legs) and the severity of scaling (S), redness (R), and plaque induration/infiltration (thickness, T) in each region, yielding an overall score of 0 for no psoriasis to 72 for the most severe disease (Fredriksson and Pettersson 1978). Severity is rated for each index (R, S, T) on a 0-4 scale (0 for no involvement up to 4 for severe involvement): 0 = none 1 = slight 2 = moderate 3 = severe 4 = very severe The body is divided into four anatomical regions comprising the head (h), upper limb (u), trunk (t), and lower limb (l). In each of these areas, the fraction of total body surface area affected is graded on a 0-6 scale (0 for no involvement; up to 6 for 90% - 100% involvement): 0 = 0% (clear) 1 = >0% to <10% 2 = 10% to <30% 3 = 30% to <50% 4 = 50% to <70%	PASI change from baseline	The composite PASI score is calculated by multiplying the sum of the individual-severity scores for each area by the weighted area-of-involvement score for that respective area, and then summing the four resulting quantities as follows: PASI = 0.1(Rh + Th + Sh)Ah + 0.2(Ru + Tu + Su)Au + 0.3(Rt + Tt + St)At + 0.4(Rl + Tl + Sl)Al Where, Rh, Ru, Rt, Rl = redness score of plaques on the head, upper limb, trunk, and lower limb, scored 0-4 respectively; Th, Tu, Tt, Tl = thickness score of plaques on the head, upper limb, trunk, and lower limb, scored 0-4 respectively; Sh, Su, St, Sl = scaliness score of plaques on the head, upper limb, trunk, and lower limb, scored 0-4 respectively; Ah, Au, At, Al = numerical value translation of % area of psoriatic involvement score for the head, upper limb, trunk, and lower limb, respectively. PASI scores are treated as a continuous score, with 0.1 increments within these values. Calculated as: observed PASI minus	If any individual score is missing, the PASI score will not be calculated, hence, missing. LOCF will be applied

Measure	Description	Variable	Derivation/Comment	Imputation approach if with missing components
	5 = 70% to <90% 6 = 90% to 100% The various body regions are weighted to reflect their respective proportion of body surface area.		baseline PASI	
Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis	SPARCC enthesitis is an index used to measure the severity of enthesitis (Maksymowych et al. 2009). The SPARCC assesses 16 sites for enthesitis using a score of "0" for no activity and "1" for activity. Sites assessed include Medial epicondyle (left/right [L/R]), Lateral epicondyle (L/R), Supraspinatus insertion into greater tuberosity of humerus (L/R), Greater trochanter (L/R), Quadriceps insertion into superior border of patella (L/R), Patellar ligament insertion into inferior pole of patella or tibial tubercle (L/R), Achilles tendon insertion into calcaneum (L/R), and Plantar fascia insertion into calcaneum (L/R).	SPARCC enthesitis change from baseline	The SPARCC is the sum of all site scores Range: 0–16, higher scores indicate more severe enthesitis. Calculated as: observed SPARCC enthesitis score – baseline SPARCC enthesitis score	If any individual component is missing, the SPARCC score will not be calculated, hence, missing. LOCF will be applied
Juvenile Spondyloarthritis Disease Activity (JSpADA) Index	The JSpADA is used to evaluate the disease activity of juvenile spondyloarthritis (Weiss et al. 2014). The range of possible scores is 0 to 8, where higher scores indicate more disease activity.	JSpADA change from baseline	The Juvenile Spondyloarthritis Disease Activity Index scores will be determined by 8 components: • Active joint count: 0 joints = 0, 1 to 2 joints = 0.5, >2 joints = 1 • Active enthesitis count: 0 entheses = 0, 1 to 2 entheses = 0.5, >2 entheses = 1 • Pain over the past week as assessed using a 0-10 NRS (0 = no pain; 10 = pain as bad as your child can imagine): 0 = 0, 1 to 4 = 0.5, 5 to 10 = 1	If any individual score is missing, the JSpADA score will not be calculated, hence, missing. LOCF will be applied

Measure	Description	Variable	Derivation/Comment	Imputation approach if with missing components
			• CRP level related to juvenile spondyloarthritis activity: normal = 0, 1 to 2 times normal = 0.5, >2 times normal = 1 • Morning stiffness >15 minutes: Absent = 0, Present = 1 • Clinical sacroiliitis (defined as the presence of ≥2 of the following: tenderness on examination, positive Patrick's test or flexion, abduction and external rotation (FABER) test, and inflammatory back pain): Absent = 0, Present = 1 • Uveitis (any uveitis including acute/symptomatic and chronic/asymptomatic disease): Absent = 0, Present =1 • Back mobility (abnormal back mobility defined as modified Schober's test <20 cm): Normal = 0, Abnormal = 1 Total score= sum of 8 domain score Calculated as: observed JSpADA score – baseline JSpADA score	

Abbreviations: LOCF – last-observation-carried-forward.

Table JAHV.6.9. Description of Analysis Period and Analysis Method of Subsets of Patients Analysis

Variable	Population	Analysis Period	Treatment Comparisons	Analysis Method	Analysis Type
PASI change from baseline	For JPsA patient in OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
	For JPsA patient in DBW	Week 16 to Week 44	No comparison	Summary statistics by treatment group	Secondary analysis
SPARCC enthesitis change from baseline	For JPsA and ERA patient in OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
	For JPsA and ERA patient in DBW	Week 16 to Week 44	No comparison	Summary statistics by treatment group	Secondary analysis
JSpADA change from baseline	For JPsA and ERA patient in OLLI2	Week 2 to Week 12	No comparison	Summary statistics	Secondary analysis
	For JPsA and ERA patient in DBW	Week 16 to Week 44	No comparison	Summary statistics by treatment group	Secondary analysis

Abbreviations: DBW = double-blind withdrawal; ERA = enthesitis-related juvenile idiopathic arthritis; Psoriasis Area and Severity Index; JSpADA = Juvenile Spondyloarthritis Disease Activity Index; OLLI = open-label lead-in; SPARCC = Spondyloarthritis Research Consortium of Canada.

6.18. Safety Analyses

The planned safety analyses are consistent with compound-level standards, which are based on various sources, including company standards, internal and external subject matter experts, and cross-industry initiatives (e.g., white papers produced by a PhUSE Computational Science Working Group [a collaboration with FDA and PhUSE], published in the PhUSE Deliverables Catalog). Descriptions of the safety analyses are provided in this SAP; however, some details are in compound-level safety standards.

The detail of analysis and corresponding period and population is defined in Section 6.1.1 and Table JAHV.6.10.

Table JAHV.6.10. Summary of Safety Analysis Population, Period, Duration of Exposure and Baseline

Population for safety analysis (Section 6.1.1)	General Safety Population	DBW Safety Population	
		Baricitinib	Placebo
Treatment period	OLLI period and Safety/PK assessment period	DBW period	
Duration of study drug exposure	date of last dose of baricitinib in safety/PK assessment or OLLI period – date of first dose of baricitinib in Safety/PK assessment or OLLI period + 1	date of last dose of treatment in DBW period – date of randomization in DBW period + 1	
Baseline for LLT used in defining treatment-emergence and change from baseline analysis	Baseline period: The start of screening and ends prior to the time of first dose of bari administration. (For labs, vital signs and growth: baseline will be all scheduled and unscheduled measurements recorded during baseline period.)		
Postbaseline for LLT used in defining treatment-emergence and change from baseline analysis	The event start time (onset time) is on or after first dose of baricitinib in Safety/PK assessment period or OLLI period, before time of randomization in DBW period (include up to 30 days off-drug follow-up time if patient did not enter DBW period). For treatment-emergent abnormal labs, vital signs and shift summaries in labs, all scheduled and unscheduled measurements will be included. For change from baseline labs, vital signs and growth, only scheduled visits will be included. The early termination visits (ETV) are considered scheduled visits.	The event start time (onset time) is on or after time of randomization in DBW period, up to 30 days off-drug follow-up time. For treatment-emergent abnormal labs, vital signs and shift summaries in labs, all scheduled and unscheduled measurements will be included. For change from baseline labs, vital signs and growth, only scheduled visits will be included. The early termination visits (ETV) are considered scheduled visits.	

Abbreviations: Bari = baricitinib; DBW = double-blind withdrawal; OLLI = open-label lead-in; JIA = juvenile idiopathic arthritis; LLT = Lowest Level Term; PK = pharmacokinetics.

Safety topics that will be addressed include the following: AEs, clinical laboratory evaluations, vital signs and physical characteristics, safety in special groups and circumstances, including adverse events of special interest (AESI) (see Section 6.18.5).

Unless otherwise specified, by-visit summaries will include planned on-treatment visits. For tables that summarize events (such as AEs, categorical lab abnormalities, shift to maximum value), post-last dose follow-up data will be included. For deaths and malignancies, all available follow-up data up to the end of the study will be included. Listings will include all safety data.

The following statistical methods will be used for safety analysis during the DBW period with the DBW safety population unless otherwise noted:

- The Fisher exact test will be used for treatment comparisons of proportions, and odds ratios with corresponding 95% confidence intervals will be provided.
- Treatment differences in mean change for continuous measurements will be assessed using an ANCOVA model fitting "baseline" as a covariate. Type 3 sums of squares will be used.

Though p-values will be provided for many of the safety analyses, they should not be overinterpreted. They correspond to data-driven hypotheses and are only useful as a flagging mechanism.

Statistical significance is designated at 2-sided p-values (rounding up to 3 decimal places) ≤0.05 for tests of treatment differences. However, p-values should not be overinterpreted.

Exposure-adjusted incidence rate (EAIR) will be provided for selected topics. The EAIR is evaluating the incidence of a first event per 100 patient-years at risk (PYR). Exposure will be calculated based on the analysis period defined as the treatment period plus up to 30 days off-drug follow-up time. Exposure time for a patient with an event will be counted up to the time of the start of event. Exposure time for a patient without an event will be censored at the end of the analysis period. For each EAIR provided, a Poisson distribution 95% CI may be calculated. Treatment group comparisons, when provided, will be provided based on the incidence rate difference (IRD) together with its 95% CI.

Not all displays described in this SAP will necessarily be included in the CSR. Any display described in this SAP 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.

6.18.1. Extent of Exposure

Duration of exposure (in days) to study drug will be summarized by treatment period with corresponding safety population, which is defined in Table JAHV.6.10.

Total PY of exposure will be reported for overall duration of exposure. Descriptive statistics (n, mean, standard deviation [SD], minimum, 1st quartile, median, 3rd quartile and maximum) will be provided for patient-days of exposure, and the frequency of patients falling into different exposure ranges will be summarized. Exposure ranges will generally be reported in weeks using the following as a general guide:

- ≥4 weeks, ≥16 weeks, ≥ 26 weeks, ≥ 52 weeks

- >0 to <4 weeks, ≥4 weeks to <16 weeks, ≥16 weeks to <26 weeks, ≥26 to <52 weeks, and ≥52 weeks

Overall exposure will be summarized in total PY which is calculated according to the following formula:

PYE = sum of duration of exposure in days (for all patients in treatment group) / 365.25

6.18.2. Adverse Events

Adverse events are recorded in the eCRFs. The planned summaries are provided in Table JAHV.6.11 and 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)]. The analysis population and period are defined in Table JAHV.6.11, and by-treatment analysis will provided for the DBW safety population in the DBW period.

If study drug is temporarily interrupted and subsequently restarted during the treatment period, the measurements taken during the temporary interruption will be included in the analysis. Where applicable, the time elapsed during the temporary interruption will also be included in analyses.

For events that are gender-specific (as defined by the MedDRA), the denominator and computation of the percentage will include only patients from the given sex.

Table JAHV.6.11. Summary Tables Related to Adverse Events

Analysis
An overview table, with the number and percentage of patients in the safety set with death, an SAE, any TEAE, discontinuation from the study due to an AE, permanent discontinuation from study drug due to an AE, or a severe TEAE
The number and percentages of patients with TEAEs will be summarized using MedDRA Preferred Term nested within System Organ Class.
The number and percentages of patients with TEAEs will be summarized using MedDRA Preferred Term.
The number and percentages of patients with TEAEs will be summarized using MedDRA Preferred Term for the common TEAEs (occurring in >=1%, before rounding, of treated patients).
The number and percentages of patients with TEAEs by maximum severity will be summarized using MedDRA Preferred Term for the common TEAEs. Only counts and percentages will be included for the TEAEs by maximum severity.
A listing of all deaths will be provided.
The number and percentage of patients who experienced a serious adverse event (including deaths and SAEs temporally associated or preceding deaths) will be summarized using MedDRA Preferred Term nested within System Organ Class.
A listing of SAEs will be provided.
The number and percentage of patients who permanently discontinued from study drug due to an adverse event (including adverse events that led to death) will be summarized using MedDRA Preferred Term nested within System Organ Class.
The number and percentage of patients who temporarily interrupted study drug due to an adverse event will be summarized using MedDRA Preferred Term nested within System Organ Class.

Abbreviations: AE=adverse event; MedDRA = Medical Dictionary for Regulatory Activities; SAE = serious adverse event; TEAE = treatment emergent adverse event.

6.18.3. Clinical Laboratory Evaluation

The planned summaries for clinical laboratory evaluations are provided in Table JAHV.6.12 and are described more fully in compound-level safety standards and in the laboratory-related PhUSE white papers [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 (PhUSE 2013) and 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 (PhUSE 2015)]. The analysis population and period are defined in Table JAHV.6.12, and by-treatment analysis will provided for the DBW safety population in the DBW period.

There is one special circumstance for laboratory values to be derived. The low-density lipoprotein (LDL)/high-density lipoprotein (HDL) ratio will be derived as the ratio of LDL cholesterol to HDL cholesterol. Similarly, the ratio of HDL to LDL will be derived. There are no central laboratory reference ranges for the LDL/HDL or HDL/LDL ratio.

For the categorical laboratory analyses (shift and treatment-emergent), the analysis period is defined as the treatment period plus up to 30 days off-drug follow-up time. The analysis period for the continuous by-visit laboratory analyses including change from baseline by visit and to last observation is defined as the treatment period excluding off-drug follow-up time.

Table JAHV.6.12. Summary Tables Related to Clinical Laboratory Evaluations

Analysis
Box plots for observed values by visit and change from baseline by visit.
Tables with number and percentages of patients who shift from normal/high to low (ie, treatment-emergent low) and number and percentages of patients who shift from normal/low to high (ie, treatment-emergent high)
Listing of abnormal findings for laboratory analyte measurements, including qualitative measures

6.18.4. Vital Signs and Other Physical Findings

The planned summaries for vital signs and physical measurement (systolic blood pressure [BP], diastolic BP, pulse, weight, , height, body mass index [BMI], temperature) are provided in Table JAHV.6.13 and are described more fully in compound-level safety standards and in the vitals-related PhUSE white papers [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 (PhUSE 2013) and 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 (PhUSE 2015)].

Table JAHV.6.13. Summary Tables Related to Vital Signs

Analysis
Box plots for observed values by visit and change from baseline by visit.
Tables with number and percentages of patients who shift from normal/high to low (ie, treatment-emergent low) and percentages of patients who shift from normal/low to high (ie, treatment-emergent high). The limits are defined in the compound-level safety standards and are based on literature.

For vital signs and physical characteristics, original-scale data will be analyzed. Mean changes from baseline and as incidence of abnormal values will be summarized. The observed values at each visit (starting at OLLI baseline) and change from baseline to each scheduled visit, respectively, will be displayed in box plots for patients by treatment period with the corresponding safety population. These box plots will be used to evaluate trends over time and to assess a potential impact of outliers on central tendency summaries. The scheduled visits/measurements will be included. The unscheduled visits and the repeat measurements will be excluded.

Table JAHV.6.14. Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurements for Children and Adolescents

Age (years)		Systolic BP, mm Hg (supine or sitting forearm at heart level)	Diastolic BP, mm Hg (supine or sitting forearm at heart level)	Pulse/HR bpm (supine or sitting)
Child 2-4	Low	≤75 (low limit) and decrease from lowest value during baseline ≥15 if >75 at each baseline visit	≤40 (low limit) and decrease from lowest value during baseline ≥10 if >40 at each baseline visit	<60 (low limit) and decrease from lowest value during baseline ≥25 if ≥60 at each baseline visit
	High[a]	≥110 (high limit) and increase from highest value during baseline ≥15 if <110 at each baseline visit	≥76 (high limit) and increase from highest value during baseline ≥10 if <76 at each baseline visit	>160 (high limit) and increase from highest value during baseline ≥25 if ≤160 at each baseline visit
Child 5-9	Low	≤80 (low limit) and decrease from lowest value during baseline ≥15 if >80 at each baseline visit	≤45 (low limit) and decrease from lowest value during baseline ≥10 if >45 at each baseline visit	<60 (low limit) and decrease from lowest value during baseline ≥25 if ≥60 at each baseline visit
	High[a]	≥119 (high limit) and increase from highest value during baseline ≥15 if <119 at each baseline visit	≥78 (high limit) and increase from highest value during basline ≥10 if <78 at each baseline visit	>150 (high limit) and increase from highest value during baseline ≥25 if ≤150 at each baseline visit
Child 10-12	Low	≤85 (low limit) and decrease from lowest value during baseline ≥20 if >85 at each baseline visit	≤50 (low limit) and decrease from lowest value during baseline ≥10 if <50 at each baseline visit	<60 (low limit) and decrease from lowest value during baseline ≥25 if ≥60 at each baseline visit
	High[a]	≥126 (high limit) and increase from highest value during baseline ≥20 if <126 at each baseline visit	≥82 (high limit) and increase from highest value during baseline ≥10 if <82 at each baseline visit	>140 (high limit) and increase from highest value during baseline ≥25 if ≤140 at each baseline visit
Adolescent 13 - 17	Low	≤90 (low limit) and decrease from lowest value during baseline ≥20 if >90 at each baseline visit	≤50 (low limit) and decrease from lowest value during baseline ≥10 if >50 at each baseline visit	<50 (low limit) and decrease from lowest value during baseline ≥15 if ≥50 at each baseline visit
	High[a]	≥129 (high limit) and increase from highest value during baseline ≥20 if <129 at each baseline visit	≥86 (high limit) and increase from highest value during baseline ≥10 if <86 at each baseline visit	>120 (high limit) and increase from highest value during baseline ≥15 if ≤120 at each baseline visit

Abbreviations: BP = blood pressure; HR = heart rate.

[a] The high limit values shown in this table correspond to 95th percentile for the age group under the 2017 American College of Cardiology/American Heart Association task Force on Clinical Practice Guidelines revised criteria for hypertension. Values higher than 95th percentile are consistent with Stage 1 or Stage 2 hypertension. Under some circumstances it may be appropriate to conduct analyses considering only the change from baseline reference limit.

6.18.4.1. Standardized Growth

Weight, height, and BMI data will be merged to the Centers for Disease Control and Prevention (CDC) standard growth data (released in 2000) by age and gender in order to compare patients' growth with the standard. Z-score and standardized percentile of weight, height, and BMI at each visit will be calculated based on the 2000 CDC growth charts.

The z-score and percentile calculations are based on algorithms and data provided by the National Center for Health Statistics. The details are provided in the CDC website (https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm).

- The following summaries will be provided by treatment period with the corresponding safety population: baseline, mean change of actual measure, z-score and standardized percentile of weight, height, and BMI will be summarized.
- Patients' mean weight, height, and BMI standardized percentile will be plotted versus investigational product exposure time.

By-patient listings of actual measures, z-scores, standardized percentiles in weight, height, and BMI for each visit will be provided.

6.18.5. Special Safety Topics, including Adverse Events of Special Interest

In addition to general safety parameters, safety information on specific topics of special interest will also be presented. Additional special safety topics may be added as warranted. The topics outlined in this section include the protocol-specified AESI.

In general, for topics regarding safety in special groups and circumstances, patient profiles and/or patient listings, where applicable, will be provided when needed to allow for medical review of the time course of cases/events, related parameters, patient demographics, study drug treatment, and meaningful concomitant medication use. In addition to the safety topics for which provision or review of patient data are specified, these will be provided when summary data are insufficient to permit adequate understanding of the safety topic.

The analysis and summary will be provided by treatment period with the corresponding safety population, where the detailed definition of the population and baseline can be found in Section 6.1.1 and Table JAHV.6.10.

6.18.5.1. Abnormal Hepatic Tests

Hepatic labs include alanine aminotransferase (ALT), aspartate transaminase (AST), total bilirubin (TBL), and serum alkaline phosphatase (ALP). When criteria are met for hepatic evaluations, investigators will complete a follow-up hepatic safety eCRF. The planned summaries are provided in Table JAHV.6.15.

Table JAHV.6.15. Summary Tables Related to Hepatic Safety

Analysis
ALT and AST: The number and percentages of patients with a measurement greater than or equal to 3 times (3X), 5 times (5X), and 10 times (10X) the central lab upper limit of normal (ULN) for all patients with a postbaseline value and for subsets based on various levels of baseline value.
TBL: The number and percentages of patients with a measurement greater than or equal to 2 times (2X) the central lab ULN will be summarized for all patients with a postbaseline value and for subsets based on various levels of baseline value.

ALP: The number and percentages of patients with a measurement greater than or equal to 1.5 times (1.5X) the central lab ULN will be summarized for all patients with a postbaseline value and for subsets based on various levels of baseline value.
Plot of maximum postbaseline ALT vs. maximum postbaseline total bilirubin.
Patient profiles including demographics, disposition, information collected on the hepatic-safety CRF (where applicable) and a display of study drug exposure, adverse events, medications, blood pressure, heart rate, and the liver-related measurements over time will be provided for patients with information collected on the hepatic-safety CRF and any additional patients meeting ALT or AST measurement greater than or equal to 5X ULN (on a single measurement) or ALP measurement greater than or equal to 2X ULN (on a single measurement).
Treatment-Emergent Potential Hepatic Disorders Based on MedDRA SMQs: treatment-emergent potentially drug-related hepatic disorders are defined by using the MedDRA preferred terms contained in any of the following SMQs: • Broad and narrow terms in the Liver related investigations, signs and symptoms SMQ (20000008) • Broad and narrow terms in the Cholestasis and jaundice of hepatic origin SMQ (20000009) • Broad and narrow terms in the Hepatitis noninfections SMQ (20000010) • Broad and narrow terms in the Hepatic failure, fibrosis and cirrhosis, and other liver damage SMQ (20000013) Narrow terms in the liver-related coagulation and bleeding disturbances SMQ (20000015)

Abbreviations: ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate transaminase; CRF = case report form; MedDRA = Medical Dictionary for Regulatory Activities; SMQ = Standardized MedDRA Query; ULN = upper limit of normal; TBL = total bilirubin.

6.18.5.2. Hematologic Changes

Hematologic changes will be assessed through analysis of hemoglobin, white blood cell (leukocyte) count, absolute neutrophil count, lymphocyte count, and platelet count. Common Terminology Criteria for Adverse Events (CTCAEs) will be applied for laboratory tests potentially related to myelosuppressive events. The planned summaries are provided in Table JAHV.6.16 and are described more fully in compound-level safety standards.

Table JAHV.6.16. Summary Tables Related to Hematologic Changes

Analysis
Shift tables showing the number and percentage of patients based on baseline to maximum will be created, with baseline depicted by the most extreme CTCAE grade during the baseline period. With each shift table, a summary displaying the number and percentage of patients who decreased, increased, or stayed the same in CTCAE grade category will be presented.
The number and percentages of patients with treatment-emergent shifts at any time will be summarized, based on any increase to CTCAE Grade 1 or above, Grade 2 or above, Grade 3 or above, and Grade 4 or above.
The number and percentages of patients with treatment-emergent thrombocytosis will be summarized, defined as an increase in platelet count from a maximum baseline value ≤600 billion/L to any postbaseline value >600 billion/L. Similar analysis will use a cut-off of 400 billion/L. Planned and unplanned measurements will be included.
Listing of patients with treatment-emergent thrombocytosis

Abbreviations: CTCAE = Common Terminology Criteria for Adverse Events.

6.18.5.3. Lipid Effects

Lipid effects will be assessed through analysis of elevated total cholesterol, elevated low-density lipoprotein (LDL) cholesterol, decreased and increased high-density lipoprotein (HDL) cholesterol, and elevated triglycerides. The planned summaries are provided in Table JAHV.6.17 and are described more fully in compound-level safety standards.

Table JAHV.6.17. Summary Tables Related to Lipid Effects

Analysis
Shift tables showing the number and percentage of patients based on baseline to maximum will be created, with baseline depicted by the most extreme NCEP-based level during the baseline period. With each shift table, a summary displaying the number and percentage of patients who decreased, increased, or stayed the same in NCEP-based level will be presented.
The number and percentages of patients with treatment-emergent shifts at any time will be summarized, based on increases to various levels of NCEP-based categories.
The number and percentages of patients with treatment-emergent potential hyperlipidemia will be summarized using a predefined MedDRA list of PTs that is a subset of the narrow scope PTs in the MedDRA SMQ "Dyslipidemia" (code 200000026) (see compound-level safety standards)..

Abbreviations: MedDRA = Medical Dictionary for Regulatory Activities; NCEP = National Cholesterol Education Program; PT = Preferred Term; SMQ = Standardised MedDRA Query.

Categorical analyses will be performed using Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents (2011) as shown in the Table JAHV.6.18.

Table JAHV.6.18. Acceptable, Borderline-High, and High Plasma Lipid, Lipoprotein, and Apolipopretein Concentrations for Children and Adolescents

Category	Low (mg per dL)[a]	Acceptable (mg per dL)	Borderline-high (mg per dL)[a]	High (mg per dL)[a]
total cholesterol	--	<170	170 to 199	≥200
LDL cholesterol		<110	110 to 129	≥130
Non-HDL cholesterol		<120	120 to 144	≥145
Apoliprotein B		<90	90 to 109	≥110
Triglycerides				
0 to 9 years of age		<75	75 to 99	≥100
10 to 19 years of age		<90	90 to 129	≥130
HDL Cholesterol	<40	>45	40 to 45	
Apolipoprotein A-1	<115	>120	120	

Abbreviations: HDL= high-density lipoprotein; LDL= Low-density lipoprotein

[a] Low cut point for HDL cholesterol and apolipoprotein A-1 represent approximately the 10th percentile. The cut points for high and borderline-high represent approximately the 95th and 75th percentiles, respectively.

6.18.5.4. Renal Function Effects

Effects on renal function will be assessed through analysis of elevated creatinine. The planned summaries are provided in Table JAHV.6.19 and are described more fully in compound-level safety standards.

Table JAHV.6.19. Summary Tables Related to Effects on Renal Function

Analysis
Shift tables showing the number and percentage of patients based on baseline to maximum will be created, with baseline depicted by the most extreme CTCAE grade during the baseline period. With each shift table, a summary displaying the number and percentage of patients who decreased, increased, or stayed the same in CTCAE grade category will be presented.
The number and percentages of patients with treatment-emergent shifts at any will be summarized, based on any increase to CTCAE Grade 1 or above, Grade 2 or above, Grade 3 or above, and Grade 4 or above.

Abbreviations: CTCAE = Common Terminology Criteria for Adverse Events.

6.18.5.5. Elevations in Creatine Phosphokinase (CPK)

The planned summaries are provided in Table JAHV.6.20 and are described more fully in compound-level safety standards.

Table JAHV.6.20. Summary Tables Related to Effects on CPK

Analysis
Shift tables showing the number and percentage of patients based on baseline to maximum will be created, with baseline depicted by the most extreme CTCAE grade during the baseline period. With each shift table, a summary displaying the number and percentage of patients who decreased, increased, or stayed the same in CTCAE grade category will be presented.
The number and percentages of patients with treatment-emergent shifts at any time will be summarized, based on any increase to CTCAE Grade 1 or above, Grade 2 or above, Grade 3 or above, and Grade 4 or above.
Treatment-emergent adverse events potentially related to muscle symptoms may also be analyzed based on reported AEs. The Muscle Symptoms special search category is a predefined MedDRA search criteria list that contains the narrow scope terms from the Rhabdomyolysis/myopathy SMQ (code 20000002) plus selected terms from the Musculoskeletal SOC

Abbreviations: AE = adverse event; CPK = creatine phosphokinase; CTCAE = Common Terminology Criteria for Adverse Events; MedDRA = Medical Dictionary for Regulatory Activities; PT = Preferred Term; SOC = System Organ Class.

6.18.5.6. Infections

Infections will be defined using all the PTs from the MedDRA Infections and Infestations System Organ Class (SOC). The MedDRA terms used to identify infections considered to be opportunistic infections (OIs) are based on Winthrop and colleagues (2015) and are listed in the compound-level safety standards. The list contains narrow (more specific) and broad (less specific) PTs.

The planned summaries are provided in Table JAHV.6.21 and are described more fully in compound-level safety standards.

Table JAHV.6.21. Summary Tables Related to Infections

Analysis
The number and percentage of patients with treatment-emergent infections, serious infections, and infections resulting in permanent study drug discontinuation will be summarized using MedDRA PTs.
The number and percentage of patients with TEAEs of infections by maximum severity will be summarized using MedDRA PTs.
Listing of patients experiencing TEAE infections will be provided. The listing will include patient demographics, treatment group, treatment start and stop dates, infectious PT event, event start and stop dates, total leukocytes, total lymphocytes, absolute neutrophils, event seriousness, and event outcome.
Summary of Opportunistic Infections based on MedDRA PTs after the potential opportunistic infections are reviewed by medical and confirmed as opportunistic infections.
Listing of Opportunistic Infections based on MedDRA PTs during the study.
A summary table of herpes zoster will be provided, including event maximum severity, seriousness, whether resulting in temporary study drug interruption, whether resulting in study drug discontinuation, whether treated with antiviral medication, and event outcome.
The summary table of herpes simplex will include event maximum severity, seriousness, whether resulting in temporary study drug interruption, whether resulting in study drug discontinuation, and whether treated with antiviral medication.
A listing of patients with detectable HBV DNA will be provided.
Hepatitis B virus DNA status (not detectable, detectable but not quantifiable [ie, < lower limit of detection (LLOD)], quantifiable [ie, ≥LLOD]) will be summarized, stratified by applicable baseline HBV serology status.

Abbreviations: MedDRA = Medical Dictionary for Regulatory Activities; OI = opportunistic infection; PT = Preferred Term; TEAE = treatment-emergent adverse event; HBV = hepatitis B virus.

6.18.5.7. Allergic Reactions and Hypersensitivities

A search for relevant events related to allergic reaction and hypersensitivity will be performed using the following SMQs:

- Anaphylactic reaction SMQ (20000021)
- Hypersensitivity SMQ (20000214)
- Angioedema SMQ (20000024)

Events that satisfy the queries will be listed, by temporal order within patient ID, and will include SOC, PT, SMQ event categorization including detail on the scope (narrow or broad), reported AE term, AE onset and end dates, severity, seriousness, outcome, etc.

The summaries described in Table JAHV.6.22 will be created if there are sufficient numbers of events to warrant further examination beyond the listing specified above.

The anaphylactic reaction SMQ consists of a narrow search containing PTs that represent core anaphylactic reaction terms, a broad search that contains additional terms (signs and symptoms

possibly indicative of anaphylactic reaction) that are added to those included in the narrow search, and an algorithm.

The algorithmic approach (Algorithm 1) (which is similar to the algorithm approach defined in Sampson, et al 2006) comprises 1 or more events associated with an individual administration of study drug, where the events include:

- A narrow term from the anaphylactic reaction SMQ (Category A of the SMQ) or
- Paired terms from the anaphylactic reaction SMQ, comprising terms from at least 2 of the following 3 categories from the SMQ:
  - Category B - (Upper Airway/Respiratory signs and symptoms)
  - Category C - (Angioedema/Urticaria/Pruritus/Flush signs and symptoms)
  - Category D - (Cardiovascular/Hypotension signs and symptoms).

Within the paired terms approach, it is important to recognize that occurrence of these events should be nearly coincident and develop rapidly after exposure to an antigen; a window wherein onset or severity change of the events occur within the same calendar day will be used.

In addition, a second algorithmic approach (Algorithm 2) will be calculated similarly to the algorithm approach defined above (Algorithm 2) but includes an additional category, Category E. The paired terms according to Algorithm 2 will comprise terms from at least 2 of 4 categories (Categories B, C, D, and E). Categories B, C and D are already defined, and Category E includes any of the Gastrointestinal preferred term events (Nausea, Vomiting, Diarrhoea, and Abdominal pain). A patient's listing will be generated based on Algorithm 2 and individual cases will be examined to determine if the cases suggest anaphylaxis. Those cases suggestive of anaphylaxis will be described in CSR.

Table JAHV.6.22. Summary Tables Related to Allergic Reactions/Hypersensitivities

Analysis
Two listings for Allergic Reaction and Hypersensitivities for events that satisfy the queries defined in this section will be listed, by temporal order within patient ID, and will include SOC, PT, SMQ event categorization including detail on the scope (narrow, algorithmic, or broad), reported AE term, AE onset and end dates, severity, seriousness, outcome, etc.
The number and percentages of patients with TEAEs will be summarized using MedDRA Preferred Term for any narrow or algorithmic term in the compound-level safety standard from any one of the 3 SMQs (each SMQ and SMQs combined)
The number and percentages of patients with TEAEs will be summarized using MedDRA Preferred Term for any broad term (each SMQ separately)

Abbreviations: MedDRA = Medical Dictionary for Regulatory Activities; SMQ = Standardised MedDRA Query; TEAE = treatment-emergent adverse event.

6.18.5.8. Major Adverse Cardiovascular Events and Other Cardiovascular Events

Major Adverse Cardiovascular Events (MACE) and other cardiovascular events will be adjudicated by an independent, external adjudication committee. All confirmed events after adjudication will be used for the analysis.

The planned summaries are provided in Table JAHV.6.23 and are described more fully in compound-level safety standards.

Table JAHV.6.23. Summary Tables Related to MACE and Other Cardiovascular Events

Analysis
The number and percentage of patients with MACE, other cardiovascular events, non-cardiovascular death, and all-cause death, as positively adjudicated, will be summarized based on the categories and subcategories as defined in compound-level safety standards.
A listing of the MACE and other CV events sent for cardiovascular adjudication will be provided to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result.

Abbreviations: MACE = Major Adverse Cardiovascular Event; MedDRA = Medical Dictionary for Regulatory Activities; PT = Preferred Term.

6.18.5.9. Thromboembolic Events

Venous thromboembolic (VTE) events will be adjudicated by an independent, external adjudication committee. Venous and pulmonary artery thromboembolic events will be classified as deep vein thrombosis (DVT), pulmonary embolism (PE), or other peripheral venous thrombosis. Additionally, arterial thromboembolic (ATE) events will be adjudicated by an independent, external adjudication committee. All confirmed events after adjudication will be used for the analysis.

The planned summaries for VTE are provided in Table JAHV.6.24 and are described more fully in compound-level safety standards.

Table JAHV.6.24. Summary of Tables Related to VTE Events

Analysis
The number and percentage of patients with a VTE, DVT/PE, DVT, PE, and other peripheral venous thrombosis, as positively adjudicated, will be summarized
A listing of the VTEs sent for adjudication will be provided to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result.

Summary of Tables Related to VTE Events

Abbreviations: DVT = deep vein thrombosis; MedDRA = Medical Dictionary for Regulatory Activities; PE = pulmonary embolism; PT = Preferred Term; VTE = venous thromboembolic.

Arterial Thrombosis(ATE) Events

The planned summaries for ATE are provided in Table JAHV.6.25 and are described more fully in compound-level safety standards.

Table JAHV.6.25. Summary of Tables Related to ATE Events

Analysis
The number and percentage of patients with a positively adjudicated ATE
A listing of the ATEs sent for adjudication to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result

Abbreviations: ATE = arterial thromboembolic; MedDRA = Medical Dictionary for Regulatory Activities; PT = preferred term; VTE = venous thromboembolism.

6.18.5.10. Malignancies

Malignancies will be identified using terms from the Malignant tumors SMQ. Malignancies excluding nonmelanoma skin cancers (NMSC) and NMSC will be reported separately. All the cases identified by the Malignant tumors SMQ (SMQ code = 20000194) will be assessed through medical review to determine *confirmed* NMSC cases.

The planned summaries are provided in Table JAHV.6.26 and are described more fully in compound-level safety standards.

Table JAHV.6.26. Summary Tables Related to Malignancies

Analysis
The number and percentage of patients with treatment-emergent malignancies excluding NMSC and NMSC will be summarized.
Listing of all malignancy cases, with an NMSC flag.

Abbreviations: NMSC = nonmelanoma skin cancers.

6.18.5.11. Gastrointestinal Perforations

Potential gastrointestinal (GI) perforations will be identified using terms from the GI perforations SMQ. Potential GI perforations identified by the SMQ (SMQ code = 20000107) search will be provided as a listing for internal review by the medical safety team. Each case will be assessed to determine whether it is a GI perforation. All confirmed events after medical review will be used

for the analysis. The planned summaries are provided in Table JAHV.6.27 and are described more fully in compound-level safety standards.

Table JAHV.6.27. Summary Tables Related to Gastrointestinal Perforations

Analysis
The number and percentage of patients with treatment-emergent gastrointestinal perforations will be summarized using MedDRA PTs.
Listing of all treatment-emergent gastrointestinal perforations during the study.

Abbreviations: MedDRA = Medical Dictionary for Regulatory Activities; PT = Preferred Term.

6.19. COVID-19 Trial Impact

Patients who experience an impact to their trial participation due to quarantine and/or travel restrictions related to COVID-19 will have their type of impact summarized. COVID-19 specific impacts will be summarized by the following by-patient listings:

- Listing of study and treatment discontinuation related to COVID-19
- Listing of COVID-19 adverse events based on SMQ=20000237 using the narrow term classification.
- Summary of COVID-19 TEAE using MedDRA PT based on SMQ=20000237 with the narrow term classification

6.20. Protocol Deviations

Protocol deviations will be tracked by the clinical team, and their importance will be assessed by key team members during protocol deviation review meetings. Out of all protocol deviations identified, a subset occurring during the OLLI prior to the Week 12 with the potential to affect futility analyses, will result in exclusion from futility analysis.

Potential examples of protocol deviations include patients who receive excluded concomitant therapy, significant noncompliance with study medication, patients incorrectly enrolled in the study, and patients whose data are questionable due to significant site quality or compliance issues.

Refer to a separate document for the important protocol deviations.

The number and percentage of patients having IPD(s) will be summarized in the OLLI and the DBW periods (by treatment) within category and subcategory of deviation using the OLLI and DBW populations. Individual patient listings of IPDs will be provided. A summary of reasons patients were excluded from the futility analysis will be provided.

6.21. Interim Analyses and Data Monitoring

A Data Monitoring Committee (DMC) will oversee the conduct of this trial. The DMC will consist of members external to Lilly. This DMC will follow the rules defined in the DMC charter, focusing on potential and identified risks for this molecule and for this class of

compounds. Data Monitoring Committee membership will include, at a minimum, specialists with expertise in rheumatology, statistics, and other appropriate specialties.

The DMC will be authorized to review unblinded results of analyses by treatment group prior to final database lock, including study discontinuation data, AEs including SAEs, clinical laboratory data, vital sign data, etc. The DMC may recommend continuation of the study, as designed; temporary suspension of enrollment; or the discontinuation of a particular dose regimen or the entire study. The DMC may request to review efficacy data to investigate the benefit/risk relationship in the context of safety observations for ongoing patients in the study. However, the study will not be stopped for positive efficacy results. Hence, no alpha is spent. Details of the DMC, including its operating characteristics, are documented in the Baricitinib Pediatric Joint DMC charter, and further details are given in the Interim Analysis Plan in Section 6.21.1.

Besides DMC members, a limited number of preidentified individuals may gain access to the limited unblinded data, as specified in the unblinding plan, prior to the interim or final database lock, to initiate the final population PK/pharmacodynamic (PD) model development processes or for preparation of regulatory documents. Information that may unblind the study during the analyses will not be reported to study sites or the blinded study team until the study has been unblinded.

A futility analysis will be conducted using the PedACR(ESR) 30 response rate observed in the first 100 patients who complete the OLLI period, this will include:

- patients who participate Safety/PK assessment period and complete OLLI period with age based final dose,
- patients who start from OLLI period and take at least 1 age-based final dose

The OLLI patients (Section 6.1.1) who discontinue early will be included. The patients who have important protocol deviations will be excluded; the detail is described in Section 6.20. The missing data imputation rule will align with Section 6.3 and Table JAHV.6.4 for PedACR(ESR) 30 response imputation. The study will stop for futility if the proportion of patient who achieve PedACR(ESR) 30 response is less than 50%. The futility analysis will based on the observed response rate, no statistical inference will be conducted.

Unblinding details are given in Section 7.

6.21.1. Interim Analysis Plan

Analyses for the DMC will include listings and/or summaries of the following information:

- patient disposition, demographics, and baseline characteristics
- exposure
- adverse events, to include the following:
  - treatment-emergent adverse events

  - serious adverse events, including deaths
  - selected special safety topics
- clinical laboratory results
- vital signs
- growth parameters (for example, height, weight, assessment of bone age)

All listings will include patient ID. Summaries will include TEAEs, SAEs, special topics AEs, and treatment-emergent high and low laboratory and vital signs in terms of counts, percentages and incidence rates, where applicable. For continuous analyses, box plots of laboratory analytes will be provided by time point.

The DMC may request efficacy data if they feel there is value and to confirm a reasonable benefit/risk profile for ongoing patients in the studies. If efficacy data are requested, they will be time to disease flare and proportion of patients who are PedACR30 responders during OLLI period or proportion of patients who maintain PedACR 30 response during DBW period. Further details are given in the DMC charter.

6.22. Annual Report Analyses

Annual report analyses, such as for the Development Update Safety Report (DSUR), will be documented in a separate document.

6.23. 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 a summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and ‘Other’ AEs are summarized by treatment group in the DBW period, 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 in the DBW period, the following are provided:
  - the number of participants at risk of an event
  - the number of participants who experienced each event term
  - 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).
- Adverse event reporting is consistent with other document disclosures, for example, the CSR, manuscripts, and so forth.

Similar methods will be used to satisfy the European Clinical Trials Database (EudraCT) requirements.

7. Unblinding Plan

Refer to a separate blinding and unblinding plan document for details.

8. References

Bazso A, Consolaro A, Ruperto N, et al. Pediatric Rheumatology International Trials Organization. Development and testing of reduced joint counts in juvenile idiopathic arthritis. *J Rheumatol*. 2009;36(1):183-190. https://doi.org/10.3899/jrheum.080432

Consolaro A, Bracciolini G, Ruperto N, et al. Paediatric Rheumatology International Trials Organization. Remission, minimal disease activity, and acceptable symptom state in juvenile idiopathic arthritis: defining criteria based on the juvenile arthritis disease activity score. *Arthritis Rheum*. 2012;64(7):2366-2374. https://doi.org/10.1002/art.34373

Filocamo G, Davì S, Pistorio A, et al. Evaluation of 21-numbered circle and 10-centimeter horizontal line visual analog scales for physician and parent subjective ratings in juvenile idiopathic arthritis. *J Rheumatol*. 2010;37(7):1534-1541. https://doi.org/10.3899/jrheum.091474

Fredriksson T, Pettersson U. Severe psoriasis-oral therapy with a new retinoid. *Dermatologica.* 1978;157(4):238-244. https://doi.org/10.1159/000250839

Giannini EH, Ruperto N, Ravelli A, et al. Preliminary definition of improvement in juvenile arthritis. *Arthritis Rheum*. 1997;40(7):1202-1209. https://doi.org/10.1002/1529-0131(199707)

HealthActCHQ. Child health Questionnaire (CHQ) Scoring and Interpretation Manual. 1st ed. Boston: HealthActCHQ, Inc.; 2013.

[ICH] International Conference on Harmonisation. Harmonised Tripartite Guideline: Clinical safety data management: definitions and standards for expedited reporting. E2A. 1994;Step 4. Available at: http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html. Accessed 21 July 2017.

Maksymowych WP, Mallon C, Morrow S, et al. Development and validation of the Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index. *Ann Rheum Dis*. 2009;68(6):948-953. https://doi.org/10.3899/jrheum.140519

National Heart, Lung, and Blood Institute. "Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents: Summary report (NIH Publication No. 12)."

Nordal EB, Zak M, Aalto K, et al. Validity and predictive ability of the juvenile arthritis disease activity score based on CRP versus ESR in a Nordic population-based setting. *Ann Rheum Dis*. 2012;71(7):1122-1127. https://doi.org/10.1136/annrheumdis-2011-200237

PhUSE Computational Science Deliverables Catalog. Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated Summary Document. 2017. https://www.phuse.eu/documents//working-groups/cswhitepaper-adverseevents-v10-4442.pdf

PhUSE Computational Science Deliverables Catalog. 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 Submissin Documents. (2013).http://www.phusewiki.org/wiki/images/4/48/CSS_WhitePaper_CentralTendency_v1.0.pdf

PhUSE Computational Science Deliverables Catalog. 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). https://www.phuse.eu/documents//working-groups/deliverables/analyses-displays-associated-with-outliers-or-shifts-from-normal-to-abnormal-version-10-10-sept-15-11815.pdf

Program Safety Analysis Plan: Baricitinib (LY3009104) Version 6. Report on file, Eli Lilly and Company.

Protocol I4V-MC-JAHV:A Randomized, Double-Blind, Placebo-Controlled, Withdrawal, Safety and Efficacy Study of Oral Baricitinib in Patients from 2 Years to Less Than 18 Years Old with Juvenile Idiopathic Arthritis (JIA). Report on file, Eli Lilly and Company.

Ravelli A, Ioseliani M, Norambuena X, et al. Adapted versions of the Sharp/van der Heijde score are reliable and valid for assessment of radiographic progression in juvenile idiopathic arthritis. Arthritis Rheum. 2007;56(9):3087-3095. https://doi.org/10.1002/art.22835

Ruperto N, Ravelli A, Pistorio A, et al. Paediatric Rheumatology International Trials Organisation. Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ) in 32 countries. Review of the general methodology. *Clin Exp Rheumatol.* 2001;19(suppl 23):S1-S9.

Van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol. 2000;27(1):261-263.

Van Reenen M, Janssen B, Oppe M, et al. 2014. EQ-5D-Y user guide: basic information on how to use the EQ-5D-Y instrument. Rotterdam: EuroQoL Group.

Wallace CA, Giannini EH, Huang B, et al.Childhood Arthritis Rheumatology Research Alliance; Pediatric Rheumatology Collaborative Study Group; Paediatric Rheumatology International Trials Organisation. American College of Rheumatology provisional criteria for defining clinical inactive disease in select categories of juvenile idiopathic arthritis. *Arthritis Care Res (Hoboken).* 2011;63(7):929-936.

Weiss PF, Colbert RA, Xiao R, et al. Development and retrospective validation of the juvenile spondyloarthritis disease activity index. *Arthritis Care Res (Hoboken)*. 2014;66(12):1775-1782. https://doi.org/10.1002/acr.22411.

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/annhrheumdis-2015-207841

Leo Document ID = 27b5c7a9-d3aa-48b1-898d-be46c8296d3c

Approver: PPD

Approval Date & Time: 14-Mar-2022 19:04:47 GMT

Signature meaning: Approved