I4V-MC-JADY Statistical Analysis Plan v 3

A Phase 3 Study to Evaluate the Long-Term Safety and Efficacy of baricitinib in Patients with Rheumatoid Arthritis

NCT01885078

Approval Date: 01-Oct-2020

1. Statistical Analysis Plan: I4V-MC-JADY: A Phase 3 Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Patients with Rheumatoid Arthritis

Confidential Information

Baricitinib (LY3009104) Rheumatoid Arthritis

Study I4V-MC-JADY is a Phase 3, multicenter, outpatient long-term safety and efficacy extension study. Patients who complete a Phase 2 or Phase 3 baricitinib rheumatoid arthritis (RA) trial may be eligible to enter this trial.

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Protocol I4V-MC-JADY
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 26-July-2013
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 20-Aug-2015
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly on date provided below.

Approval Date: 01-Oct-2020 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Table of Contents

Appendix **Page**

3. Revision History

Statistical Analysis Plan (SAP) Version 1 was approved prior to unblinding the Phase 3 originating studies, on 26-July-2013.

Statistical Analysis Plan Version 2 was approved prior to unblinding of the step-down phase. The overall changes and rationale for the changes incorporated in Version 2 are as follows:

- Incorporated updates from Protocol Amendment (e)
- Updated to use Integrated Efficacy Analysis Plan (IEAP) Version 4 and Program Safety Analysis Plan (PSAP) Version 4 as reference documents
- Updated analyses to refer to interim efficacy and safety analyses and a final analysis
- Updated to include analyses performed in interim analysis

Statistical Analysis Plan Version 3 was approved prior to the final database lock. The overall changes and rationale for the changes incorporated in Version 3 are as follows:

Section	Amendment or update	Reason
Across sections	Removed the language related to the integrated safety analyses that will be covered in a separate SAP; Removed the language on the interim analyses that were completed and designed to support the initial bari RA submission.	Need to clarify the analyses that are needed and appropriate for the final JADY Clinical Study Report
5	Updated the study design diagram	Incorporated updates from the latest Protocol Amendment (i)
6.1.1	Updated the analysis populations	Added definition for the step-down population and updated the safety population to those who didn't discontinue the study for the reason 'lost to follow-up' at the first postbaseline visit in Study JADY
6.1.2	Removed language on interim analysis and clarified baseline definition for the final analyses	Clarify the baseline definition for different efficacy analysis sets
6.3.1.1	Clarified that the missing data imputation will be applied for the efficacy and health outcomes endpoints in stepdown analyses and structural analyses	Due to lack of randomization in Study JADY, there are no treatment comparisons planned for the general analyses except for the step-down substudy and structural analyses. So, the missing data imputation will only be applied to these analyses
6.10	Clarified the treatment compliance will be calculated for the entire study period	Provided more details
6.11	Dropped Hybrid American College of Rheumatology Response Measure	This efficacy measure is not commonly used any more and, therefore, is not of interest for this study

Section	Amendment or update	Reason
6.14	Updated analysis period to include both the treatment period and off-drug follow-up time	To be consistent with the analyses performed in the previous RA submissions
6.14	Clarified the safety analyses needed for the JADY CSR and moved the integrated safety analyses to a different analysis plan	Updated to provide clarity on what is planned for analysis of JADY data alone. Safety data are more thoroughly assessed in the context of integrated data, in which the data from originating studies are included. Having multiple safety assessments on JADY alone without the data from the originated studies does not correspond with any safety question of interest or estimate of interest. The integrated data will be presented in safety updates.
6.15	Added Clinical Trial Registry required Analyses	New information
6.16	Added language for patients impacted by COVID-19	New information

Abbreviations: bari = baricitinib; COVID-19 = coronavirus disease 2019; CSR = Clinical Study Report; RA = rheumatoid arthritis; SAP = statistical analysis plan.

4. Study Objectives

The following objectives are replicated from Amendment (i) of Study I4V-MC-JADY (JADY) protocol, dated 10-Dec-2019.

4.1. Primary Objective

The primary objective of the study is to evaluate the long-term safety and tolerability of baricitinib. Safety and tolerability assessments will include:

- Treatment-emergent adverse events (TEAEs), adverse events of special interest (AESIs), and serious adverse events (SAEs)
- Temporary investigational product (IP) interruptions and permanent IP discontinuations
- Vital signs and laboratory evaluations (including chemistry and hematology)

To evaluate the long-term safety and tolerability of baricitinib, the clinical study report (CSR) will include the following analyses explained in further detail in Section 6.13:

- listing of SAEs
- listing of adverse events (AEs) leading to permanent discontinuation of IP
- summary of SAEs, and
- summary of AEs leading to permanent discontinuation of IP.

Rationale for not including integrated safety data

Safety data are more thoroughly assessed in the context of integrated data, in which the data from originating studies are included. Having safety assessments on JADY alone without the data from the originated studies does not correspond with safety questions of interest or estimates of interest. Including integrated data within the JADY CSR was deemed inappropriate.

Integrated safety data arising from Study JADY will be presented with appropriate context elsewhere. More details regarding the analysis of the integrated safety data will be presented in the integrated safety analysis plan (ISAP) or other integrated planning documents.

4.2. Secondary Objectives

The secondary objectives are to evaluate the effect of long-term administration of baricitinib in patients initially randomized to receive baricitinib in the originating study as follows:

- Proportion of patients who maintain an American College of Rheumatology (ACR)20, ACR50, and ACR70 through each 12 months of treatment

- Proportion of patients who maintain a Disease Activity Score modified to include the 28 diarthrodial joint count (DAS28)-high-sensitivity C-reactive protein (hsCRP) ≤3.2, DAS28-erythrocyte sedimentation rate (ESR) ≤3.2, DAS28-hsCRP <2.6, and DAS28-ESR <2.6; Clinical Disease Activity Index (CDAI) ≤10 and CDAI ≤2.8; Simplified Disease Activity Index (SDAI) ≤11 and SDAI ≤3.3; ACR/European Union League Against Rheumatism (EULAR) remission according to the Boolean-based definition; and Health Assessment Questionnaire-Disability Index (HAQ-DI) improvement of ≥0.22 and ≥0.3 from Month 6 (of the originating study) through each 12 months of treatment
- Change from baseline of originating study through each 12 months of treatment up to 5 years in structural joint damage as measured by modified Total Sharp Score (mTSS) (van der Heijde method)
- Proportion of patients with mTSS change ≤0 from baseline of originating study through each 12 months of treatment up to 5 years
- Change from baseline of originating study through each 12 months of treatment up to 5 years in Joint Space Narrowing (JSN) and bone erosion score (ES)
- Change from baseline of originating study in duration of morning stiffness through each 12 months of treatment
- Change from baseline in European Quality of Life-5 Dimensions-5 Level (EQ-5D-5L) scores through each 12 months of treatment
- Evaluation of healthcare resource utilization through 12 months of treatment
- To determine if treatment with baricitinib 2 mg once daily (QD) maintains the low disease activity level achieved with the 4-mg QD dose (ie, step-down dosing) on the following outcomes:
    - Proportion of patients who maintain a CDAI score of ≤10 from Studies I4V-MC-JADV (JADV), I4V-MC-JADW (JADW), and I4V-MC-JADX (JADX) after 3 months of treatment with baricitinib 2 mg QD and with patients continuing treatment with the 4-mg QD dose
    - Time to relapse (where relapse is defined as a CDAI score >10 from Studies JADV, JADW, and JADX) after randomization to the baricitinib 2-mg and 4-mg QD doses

Similar analyses will be conducted on patients who initiated treatment with baricitinib as rescue therapy at some point during the originating study.

4.3. Exploratory Objectives

The exploratory objectives are to:

- To determine if treatment with baricitinib 2 mg QD maintains the low disease activity level or remission achieved with the 4-mg QD dose (ie, step-down dosing) on the following outcomes:

  - Proportion of patients who maintain a CDAI score of ≤2.8 from Study I4V-MC-JADZ (JADZ) after 3 months of treatment with baricitinib 2 mg QD and with patients continuing treatment with the 4-mg QD dose
  - Time to relapse after randomization to the baricitinib 2-mg and 4-mg QD doses where relapse is defined as:
    - a CDAI score >2.8 from Study JADZ
    - loss of CDAI categorization at randomization (a CDAI score >2.8 if CDAI score at randomization is ≤2.8 or a CDAI score >10 if CDAI score at randomization is >2.8 and ≤10)
    - having been rescued

- To describe the clinical course of patients initiating baricitinib at the time of enrollment in Study JADY as assessed by the DAS28-hsCRP scale, the DAS28-ESR scale, and CDAI at 3, 6, 12, and at each 12 months of baricitinib treatment by initial treatment allocation in the originating study
  - From placebo in Study JADW or Study JADX to baricitinib in Study JADY
  - From methotrexate (MTX) in Study JADZ to baricitinib in Study JADY
  - From adalimumab in Study JADV to baricitinib in Study JADY
- To describe the clinical course of patients switching from MTX + baricitinib in Study JADZ to baricitinib monotherapy in Study JADY as assessed by DAS28-hsCRP, DAS28-ESR, and CDAI at 3, 6, 12, and at each 12 months of treatment.

5. Study Design

5.1. Summary of Study Design

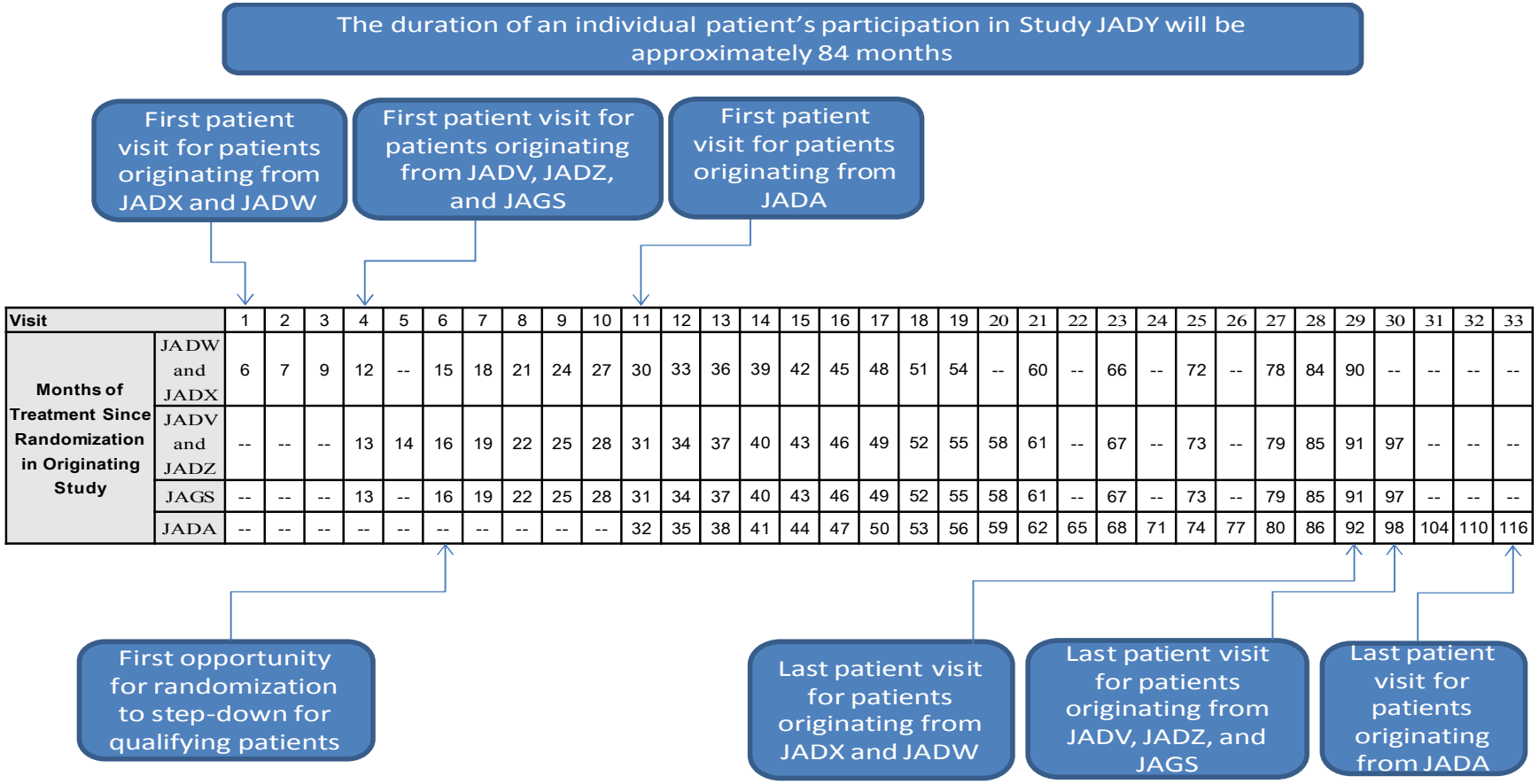
Study JADY is a Phase 3, multicenter, long-term extension study evaluating the safety and efficacy of baricitinib (2 mg QD and 4 mg QD) in patients with rheumatoid arthritis (RA). Patients who completed an originating study (Studies JADA, JADZ, JADV, JADX, JADW, or I4V-MC-JAGS [JAGS]) may be eligible for enrollment into Study JADY. Planned enrollment will be approximately 2400 to 3500 patients.

Study JADY will consist of 3 parts:

- Screening: screening should occur during the last visit of the originating study. However, in particular circumstances, this duration may be extended after consultation with the sponsor.
- Part A: treatment period lasting up to 84 months from enrollment into Study JADY.
- Part B: posttreatment follow-up period of 28 days.

Patients may continue to receive the background non-investigational, open-label conventional disease-modifying antirheumatic drugs (cDMARDs), nonsteroidal anti-inflammatory drugs (NSAIDs), corticosteroids, and other analgesic therapies they were receiving at completion of the originating study. Patients with renal impairment at baseline of the originating study will not receive doses higher than 2 mg baricitinib QD.

Figure JADY.5.1 illustrates the study design.



Visit		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Months of Treatment Since Randomization in Originating Study	JADW and JADX	6	7	9	12	--	15	18	21	24	27	30	33	36	39	42	45	48	51	54	--	60	--	66	--	72	--	78	84	90	--	--	--	--
	JADV and JADZ	--	--	--	13	14	16	19	22	25	28	31	34	37	40	43	46	49	52	55	58	61	--	67	--	73	--	79	85	91	97	--	--	--
	JAGS	--	--	--	13	--	16	19	22	25	28	31	34	37	40	43	46	49	52	55	58	61	--	67	--	73	--	79	85	91	97	--	--	--
	JADA	--	--	--	--	--	--	--	--	--	--	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74	77	80	86	92	98	104	110	116

Figure JADY.5.1. Illustration of study design for Clinical Protocol I4V-MC-JADY – patients enrolled after completing Studies JADA, JADV, JADZ, JADX, JADW, or JAGS.

5.1.1. Screening Period

Patients will be screened for study eligibility during the last visit of the originating study, and eligible patients will continue to Part A of Study JADY.

As Studies JADA, JADV, JADW, JADX, JADZ, and JAGS vary in duration, patients enrolling in Study JADY will enter the study at different visits. Timing of assessments in Study JADY is described in terms of 28-day months of treatment since randomization in the originating study. This study construct allows the activities required at each visit to be the same for all patients enrolled in Study JADY regardless of the originating study.

- Patients completing Studies JADX and JADW will have had approximately 6 months of treatment with IP (baricitinib or placebo) and will enter Study JADY at Visit 1.
- Patients completing Studies JADV, JADZ, and JAGS will have had approximately 12 months of treatment with IP (baricitinib, adalimumab [JADV], or MTX [JADZ]) and will enter Study JADY at Visit 4.
- Patients completing Study JADA will have had approximately 30 months of treatment with baricitinib and will enter Study JADY at Visit 11.

5.1.2. Part A: Seven-Year Treatment Period

Studies JADA, JADV, JADW, JADX, JADZ, and JAGS vary by whether patients are receiving open-label baricitinib or blinded IP at the time of study completion. Therefore, treatment assignments and blinding in Study JADY will vary depending on the originating study. In all cases, however, patients and investigators will remain blind to their original (initial) blinded treatment assignment in the originating study.

- Patients completing Study JADA will have been receiving open-label baricitinib 4 mg QD and will receive open-label baricitinib 4 mg QD in Study JADY.
- Patients enrolled in Study JADW and Study JADX were initially randomized to baricitinib 2 mg QD, 4 mg QD, or placebo and had the option of rescue therapy beginning at Week 16. Patients completing these studies without requiring rescue therapy are still blinded to treatment allocation. To maintain the study blinds, these patients will receive blinded therapy in Study JADY. Patients who had been receiving baricitinib 2 mg QD will receive the 2-mg dose in Study JADY. Patients who had been receiving either baricitinib 4 mg QD or placebo will receive the 4-mg dose. Patients who received rescue therapy during Study JADW and Study JADX will be receiving baricitinib 4 mg QD at study completion. These patients will continue to receive baricitinib 4 mg QD in Study JADY. Patients with renal impairment will receive baricitinib 2 mg QD.

- Patients enrolled in Study JADV were initially randomized to baricitinib 4 mg QD, adalimumab 40 mg subcutaneous (SC) every 2 weeks, or placebo (both placebo tablet and SC injection). Patients had the option of rescue therapy beginning at Week 16. Patients initially randomized to placebo received baricitinib 4 mg QD beginning at Week 24. Patients completing the study without rescue therapy are still blinded to treatment allocation. To maintain the study blind and to avoid administering a SC placebo injection for an extended period, these patients will receive baricitinib 4 mg QD, and adalimumab IP (ie, adalimumab and matching adalimumab placebo) will be discontinued. Patients who had received rescue therapy during Study JADV will be receiving open-label 4 mg QD baricitinib at study completion. These patients will continue to receive baricitinib 4 mg QD in Study JADY. Patients with renal impairment will receive baricitinib 2 mg QD.
- Patients enrolled in Study JADZ were initially randomized to baricitinib 4 mg QD, MTX monotherapy, or baricitinib plus MTX. Patients had the option of rescue therapy beginning at Week 24. Patients completing the study without rescue therapy are still blinded to treatment allocation. To maintain the study blind, these patients will receive baricitinib 4 mg QD and MTX IP (ie, MTX or matching MTX placebo) will be discontinued. Patients who had received rescue therapy during Study JADZ will be receiving open-label baricitinib 4 mg QD and open-label MTX at study completion. These patients will continue to receive baricitinib 4 mg QD, and open-label MTX will be discontinued. Patients with renal impairment will receive 2 mg baricitinib QD.
- Patients enrolled in Study JAGS were initially randomized to baricitinib 4 mg QD or placebo. At Week 24, all patients entered the open-label period and received a dose of baricitinib 4 mg QD. Patients will continue to receive baricitinib 4 mg QD in Study JADY. Patients with renal impairment will receive baricitinib 2 mg QD.

5.1.2.1. Dose Titration

An exploratory objective of this study is to evaluate the effectiveness of a reduced dose of baricitinib (step-down from baricitinib 4 mg QD to baricitinib 2 mg QD) in patients who achieve a sustained (at least 3 months in Study JADY) low disease activity level (CDAI score ≤10; for patients originating in Studies JADV, JADW, JADX, and JAGS) or a sustained remission (CDAI score ≤2.8; for patients originating in Study JADZ). For patients with early RA originating in Study JADZ, more stringent eligibility criteria for step-down dosing are applied, as the goal of therapy in these patients is to achieve remission. For patients originating in Studies JADV, JADW, JADX, and JAGS, less stringent criteria are applied as a more reasonable goal of treatment would be attainment of low disease activity. Patients achieving these disease activity criteria will be randomized following a 1:1 ratio allocation to continue baricitinib 4 mg QD or receive the 2-mg QD dose in a blinded fashion. Patients eligible for randomization to step-down must have received at least 15 months of treatment with baricitinib 4 mg QD and have not received rescue therapy in the originating study or JADY. Fifteen months of treatment with baricitinib prior to step-down was selected to allow adequate study duration (>1 year) on stable doses of baricitinib 4 mg QD to thoroughly assess the benefit/risk profile of the dose regimen, including assessment of key endpoints such as radiographic progression of structural joint damage at 12 months. Allowing step-down after a minimum of 15 months treatment with baricitinib also ensures that patients have been on stable doses of baricitinib for at least 3 months

in the context of Study JADY prior to step-down. Both the investigator and patient will be blinded to treatment if or when randomization to step-down dosing occurs. Patients from Study JADA are not eligible for participation in step-down dosing.

If a patient experiences worsening of disease symptoms (either CDAI score >10 or CDAI score >2.8, depending on originating study) following step-down, a change in analgesic/NSAID dosing or the addition of an analgesics/NSAIDs may be considered to manage transient flares. An unscheduled study visit may be needed to assess response to NSAIDs/analgesics and need for rescue therapy. If the patient fails to achieve low disease activity or remission, they may be returned to baricitinib 4 mg QD, and an alteration in cDMARD or corticosteroid therapy may be considered. Patients will be eligible for step-down dosing only once. If after rescue therapy and adjustment of concomitant medications, in the opinion of the investigator, the patient is unable to derive adequate benefit (eg, achieving a CDAI score ≤10 [patients originating in Studies JADV, JADX, JADW, and JAGS] or CDAI score ≤2.8 [patients originating in Study JADZ] from treatment with baricitinib over an appropriate time period (eg, 3 months), the patient should be discontinued from the study.

5.1.3. Part B: Posttreatment Follow-up Period

Patients who discontinue the study or patients who complete Part A of the study will have a Follow-up Visit (Visit 801) approximately 28 days after the last dose of study drug.

Infrequently, patient and investigator availability may be such that the Early Termination Visit and the Follow-up Visit may occur on or about the same date. In this instance, the visits may be combined and should occur approximately 28 days after the last dose of study drug. All activities required for the Early Termination Visit should be completed according to the protocol study schedule.

5.2. Determination of Sample Size

No formal sample size calculation was required for this extension study. All patients who complete any of the originating studies (Studies JADA, JADZ, JADV, JADX, JADW or JAGS) may be eligible for enrollment into Study JADY. It is expected that 80% of patients will complete Studies JADV, JADZ, JADX, JADW, JADA, or JAGS; therefore, Study JADY will enroll approximately 2400 to 3500 patients.

5.3. Method of Treatment Assignment

In this extension study, patients who meet all criteria for enrollment and who have completed Studies JADV, JADW, JADX, JADZ, or JAGS will be assigned to baricitinib treatment of 4 mg QD or 2 mg QD. Patients receiving baricitinib during the originating study will continue on the baricitinib dose administered at the end of that study. Patients who meet all criteria for enrollment and who have completed Study JADA will be assigned to receive open-label baricitinib 4 mg QD for the duration of Study JADY. Patients with renal impairment in Studies JADW, JADX, JADV, JADZ, and JAGS will receive baricitinib 2 mg QD.

Patients eligible for step-down will be assigned to step-down treatment by a computer-generated random sequence using an interactive voice-response system (IVRS) or interactive web-response system (IWRS). Randomization in the step-down phase to baricitinib 2 mg QD or 4 mg QD will follow a 1:1 ratio allocation and will be stratified by region and originating study. The IVRS will be used to assign packages containing double-blind IP to each patient. Site personnel will confirm that they have located the correct packages by entering a confirmation number found on the packages into the IVRS before dispensing to the patient.

This study will be conducted internationally at many sites. Region in Study JADY will be defined according to the geographical region defined in the originating studies except for 2 cases: all European countries will be treated as one region called Europe, and Asia will include Japan. Therefore, regions will be categorized into 5 categories for the randomization in the step-down phase (see Table JADY.5.1).

Table JADY.5.1. Countries and their Geographical Regions

Geographical Region	Country or Countries
USA and Canada	United States (including Puerto Rico), Canada
Central and South America and Mexico	Argentina, Brazil, Mexico
Asia	China, S Korea, Taiwan, Japan
Europe	Austria, Belgium, Croatia, Czech Republic, Denmark, France, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, UK
Rest of World	Australia, India, Israel, Russia, South Africa, Turkey

5.3.1. Treatment Administered

Patients in this long-term extension study will receive baricitinib doses of 2 mg or 4 mg administered orally QD according to the study they completed before entering Study JADY and as shown in Table JADY.5.2. Patients with renal impairment who received a baricitinib dose of 2 mg QD in the originating study will continue to receive a 2 mg QD dose of baricitinib in Study JADY.

Table JADY.5.2. Treatment Regimens

Originating Study	Treatment at end of Originating Study	Treatment at Start of Study JADY
JADV (12-month duration)	Baricitinib 4 mg QD	Baricitinib 4 mg QD
	Adalimumab 40 mg SC biweekly	
JADZ (12-month duration)	Baricitinib 4 mg QD + MTX weekly	Baricitinib 4 mg QD
	Baricitinib 4 mg QD monotherapy	
	MTX monotherapy weekly	
JADX (6-month duration)	Baricitinib 4 mg QD	Baricitinib 4 mg QD
	Baricitinib 2 mg QD	Baricitinib 2 mg QD
	Placebo QD	Baricitinib 4 mg QD
JADW (6-month duration)	Baricitinib 4 mg QD	Baricitinib 4 mg QD
	Baricitinib 2 mg QD	Baricitinib 2 mg QD
	Placebo QD	Baricitinib 4 mg QD
JADA (30-month duration)	Baricitinib 4 mg QD	Baricitinib 4 mg QD
JAGS (12-month duration)	Baricitinib 4 mg QD	Baricitinib 4 mg QD

Abbreviations: MTX = methotrexate; QD = once daily.

The investigator or his/her designee will be responsible for explaining the correct use of the IP to the patient, verifying that instructions are followed properly, maintaining accurate records of IP dispensing and collection, and returning all unused medication to Lilly or its designee at the end of the study.

In some cases, sites may destroy the material if, during the investigator site selection, the evaluator has verified and documented that the site has appropriate facilities and written procedures to dispose of clinical trial (CT) materials.

Patients will be instructed to contact the investigator as soon as possible if they have a complaint or problem with the IP so that the situation can be assessed.

5.3.2. Selection and Time of Doses

Each patient will receive IP beginning on the first visit of Study JADY. Baricitinib dosing will continue daily for 84 months. Oral IP should be taken at home at approximately the same time each day.

5.3.3. Rescue Therapy

For patients originating in Studies JADA, JADV, JADW, JADX, and JAGS, rescue therapy may be provided to any patient who has a CDAI score >10 at or after 3 months following enrollment into Study JADY. Patients receiving baricitinib 2 mg QD and who have normal renal function (defined as estimated glomerular filtration rate [eGFR] ≥60 mL/min/1.73 m^2) will be rescued to baricitinib 4 mg QD. Patients receiving baricitinib 4 mg QD and patients receiving a baricitinib 2-mg QD dose due to renal impairment will continue to receive these doses. Addition or increase in dose of cDMARDs may occur only after rescue.

Furthermore, addition or increase in dose of cDMARDs may occur only after rescue.

For patients originating in Study JADZ, MTX or other cDMARDs may be prescribed at the discretion of the investigator and according to local clinical practice at any time point during Study JADY (including prior to 3 months since enrollment).

6. A Priori Statistical Methods

6.1. General Considerations

Study JADY was terminated early as it met the study objectives, and the projected exposure to baricitinib that would accrue during the remaining study period would not add significant scientific value. The analyses included in this plan are mainly for the final CSR of Study JADY.

This plan describes *a priori* statistical analyses for efficacy, health outcomes, and safety for all patients who participated in Study JADY. Study JAGS may have not been listed in the secondary study objectives due to its data availability at the initial submission time, but it will be included in the final analysis of Study JADY.

The statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly).

In general, the collected data will be presented using summary tables by study of origin for efficacy and health outcomes analyses and pooled studies of origin for the safety analyses. Since no randomization occurred at the study entry and all patients took baricitinib during Study JADY, no hypothesis tests for difference between treatment groups will be performed, and all analyses will be descriptive unless otherwise specified.

Continuous data will be summarized using the sample size, mean, median, standard deviation (SD), standard error of the mean, 1st quartile, 3rd quartile, and range (minimum and maximum). Categorical data will be summarized in terms of sample size, frequency count, and percentage. Graphical presentations may be used to support the presentation of selected data using, but not limited to, Kaplan-Meier time-to-event analyses (eg, time to relapse).

When statistical analyses are performed, a 2-sided test at a significance level of 0.05 will be used. Unless otherwise stated, for categorical variables, the 95% confidence interval (CI) of treatment difference in response rate will be calculated using the Newcombe-Wilson method (Newcombe 1998) without continuity correction. Fisher's exact test will be used for the between-treatment comparisons unless otherwise specified. For continuous variables, within-treatment and between-treatment changes from baseline along with the 95% CI will be analyzed using the t-test.

All p-values will be rounded up to 3 decimal places. For example, any p-value strictly >0.049 and ≤0.05 will be displayed as 0.050. This guarantees that on any printed statistical output, the unrounded p-value will always be less than or equal to the displayed p-value. A displayed p-value of 0.001 will always be understood to mean ≤0.001. Likewise, any p-value displayed as 1.000 will be understood to mean >0.999 and ≤1.

Last visit data from the originating studies will be used as first visit data for Study JADY. For by-visit summaries, only visits in which a measure was scheduled to be collected will be summarized. Unscheduled visit data will be included within the patient-level listings.

6.1.1. Analysis Populations

6.1.1.1. Modified Intent-to-treat (mITT) Population

This analysis set will include all enrolled patients treated with at least 1 dose of the IP in Study JADY.

Analyses of structural progression will be conducted on a subset of the JADY modified intent-to-treat (mITT) population consisting of patients with available baseline and at least 1 postbaseline x-ray assessment that is collected after 2 years since the randomization in the originating studies.

Analyses during the step-down period will be conducted on a subset of the JADY mITT population who have been randomized and took at least 1 dose of the IP during the step-down period.

In addition, a subset of the JADY mITT population who have been rescued in the originating studies will be combined and analyzed separately.

6.1.1.2. Safety Population

This analysis set is defined as all enrolled patients who received at least 1 dose of the IP in the study JADY and did not discontinue the study for the reason 'Lost to Follow-up' at the first postbaseline visit in study JADY.

6.1.2. Definition of Baseline

The following analyses will be performed for patients in Study JADY:

- General efficacy analyses during Study JADY alone will include
    - maintenance of response observed at the beginning of Study JADY, and
    - change from the originating study baseline.
- Step-down analyses will include data from the randomized step-down phase in Study JADY.

In the efficacy analyses of maintenance of response, the baseline value will be defined as the data collected at the first visit in Study JADY. JADY baseline, in most cases, is the value obtained at the last visit in the originating study. However, the following endpoints use the baseline of the originating study:

- ACR20, ACR50, and ACR70
- HAQ-DI improvement ≥0.22
- HAQ-DI improvement ≥0.3.

In the step-down analyses, the baseline value will be defined as the data collected at the first dose date of rerandomized study drug in the step-down phase (step-down baseline).

6.1.3. Derived Data

- Age (year) = (Date of informed consent – (year of birth, July, 1) +1) / 365.25, and then truncated to a whole-year (integer) age
- Age group (<65 years old, ≥65 years old)
- Age group (<65 years old, 65-<75 years old, 75-<85 and ≥85)
- BMI (kg/m2) = Weight (kg)/((Height (cm)/100)2)
- Change from baseline = postbaseline measurement at Visit x – baseline measurement
- Percent change from baseline at Visit x:
    - ((Postbaseline measurement at Visit x – baseline measurement)/baseline measurement)*100.
- Weight (kg) = weight (lbs) * 0.454.
- Height (cm) = height (in) * 2.54.
- Baseline disease activity based on the DAS28 (calculated separately for DAS28-ESR and DAS28-hsCRP)
    - Low disease activity: DAS28 ≤3.2,
    - Moderate disease activity: DAS28 >3.2 to ≤5.1,
    - High disease activity: DAS28 >5.1
- CDAI, SDAI, and EULAR Responder Index categories along with other efficacy and health outcomes questionnaires scoring are included in Appendix 2.

6.2. Adjustment for Covariates

No adjustment of covariates are employed unless otherwise specified.

6.3. Handling of Dropouts or Missing Data

In general, the efficacy and health outcomes endpoints in Study JADY will be summarized descriptively using observed data. No missing data imputation will be applied.

However, for a subset of analyses, including structural joint damage and step-down analyses, where treatment comparison and statistical inferences will be performed, the missing data imputation will be applied. In accordance with precedent set with other Phase 3 RA trials (Keystone et al. 2004, 2008, 2009; Cohen et al. 2006; Smolen et al. 2008, 2009), the following methods for imputation of missing data will be used accordingly.

6.3.1. Non-responder Imputation (NRI)

Analysis of ordinal categorical efficacy measures (including responder/non-responder) will be assessed using a non-responder imputation (NRI) method. Patients will be considered as non-responders for the NRI analysis if they

- do not meet the clinical response criteria at an analysis time point,
- have discontinued the study or permanently discontinued from study drug for any reason,
- have received rescue treatment at any time prior to the analysis time point per IWRS records, or
- have missing data that are not subject to any other imputation method at the analysis time point.

6.3.2. Modified Last Observation Carried Forward (mLOCF)

The modified last observation carried forward (mLOCF) method will be a general approach to impute missing data for continuous measures when single time point analysis methods are utilized. For patients who receive rescue therapy, the last nonmissing postbaseline observation at or before rescue will be carried forward to subsequent time points for evaluation. For all other patients who discontinue from the study or permanently discontinue from study drug for any reason, the last nonmissing postbaseline observation before discontinuation will be carried forward to subsequent time points for evaluation.

6.3.3. Linear Extrapolation (LE)

The linear extrapolation (LE) method will be used to impute missing data for analysis of the joint structural damage progression endpoints (mTSS, ES, and JSN). For patients who discontinue the study or patients who miss a radiograph for any reason, the most recent two radiographic data prior to discontinuation or the missed radiograph, adjusted for time, will be used for LE to impute missing data at subsequent time points for the individual patient. This imputation will only be performed when both of these conditions are met:

- The timing between all three time points are within a 2-year time-frame.
- The timing between the last non-missing data time point and the imputed time point is <1 year.

The LE method has been established as an appropriate missing data imputation method in other Phase 3 RA trials (Keystone et al. 2004, 2008, 2009; Cohen et al. 2006; Smolen et al. 2008, 2009).

Methods for between-reader adjudication, score derivations, and other details for the bone ES, JSN scores and mTSS are provided in the BioClinica imaging charter and in Appendix 2 and Appendix 3.

6.4. Multiple Comparisons

No adjustments for multiple comparisons will be utilized in the statistical analyses for this study.

6.5. Patient Disposition

Patient disposition will be summarized using the JADY mITT population. Frequency counts and percentages of all patients who are enrolled, discontinued, or ongoing will be presented. Frequency counts and percentages of patients will also be summarized separately by treatment

group for patients who participated in the step-down phase and for patients who did not participate in the step-down phase, as well as for patients who are rescued and for patients who are not rescued, along with their reasons for study discontinuation and permanent study drug discontinuation.

A listing of patient disposition will be provided for the JADY mITT population, with the extent of their participation in the study and the reason for discontinuation. A listing of randomization at step-down will also be provided.

6.6. Protocol Deviation

Protocol deviations will be tracked by the clinical team, and the importance will be assessed by key team members during the protocol deviation review meetings.

Potential examples of deviations include patients who receive excluded concomitant antirheumatic therapy, significant non-compliance with study medication (<80% of assigned doses taken, including failure to take study medication and taking incorrect study medication), patients incorrectly enrolled in the study, and patients whose data are questionable because of significant site quality or compliance issues.

Important protocol deviations are the following:

- Entry criteria
  - Are unable to read, understand and give written informed consent
  - Had IP permanently discontinued during a previous baricitinib study
- Use of prohibited concomitant medication during study
  - New or increased dose of cDMARD within the first 3 months of the study (except for patients from Study JADZ)
  - Use of any biologic DMARD
- Study drug administration and compliance
  - Drug dispensing error (eg, patient dispensed wrong package and received incorrect dose or drug)
  - Patient received drug that was declared “not Fit for Use”
  - Use of expired CT material
  - Significant non-compliance as defined per protocol
- Protocol non-compliance
  - Patient met temporary or permanent IP discontinuation criteria and IP not discontinued per protocol
  - Inadvertent unblinding
  - Fraud

- Other significant Good Clinical Practice (GCP) issues

A summary of the number and percentage of patients with an important protocol deviation by type of deviation will be provided. Individual patient listings of important protocol deviations will be provided.

In addition, information of patients with important protocol deviations due to coronavirus disease 2019 (COVID-19) will be provided.

6.7. Patient Characteristics

The following baseline demographics, RA clinical characteristics, and RA history will be summarized descriptively for the JADY mITT population but will not be subset by treatment group or originating study. For the step-down patients, data will be summarized descriptively by treatment groups and will be presented separately for patients originating from Study JADZ and from Studies JADV, JAGS, JADX, and JADW combined. Unless otherwise specified, the JADY study baseline and step-down baseline will be used for the JADY mITT population and the step-down patients, respectively. The summaries will include descriptive statistics such as the number of patients (n), mean, SD, minimum, 1st quartile, median, 3rd quartile, and maximum for continuous measures, and frequency counts and percentages for categorical measures. No formal statistical comparisons will be made.

- Age
- Age group (<65 vs ≥65; <75 vs ≥75; <85 vs ≥85)
- Age group (<65, 65 to <75, 75 to <85 and ≥85)
- Gender (female, male)
- Race (American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian or other Pacific Islander, White, and multiple)
- Ethnicity (for patients from United States/Puerto Rico only: Hispanic or Latino, Non-Hispanic and non-Latino)
- Region (USA and Canada [including Puerto Rico], Central and South America and Mexico, Asia [excluding Japan], Japan, European Union, and all remaining countries pooled together [Rest of World]).
- Tender joint count (TJC) based on 28 joints
- Swollen joint count (SJC) based on 28 joints
- TJC based on 68 joints
- SJC based on 66 joints
- HAQ-DI
- Physician's Global Assessment of Disease Activity (0-100 mm)
- Patient's Global Assessment of Disease Activity (0-100 mm)

- Patient's assessment of pain (0-100 mm)
- hsCRP (mg/L)
- DAS28-hsCRP
- DAS28-hsCRP category (≤3.2, >3.2 to ≤5.1, >5.1)
- DAS28-ESR
- DAS28-ESR category (≤3.2, >3.2 to ≤5.1, >5.1)
- CDAI score
- SDAI score
- ESR (mm/hr)
- Rheumatoid factor (RF)
- RF status (positive, negative)
- Anti-citrullinated peptide antibody (ACPA)
- ACPA status (positive, negative)
- Seropositivity status (RF negative/ACPA negative, RF positive/ACPA negative, RF negative/ACPA positive, RF positive/ACPA positive)
- Time from symptom onset of RA (years)
- Time from symptom onset of RA category (<1 year, ≥1 to <5 years, ≥5 years)
- Time from RA diagnosis (years)
- Current use of corticosteroid (yes, no)
- Daily dose of corticosteroid (mg/day)
- MTX average weekly dose (mg/week)
- Number of cDMARDs previously used (0, 1, 2, ≥3) (based on the baseline in the originating studies)
- cDMARDs previously used (based on the baseline in the originating studies)
- Number of cDMARDs currently used (0, 1, 2, ≥3)
- Type of cDMARDs currently used (MTX only, MTX + 1 other cDMARD, 1 non-MTX cDMARD only, 2 non-MTX cDMARDs only, 3 or more cDMARDs).

In the step-down summary, the RF and ACPA will be based on the last observation collected before or at randomization in the step-down phase.

In addition, the following data will be summarized for the analysis sets, which include the Study JADW subset. The previous biologic disease-modifying anti-rheumatic drugs (bDMARDs) are based on the baseline in the originating JADW study.

- Number of previous bDMARDs used: (0, 1, 2, ≥3)
- bDMARDs previously used: adalimumab, etanercept, infliximab, etc.
- Number of previous tumor necrosis factor inhibitors (TNFi) used (0, 1, 2, ≥3)
- Number of previous non-TNFi used (0, 1, 2, ≥3)

6.8. Concomitant Medication

Concomitant therapy is defined as the therapy that starts before, on, or after the first dose of study treatment and continues into the treatment period. Concomitant therapy used during Study JADY will be summarized using frequency counts and percentages by World Health Organization (WHO) drug Anatomical Therapeutic Chemical (ATC) classification.

The summaries for the following categories will be provided for the safety population with treatment groups and originating studies combined:

- concomitant medications used for RA
- concomitant medications – statins
- concomitant medications excluding RA and statin therapies

6.9. Historical Illnesses and Preexisting Conditions

Since historical illnesses and preexisting conditions are described in the originating studies, no summary will be created for Study JADY.

6.10. Treatment Compliance

Patient compliance with baricitinib will be assessed from Visit 2 of Study JADY to the scheduled last visit or the Early Termination Visit during the treatment period (up to but not including the Posttreatment Follow-up Period, Part B). Summaries and listings of compliance that are calculated for the entire study period will be presented.

Patients are instructed to take between one to two tablets daily depending on the originating study; treatment assigned and rescue status in order to maintain the blind of the originating study and to allow blinded step-down in Study JADY. Patients originating from Study JADA will receive one open-label bottle and take one 4-mg tablet for the duration of Study JADY. Patients originating from Studies JADV, JADZ, JADX, JADW, or JAGS and who were not previously rescued will receive either a 4-mg or 2-mg matching placebo tablet to maintain the blind of the step-down. Patients who were rescued in the originating study or those with renal impairment will receive matching placebo tablets through Visit 5 in Study JADY, after which they will receive open-label treatment and only take one tablet. Patients who were rescued in Study JADY will receive open-label treatment after rescue and only take one tablet. A patient is considered noncompliant if he or she misses >20% of the prescribed doses during the study, unless the patient's study drug is withheld by the investigator.

For patients who had their treatment temporarily interrupted by the investigator, the period of time that the dose was withheld will be considered in the compliance calculation.

Compliance will be calculated as the difference in the number of tablets dispensed and returned divided by the expected number of tablets to be administered for the patient's treatment regimen minus any tablets not required to be administered due to study drug being temporarily discontinued by the investigator, as follows:

$$\frac{\text{actual total \# of tablets used}}{\text{expected total \# of tablets used}} * 100$$

Where:

Actual total # of tablets used = total number dispensed – total number returned in the period of interest,

*Expected total # tablets used = 2 * number of days in the period of interest where patient takes 2 tablets + 1 * number of days in the period of interest where patient takes 1 tablet.*

If a patient had his/her treatment temporarily interrupted by the investigator during the period, the number of days that drug was withheld will be deducted from the total number of days in computing the expected total number of tablets used.

6.11. Efficacy Analyses

The efficacy analyses will be performed using the mITT Population defined in Section 6.1.1.1. Refer to Appendix 2 for a detailed description of the efficacy and health outcomes measures. For patients who were rescued in the originating studies, similar efficacy analyses will be performed as appropriate.

6.11.1. ACR20, ACR50, and ACR70 Indices

The proportion of patients who maintain an ACR20, ACR50, and ACR70 response during the JADY study based on the change in ACR core set measures (see Appendix 2) from the baseline of the originating study will be summarized for the JADY mITT population. The summary will be presented by treatment groups and originating studies.

Refer to Appendix 1 for the algorithm to calculate ACR response.

6.11.2. Disease Activity Score-Erythrocyte Sedimentation Rate (DAS28-ESR) and Disease Activity Score-High-Sensitivity C-Reactive Protein (DAS28-hsCRP)

The proportion of patients who maintain a DAS28-hsCRP ≤3.2 and DAS28-hsCRP <2.6 during the JADY study will be summarized for the JADY mITT population. The summary will be presented by treatment groups and originating studies.

The proportion of patients who achieve a DAS28-hsCRP ≤3.2 and DAS28-hsCRP <2.6 during the step-down period will be summarized for the step-down patients.

Descriptive statistics (n, mean, SD, minimum, 1st quartile, median, 3rd quartile, and maximum) of DAS28-hsCRP and the change from the originating study baseline will be summarized for the

JADY mITT population during the JADY study. Changes in DAS28-hsCRP will also be presented using the step-down baseline for the step-down patients.

Similar analyses for DAS28-ESR will be provided.

Refer to Appendix 2 for the calculation of DAS28-hsCRP and DAS28-ESR.

6.11.3. European League Against Rheumatism Responder Index (EULAR)

Assessments of patients with RA by EULAR response criteria will be used to categorize patients as non-responders, moderate responders, good responders, or responders (moderate + good responders) according to van Gestel et al. 1998.

The proportion of patients who maintain a EULAR responses (moderate and good EULAR responses) during the JADY study will be summarized for the JADY mITT population. The summary will be presented by treatment groups and originating studies.

Refer to Appendix 2 for the calculation of EULAR response.

6.11.4. van der Heijde Modified Total Sharp Score (mTSS), Joint Space Narrowing and Bone Erosions

X-rays of the hands/wrists and feet will be scored using structural progression as measured using the van der Heijde mTSS. This methodology quantifies the extent of bone erosions and JSN for 44 and 42 joints, respectively, with higher scores representing greater damage. See Appendix 2 for more details.

The final analysis will be based on a 5-year reading campaign which includes all patients who have x-ray data at baseline and at least one x-ray data after 2 years of treatment from the initial randomization in the originating studies. It will include evaluation of:

- Change from baseline of the originating study to 2, 3, 4, and 5 years in structural joint damage as measured by the mTSS [van der Heijde method], JSN, and bone ES.
- Proportion of patients with mTSS change ≤ 0, ≤ 0.5, and less than or equal to the smallest detectable change (SDC) from baseline of the originating study to 2, 3, 4, and 5 years.

Descriptive statistics (n, mean, SD, minimum, 1st quartile, median, 3rd quartile, and maximum) of actual mTSS, JSN, and bone ES and the change from the originating study baseline will be summarized by the originating studies where x-ray data are collected (excluding patients from Study JADA due to too few patients enrolled into JADY). The LE method described in Section 6.3.3 will be used for data imputation.

For the analysis of continuous structural joint damage data, treatment comparisons will be made using a mixed model for repeated measures (MMRM) with treatment, baseline value, visit, and the interactions of baseline-by-visit and treatment-by-visit as fixed factors. The covariance structure to model the between- and within-patient errors will be unstructured. If this analysis fails to converge, other structures will be tested such as heterogeneous autoregressive [ARH(1)],

heterogeneous compound symmetry (CSH) or heterogeneous Toeplitz (TOEPH). The variance-covariance structure that results in the smallest Akaike information criterion (AIC) will be used. The Kenward-Roger method will be used to calculate the degrees of freedom. Type III sums of squares for the least squares (LS) means will be used for the statistical comparison; the 95% CI will also be reported. For each treatment comparison at a visit, the LS mean for each treatment group, an LS mean estimate of the difference between treatments, corresponding 95% CI, and p-value will be presented.

For the analysis of categorical structural joint damage data, treatment comparisons will be made using a logistic regression model with treatment included.

6.11.5. Clinical Disease Activity Index (CDAI)

The CDAI is a tool for measurement of disease activity in RA that integrates measures of physical examination, patient self-assessment, and evaluator assessment. See Appendix 2 for more calculation details.

The proportion of patients who maintain a CDAI score ≤10 and CDAI score ≤2.8 during the JADY study will be summarized for the JADY mITT population. The summary will be presented by treatment groups and originating studies.

The proportion of patients who achieve a CDAI score ≤10 and CDAI score ≤2.8 during the step-down period will be summarized for the step-down patients. In addition, the proportion of patients who maintain a CDAI response status based on the step-down criteria during the step-down period will be summarized. A time to relapse (defined as losing step-down criteria, losing baseline CDAI response status, rescue) for step-down patients will be performed.

Descriptive statistics (n, mean, SD, minimum, 1st quartile, median, 3rd quartile, and maximum) of actual CDAI scores and the change from the originating study baseline will be summarized for the JADY mITT population during the JADY study. Changes from the step-down baseline will also be summarized for the step-down patients.

6.11.6. Simplified Disease Activity Index (SDAI)

The SDAI score will be analyzed similarly to the analyses planned for the CDAI score.

6.11.7. Tender Joint Count and Swollen Joint Count

The number of tender and painful joints (TJC) and swollen joints (SJC) is determined by examination of 34 joints on each side of the patient's body. For additional information on TJC and SJC as used in the calculation of ACR, refer to Appendix 2.

Descriptive statistics (n, mean, SD, minimum, 1st quartile, median, 3rd quartile, and maximum) of TJC and the change from the step-down baseline will be summarized for the step-down patients.

Similar analyses will be performed for the SJC.

6.11.8. Patient's Assessment of Pain, Physician's Global Assessment of Disease Activity, Patient's Global Assessment of Disease Activity

Patient's assessment of pain, Physician's Global Assessment of Disease Activity, and Patient's Global Assessment of Disease Activity are ACR components, and descriptions are given in Appendix 2.

Descriptive statistics (n, mean, SD, minimum, 1st quartile, median, 3rd quartile, and maximum) of the measures and the change from the step-down baseline will be summarized for the step-down patients.

6.11.9. High-Sensitivity C-Reactive Protein (hsCRP)

High-sensitivity c-reactive protein is an ACR core set measure of acute phase reactant.

Descriptive statistics (n, mean, SD, minimum, first quartile, median, third quartile, and maximum) of hsCRP and the change from the step-down baseline will be summarized for the step-down patients.

6.11.10. ACR/EULAR Rheumatoid Arthritis Remission

The proportion of patients who maintain an ACR/EULAR remission according to the Boolean-based definition during the JADY study will be summarized for the JADY mITT population. The summary will be presented by treatment groups and originating studies.

6.12. Pharmacokinetic/Pharmacodynamic Analysis (PK/PD)

No PK/PD analyses will be performed in Study JADY.

6.12.1. Health Outcomes Analyses

6.12.1.1. Duration of Morning Joint stiffness

The duration of morning joint stiffness is a patient-administered item that allows for the patients to enter the length of time in minutes that their morning joint stiffness lasted the day prior to the visit (using an electronic patient-reported outcomes [ePRO] tablet).

Duration of morning stiffness will be recorded in minutes. Durations recorded as longer than 12 hours (720 minutes) will be censored at 720 minutes in the analyses.

Descriptive statistics (n, mean, SD, minimum, 1st quartile, median, 3rd quartile, and maximum) of the duration of morning joint stiffness and the change from the originating study baseline will be summarized for the JADY mITT population during the JADY study. The summary will be presented by treatment groups and originating studies.

6.12.1.2. Health Assessment Questionnaire – Disease Activity (HAQ-DI)

Patients will be asked to complete the HAQ-DI to measure the impact of arthritis on their physical functioning. The HAQ-DI score ranges from 0 to 3, with lower scores indicating better physical functioning. See Appendix 2 for the algorithm to calculate the HAQ-DI score.

The proportion of patients who maintain a HAQ-DI improvement ≥0.22 and ≥0.3 during the JADY study will be summarized for the JADY mITT population. The summary will be presented by treatment groups and originating studies.

The proportion of patients who achieve a HAQ-DI improvement ≥0.22 and ≥0.3 during the step-down period will be summarized for the step-down patients.

Descriptive statistics (n, mean, SD, minimum, 1st quartile, median, 3rd quartile, and maximum) of actual HAQ-DI scores and the change from the originating study baseline will be summarized for the JADY mITT population during the JADY study. Changes from the step-down baseline will also be summarized for the step-down patients.

6.12.1.3. European Quality of Life-5 Dimensions-5 Level (EQ-5D-5L)

The EQ-5D-5L is a standardized measure of health status of the patient at the visit (same day) that provides a simple, generic measure of health for clinical and economic appraisal. The EQ-5D-5L consists of 2 components:

- a descriptive system of the respondent's health and
- a rating of his or her current health state using a 0 to 100 mm visual analog scale (VAS).

The descriptive system comprises the following 5 dimensions:

- mobility,
- self-care,
- usual activities,
- pain/discomfort, and
- anxiety/depression.

Each dimension has 5 levels:

- no problems,
- slight problems,
- moderate problems,
- severe problems, and
- extreme problems.

The respondent is asked to indicate his/her health state by ticking (or placing a cross) in the box associated with the most appropriate statement in each of the 5 dimensions. It should be noted that the numerals 1 to 5 have no arithmetic properties and should not be used as a cardinal score. The VAS records the respondent's self-rated health on a vertical VAS in which the endpoints are labeled "best imaginable health state" and "worst imaginable health state." This information can be used as a quantitative measure of health outcome. The EQ-5D-5L health states, defined by the EQ-5D-5L descriptive system, may be converted into a single summary index by applying a

formula that essentially attaches a value (also called weights) to each of the levels in each dimension (Brooks 1996; EuroQol Group 2011 [WWW]; Herdman et al. 2011).

The change from the originating study baseline in EQ-5D-5L scores will be summarized for the JADY mITT population during the JADY study.

6.12.1.4. Healthcare Resource Utilization

The healthcare resource utilization data will be collected by site staff regarding the number of visits to medical care providers such as general practitioners, specialists, physical or occupational therapists, and other nonphysical care providers for services outside of the clinical study; emergency room admissions; hospital admissions; and concomitant medications related to the treatment of RA. These data will be collected to support economic evaluations of treatment. The evaluation of healthcare resource utilization will be presented for JADY mITT population during the JADY study at the time points where data are collected based on the schedule.

6.13. Safety Evaluations and Analysis

Safety data from Study JADY will be more thoroughly assessed in the context of combining the JADY safety data with the safety data from the originating studies. The analysis details will be included in the ISAP or other integrated planning documents.

For the purpose of the Study JADY CSR, the following are planned to be analyzed for the JADY study alone with treatment and follow-up period combined:

- Summary of SAEs by Preferred Term (PT) nested within System Organ Class (SOC)
- Listing of SAEs including deaths
- Summary of AEs leading to permanent study drug discontinuation by PT nested within SOC
- Listing of AEs leading to permanent study drug discontinuation

6.13.1. Adverse Events (AEs)

Adverse events are recorded in the electronic Case Report Forms (eCRFs). Each AE will be coded to SOC and PT using the MedDRA version that is current at the time of database lock. Severity of AEs are reported as mild, moderate, or severe.

6.13.2. Serious Adverse Events (SAEs)

An SAE is any AE from this study that results in one of the following outcomes:

- Death
- Initial or prolonged inpatient hospitalization
- A life-threating experience (ie, immediate risk of dying)
- Persistent or significant disability/incapacity
- Congenital anomaly/birth defect

Important medical events that may not be immediately life-threatening or result in death or hospitalization but may jeopardize the patient or may require intervention to prevent one of the other outcomes listed in the definition above should be considered as serious. See examples in the ICH E2A guideline Section 3B.

The number and percentage of patients who experienced an SAE will be summarized using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the baricitinib group within decreasing frequency of SOC.

A listing of all SAEs that started during Study JADY will be provided.

6.13.3. Other Significant Adverse Events

The number and percentage of patients who permanently discontinued study drug because of an AE or death will be summarized using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the baricitinib group within decreasing frequency SOC.

In addition, a listing of all AEs leading to permanent discontinuation from the study will be provided.

6.14. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements. Analyses will 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, by MedDRA PT.

- An AE is considered 'Serious' whether or not it is a TEAE.
- An AE is considered in the 'Other' category if it is both a TEAE and is not serious.
- For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term
    - the number of events experienced.
- Consistent with ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included.

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

A TEAE is defined as an event that either first occurred or worsened in severity after the first dose of study treatment in study JADY and on or prior to the last visit date during the analysis period. The analysis period is defined as the treatment period and off-drug follow-up time. Adverse events are classified based upon the MedDRA PT. The MedDRA Lowest Level Term (LLT) will be used in defining which events are treatment-emergent. The maximum severity for each LLT during the baseline period up to the first dose of study medication in Study JADY will

be used as baseline. If an event with missing severity is preexisting during the baseline period and persists during the treatment period, then the baseline severity will be considered mild for determining TEAE (ie, the event is treatment-emergent if the severity is coded moderate or severe postbaseline and not treatment-emergent if the severity is coded mild postbaseline). If an event occurring postbaseline has a missing severity rating, then the event is considered treatment-emergent unless the baseline is severe, in which case the event is not treatment-emergent. The day and time for events where onset is on the day of the first dose of study treatment will both be used to distinguish between pretreatment and posttreatment in order to derive treatment-emergence. Should there be insufficient data for AE start date to make this comparison (eg, the AE start year is the same as the treatment start year, but the AE start month and day are missing), the AE will be considered treatment-emergent.

6.15. COVID-19 Impact

Study JADY was ongoing during the COVID-19 pandemic. Listings will be provided, such as listing of patients impacted by COVID-19, listing of important protocol deviations due to COVID-19, listing of COVID-19 AEs or deaths (based on SMQ code 20000237), etc.

7. Unblinding Plan

Early termination of Study JADY

Study JADY was terminated early as it met the study objectives and the projected exposure to baricitinib that would accrue during the remaining study period would not add significant scientific value.

Unblinding of study team members at previous database lock

The Study JADY has been locked a few times in the past for the purposes of initial baricitinib submission, resubmission for the RA indication, and safety updates to regulatory agencies. Study team members who were involved in previous baricitinib submissions and database locks have been unblinded to

- treatment information in Study JADY for all patients who participated in Study JADY and
- randomized treatment information during the step-down period for all patients who entered the step-down substudy at the time of the database locks.

Unblinding of study team members at final database lock

Study team members are blinded to treatment information for patients who were randomized to the step-down substudy after the most recent database lock. Study team members will be unblinded to this information after final database lock for the analysis purpose.

Blinding of investigators and patients

Investigators and patients are blinded to the initial randomized treatment assignment during the originating studies and the step-down period during the entire JADY study.

8. References

Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. *Clin Exp Rheumatol*. 2005;23(5 suppl 39):S100-S108.

Brooks R. EuroQol: The current state of play. *Health Policy*. 1996;37:53–72.

Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, Keystone EC, Loveless JE, Burmester GR, Cravets MW, Hessey EW, Shaw T, Totoritis MC. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy. Results of a multicenter, randomized, double-blind, placebo-controlled phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum*. 2006;54(9):2793-2806.

The EuroQol Group. EQ-5D-5L User Guide. Version 1.0. April 2011. Available at: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/UserGuide_EQ-5D-5L.pdf. Accessed 01 Jul 2012.

Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, Aletaha D, Allaart CF, Bathon J, Bombardieri S, Brooks P, Brown A, Matucci-Cerinic M, Choi H, Combe B, de Wit M, Dougados M, Emery P, Furst D, Gomez-Teino J, Hawker G, Keystone E, Khanna D, Kirwan J, Kvien TK, Landewé R, Listing J, Michaud K, Martin-Mola E, Montie P, Pincus T, Richards P, Siegel JN, Simon LS, Sokka T, Strand V, Tugwell P, Tyndall A, van der Heijde D, Verstappen S, White B, Wolfe F, Zink A, Boers M. American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Ann Rheum Dis*. 2011;70:404-413.

Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736.

Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, Fischkoff SA, Chartash EK. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti–tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum.* 2004;50(5):1400-1411.

Keystone E, Heijde D, Mason D Jr, Landewé R, Vollenhoven RV, Combe B, Emery P, Strand V, Mease P, Desai C, Pavelka K. Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum.* 2008;58(11):3319-3329.

Keystone E, Emery P, Peterfy CG, Tak PP, Cohen S, Genovese MC, Dougados M, Burmester GR, Greenwald M, Kvien TK, Williams S, Hagerty D, Cravets MW, Shaw T. Rituximab inhibits structural joint damage in patients with rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitor therapies. *Ann Rheum Dis*. 2009;68(2):216-221.

Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med*. 1998;17:873-890. Erratum in *Stat Med*. 1999;18:1293.

Smolen JS, Beaulieu A, Rubbert-Roth A, Ramos-Remus C, Rovensky J, Alecock E, Woodworth T, Alten R, OPTION investigators. Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial. *Lancet.* 2008;371(9617):987–997.

Smolen J, Landewé RB, Mease P, Brzezicki J, Mason D, Luijtens K, van Vollenhoven RF, Kavanaugh A, Schiff M, Burmester GR, Strand V, Vencovskỳ J, van der Heijde D. Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the RAPID 2 study. A randomised controlled trial. *Ann Rheum Dis.* 2009;68(6):797-804.

van Gestel AM, Haagsma CJ, van Riel PL. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. *Arthritis Rheum*. 1998;41(10):1845-1850.

Wilson EB. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*. 1927;22:209-212.

9. Appendices

Appendix 1. Algorithm for Determining ACR Response

Details presented in this appendix will use "x" as a generic symbol, and the appropriate number (either 20, 50, or 70) is to be filled in when implementing in dataset programming code.

ACR*x* response is defined as ≥*x*% improvement from baseline of the originating studies in tender joint count (68 joint counts), and ≥*x*% improvement in swollen joint count (66 joint counts), and ≥*x*% improvement in at least three of the following five items:

Patient's assessment of pain,

Patient's Global Assessment of Disease Activity,

Physician's Global Assessment of Disease Activity,

HAQ-DI,

hsCRP.

The following abbreviations will be used throughout this appendix to refer to the items needed in the algorithm definitions:

Parameter	Abbreviation for the Parameter
% improvement in tender joint count	TJC68
% improvement in swollen joint count	SJC66
% improvement in patient's assessment of pain	PATPAIN
% improvement in Patient's Global Assessment of Disease Activity	PATGA
% improvement in Physician's Global Assessment of Disease Activity	PHYGA
% improvement in HAQ-DI	HAQ
% improvement in hsCRP	hsCRP

For all seven parameters mentioned above, % improvement at a visit is calculated as:

(baseline value of the originating study – value at visit) * 100 / baseline value of the originating study.

To calculate the observed ACR*x* response at a visit:

- Step 1: If the patient discontinued from the study prior to reaching the visit, then STOP – assign ACR*x* response as blank (ie, missing). Otherwise, calculate the % improvement at the visit for all seven parameters as described above.
- Step 2:
    - If TJC68 and SJC66 are BOTH ≥*x*%, then proceed to step 3.
    - If both are non-missing but one or both is <*x*%, then STOP – assign the patient as a non-responder for ACR*x*.
    - If either or both are missing, proceed as follows:

a. If both are missing, then STOP – assign ACR*x* response as blank (ie, missing).

b. If one of TJC68 or SJC66 is missing and the non-missing value is <*x*%, then STOP – assign the patient as a non-responder for ACR*x*.

c. If one of TJC68 or SJC66 is missing and the non-missing value is ≥*x*%, then STOP – assign ACR*x* response as blank (ie, missing).

- Step 3: Consider the following five variables: PATPAIN, PATGA, PHYGA, HAQ, and hsCRP.
    - If three or more items are missing, then STOP – assign ACR*x* response as blank (ie, missing).
    - If three or more items are non-missing, then proceed with the following order:
        a. If at least three items are ≥*x*%, then STOP – assign the patient as a responder for ACR*x*.
        b. If at least three items are <*x*%, then STOP – assign the patient as a non-responder for ACR*x*.
        c. If less than three items are ≥*x*%, then STOP – assign ACR*x* response as blank.

To calculate the ACR*x* response at postbaseline visits using NRI:

- Step 1: Calculate the % improvement at the visit for all seven parameters as described above.
- Step 2: If all seven parameters are missing at the visit, or if prior to the visit the patient is considered discontinued from the study for imputation purposes (ie, previously had discontinued from study, permanently discontinued study drug, or had started a temporary drug interruption that does not end prior to the end of the study), then STOP – assign the patient as a non-responder for ACR*x*.
- Step 3: If at least one of the seven parameters is non-missing at the visit and the patient is still enrolled in the study, then use last observation carried forward (LOCF) to fill in any missing values.
- Step 4: If TJC68 AND SJC66 are BOTH ≥*x*%, then proceed to step 5.
    - If both are non-missing but one or both is <*x*%, then STOP – assign the patient as a non-responder for ACR*x*.
    - If either or both are missing, then STOP – assign the patient as a non-responder for ACR*x*.
- Step 5: Consider the following five variables: PATPAIN, PATGA, PHYGA, HAQ, hsCRP.
    - If three or more of the five items are non-missing, then:

- If three or more of the five items are ≥*x*%, then STOP – assign the patient as a responder for ACR*x*.
- If less than three of those five items are ≥*x*%, then STOP – assign the patient as a non-responder for ACR*x*.

- If three or more of the five items are missing, then STOP – assign the patient as a non-responder for ACR*x*.

Appendix 2. Details of Efficacy and Health Outcome Measures

van der Heijde Modified Total Sharp Score

X-rays of the hands/wrists and feet will be scored (according to the BioClinica predefined imaging charter) for structural progression as measured using the mTSS (van der Heijde 2000). This methodology quantifies the extent of bone erosions for 44 joints and joint space narrowing for 42 joints with higher scores representing greater damage.

In the 5-year x-ray read campaign, radiographs (ie, a single posteroanterior view of each hand and a single dorsoplantar view of each foot) obtained at the originating study baseline and at year 2, 3, 4, and 5 for Studies I4V-MC-JADX, I4V-MC-JADV, I4V-MC-JADZ and at year 2 and 3 for Study I4V-MC-JAGS will be analyzed by the 2 independent readers with x-rays being presented with time points blinded to the chronological order of the visits. For patients who discontinue prematurely from the study, x-rays will be performed at the early termination visit only if the previous x-rays were taken more than 12 weeks earlier.

Bone Joint Erosion Score

The joint erosion score is a summary of erosion severity in 32 joints of the hands and 12 joints of the feet. The maximum erosion score for a hand joint is 5 and for a foot joint is 10. Thus, the maximal erosion score is 280 for a timepoint (160 for both hands/ wrists and 120 for both feet). Each joint is scored according to the surface area involved from 0 to 5 for hand joints and 0 to 10 for the foot joints. The highest score (5 for the hand and 10 for the foot) indicates extensive loss of bone from more than one half of the articulating bone. A score of 0 in either the hand or foot joints indicates no erosion.

Erosion for each hand joint (16 joints per hand) is graded on the following scale.

0 – Normal (no erosion)

1 – Discrete small erosion

2 – Two discrete erosions or one large erosion not passing the joint middle line

3 – One large erosion passing the joint middle line of the bone or combination of the above

4 – Combinations of discrete and/or large erosions adding up to 4

5 – Combinations of discrete and/or large erosions adding up to 5 or more

NA - Not assessable due to poor radiographic depiction

S - Surgically modified (eg, joint replacement, amputation or other surgery)

Discrete erosions will be graded 1 if small and 2 or 3 if larger, with a score of 3 if the erosion is large and extends across the imaginary middle of the bone. Discrete erosions of each grade will

then be summed over both sides of the joint to a maximum of 5 to give a total score for each location in the hands. When erosions in the carpal bones of the wrist are confluent rather than discrete and focal, the percent surface eroded will be scored from 0 to 5 in approximately 20% intervals. A score of 'NA' will be assigned in the case that the location is not evaluable due to advanced subluxation or poor radiographic depiction. A score of 'S' will be assigned in the case of joint replacement, amputation or other surgery.

Erosion for each side of a foot joint (6 joints per foot) is graded on the following scale.

0 – Normal (no erosion)

1 – Discrete small erosion

2 – Two discrete erosions or one large erosion not passing the joint middle line

3 – One large erosion passing the joint middle line

4-10 – Combination of the above conditions.

NA - Not assessable due to poor radiographic depiction

S - Surgically modified (eg, joint replacement, amputation or other surgery)

Since both sides of each foot joint are graded, the maximum score for a foot joint is 10 (a score of 5 on each side).

Joint space Narrowing (JSN) Score

The JSN score summarizes the severity of JSN in 30 joints of both hands and 12 joints of both feet. Assessment of JSN for each hand (15 joints per hand) and foot (6 joints per foot), including subluxation, is scored from 0 to 4, with 0 indicating no (normal) JSN and 4 indicating complete loss of joint space, bony ankylosis or luxation. Thus, the maximum JSN score is 168 (120 for both hands/wrists and 48 for both feet) at a time point. JSN for each hand and foot joint is graded on the following scale:

0 – Normal or no narrowing

1 – Asymmetrical and/or minimal narrowing - up to 25%

2 – Definite narrowing with loss of up to 50% of the normal space

3 – Definite narrowing with loss of 50-99% of the normal space or subluxation

4 – Absence of a joint space, presumptive evidence of ankylosis or complete luxation

NA - Not assessable due to poor radiographic depiction

S - Surgically modified (eg, joint replacement, amputation or other surgery)

Total Modified Sharp Score (mTSS)

The total mTSS score at a time point is the sum of the erosion (maximum of 280) and JSN (maximum of 168) scores, for a maximum score of 448.

Handling of Missing Joint Data

The handling of a missing or inadequate score for a particular joint of a segment at a timepoint is dependent upon whether it qualifies for imputation. This depends on the number of missing or inadequate joints in a segment relative to the number of adequate joints in that segment.

Refer to the X-ray Score Derivations Process in Appendix 4 for a detailed description of imputation of missing joint data.

Adjudication Process

The read of x-ray images will be performed by 2 independent readers. An adjudication of results will be in place to show any discrepancy between the two independent readers. The adjudication process will involve re-reading by a third independent reader. Cases that require adjudication will be identified once the two primary independent reviews are completed. To ensure a consistent read by the two independent readers, an inter-/intra-reader variability assessment will be evaluated. Refer to the BioClinica imaging charter for more details.

Joint Assessment

Each of 28 or 68 joints will be evaluated for tenderness and 28 or 66 joints will be evaluated for swelling (hips are excluded for swelling) at the specified visits as shown in the schedule of events of the protocol. The 68/66 joint count will be performed at the visits, end of treatment and at the follow-up visit. For scores using 28 joints, the following subset will be assessed (both right and left side):

- Shoulder
- Elbow
- Wrist
- Metacarpophalangeal (MCP) I, II, III, IV, and V
- Thumb interphalangeal
- Proximal interphalangeal (PIP) II, III, IV, and V
- Knee

Joints will be assessed for tenderness by pressure and joint manipulation on physical examination. Any positive response on pressure, movement, or both will then be translated into a single tender-versus non-tender dichotomy. Joints will be classified as either swollen or not swollen. Swelling is defined as palpable fluctuating synovitis of the joint. Swelling secondary to osteoarthrosis will be assessed as not swollen, unless there is unmistakable fluctuation.

The following joint count imputation rules will be applied.

1. Joints that had any of the following procedures or disease conditions that occur at the screening (Visit 1) up to and including baseline (Visit 2) of the originating study will be imputed as follows:
    - **Arthroplasty, fusion, synovectomy, ankylosis, amputation, injury, fracture, infection, and other condition:** these will be considered "non-evaluable" from baseline of the originating study up to end of the study in JADY. Joints that include "non-evaluable" joints will be prorated from baseline of the originating study up to end of the study in JADY. For example, the swollen joint count score for 6 swollen plus 2 "non-evaluable" joints for DAS28 is calculated as [6/(28-2)]*28 = 6.46.
    - **Injection and other procedure:** joints that have received intra-articular and bursal injections will be collected in the data but set to tender and swollen after the injection for 6 months.
2. Joints that had any of the following procedures or disease conditions that occur during Study JADY (all visits are postbaseline of the originating study) will be imputed as follows:

- **Amputation:** amputation that occurs postbaseline of the originating study will be considered "non-evaluable" from the visit after amputation up to end of Study JADY. Joints that include "non-evaluable" joints will be prorated from the visit after amputation up to end of the study. For example, the swollen joint count score for 6 swollen plus 2 "non-evaluable" joints for DAS28 is calculated as [6/(28-2)]*28 = 6.46.
- **Arthroplasty, fusion, synovectomy** and **ankylosis:** if any of these procedures occur postbaseline of the originating study, joints will be set to tender and swollen from the visit after the procedure up to end of the study.
- **Infection, injury, fracture, and other condition:** if any of these conditions occur postbaseline of the originating study, joints will be set to tender and swollen from the visit after the condition until the condition is resolved.
- **Injection and other procedure:** joints that have received intra-articular and bursal injections postbaseline will be collected in the data but set to tender and swollen after the injection for 6 months or until the end of study (whichever comes first).

The number of tender and swollen joints will be calculated by summing all joints. For patients who have an incomplete set of joints evaluated, the joint count will be adjusted to a 28- or 68-joint count for tenderness and a 28- or 66-joint count for swelling by dividing the number of affected joints by the number of evaluated joints and multiplying by 28 or 68 for tenderness and 28 or 66 for swelling.

ACR Core Set

The individual components that make up the ACR Core Set of measures for RA are:

Tender Joint Count (TJC)

For ACR measures, the number of tender and painful joints will be determined by examination of 68 joints (34 joints on each side of the patient's body). These 68 joints will be assessed and classified as tender, not tender, or not evaluable. The number of tender joints ranges from 0 to 68. The algorithm to calculate TJC is described in the Joint Assessment section of this appendix.

Swollen Joint Count (SJC)

For ACR measures, the number of swollen joints will be determined by examination of 66 joints (33 joints on each side of the patient's body). These 66 joints will be assessed and classified as swollen, not swollen, or not evaluable. The number of swollen joints ranges from 0 to 66. The algorithm to calculate SJC is described in the Joint Assessment section of this appendix.

Patient's Assessment of Pain (VAS)

The patient will be asked to assess his or her current level of pain by marking a vertical tick on a 100-mm horizontal VAS (range 0 to 100 mm, with higher scores indicating more severe pain) with the left end marked as "no pain" and the right end marked as "worst possible pain." Results will be expressed in millimeters measured between the left end of the scale and the crossing point of the vertical line of the tick. This procedure is applicable for all VAS scales used in the trial.

Patient's Global Assessment of Disease Activity (VAS)

The patients will be asked to give an overall assessment of how their rheumatoid arthritis is affecting them at present by marking a vertical tick on a VAS (range 0 to 100 mm with higher scores indicating poorer status or more active arthritis) from "very well" to "very poor".

Physician's Global Assessment of Disease Activity (VAS)

The physician's assessment of a patient's disease activity assessed at the visit will be recorded using the VAS (range 0 to 100 mm with higher scores indicating poorer status or more active arthritis) from "none" to "extremely active arthritis".

Patient's Assessment of Physical Function (HAQ-DI)

The patient will be asked to complete the HAQ-DI to measure the impact of arthritis on their ability to function in daily life. The HAQ-DI functional disability index score ranges from 0 to 3, with lower scores indicating better physical functioning. The algorithm to calculate the HAQ-DI functional disability index score is described in the HAQ-DI section of this Appendix.

High-sensitivity C-Reactive Protein (hsCRP)

hsCRP will be the ACR Core Set measure of acute phase reactant. It will be measured at the central laboratory.

Disease Activity Score-Erythrocyte Sedimentation Rate (DAS28-ESR) and Disease Activity Score-High-Sensitivity C-Reactive Protein (DAS28-hsCRP)

The DAS28 is a measure of disease activity in 28 joints that consists of a composite numeric score of the following variables: TJC, SJC, hsCRP ***or*** ESR, and Patient's Global Assessment of Disease Activity (Vander Cruyssen et al. 2005). The 28 joints to be examined and assessed as tender or not tender for TJC and as swollen or not swollen for SJC include 14 joints on each side of the patient's body: the 2 shoulders, the 2 elbows, the 2 wrists, the 10 metacarpophalangeal joints, the 2 interphalangeal joints of the thumb, the 8 proximal interphalangeal joints, and the 2 knees (Smolen et al. 1995). The following equation will be used to calculate the DAS28-hsCRP (Vander Cruyssen et al. 2005):

$$DAS28\text{-}hsCRP = 0.56(TJC28)^{1/2} + 0.28(SJC28)^{1/2} + 0.36(\ln(hsCRP + 1)) + 0.014(VAS) + 0.96$$

The hsCRP value in mg/L must be used. In order to calculate DAS28-hsCRP data must be present for all 4 components: TJC28, SJC28, patient's global assessment (VAS) and hsCRP. Component-level imputation will be performed for the calculation of DAS28-hsCRP for the imputed visits summaries. If at least one of the 4 components is nonmissing at the visit and the patient is still enrolled in the study, then use LOCF to fill in any missing values.

Based on the DAS- hsCRP, the following will be calculated:

- DAS28- hsCRP ≤ 3.2
- DAS28- hsCRP < 2.6

Similarly,

$$DAS28\text{-}ESR = 0.56(TJC28)^{1/2} + 0.28(SJC28)^{1/2} + 0.70\ln(ESR) + 0.014(VAS)$$ [If ESR = 0 in analysis database, set ln(ESR)=0]

The ESR (erythrocyte sedimentation rate) in mm/first hour must be used. In order to calculate DAS28-ESR data must be present for all 4 components: TJC28, SJC28, patient's global assessment (VAS) and ESR. Component-level imputation will be performed for the calculation of DAS28-ESR for the imputed visit summaries. If at least one of the 4 components is nonmissing at the visit and the patient is still enrolled in the study, then use LOCF to fill in any missing values.

Based on the DAS-ESR, the following will be calculated:

- DAS28- ESR ≤ 3.2
- DAS28- ESR < 2.6

A NRI rule for the DAS28 binary endpoints will be applied for each visit as follows: if the DAS28 is missing at the visit or if prior to the visit the patient is considered discontinued from the study for imputation purposes (ie, previously had discontinued from study, permanently discontinued study drug, or had started a temporary drug interruption that does not end prior to

the end of the study), then assign the patient as DAS28 non-responder. If component-level imputation was performed for the continuous measures, the continuous imputed values will be used for the dichotomous endpoint for the NRI summaries.

Simplified Disease Activity Index (SDAI)

The SDAI is a tool for measurement of disease activity in RA that integrates measures of physical examination, acute phase response, patient self-assessment, and evaluator assessment. The SDAI is calculated by adding together scores from the following assessments:

- number of swollen joints (0 to 28),
- number of tender joints (0 to 28),
- hsCRP in mg/dL (0.1 to 10.0),
- patient global assessment of DAS on VAS (0 to 10.0, measured in cm), and
- evaluator or physician global assessment of DAS on VAS (0 to 10.0, measured in cm) (Aletaha and Smolen 2005)

Component-level imputation will be performed for the calculation of SDAI. If at least one of the 5 components is nonmissing at the visit and the patient is still enrolled in the study, then use LOCF to fill in any missing values.

Disease remission according to the ACR/EULAR index-based definition of remission is defined as an SDAI score of ≤3.3 (Felson et al. 2011).

Low disease activity according to SDAI is defined as a SDAI score ≤11 (Aletaha and Smolen 2005).

Clinical Disease Activity Index (CDAI)

The CDAI is similar to the SDAI, but it allows for immediate scoring because it does not use a laboratory result. The CDAI is calculated by adding together scores from the following assessments:

- number of swollen joints (0 to 28),
- number of tender joints (0 to 28),
- patient global assessment of DAS on VAS (0 to 10.0, measured in cm), and
- evaluator or physician global assessment of DAS on VAS (0 to 10.0, measured in cm) (Aletaha and Smolen 2005)

Component-level imputation will be performed for the calculation of CDAI. If at least one of the 4 components is nonmissing at the visit and the patient is still enrolled in the study, then use LOCF to fill in any missing values.

Remission according to CDAI is defined as a CDAI score ≤2.8 (Felson et al. 2011).

Low disease activity according to CDAI is defined as a CDAI score ≤10 (Aletaha and Smolen 2005).

European League Against Rheumatism Responder Index

Assessments of patients with RA by the EULAR Responder Index based on the 28-joint count will be used to categorize patients as non-responders, moderate responders, good responders, or responders (moderate + good responders) according to van Gestel et al. 1998.

Table APP.2.1 Categorization of Patients as Non-responders, Moderate Responders, or Good Responders

Postbaseline Level of DAS28	Improvement Since Baseline in DAS28 (Decline in DAS28)		
	>1.2	≤1.2 and >0.6	≤0.6
DAS28 ≤3.2	Good response		
3.2 < DAS28 ≤5.1		Moderate response	
DAS28 >5.1			No response

Abbreviation: DAS28 = Disease Activity Score modified to include the 28 diarthrodial joint count used for both hsCRP and ESR.

There are 3 categories, no improvement (or no response), moderate improvement (or moderate response), and good improvement (highest kind of improvement or good response). The category of no improvement also includes a worsening of RA. Hence, relative to baseline, a patient can only have a "0" change in categories, a change of "1" category or a change of "2" categories.

ACR/EULAR Rheumatoid Arthritis Remission

Two new ACR/EULAR definitions of RA remission will be evaluated, a "Boolean-based definition" and an "index-based definition" (Felson et al. 2011).

- Boolean-based Definition of Remission: All 4 criteria below must be met (at the same visit):
    - TJC28 ≤1
    - SJC28 ≤1
    - hsCRP ≤1 mg/dL (10 mg/L)
    - Patient Global Assessment of Disease Activity (on a 0-100 mm scale) / 10 ≤1
- Index-based Definition of Remission:
    - Simplified Disease Activity Index (SDAI) ≤3.3

PtGA = Patient Global Assessment of Disease Activity / 10 in order to convert from a 0-100 mm scale to a 0-10 cm scale.

The hsCRP value converted to mg/dL will be used.

European Quality of Life-5 Dimensions-5 Level (EQ-5D-5L) Scores:

The European Quality of Life-5 dimensions (EQ-5D) questionnaire is a widely used, generic questionnaire that assesses health status (EuroQol Group 2011 [WWW]/Herdman et al. 2011). The questionnaire consists of 2 parts:

The first part assesses 5 dimensions: mobility, self-care, usual activities, pain/ discomfort and anxiety/ depression. The levels of response for mobility, self-care and usual activities are "no problems", "slight problems", "moderate problems", "severe problems" or "unable to". For pain/ discomfort, the levels of response are "no pain/discomfort", "slight pain/discomfort", "moderate pain/discomfort", "severe pain/discomfort" and "extreme pain/discomfort". Lastly, for anxiety/ depression, the levels of response are "not anxious or depressed", "slightly anxious or depressed", "moderately anxious or depressed", "severely anxious or depressed", and "extremely anxious or depressed". This part of the EQ-5D can be used to generate a health state index score, which is often used to compute quality-adjusted life years (QALY) for utilization in health economic analyses. The health state index score is calculated based on the responses to the 5 dimensions, providing a single value on a scale from less than 0 (where zero is a health state equivalent to death; negative values are valued as worse than dead) to 1 (perfect health), with higher scores indicating better health utility.

The second part of the questionnaire consists of a visual analog scale (VAS) on which the patient rates their perceived health state from 0 (worst imaginable health state/the worst health you can imagine) to 100 (best imaginable health state/the best health you can imagine).

Health Assessment Questionnaire –Disability Index (HAQ-DI)

The HAQ-DI is a patient-reported questionnaire that is commonly used in RA to measure disease-associated disability (assessment of physical function). It consists of 24 questions referring to 8 domains: dressing/grooming, arising, eating, walking, hygiene, reach, grip, and activities (Fries et al. 1980, 1982; Ramey et al. 1996).

1. Dressing and grooming (C1. Dress yourself, including tying shoelaces and doing buttons, C2. Shampooing your Hair)
2. Includes 2 component questions, 1 device checkbox (devices used for dressing), 1 help checkbox.
3. Arising (C1. Stand up from straight chair, C2. Get in and out of bed)
4. Includes 2 component questions, 1 device checkbox (built-up or special chair), 1 help checkbox.
5. Eating (C1. Cut your meat, C2. Lift a cup or glass to your mouth, C3. Open a new carton of milk)
6. Includes 3 component questions, 1 device checkbox (built-up or special utensils), 1 help checkbox.
7. Walking (C1. Walk outdoors on flat ground, C2. Climb up five steps)
8. Includes 2 component questions, 4 device checkboxes (cane, walker, crutches, wheelchair), 1 help checkbox.
9. Hygiene (C1. Wash and dry your body, C2. Take a tub bath, C3. Get on and off the toilet)
10. Includes 3 component questions, 4 device checkboxes (raised toilet seat, bathtub seat, bathtub bar, long-handled appliances in bathroom), 1 help checkbox.
11. Reach (C1. Reach and get down a 5-pound object (such as a bag of sugar) from just above your head, C2. Bend down to pick up clothing from the floor)
12. Includes 2 component questions, 1 device checkbox (long-handled appliances for reach), 1 help checkbox.
13. Grip (C1. Open car doors, C2. Open jars which have been previously opened, C3. Turn faucets on and off)
14. Includes 3 component questions, 1 device checkbox (jar opener), 1 help checkbox.
15. Activities (C1. Run errands and shop, C2. Get in and out of a car, C3. Do chores such as vacuuming or yard work)
16. Includes 3 component questions, 1 help checkbox.

In order to compute the HAQ-DI (Standard Disability Index) score, the following scores are assigned to the responses:

Without any difficulty = 0

With some difficulty = 1

With much difficulty = 2

Unable to do = 3.

The disability section of the questionnaire scores the patient's self-perception on the degree of difficulty (0 = without any difficulty, 1 = with some difficulty, 2 = with much difficulty, and 3 = unable to do) when dressing and grooming, arising, eating, walking, hygiene, reach, grip, and performing other daily activities. The reported use of special aids or devices and/or the need for assistance of another person to perform these activities is also assessed. The scores for each of the functional domains will be averaged to calculate the functional disability index.

Calculating the HAQ-DI:

The patient must have a score for at least 6 of the 8 categories. If there are less than six categories completed, a HAQ-DI cannot be computed.

- A category score is determined from the highest score of the sub-categories, or components, in that category. (For example, in the category ARISING there are three sub-category items. If a patient responds with a 1, 2, and 0, respectively; the category score is 2.)
- Adjust for use of aids/devices and/or help from another person when indicated :
    - When there are no aids or devices or help indicated for a category, the category's score is not modified.
    - When aids or devices or help ARE indicated by the patient, adjust the score for a
    - category by increasing a zero or a one to a two. If a patient's highest score for that sub-category is a two it remains a two, and if a three, it remains a three.
    - Sum the eight category scores
    - Divide the sum by the number of categories answered (range 6-8)

The scale is not truly continuous but has 25 possible values (ie, 0, 0.125, 0.250, 0.375 … 3).The mapping of the aids or devices to the categories is the following:

HAQ-DI Category	Companion aids or devices item
Dressing and Grooming	Devices used for dressing (button hook, zipper pull, long handled shoehorn, etc.)
Arising	Built up or special chair
Eating	Built up or special utensils
Walking	Cane, walker, crutches
Hygiene	Long handled appliances in bathroom
Reach	Long handled appliances for reach
Grip	Jar opener (for jars previously opened)

Minimum Clinically Important Differences:

The summary and analyses of the HAQ-DI Score will include the number of patients achieving a MCID, which is defined as: MCID ≥0.22 reduction in HAQ-DI Score. Also, the number of patients having ≥0.3 reduction in HAQ-DI Score will be assessed.

References

Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. *Clin Exp Rheumatol*. 2005;23(5 suppl 39):S100-S108.

Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, Aletaha D, Allaart CF, Bathon J, Bombardieri S, Brooks P, Brown A, Matucci-Cerinic M, Choi H, Combe B, de Wit M, Dougados M, Emery P, Furst D, Gomez-Teino J, Hawker G, Keystone E, Khanna D, Kirwan J, Kvien TK, Landewé R, Listing J, Michaud K, Martin-Mola E, Montie P, Pincus T, Richards P, Siegel JN, Simon LS, Sokka T, Strand V, Tugwell P, Tyndall A, van der Heijde D, Verstappen S, White B, Wolfe F, Zink A, Boers M. American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Ann Rheum Dis*. 2011;70:404-413.

Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum*. 1980;23(2):137-145.

Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol*. 1982;9(5):789-793.

Ramey DR, Fries JF, Singh G. The Health Assessment Questionnaire 1995: status and review. In: Spiker B, editor. Quality of life and pharmacoeconomics in clinical trials. 2nd ed. Philadelphia: Lippincott-Raven, 1996: p. 227-237.

Smolen JS, Breedveld FC, Eberl G, Jones I, Leeming M, Wylie GL, Kirkpatrick J. Validity and reliability of the twenty-eight joint count for the assessment of rheumatoid arthritis activity. *Arthritis Rheum.* 1995;38(1):38-43.

Vander Cruyssen B, Van Looy S, Wyns B, Westhovens R, Durez P, Van den Bosch F, Veys EM, Mielants H, De Clerck L, Peretz A, Malaise M, Verbruggen L, Vastesaeger N, Geldhof A, Boullart L, De Keyser F. DAS28 best reflects the physician's clinical judgment of response to infliximab therapy in rheumatoid arthritis patients: validation of the DAS28 score in patients under infliximab treatment. *Arthritis Res Ther*. 2005;7(5):R1063-1071.

van der Heijde, D. How to read radiographs according to the Sharp/van der Heijde method. *Journal of Rheumatology*. 2000;27(1): 261-263.

Appendix 3. X-ray Missing Data Imputation Plan

This document focuses on how to handle missing scores for individual joints. If after applying the following imputation method, the total score for modified Total Sharp Score (mTSS), erosion score (ES), and joint space narrowing (JSN) score are still missing, additional missing data imputation methods defined in Section 6.3.1 will be applied.

Sequence:

Two sets of scores will be generated by 2 independent readers at BioClinica for all patients and provided to Lilly.

1. For each reader, identify missing joint scores and impute values where applicable
2. For each reader, identify any missing timepoints
3. If an adjudicated read is present, identify which reader pair will be used to derive scores
4. Derive mean scores from the reader pair selected

Rules:

- Missing individual joint scores for a given reader:
    - Missing joint score will be defined as joints scored as "S" (Surgically modified) or "NA" (not assessable due to poor radiographic depiction), as outlined in the BioClinica imaging charter, or no score is provided for a joint because of missing x-ray.
    - For baseline:
        - If an individual joint for a patient has missing score at baseline, this joint will be set to be missing across all study visits and will not be used in calculation of mTSS, ES, and JSN total scores or change from baseline of these total scores during the entire study for this patient. If >50% of the joint scores at baseline are missing, the baseline will be set to missing and this patient will not be included in the analysis.
    - For postbaseline:
        - If >50% of the joint scores at a given timepoint are missing (including joints that have been set to missing as a result of missing baseline), this will be taken as an invalid/missing timepoint (visit), the mTSS, ES, and JSN total score will be set to missing for the timepoint. Additional imputation methods specified in Study SAP Section 6.3.1 will be applied to impute missing total score (see below for inter-reader discordance on missing timepoints).

        - If ≤50% of the joint scores at a given timepoint are missing and an individual joint score becomes missing postbaseline, the score for this joint will be imputed based on an LOCF approach, using the last previous timepoints (including baseline) for which a non-missing score is available for the joint.

- Missing timepoints
    - If a time point is set to invalid/missing based on both independent readers, this timepoint will be taken as a missing timepoint.
    - If adjudication occurs because a timepoint was set as missing for only one of the two initial independent readers, then the reader pair selected and handling of the timepoint will depend on whether the adjudication reader's score for the timepoint concerned is set as missing or not missing:
        - If the adjudication reader's score for the timepoint concerned is not set as missing, the pair used will include the adjudication reader and the other independent reader for whom that timepoint was not set as missing. This will not be handled as a missing timepoint.
        - If the adjudication reader's score for the timepoint concerned is set as missing, the pair will include the adjudication reader and the other independent reader for whom that timepoint was set as missing. This will be handled as a missing timepoint and the mTSS, ES, and JSN total score will be set to missing. Additional imputation methods specified in Study SAP Section 6.3.1 will be applied to impute missing total score.
- General approach to scoring
    - For description of static scores at a given timepoint, the mean of the scores generated by the two independent readers at that timepoint (following imputation of missing scores if applicable) will be used.
    - For calculation of changes in mTSS, ES, and JSN score, the change from baseline will first be calculated for the 2 independent readers separately (following imputation of missing scores if applicable). The change score will then be taken as the mean of these two independent change scores.
    - In the case of adjudicated patient scores, a third set of scores generated by an independent adjudication reader will be provided (as outlined in the BioClinica Imaging Charter).
        - Where an adjudication reader is used, Lilly will use scores generated by the adjudication reader and the other independent reader whose score for change from baseline in mTSS to the primary time point for structure data analysis is closest to that generated by the adjudication reader (following imputation of missing scores if applicable) as a general principle.

- If the smallest difference at the primary time point for structure data analysis cannot be determined, then move back to the previous timepoint to determine the pair.

Footnotes:

a. In such cases, the last valid (non-imputed) score for that individual joint will be used for imputation, even if it is obtained from a timepoint set as missing as a result of >50% missing joints.

Leo Document ID = a5e7dc00-2400-4822-b52e-1143125e0a58

Approver: PPD

Approval Date & Time: 01-Oct-2020 20:20:02 GMT

Signature meaning: Approved