CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

207924Orig1s000

STATISTICAL REVIEW(S)

DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF BIOSTATISTICS

Statistical Review

CLINICAL STUDIES

NDA Number:	207924
Product Name:	Olumiant (baricitinib)
Formulation:	Two dose strengths: 2 mg and 4 mg tablets once daily
Sponsor:	Eli and Lilly Co.
Indications:	Rheumatoid Arthritis
Safety Issue:	Venous Thrombosis
Project Manager:	Jessica Lee
Clinical Review Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical reviewer:	Raj Nair, MD (Medical Officer)
Date:	Consult received on December 8, 2017
Biometrics Division:	Division of Biometrics VII
Statistical Reviewer:	Changming (Sherman) Xia, Ph.D.
Concurring Reviewers:	Eugenio Andraca-Carrera, Ph.D. (Team Leader) Mark Levenson, Ph.D. (Division Director)

Keywords: Venous thrombosis, rheumatoid arthritis, JAK inhibitor, safety

Table of Contents

1 INTRODUCTION AND BACKGROUND

NDA 207924 baricitinib was first submitted to the Food and Drug Administration (FDA) on January 15, 2016 for the indication of treatment of adult patients with moderately to severely active rheumatoid arthritis (RA) who have had an inadequate response or intolerance to methotrexate.

During the review of this original submission, a numerical imbalance of venous thrombosis events (VTE) was observed associated with baricitinib relative to placebo. An imbalance in combined reports of VTE was found in the placebo-controlled period of Phase 2 and 3 clinical trials comparing baricitinib 4mg to placebo (5/997 events for 4mg vs. 0/1070 events for placebo in 0-16 weeks, and 6/997 vs. 0/1070 in 0-24 weeks[1], and no events from the 2mg arm[2]). Additional events were observed in patients treated with baricitinib 2mg and 4mg after 24 weeks of exposure. Note that the placebo-controlled clinical trials in the baricitinib development program allowed subjects to switch between randomized treatments based on their disease activity after certain amount of time per protocol. To evaluate the risk of VTE associated with baricitinib, the applicant conducted retrospective analyses of VTE using data collected during both the placebo-controlled periods and post cross-over periods in Phase 1-3 trials designed to show efficacy of baricitinib. For a detailed description of the individual clinical trials in the baricitinib program as well as a statistical review of the initial NDA submission, refer to the review authored by Dr. Robert Abugov. For clinical reviews of the program, refer to reviews by Dr. Raj Nair.

The application received a Complete Response Letter (CRL) on April 12, 2017. The CRL identified the following deficiency regarding VTE:

> *Thrombosis is a unique risk for baricitinib, not seen with biologics DMARDs or with tofacitinib. Deep vein thrombosis and pulmonary embolism occurred with both baricitinib 2 mg and 4 mg at comparable rates and at rates higher than placebo. There were also few cases of arm and leg artery thrombosis with baricitinib. The thrombosis findings are of particular concern because these events are not predictable, and some were associated with death. As for laboratory parameters, it is worth noting that 2 patients were withdrawn from the studies for meeting platelet threshold criteria for withdrawal, both were from baricitinib 4 mg dose. Lilly argues against the thrombosis risk by comparing to population data. Comparison to population data is not relevant because the risk with baricitinib was seen in controlled clinical studies.*

The applicant re-submitted this NDA application on December 4, 2017 and included additional analyses of VTE.

In December, 2017, the Division of Pulmonary, Allergy, and Rheumatology Products (DPARP) consulted the Division of Biometrics VII to help describe and analyze the risk of VTE based on

[1] One additional VTE case was identified in the 4mg arm during 0-16 weeks by the applicant during the resubmission.
[2] Based on CDTL Memo dataset on dose as randomized, which is discussed in Section 4.3.4.

different analysis sets discussed in the NDA resubmission. Analysis sets were defined based on different subsets of trials and comparisons of interest (baricitinib vs placebo, or baricitinib 4 mg vs 2 mg). This document presents a statistical review to examine the risk of VTE associated with baricitinib based on a retrospective evaluation of VTE data focused on different analysis sets discussed in the NDA resubmission. For statistical and clinical reviews of this resubmission please refer to additional reviews authored by Dr. Robert Abugov and Dr. Raj Nair.

1.1 Resubmission Update

This review will focus on the analyses of VTE described in the Resubmission Safety Update (RSU) submitted on December 4, 2017. The applicant noted that the following new information was included:

1.1.1 Additional exposure data during the uncontrolled period in Trial JADY

In the *All BARI RA* analysis set, a total of 3492 patients with RA (total exposure of 7860.3 patient years) were exposed to any dose of baricitinib as of 01 April 2017, an increase of 28 patients and 3646.2 PY compared to the original submission. The *All BARI RA* set includes all subjects in the baricitinib program ever exposed to any dose of baricitinib as described in Section 4.3.1. Trial JADY is a long-term extension trial consisting of eligible patients who completed an originating trial (Trials JADZ, JADV, JADX, JADW, JADA, or JAGS).

1.1.2 New clinical trials: JAGS, JAHG

There were 2 newly completed trials, JAGS and JAHG, with 278 and 75 patients exposed to baricitinib respectively. No thrombosis events were reported in these trials. Subjects in trial JAGS were eligible to continue to be followed in the extension study JADY; however, at the time of integration of the safety database (01 April 2017), JAGS had not achieved final database lock and therefore the resubmission did not include follow-up data from these subjects in JADY. Study JAHG was designed to target atopic dermatitis instead of RA patients. Trials JAGS and JAHG were described in the resubmission but the applicant did not include them in the analysis of VTE in RA patients.

1.1.3 A new analysis set

A new analysis set, *FDA Ext BARI 2mg/4mg RA*, was presented. This analysis set consists of 7 Phase 2 and 3 trials with either 2 or 4mg baricitinib randomized arms and their JADY extension. The different analysis sets utilized to assess the risk of VTE are described in more detail in Section 4.3.

Reviewer Comment*: The additional exposure data in the resubmission was mostly limited to uncontrolled periods of clinical trials in the baricitinib program. The available placebo-controlled data have not changed since the original submission.*

2 OBJECTIVES

Based on input provided by DPARP, this review has the following objectives:

1) Evaluate the risk of VTE comparing baricitinib to placebo.
2) Evaluate the risk of VTE comparing the two proposed doses (2mg, 4mg) of baricitinib in the completed trials.
3) Estimate the VTE incidence rates (IR) among RA subjects treated with baricitinib.

3 DATA SOURCES

Data for VTE analyses in the resubmission were extracted from the following datasets from the applicant's IR response on Feb 1st, 2018:

- **ADAE_IDB4.xpt** – VTE events from initial submission ADAE.
- **ADSI_IDB4.xpt** – VTE events from initial submission, with exposure censored at the time of first event.
- **ADAE_4MSU.xpt** – VTE events from 4-Month Safety Update (4MSU) ADAE.
- **ADSI_4MSU.xpt** - VTE events from 4MSU, with exposure censoring at event
- **ADSL_19MU.xpt** - Subject-Level Analysis
- **ADAE_19MU.xpt** - VTE events from 19 Month Update (19MU) ADAE
- **ADSI_19MU.xpt** - VTE events from 19MU, with exposure censored at the time of first event
- **ADSI_19MU_FDA7.xpt** - VTE events from 19MU for 7 studies, with exposure censored at the time of first event

4 TRIALS AND ANALYSIS SETS

4.1 List of Trials

Data from the following 9 trials from the baricitinib program were utilized for the integrated analysis of VTE. JADY is an extension study of eligible subjects who have completed the final active treatment study visit in an originating baricitinib RA study (Studies JADA, JADZ, JADV, JADX, JADW, or JAGS). Two new trials (JAGS, JAGH) were submitted separately during the resubmission, and no thrombosis events were reported in either of these two trials. They are not included in the integrated analysis.

Table 1 List of Trials

Study ID ***Phase***	**Randomized Treatment (in the initial period)**	**N (randomized and treated)**	**Rescue, switch, or dose tapering**	**Population**
JADA *2*	Bar 1 mg QD Bar 2 mg QD Bar 4 mg QD Bar 8 mg QD Placebo, up to wk 12 (Part B explored BID Dosing)	49 52 52 50 98 Total: **301**	Evaluated BID and QD of different doses after wk 12 (no placebo after wk 12). Double-blind, randomized for a total of 24 weeks and then open-label up to wk 128.	Regular use of a stable dose of MTX, but no prior bDMARD

JADB *1*	Open-label (Bar 5mg BID, 10mg QD, 15mg QD) for 28 days	53	Total 4 weeks.	RA subjects who had been receiving MTX
JADC *2*	Bar 4 mg QD Bar 7 mg QD Bar 10 mg QD Placebo, up to wk 12	31 32 31 31 Total: **124**	Placebo group re-randomized to Bar 7 or 10 at wk 12. Open-label from wk 12-24.	Inadequate response to any DMARD therapy
JADN *2*	Bar 1 mg QD Bar 2 mg QD Bar 4 mg QD Bar 8 mg QD Placebo, up to wk 12	24 24 24 24 49 Total: **145**	After wk 12, all Placebo subjects were re-randomized to 4 or 8mg, all other arms remain on initial treatment for 52 weeks.	Japanese with regular use of MTX
JADV *3*	Bar 4 mg QD Adalimumab Placebo, up to wk 16	487 330 488 Total: **1305**	Rescue at wk 16. PC 24 weeks. All Placebo subjects switch to 4mg at wk 24. Total 52 weeks.	Inadequate response to MTX; patients continued to take background MTX during the study; no prior bDMARD.
JADW *3*	Bar 2 mg QD Bar 4 mg QD Placebo, up to wk 16	174 177 176 Total: **527**	Rescue at wk 16, PC/total 24 weeks.	Inadequate response or tolerance to ≥ 1 bDMARD TNF-α inhibitor and regular use of a stable dose of cDMARD therapy
JADX *3*	Bar 2 mg QD Bar 4 mg QD Placebo, up to wk 16	229 227 228 Total: **684**	Rescue at wk 16; PC/total 24 weeks.	Insufficient response or were intolerant to cDMARDs
JADZ *3*	MTX Bar 4 mg QD MTX + Bar 4 mg QD, up to wk 24	210 159 215 Total: **584**	Rescue at wk 24; total 52 weeks.	Limited or no treatment with MTX and no prior cDMARD or bDMARD therapy
JADY *3*	Bar 2 mg QD (patients from JADX and JADW) Bar 4 mg QD	297 2242 Total: **2539**	Placebo or active-control subjects were switched to 4mg; 4mg subjects who achieved sustained low disease activity or remission were re-randomized to 2 or 4mg; if response was not maintained, rescued to 4mg. Total	Long-term extension study for JADA, JADZ, JADV, JADX, JADW and JAGS

			48 months (data cutoff 4/1/2017)	

Source: Summarized by FDA statistical reviewer from clinical study reports by the applicant.
Abbreviations: MTX=Methotrexate, bDMARD=biological disease-modifying anti-rheumatic drug, cDMARD=conventional disease-modifying anti-rheumatic drug.

Reviewer Comment*: Dose changes were permitted per protocol and were common in these trials (rescue medication, dose switch or dose tapering), which makes it difficult to compare the risk of VTE associated with baricitinib relative to placebo, or to compare different doses of baricitinib, especially when the data are outside of the initial 16-week placebo-controlled period prior to dose change or rescue. Also, note that the populations of these trials were different, which can further limit the interpretability of integrated analyses.*

4.2 List of Analysis Sets

The safety population consists of all randomized subjects who received at least one dose of the study treatment. Various analysis sets in the safety population were proposed to serve different purposes based on interactions between FDA and the applicant:

Table 2 List of Analysis Sets

Analysis Set	Trials	N (PYE)	Description	Purpose
All BARI RA	Phase 1: JADB Phase 2: JADA/Y, JADC, JADN Phase 3: JADV/Y, JADW/Y, JADX/Y, JADZ/Y	Total: **3492** (7860 PY, can be up to 7993 PY depending on safety outcome)	8 trials and extension trial JADY with any dose of baricitinib; only subjects who have ever been exposed to baricitinib were included	IR estimate for Bar-treated subjects
FDA Ext BARI 2mg/4mg RA	Phase 2: JADA/Y, JADC, JADN Phase 3: JADV/Y, JADW/Y, JADX/Y, JADZ/Y	2mg: 479 (615 PY, censored at dose change) 4mg: 1371 (2936 PY, censored at dose change) 2/4mg: 1850 (4575 PY, not censored at dose change) Total: **1850**	7 trials and extension trial JADY with either 2mg or 4mg randomized arm	IR estimate for Bar 2/4mg subjects
Lilly Ext BARI 2mg vs. 4mg RA (PC 0-16 weeks)	Phase 2: JADA/Y, JADN Phase 3: JADW, JADX	2mg: 479 (137.3 PY) 4mg: 479 (137.6 PY) Placebo: 551 (150 PY) Total: **1509**	4 trials with both 2mg and 4mg randomized arms, as well as placebo	Compare Bar 4mg vs. 2mg vs. placebo
CDTL Memo (PC 0-16 weeks)	Phase 2: JADA, JADC, JADN, Phase 3: JADV, JADW, JADX	2mg: 479 (140 PY) 4mg: 997 (298 PY) Placebo: 1070 (308 PY) Total: **2546**	6 placebo-controlled trials with either 2mg or 4mg randomized arm.	Compare Bar 4mg vs. placebo

Division Director's Review	Phase 3: JADV/Y, JADX/Y, JADW/Y, JADZ/Y	2mg: 403 4mg: 1265 Placebo: 892 Total: **2560**	4 pivotal trials and extension	Not a focus of resubmission

Source: Summarized by FDA statistical reviewer from RSU, Director's Review and CDTL Memo.

4.3 Analysis Sets in Resubmission

The first four analysis sets in Table 2 were emphasized during the resubmission, and this review will focus on VTE analysis based on these 4 analysis sets: *All BARI RA, FDA Ext BARI 2mg/4mg RA, Lilly Ext BARI 2mg vs. 4mg RA and CDTL Memo*. For details on analyses based on the *Division Director's Review* analysis set, refer to the review by Dr. Badrul Chowdhury. Below is a description of these four analysis sets and how they were used to analyze VTE in this review.

4.3.1 All BARI RA

The *All BARI RA* analysis set consists of data from 8 originating baricitinib trials for RA and their extension in Trial JADY, including all subjects ever exposed to any dose of baricitinib (the doses in this set range from 1 mg to 10 mg once daily). This analysis set includes all follow-up time in these trials, from the first dose of baricitinib, to the last recorded observation or the date of the first adverse event of interest, whichever is earlier, regardless of dose changes, tapering, and rescue, and regardless of treatment adherence. Subjects who were randomized to placebo but switched treatment to baricitinib contribute data to this analysis set from the time of their first recorded dose of baricitinib. Because this analysis set considers the longest available follow-up time and exposure to baricitinib, it was used to evaluate the overall safety profile of baricitinib and to calculate IR of adverse events associated with baricitinib. Because the trials in this analysis set allowed crossovers from placebo to baricitinib as well as dose switches/tapering/rescue, comparisons between different doses of baricitinib and between baricitinib and placebo are not feasible.

4.3.2 FDA Ext BARI 2mg/4mg RA

The *FDA Ext BARI 2mg/4mg RA* analysis set consists of 7 trials that included either a baricitinib 2mg or 4mg randomized treatment arm, but not necessarily both, and their extension in Trial JADY. The data include all follow-up time until the last observation or the date of the first adverse event of interest, whichever is earlier, and were not censored by dose change or rescue. This analysis set was used to evaluate the risk profile and to calculate IR of adverse events associated with the combined baricitinib 2/4mg doses. Note that because patient populations varied across different trials, the 4mg arm from one trial cannot be directly compared to the 2mg from another trial. Also, this analysis set included study JADZ, which did not have a randomized placebo arm. Therefore, this analysis set was not used to compare baricitinib 2 mg vs 4 mg, or to compare baricitinib to placebo.

4.3.3 Lilly Ext BARI 2mg vs. 4mg RA

The *Lilly Ext BARI 2mg vs. 4mg RA* analysis set contains only trials that included placebo, 2mg and 4mg randomized treatment arms. This set was used to compare the safety of baricitinib 2 mg vs. 4 mg, as well as baricitinib 2/4mg vs. placebo during the placebo-controlled period of the trials. The main limitation of this set is that it has the shortest total patient-year exposure and therefore very few events of VTE were observed. Data are from randomization, censored at dose change or rescue.

4.3.4 CDTL Memo

The *CDTL Memo* analysis set contains 6 trials that have both 4mg and placebo randomized arms. It has more patient exposure than the *Lilly Ext BARI 2mg vs. 4mg RA* set and it is used to compare the 4mg arm to placebo during the randomized controlled period. Study JADZ was not included in this set because it did not have a placebo arm. Data are from randomization, censored at dose change or rescue.[3]

5 OUTCOME

The primary outcome of interest is venous thrombosis (VTE) that occurred after at least one dose of randomized treatment in the trials.

5.1 Definition and Identification of Venous Thrombosis

The signal for VTE was first identified based on the Medical Dictionary for Regulatory Activities (MedDRA) Preferred Terms (PT) 'deep vein thrombosis' (DVT) and 'pulmonary embolism' (PE).

In order to identify additional events that might have been missed in the original submission, during the resubmission, the applicant searched the clinical trial and safety database (Lilly Safety System [LSS]) for all events reported in the MedDRA SMQs of embolic and thrombotic events, venous; embolic and thrombotic events, vessel type unspecified; mixed venous and arterial events; and thrombophlebitis (narrow terms only). All AEs identified in this expanded review were medically reviewed and those that represented DVT or PE were included in a reanalysis of the initial submission safety data.

5.2 Exposure and Incidence Rate

To adjust for the different observation times across dose levels, patient-year-adjusted IRs were reported to compare event rates among different treatment groups. For AEs of special interest, including VTE, the IRs were calculated as the number of patients with an event per 100 PY of observation time, including any post-drug follow-up time, with observation time censored at the first event date[4].

[3] Different censoring methods were utilized for the same subset of trials in the CDTL Memo on Jan 5, 2017, but only the data prior to dose change or rescue are considered for this analysis set in this review.
[4] Page 23, Resubmission Safety Report by the applicant.

6 STATISTICAL ANALYSIS

6.1 Analyses of Interest

The comparisons of interest for VTE include the following: (1) Placebo vs. baricitinib 2mg and baricitinib 2mg vs. baricitinib 4mg; (2) Placebo vs. baricitinib 4mg.

We are also interested in estimating the VTE incidence rate for (1) subjects who have taken any dose of baricitinib; (2) subjects who have taken either baricitinib 2mg or 4mg.

6.2 Multiplicity

No adjustments for multiplicity were made.

6.3 Baricitinib 4mg vs. Baricitinib 2mg and Baricitinib 2mg vs. Placebo

As described in Section 4.3, the *Lilly Ext BARI 2mg vs. 4mg RA* analysis set during the placebo-controlled period is the most relevant set for this comparison of interest. Outside of the placebo-controlled period, the randomization could be violated by allowing dose switching and treatment comparisons would be difficult to interpret.

Table 3 VTE Rates: Placebo vs. Bar 2mg and Bar 2mg vs. Bar 4mg (PC Week 0-16)

	PBO	**2mg**	**4mg**
	N=551	**N=479**	**N=479**
VTE	0	0	2
PYE	150	137.3	137.6
IR (*per 100PY*) [95% CI]	0 [0, 2]	0 [0, 2.2]	1.5 [0.4, 5.7]

Source: Applicant's RSU, Tables 5.2 and 5.32 using the *Lilly Ext BARI 2mg vs. 4m0, g RA* analysis set. Data were censored at dose change. Confidence intervals (CI) for IR were calculated based on a log-normal distribution for 2/4 mg, and a Bernoulli distribution for placebo.

There were 2 VTE events observed in the baricitinib 4mg arm and no events were reported in the placebo or baricitinib 2mg arms in the placebo-controlled (PC) 16 weeks period using the *Lilly Ext BARI 2mg vs. 4mg RA* set.

6.4 Baricitinib 4mg vs. Placebo

The *CDTL Memo* data set contains 6 studies of both placebo and baricitinib 4mg arms and it provides more patient years to compare baricitinib 4mg to placebo.

Table 4 VTE Rates: Placebo vs. Bari 4mg (PC Week 0-16)

	PBO	**4mg**
	N=1070	**N=997**

VTE	0	5
PYE	300.4	292.2
IR (*per 100PY*) [95% CI]	0 [0, 1]	1.7 [0.7, 4.1]

Source: FDA statistical reviewer summarized from applicant's RSU, Tables 5.2 and 5.32 using the *CDTL Memo* analysis set. Data were censored at dose change. Confidence intervals (CI) for IR were calculated based on a log-normal distribution for 4mg, and a Bernoulli distribution for placebo.

There is a numerical imbalance between placebo and baricitinib 4mg during the first 16 weeks of placebo-controlled period using the *CDTL Memo* set, with 5 events in the baricitinib 4mg arm and no events in the placebo arm.

6.5 VTE Incidence Rate Among Baricitinib Subjects

The *All BARI RA* dataset included all doses of baricitinib, while the *FDA Ext BARI 2mg/4mg RA* dataset included subjects who were randomized to baricitinib 2mg or 4mg. The extension period from Trial JADY was included with a data cutoff on April 1, 2017. The estimated rates are 0.53/100PY for any dose of baricitinib using the *ALL BARI RA* set, and 0.37/100PY for the combined 2/4mg of baricitinib using the *FDA Ext BARI 2mg/4mg RA* set.

Table 5 VTE IR Estimate (Including Follow-up Data in JADY)

	All BARI RA	**FDA Ext BARI 2mg/4mg RA**
	N=3492	**N=1850**
# of VTE	42	17
PYE	7948.6	4575.1
IR (*per 100PY*) [95% CI]	0.53 [0.39, 0.71]	0.37 [0.23, 0.60]

Source: FDA statistical reviewer using adsi19mu.xpt and adsi19_7.xpt, and Table APP.8.62 on Page 896 of RSU. Data were not censored at dose change. Exposure was calculated up to first of VTE for subjects with events (ASTDT), or Analysis Period End Date for subjects without events until the end of the analysis period. Confidence intervals (CI) for IR were calculated based on a log-normal distribution.

7 LIMITATIONS

Some limitations of these analyses are described below. These limitations should be considered when interpreting the strength of the findings.

(1) VTE was an unexpected safety signal and was not prospectively and proactively collected during the baricitinib program. Due to the limited number of events in the controlled period, there is not sufficient evidence to either confirm the VTE signal or to rule out a risk associated with baricitinib.
(2) Outside of the short, 16-week placebo-controlled period prior to dose change or rescue, these trials allowed dose switching, tapering, and rescue per protocol. Thus, it was difficult

to attribute VTE to specific dose levels of baricitinib after this time, or to evaluate the association between length of exposure to baricitinib and the risk of VTE.

8 CONCLUSIONS

In this resubmission, the applicant provided additional data and analyses to further evaluate the VTE risk for the baricitinib program. The additional data from the extension study JADY were utilized to provide a better estimate of the overall IR for RA patients exposed to baricitinib. However, no additional well-controlled data (prior to dose switching) were provided for a reliable comparison between the different doses of baricitinib, or between baricitinib and placebo. Therefore, our conclusion remains unchanged from the original submission[5]. Baricitinib, especially the 4mg dose, shows a numerically higher VTE risk than placebo. More data are needed to either confirm or rule out this potential important safety signal.

[5] Refer to FDA's CRL on 4/12/2017

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

CHANGMING N XIA
05/22/2018

EUGENIO ANDRACA-CARRERA
05/22/2018

MARK S LEVENSON
05/23/2018

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

Statistical Review Addendum

CLINICAL STUDIES

NDA / Sequence Number:	NDA 207924 / Seq 0048
Drug Name:	Olumiant (baricitinib)
Proposed Indication:	Rheumatoid arthritis (RA)
Applicant:	Eli Lilly and Company
Date(s):	Received: December 4, 2017
Review Priority:	Priority Review
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Robert Abugov, Ph.D.
Concurring Reviewer:	Gregory Levin, Ph.D.
Statistics Supervisor:	Mark Rothmann, Ph.D. (Division Director)
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Raj Nair M.D. (Medical Officer) Nikolay Nikolov, M.D. (Medical Team Leader) Badrul A Chowdhury, M.D., Ph.D. (Division Director)
Project Manager:	Jessica Lee

Keywords: NDA review, Clinical Studies

Contents

Tables

Figures

1 SUMMARY

In an NDA submitted on January 15, 2016, Eli Lilly and Company submitted results from four randomized, double-blind, confirmatory trials which demonstrated effectiveness of baricitinib 2 mg or 4 mg tablets once daily (QD) for the treatment of rheumatoid arthritis. However, safety concerns resulted in a Complete Response (CR), communicated to the applicant on April 12, 2017.

While the 4 mg dose was efficacious, it appeared to be associated with thrombotic risks which suggested a potentially unfavorable benefit-risk profile. The 2 mg dose was similarly efficacious, however there were thrombotic events on that dose as well, and there were concerns that patient exposure to the 2 mg dose may have not been adequate to reliably evaluate benefit-risk. To support approval of baricitinib, the Complete Response letter noted that additional clinical data would be necessary to identify a safe and effective dose, including adequate safety data to evaluate risk of thrombosis at the final proposed dose. The CR letter also noted that there was not a consistent finding of greater efficacy for baricitinib 4 mg over 2 mg.

In this submission, the applicant has recommended an alternative dosing regimen of 2 mg once daily, with 4 mg once daily recommended only for patients with an inadequate response or intolerance to more than one disease modifying anti-rheumatic drug (DMARD). To support this dosing regimen, the applicant provided a post-hoc analysis which purported to demonstrate an increased benefit of the 4 mg dose over the 2 mg dose for this subpopulation. Implicit in the applicant's argument is an improved benefit-risk profile for 4 mg in these patients.

The subgroup analyses of study JADX provided in this submission are inadequate to support increased benefit of the 4 mg dose over the 2 mg dose in the subpopulation of patients with an inadequate response or intolerance to more than one DMARD. First, the analyses were designed and conducted after examination of the data; such post-hoc analyses are suspect because, in any submission, under the null hypothesis of no difference between treatment arms, roughly half of all subgroup analyses for any given endpoint would be expected to show differences in the direction of improvement (as in the subgroup examined in this particular submission). While such analyses are useful to generate hypotheses, to be tested in subsequent preplanned clinical trials, they are inadequate to confirm true differences. Second, in the complementary subgroup of patients, with inadequate response to one or fewer DMARDs, there were similar trends, but toward superior efficacy of 2 mg compared to 4 mg, a conclusion which seems implausible and which suggests that observed differences across subgroups may have been chance findings. Further, interaction tests conducted in this review, between prior DMARD status and treatment, indicate that the evidence is insufficient to conclude that prior DMARD status influences the relative efficacy of the two baricitinib doses.

The best available estimate for the difference in efficacy between the 4 mg and 2 mg doses was provided by the applicant on July 21, 2016, in response to an information request submitted by FDA for an integrated analysis combining results from two earlier dose-ranging studies as well as results from phase 3 studies JADX and JADW. That analysis suggests that, while a difference in efficacy likely does exist between the 4 mg and 2 mg doses, it is small, e.g., for ACR20 at week 12, the estimated response rate differs between the two doses only by 2 percentage points.

In summary, the 2 mg and 4 mg doses likely have very similar efficacy. There is some concern that the benefit-risk profile for the 4 mg dose is unfavorable. Further, whether the 2 mg dose is associated with an improved benefit-risk profile compared to the 4 mg dose has not been established.

2 STATISTICAL EVALUATION

2.1 Data and Analysis Quality

All analyses of study JADX provided by the applicant were post-hoc. As such, any conclusions drawn are exploratory, suitable only for development of future confirmatory trials.

2.2 Evaluation of Efficacy

2.2.1 Study Design and Endpoints

The original submission provided results from four confirmatory studies. Studies JADX and JADW, respectively enrolled patients refractory to treatment with conventional disease modifying antirheumatic drugs (cDMARDS) and tumor necrosis factor (TNF) inhibitors, randomizing them to baricitinib 4 mg oral tablets administered QD (B4), baricitinib 2 mg oral tablets administered QD (B2), or placebo (Pbo) as add-on therapies to cDMARDS. Both studies demonstrated statistically significant effects of B4 and B2 compared to Pbo for the proportion of patients exhibiting a positive American College of Rheumatology 20 (ACR20) response. A third study, JADV, which randomized bDMARD naive patients refractory to treatment with methotrexate (MTX) to B4, adalimumab 40 mg SC Q2W (A), or Pbo as add-on therapies to MTX, showed statistically significant effects of B4 for ACR20 response compared to both Pbo and A. And finally, study JADZ, which randomized biologic disease modifying antirheumatic drugs (bDMARD) and cDMARD naive patients to B4, B4 administered concurrently with MTX (B4MTX), or MTX, showed improved ACR20 response rates for either B4 or B4MTX compared to MTX monotherapy. For all doses tested, the four confirmatory studies consistently demonstrated efficacy for the primary endpoint, ACR20 response, as well as for multiple

secondary endpoints. Additional details on the design and results of these studies can be found in my statistical review of the original NDA submission dated November 17, 2016.

The present submission provides further results comparing efficacy of B4 and B2 from trials JADX and JADW.

Study JADX (Table 1) was a parallel-group, double-blind, double-dummy, placebo-controlled trial randomizing 660 adult patients with moderately to severely active RA who were bDMARD naive with inadequate response to cDMARDs, to B4, B2, or Pbo. Randomization was in a 1:1:1 ratio stratified by region and presence or absence of joint erosion. All treatments were added to therapy with cDMARDs. The Pbo group was continued to W24, with the possibility of rescue to B4 offered at W16 to patients in the Pbo and B2 trial arms. The primary endpoint was W12 ACR20 response on randomized treatment without rescue.

Study JADW (Table 2) was a parallel-group, double-blind, double-dummy, placebo-controlled trial randomizing 525 adult patients with moderately to severely active RA who were anti-TNF inadequate responders and who were on stable doses of cDMARDs, to B4, B2, or Pbo. Randomization was in a 1:1:1 ratio stratified by region and presence of joint erosion. All treatments were added to therapy with cDMARDs. The Pbo group was continued to W24, with rescue to B4 offered at W16 to patients in the Pbo and B2 trial arms. The primary endpoint was W12 ACR20 response on randomized treatment without rescue.

In both studies, rescued patients without satisfactory improvement following at least four weeks of rescue treatment were discontinued from treatment. Patient rescued at week 16 were discontinued from treatment if they met nonresponder criteria at Week 20.

Table 1. Design for Study JADX

Study	Design	Population	Endpoints
JADX	B4 + cDMARD	Mod to severe active RA	Primary (W12):
	B2 + cDMARD	Age ≥ 18	ACR20 on treatment
(III)	Pbo + cDMARD	cDMARD inadequate resp	
		On cDMARD (incl MTX)	Major Secondary (W12):
	DB, DD, PG	bDMARD naive	B4 vs P
		TJC and SJC ≥ 6	ΔHAQ-DI
	Rescue[a]:	CRP ≥ 1.2 ULN	DAS28
	Pbo, B2 from W16		SDAI ≤ 3.3
		exclusions:	Morning stiffness dur
	End W24	> 10 mg/day predn	B2 vs P
		MTX and HCQ and SSZ	ACR 20
			Worst Tiredness
		N = 220/220/220	B4 vs P
		Strat: region,	Morning stiffness sev
		joint erosion (Y/N)	Worst tiredness
			Worst pain
			B2 vs P
			HAQ-DI
			DAS28
			SDAI ≤ 3.3

source: reviewer

Abbreviations: B4 and B2 baricitinib tablets 2 mg or 4 mg QD, cDMARD conventional DMARD including methotrexate, DB double blind, DD double dummy, PG parallel group, W12 W16 W24 weeks 12, 16, and 24 respectively, TNF tumor necrosis factor, cDMARD conventional disease modifying antirheumatic drug, bDMARD biologic DMARD, TJC tender joint count, SJC swollen joint count, CRP C-reactive protein, hs high sensitivity, predn prednisone, ULN upper limit of normal, MTX methotrexate, HCQ hydroxychloroquine, SSZ sulfasalazine, ACR20 20% improvement in American College of Rheumatology criteria, Δ = change from baseline, HAQ-DI health assessment questionnaire disability index, mTSS modified total Sharp score, DAS28 Disease Activity Score 28 using high-sensitivity C-reactive protein, SDAI simplified disease activity index

[a] Rescue to B4 if lack of improvement of at least 20% in both TJC and SJC at both week 14 and week 16 compared to baseline. At week 20, rescue therapy offered to patients at the discretion of the investigator based on TJCs and SJCs.

Note: The dose for renally impaired patients randomized or rescued to baricitinib is B2 for all studies in this submission

Table 2. Design for Study JADW

Study	Design	Population	Endpoints
JADW	B4 + cDMARD	Mod to severe active RA	Primary (W12):
	B2 + cDMARD	Age ≥ 18	ACR20 on trt
(IV)	Pbo + cDMARD	anti-TNF inadequate resp	
		On cDMARD	Major Secondary (W12):
		TJC and SJC ≥ 6	Δ HAQ-DI
	DB, DD, PG	CRP or hsCRP ≥ 1.0 ULN	DAS28-hsCRP
			SDAI ≤ 3.3
	Rescue[a]:	exclusions:	
	Pbo, B2 from W16	> 10 mg/day predn	
		MTX and HCQ and SZZ	
	End W24		
		N = 175/175/175	
		Strat: region, bDMARD (<3, ≥3 previous)	

source: reviewer

2.2.2 Statistical Methodologies

Categorical efficacy variables were analyzed using logistic regression, with independent factors treatment and any stratification variables included in the randomization (see Table 1 and Table 2). Continuous efficacy variables were analyzed using ANCOVA, with independent factors treatment, baseline, and any stratification variables included in the randomization.

Efficacy analyses were generally conducted on the modified-intent-to-treat (mITT) population, defined as patients receiving at least one dose of the study drug. In the original submission, type 1 error rates in the face of multiple endpoints and doses, was controlled at the .05 level of significance using analysis hierarchies defined graphically as in Bretz et al.[1]

Non-response was recorded for binary response data missing or collected after permanent treatment discontinuation or escape. Therefore, we consider these variables to be composite endpoints defined by remaining on randomized treatment through the time point of interest and meeting the binary response criteria at the time point of interest. For key secondary endpoints, missing continuous data was imputed using modified baseline observation carried forward (mBOCF), with BOCF used after patients discontinued the study or study treatment due to an adverse event, and last observation carried forward (LOCF) used after patients who discontinued the study or study treatment due to other reasons. For most other secondary endpoints, missing continuous data was imputed using modified last observation carried forward (mLOCF), in which data for patients who received rescue therapy was imputed using LOCF from the time at which rescue occurred and in which, for all other patients who discontinued the study for any reason, LOCF was conducted by carrying forward the last post-baseline observation.

2.2.3 Patient Disposition, Demographic and Baseline Characteristics

There were no obvious differences between treatments for baseline characteristics in the submitted studies (Appendix A; Table 22, Table 23). The mean age of patients in the treatment arms in these studies ranged from 49 to 56 years, and a majority were female and white or Asian. Patterns of patient disposition did not contradict efficacy of B4 or B2 (Table 3 through Table 6). Compared to control, baricitinib treated patients tended to have lower early discontinuation rates,

---

[1] Bretz, F., Maurer, W., Brannath, W. and Posch, M. (2009) A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine 28(4), 586-604

higher adverse event rates, and lower discontinuation rates due to lack of efficacy.

Table 3. Patient Disposition. Week 12, mITT Population, Study JADW

	Pbo N (%)	**B2 N (%)**	**B4 N (%)**
Randomized	176	174	177
Completed W12	154 (88%)	160 (92%)	166 (94%)
Early discontinuation	22 (12%)	14 (8%)	11 (6%)
Adverse event	4 (2.3%)	6 (3.4%)	7 (4.0%)
Lack of efficacy	12 (6.8%)	3 (1.7%)	2 (1.1%)
Entry criteria not met	0 (0%)	0 (0%)	0 (0%)
Patient withdrawal	5 (2.8%)	5 (2.9%)	1 (0.6%)

Source: Sequence 0006 Regulatory Response Table 2

Table 4. Patient Disposition. Week 24, mITT Population, Study JADW

	Pbo N (%)	**B2 N (%)**	**B4 N (%)**
Randomized	176	174	177
Rescued	56 (32%)	38 (22%)	33 (19%)
Completed W24	89 (50%)	119 (68%)	126 (71%)
Post-rescue	55 (31%)	38 (22%)	32 (18%)
Early discontinuation (w/out rescue)	31 (18%)	17 (10%)	18 (10%)
Adverse event	7 (4.0%)	7 (4.0%)	10 (5.6%)
Death	0	0	1 (0.6%)
Lack of efficacy	15 (8.5%)	4 (2.3%)	3 (2.8%)
Entry criteria not met	0	0	0
Patient withdrawal	7 (4.0%)	6 (3.4%)	1 (0.6%)

Source: Sequence CSR Table JADW 10.1

Table 5. Patient Disposition. Week 12, mITT Population, Study JADX

	Pbo N (%)	**B2 N (%)**	**B4 N (%)**
Randomized	228	229	227
Completed W12	206 (90%)	216 (94%)	212 (93%)
Early discontinuation	22 (10%)	13 (6%)	15 (7%)
Adverse event	5 (2.2%)	7 (3.1%)	6 (2.6%)
Lack of efficacy	4 (1.8%)	3 (1.3%)	0 (0%)
Entry criteria not met	0 (0%)	0 (0%)	0 (0%)
Patient withdrawal	10 (4.4%)	3 (1.3%)	6 (2.6%)

Source: Sequence 0006 Regulatory Response Table 3

Table 6. Patient Disposition. Week 24, mITT Population, Study JADX

	Pbo N (%)	**B2 N (%)**	**B4 N (%)**
Randomized	228	229	227
Rescued	55 (24%)	21 (9%)	15 (7%)
Completed W24	145 (64%)	189 (83%)	189 (83%)
Post-rescue	54 (24%)	20 (9%)	14 (6%)
Early discontinuation (w/out rescue)	28 (12%)	19 (8%)	23 (11%)
Adverse event	2 (0.9%)	0	0
Death	2 (0.9%)	0	0
Lack of efficacy	7 (3.1%)	3 (1.3%)	0
Patient withdrawal	11 (4.8%)	5 (2.2%)	8 (3.5%)

Source: Sequence CSR Table JADX 10.1

2.2.4 Results and Conclusions

2.2.4.1 Primary Endpoint: ACR20

Studies JADX and JADW both showed statistically significant differences between baricitinib and placebo (Table 7 and Table 8). Differences at week 12 between B4 and B2 were not statistically significant and did not trend in consistent directions. Results at week 24 were similar to those at week 12.

Table 7. ACR20 and Remain on Randomized Treatment, Study JADX

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2[a]
12	62 (140/227)	66 (151/229)	39 (90/228)	2.5 (<.001) (1.7, 3.7)	3.0 (<.001) (2.0, 4.4)	0.8 (.4) (0.6, 1.2)
24[a]	65 (148/227)	61 (140/229)	42 (96/228)	2.6 (<.001) (1.8, 3.9)	2.2 (<.001) (1.5, 3.2)	1.2 (0.3) (0.8, 1.8)

source: reviewer programs ACR JADX W12 &24 2016 03 14.sas, CSR Table 4, JADX 11.9, 11.18

[a] Exploratory endpoint, p-values nominal only

Table 8. ACR20 and Remain on Randomized Treatment, Study JADW

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2[a]
12	55 (98/177)	49 (85/174)	27 (48/176)	3.4 (<.001) (2.2, 5.4)	2.7 (<.001) (1.7, 4.2)	1.3 (0.3) (0.8, 2)
24[a]	46 (82/177)	45 (78/174)	27 (48/176)	2.4 (<.001) (1.5, 3.7)	2.3 (<.001) (1.5, 3.6)	1.0 (.9) (0.7, 1.6)

source: reviewer program ACR JADX W12 &24 2016 03 14.sas, CSR Table JADW.11.9, 11.14

[a] Exploratory endpoints, p-values nominal only

For study JADX, ACR components at week 12 did not trend in favor of B4 or B2 (Table 9). However, for study JADW, although none of the differences were statistically significant, there appeared to be an efficacy trend favoring B4 over B2 (Table 10).

Table 9. ACR20 and ACR Components, Study JADX, Week 12

Endpoint	**B4**	**B2**	**Placebo**	**B4 vs B2**		**(95% CI)**
ACR20	62%	66%	39%	Odds Ratio	0.8	(0.6, 1.2)
ΔHAQ-DI	-0.56	-0.57	-0.36	Mean Difference	0.01	(-0.08, 0.11)
ΔTJC	-13	-13	-10		0	(-1, 1)
ΔSJC	-9	-9	-6		0	(-2, 2)
ΔPain	-23	-25	-16		2	(-2, 6)
ΔPaGA	-26	-25	-17		-1	(-5, 4)
ΔPhGA	-34	-32	-22		-3	(-6, 1)
ΔCRP	-9	-9	0		0	(-3, 2)

source: FDA statistical review for NDA 207924 Seq 0001

Table 10. ACR20 and ACR Components, Study JADW, Week 12

Endpoint	**B4**	**B2**	**Placebo**	**B4 vs B2**		**(95% CI)**
ACR20	55%	49%	27%	Odds Ratio	1.3	(0.8, 2.0)
ΔHAQ-DI	-0.41	-0.37	-0.17	Mean Difference	-0.03	(-0.14, 0.07)
ΔTJC	-14	-12	-9		-2	(-5.1, 0.37)
ΔSJC	-9	-7	-5		-2	(-3, 0.1)
ΔPain	-22	-17	-9		-5	(-10, -0.3)
ΔPaGA	-23	-20	-9		-3	(-7, 2.0)
ΔPhGA	-35	-31	-17		-5	(-9, 0.4)
ΔCRP	-9	-5	1		-4	(-8, 0.0)

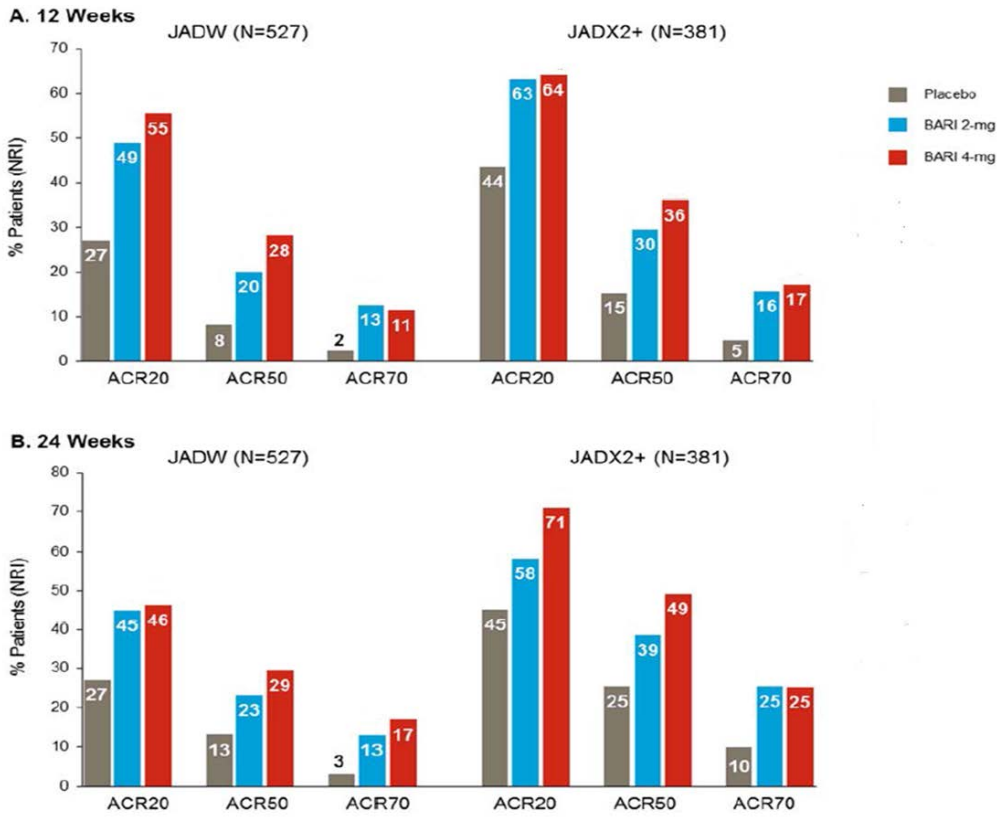
source: FDA statistical review for NDA 207924 Seq 0001

2.2.4.2 Exploratory Subgroup Analyses

Noting the positive, albeit not statistically significant trend in study JADW, the applicant conducted exploratory subgroup analyses of patients in study JADX who failed two or more DMARDS. The applicant also extended the exploratory analyses to week 24, a timepoint which was not prespecified as a secondary endpoint in the protocol but which seemed to show stronger trends in favor of B4 over B2 (Figure 1).

Note here that the trends in Figure 1 and other graphical comparisons provided in the applicant's clinical study report are exaggerated because the y-axes are truncated before reaching 100%.

Figure 1. Exploratory Comparison, B4 vs B2, Patients with at least 2 Prior DMARDs



source: Figure 3.1 of applicant report 5-3-5-3-2dmardir.pdf, NDA submission 207924 Seq 0048

The applicant's clinical study report provided graphs similar to those in Figure 1 for a broad range of efficacy endpoints. Reported here are statistical analyses of a broad sample of responder endpoints as well as for the continuous endpoints HAQ-DI and DAS28-hsCRP. For responder endpoints in the overall JADX study population, there were no trends favoring B4 over B2 at week 12 (Table 11) or week 24 (Table 12). There was a trend toward slightly greater efficacy for B4 with respect to DAS28, but not for HAQ-DI (Table 13 and Table 14). For all of these endpoints, the confidence intervals included the null hypothesis, indicating that none of the differences between B4 and B2 were statistically significant.

Table 11. Efficacy of B4 vs B2, Responder Endpoints, Study JADX, Preplanned Week 12

Endpoint	B4 n=227	B2 n=229	Placebo n=228	OR B4:B2	OR 95% CI
ACR20	62%	66%	39%	0.8	(0.6, 1.2)
ACR50	33%	34%	13%	1	(0.7, 1.5)
ACR70	18%	18%	3%	1	(0.7, 1.7)
DAS28CRP ≤ 2.6	26%	26%	9%	1	(0.7, 1.6)
DAS28CRP ≤ 3.2	39%	36%	17%	1.2	(0.8, 1.8)
DAS28ESR ≤ 2.6	9%	11%	2%	0.9	(0.5, 1.6)
DAS28ESR ≤ 3.2	22%	21%	7%	1	(0.7, 1.6)
EULAR Resp	78%	72%	54%	1.4	(0.9, 2.1)
CDAI ≤ 2.8	9%	10%	2 %	1.1	(0.5, 1.8)
SDAI ≤ 3.3	15%	17%	4%	0.9	(0.5, 1.5)
Δ HAQ-DI ≤ -3.3	55%	58%	37%	0.9	(0.6, 1.3)

source: reviewer program ACR JADX W12 &24 2018 02 23 all.sas

Table 12 Exploratory Analysis, B4 vs B2, Responder Endpoints, Study JADX, Week 24

Endpoint	B4 n=227	B2 n=229	Placebo n=228	OR B4:B2	OR 95% CI
ACR20	65%	61%	42%	1.2	(0.8, 1.8)
ACR50	44%	41%	21%	1.1	(0.8, 1.6)
ACR70	24%	25%	8%	1	(0.6, 1.5)
DAS28CRP ≤ 2.6	33%	31%	11%	1.1	(0.8, 1.7)
DAS28CRP ≤ 3.2	52%	46%	24%	1.2	(0.8, 1.8)
DAS28ESR ≤ 2.6	16%	14%	4%	1.2	(0.7, 1.9)
DAS28ESR ≤ 3.2	32%	29%	10%	1.1	(0.8, 1.7)
EULAR Resp	78%	72%	54%	1.4	(0.9, 2.1)
CDAI ≤ 2.8	15%	15%	4%	1	(0.6, 1.7)
SDAI ≤ 3.3	15%	17%	4%	0.9	(0.5, 1.5)
Δ HAQ-DI ≤ -3.3	55%	58%	37%	0.9	(0.6, 1.3)

source: reviewer program ACR JADX W12 &24 2018 02 23 all.sas

Table 13. Exploratory Analysis, B4 vs B2, Mean Change from Baseline of Continuous Endpoints, Study JADX, Week 12

Endpoint	B4 (n)	B2 (n)	Placebo (n)	Difference B4-B2 (95% CI)
DAS28-hsCRP	-1.92 (227)	-1.83 (229)	-1.08 (228)	-0.09 (-0.31, 0.13)
HAQ-DI	-0.56 (222)	-0.57 (228)	-0.36 (220)	0.01 (-0.08, 0.11)

source: reviewer

Table 14. Exploratory Analysis, B4 vs B2, Mean Change from Baseline of Continuous Endpoints, Study JADX, Week 24

Endpoint	B4 (n)	B2 (n)	Placebo (n)	Difference B4-B2 (95% CI)
DAS28-hsCRP	-2.2 (227)	-2.01 (229)	-1.23 (228)	-0.19 (-0.44, 0.06)
HAQ-DI	-0.62 (222)	-0.62 (228)	-0.38 (220)	0 (-0.1, 0.11)

source: reviewer

The applicant's revised dosage recommendations in the present submission,

> "2 mg once daily. For patients with an inadequate response or intolerance to more than one DMARD, a dose of 4 mg once daily is recommended."

implicitly suggest that the 4 mg dose is superior to the 2 mg dose specifically in the subpopulation of patients with inadequate response or intolerance to two or more DMARDs.

In study JADX at week 12, for the patient subpopulation who had prior inadequate response or intolerance to two or more DMARDs, in 7 of 11 endpoints examined, response was numerically greater in B4 over B2 (Table 15). None of the differences were statistically significant for any of the endpoints. At week 24, in the same patient subpopulation, the numerical trend favored B4 over B2 in 9 of 11 endpoints, again with none of the differences between B4 and B2 statistically significant (Table 16).

That a majority of the endpoints in Table 15 and Table 16 tended to favor B4 over B2 suggests a large 'weight of evidence' in favor of B4 over B2. Such an argument could be compelling if these endpoints were independent of one another. However, it is likely that they are correlated as many of them share components and they are all indicators of the inflammation of the joint.

For study JADX at weeks 12 and 24, the responder analyses were further explored by adding prior DMARD (<2, ≥2) and prior DMARD by treatment interaction to the statistical models. None of the interactions were significant (all p-values > 0.10), indicating that there was no evidence that number of prior DMARDS impacted whether the appropriate dose was 2 or 4 mg.

I also evaluated change from baseline for two continuous endpoints, DAS28-hsCRP and HAQ-DI, which are generally evaluated in all RA submissions for new drugs. At weeks 12 and 24, for DAS28-hsCRP and HAQ-DI, the 95% confidence intervals for differences between B4 and B2 all overlapped zero, regardless of prior DMARD status (Table 17 and Table 18). There was a trend toward slightly greater efficacy for B4 with respect to DAS28 in the subgroup of patients with at least two prior DMARDs, but the mean difference (-0.27 at week 12) was small and a similar trend was not observed for HAQ-DI.

Table 15 Exploratory Subgroup Analysis of Responder Endpoints, B4 vs B2, Study JADX, Week 12, ≥ Two Prior DMARDs

Endpoint	**B4 n=128**	**B2 n=122**	**Placebo n=131**	**OR B4:B2**	**OR 95% CI**
ACR20	64%	63%	44%	1	(0.6, 1.8)
ACR50	36%	30%	15%	1.4	(0.8, 2.4)
ACR70	17%	16%	5%	1.2	(0.6, 2.4)
DAS28CRP ≤ 2.6	26%	24%	9%	1.2	(0.6, 2.1)
DAS28CRP ≤ 3.2	43%	34%	18%	1.5	(0.9, 2.5)
DAS28ESR ≤ 2.6	8%	11%	2%	0.7	(0.3, 1.8)
DAS28ESR ≤ 3.2	2%	20%	7%	1	(0.6, 2)
EULAR Resp	84%	76%	58%	1.7	(0.9, 3.2)
CDAI ≤ 2.8	9%	7%	2%	1.2	(0.4, 3.2)
SDAI ≤ 3.3	7%	7%	1%	1.1	(0.4, 3)
Δ HAQ-DI ≤ -3.3	55%	57%	48%	0.9	(0.6, 1.5)

source: reviewer program ACR JADX W12 &24 2018 02 24 2dmard.sas

Table 16 Exploratory Subgroup Analysis of Responder Endpoints, B4 vs B2, Study JADX, Week 24, ≥ Two Prior DMARDs

Endpoint	B4 n=128	B2 n=122	Placebo n=131	OR B4:B2	OR 95% CI
ACR20	71%	58%	45%	1.7	(1, 2.9)
ACR50	49%	39%	25%	1.5	(0.9, 2.5)
ACR70	25%	25%	10%	0.9	(0.5, 1.7)
DAS28CRP ≤ 2.6	37%	30%	9%	1.4	(0.8, 2.3)
DAS28CRP ≤ 3.2	55%	46%	24%	1.4	(0.8, 2.3)
DAS28ESR ≤ 2.6	19%	12%	3%	1.6	(0.8, 3.3)
DAS28ESR ≤ 3.2	36%	30%	10%	1.2	(0.7, 2.1)
EULAR Resp	82%	74%	57%	1.6	(0.9, 3.0)
CDAI ≤ 2.8	17%	11%	5%	1.7	(0.8, 3.5)
SDAI ≤ 3.3	14%	13%	4%	1.1	(0.5, 2.2)
Δ HAQ-DI ≤ -3.3	55%	55%	39%	0.9	(0.6, 1.5)

source: reviewer program ACR JADX W12 &24 2018 02 24 2dmard.sas

Table 17. Exploratory Subgroup Analysis of Mean Change from Baseline of Continuous Endpoints, B4 vs B2, Study JADX, Week 12, ≥ Two Prior DMARDs

Endpoint	B4 (n)	B2 (n)	Placebo (n)	Difference B4-B2 (95% CI)
DAS28-hsCRP	-2.02 (128)	-1.76 (122)	-1.12 (131)	-0.27 (-0.56, 0.03)
HAQ-DI	-0.55 (126)	-0.55 (122)	-0.38 (127)	0 (-0.13, 0.13)

source: reviewer

Table 18. Exploratory Subgroup Analysis of Mean Change from Baseline of Continuous Endpoints, B4 vs B2, Study JADX, Week 24, ≥ Two Prior DMARDs

Endpoint	B4 (n)	B2 (n)	Placebo (n)	Difference B4-B2 (95% CI)
DAS28-hsCRP	-2.3 (128)	-1.99 (122)	-1.26 (131)	-0.31 (-0.63, 0.02)
HAQ-DI	-0.59 (126)	-0.60 (122)	-0.39 (127)	0.01 (-0.13, 0.14)

source: reviewer

It is instructive to apply the applicant's exploratory methods to evaluate the complement subpopulation of JADX, patients who had prior inadequate response or intolerance to fewer than 2 DMARDs. At week 24, in 9 of 11 responder endpoints examined, effectiveness was numerically greater in B2 rather than in B4 (Table 19). Mean changes from baseline for continuous endpoint DAS28 and HAQ-DI were similar between the doses in this patient subgroup (Table 20).

Table 19 Exploratory Subgroup Analysis of Responder Endpoints, B4 vs B2, Study JADX, Week 24, < Two Prior DMARDs

Endpoint	B4 n=99	B2 n=107	Placebo n=97	OR B4:B2	OR 95% CI
ACR20	58%	64%	38%	0.8	(0.4, 1.4)
ACR50	37%	45%	16%	0.7	(0.4, 1.3)
ACR70	23%	25%	5%	0.9	(0.5, 1.7)
DAS28CRP ≤ 2.6	28%	32%	12%	0.8	(0.4, 1.5)
DAS28CRP ≤ 3.2	46%	47%	23%	1	(0.6, 1.7)
DAS28ESR ≤ 2.6	13%	17%	6%	0.7	(0.3, 1.6)
DAS28ESR ≤ 3.2	26%	28%	10%	0.9	(0.5, 1.7)
EULAR Resp	73%	70%	48%	1.2	(0.6, 2.1)
CDAI ≤ 2.8	13%	20%	3%	0.6	(0.3, 1.4)
SDAI ≤ 3.3	16%	21%	4%	0.7	(0.4, 1.6)
Δ HAQ-DI ≤ -3.3	56%	62%	35%	0.8	(0.5, 1.4)

source: reviewer program ACR JADX W12 &24 2018 03 03 le1dmard.sas

Table 20. Exploratory Subgroup Analysis of Mean Change from Baseline of Continuous Endpoints, B4 vs B2, Study JADX, Week 24, < Two Prior DMARDs

Endpoint	**B4 (n)**	**B2 (n)**	**Placebo (n)**	**Difference B4-B2 (95% CI)**
DAS28-hsCRP	-2.07 (99)	-2.07 (107)	-1.19 (97)	-0.01 (-0.4, 0.39)
HAQ-DI	-0.68 (96)	-0.68 (106)	-0.38 (93)	0 (-0.16, 0.16)

source: reviewer

To summarize thus far, the applicant provided analyses which purport to demonstrate superior efficacy of B4 over B2 in patients with prior inadequate response or intolerance to at least 2 DMARDs. However, the subgroup analyses were post-hoc, there was not evidence of an interaction between prior DMARD use and treatment effect, differences between doses within subgroups were not statistically significant, and magnitudes of estimated differences were generally small. Furthermore, similar analyses and interpretations in the complementary subgroup might lead to a likely implausible conclusion, i.e. that the lower dose B2 is superior to the higher dose B4 among patients with prior inadequate response or intolerance to fewer than 2 DMARDs. Therefore, the applicant's subgroup analyses are considered exploratory and hypothesis-generating rather than confirmatory.

2.2.4.3 Integrated Analyses Comparing B4 to B2

The most reliable estimates comparing efficacy of B4 and B2 for the treatment of RA were provided by the applicant in response to an information request by FDA for an integrated analysis of available RA studies which randomized patients to both B4 and B2. While the integrated analysis was only exploratory, its sole purpose was to use all available data to increase the precision of estimated differences between doses. In addition to studies JADX and JADW, the integrated analysis included data from dose-ranging studies JADA and JADN, which enrolled patients with active RA with inadequate response or intolerance to methotrexate and which provided randomized treatment as an add-on to methotrexate.

For proportion of ACR20 responders, the integrated analysis trends toward greater efficacy of B4 over B2 (Table 21). However, the advantage of B4 over B2 in response rate appears to trend downward over time, from 9% at week 2 to 2% at week 12. For changes from baseline DAS28(CRP) and HAQ-DI, the advantage of B4 over B2 was minimal considering commonly used estimates of minimally important clinical differences are approximately 0.6 and 0.22 for change from baseline DAS28(CRP) and HAQ-DI respectively.

Table 21. B4 vs B2, Integrated Analysis of Studies JADX, JADW, JADA, and JADN

Endpoint	Week	B4	B2	Placebo	B4-B2 (95% CI)
ACR20 Response (%)	2	40%	31%	19%	9% (3%, 15%)
	4	54%	45%	28%	8% (2%, 15%)
	8	59%	53%	34%	7% (0%, 13%)
	12	61%	59%	35%	2% (-4%, 8%)
Mean Change DAS28 CRP	2	-1.26	-0.99	-0.60	-0.27 (-0.39, -0.15)
	4	-1.59	-1.33	-0.75	-0.27 (-0.40, -0.13)
	8	-1.86	-1.59	-0.95	-.0.26 (-0.41, -0.12)
	12	-1.97	-1.73	-1.02	-0.24 (-0.40, -0.09)
Mean Change HAQ-DI	2	-0.30	-0.23	-0.17	-0.06 (-0.11, -0.01)
	4	-0.37	-0.30	-0.20	-0.06 (-0.12, -0.01)
	8	-0.44	-0.39	-0.23	-0.05 (-0.11, -0.01)
	12	-0.47	-0.43	-0.24	-0.03 (-0.10, -0.03)

source: Tables 4.5, 4.6, and 4.7 of applicant response, NDA 207924 Seq 0048

3 LABELING RECOMMENDATIONS

This submission proposes that the Dosage and Administration section of the label be changed, from:

> "The recommended dose of Olumiant is 4 mg once daily. A dose of 2 mg once daily may be considered for patients who have achieved sustained control of disease activity with 4 mg once daily and are eligible for dose tapering."

to

> "The recommended dose of OLUMIANT is 2 mg once daily. For patients with an inadequate response or intolerance to more than one DMARD, a dose of 4 mg once daily is recommended."

However, the additional evidence provided in this submission is not adequate to support clinically relevant superior efficacy of the 4 mg dose over the 2 mg dose among patients with an inadequate response or intolerance to more than one DMARD.

4 APPENDICES

4.1 Baseline Demographic Characteristics

Table 22. Baseline Demographics, mITT Population, Study JADW

Category		Pbo	B2	B4
Randomized		176	174	177
Age (mean)		56	55	56
	< 65 years	77%	80%	77%
	< 75 years	98%	97%	97%
Male (%)		18%	21%	16%
Race N (%)				
	White	147 (83.5%)	144 (82.8%)	144 (82.8%)
	Black	8 (4.5%)	9 (5.2%)	7 (4.0%)
	Asian	11 (6.3%)	9 (5.2%)	12 (6.9%)
	Other	1 (0.6%)	0 (0%)	0 (0%)
Weight (kg)		82	83	81

Source: CSR Table JADW.11.1

Table 23. Baseline Demographics, mITT Population, Study JADX

Category		Pbo	B2	B4
Randomized		228	229	227
Age (mean)		51	52	52
	< 65 years	86%	86%	86%
	< 75 years	98%	98%	97%
Male (%)		17%	20%	18%
Race N (%)				
White		153 (67.1%)	156 (68.1%)	148 (65.5%)
	Black	10 (4.4%)	9 (3.9%)	9 (4.0%)
	Asian	60 (26.3%)	61 (26.6%)	59 (26.1%)
	Other	2 (0.8%)	1 (0.4%)	1 (0.4%)
Weight (kg)		76	75	78

Source: CSR Table JADX.11.1

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

ROBERT ABUGOV
05/21/2018

GREGORY P LEVIN
05/21/2018

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

SECONDARY STATISTICAL MEMORANDUM

CLINICAL STUDIES

NDA/BLA #:	NDA 207924
Drug Name:	Olumiant (baricitinib)
Indication:	Adult patients with moderately to severely active rheumatoid arthritis who have had an inadequate response or intolerance to methotrexate
Applicant:	Ely Lilly and Company
PDUFA Date:	January 15, 2017
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Gregory Levin
Concurring Reviewers:	Robert Abugov
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Raj Nair, Janet Maynard
Project Manager:	Jessica Lee

Keywords: dose-response, integrated efficacy, effect size, benefit-risk

SUMMARY

The efficacy of Olumiant (baricitinib) was evaluated in detail in the primary statistical review by Dr. Robert Abugov, and I agree with the key conclusions of Dr. Abugov's review. In particular, there is convincing statistical evidence that baricitinib, at both the 2 mg and 4 mg doses proposed for marketing, is effective for treatment of patients with moderately to severely active rheumatoid arthritis who have had an inadequate response or intolerance to methotrexate. The reader is referred to Dr. Abugov's review dated November 17, 2016 for additional details on the design and results of the phase 3 clinical trials supporting the effectiveness of baricitinib.

This memorandum seeks to elaborate on a specific aspect of the statistical review: a comparison of the efficacy of the 2 and 4 mg doses to help inform dosing recommendations. The primary statistical review focused on the two phase 3 studies JADX and JADW that contained both 2 mg and 4 mg treatment arms. Results from these two studies did not show consistent separation between the doses with respect to the primary efficacy endpoint ACR20, but there were trends toward slightly greater benefit on baricitinib 4 mg as compared to baricitinib 2 mg with respect to the continuous components of the composite ACR20 outcome in Study JADW. Similar trends were not seen in Study JADX (Table 1). The primary statistical review also noted that the long-term extension Study JADY included an evaluation of step-down dosing by randomizing patients with low disease activity on baricitinib 4 mg to remain on 4 mg or step down to 2 mg. There was no primary endpoint specified or any planned approach to control the type I error probability across the multiple endpoints evaluated, somewhat limiting the interpretability of results. However, there were relatively consistent trends toward better outcomes among patients remaining on the 4 mg treatment arm across several typical relevant efficacy endpoints (Table 2). Finally, Dr. Abugov's review noted that while there was convincing evidence of an effect of baricitinib 4 mg on radiographic progression, there was not sufficient evidence to support an effect of baricitinib 2 mg on radiographic progression. The totality of the available data from analyses of the two individual phase 3 studies JADX and JADW therefore provides some suggestion of slightly greater benefit with the 4 mg dose compared to the 2 mg dose, but the evidence is not consistent across both studies or across all endpoints.

It also of note that additional supportive integrated analyses were requested from the applicant during the review cycle to help compare the efficacy of the two doses. On June 24, 2016, the statistical review team, with concurrence from the clinical review team, asked the applicant to conduct efficacy analyses based on integrated patient-level data from Studies JADA, JADN, JADX, and JADW to further explore comparative efficacy between the two doses with respect to ACR20 and continuous secondary endpoints. The goal of requesting these supportive integrated analyses was to provide more precision in estimated comparisons between the two doses, given that the individual studies were not powered to detect potential differences between doses. The four selected studies were the phase 2/3 studies from the applicant's integrated safety analyses that contained both 2 mg and 4 mg treatment arms. Secondary endpoints such as DAS28 and HAQ-DI were selected because they are important continuous endpoints that may be more sensitive to differences than the binary primary endpoint ACR20.

The applicant submitted the requested results on July 21, 2016. Results showed trends toward slightly greater benefit on baricitinib 4 mg than 2 mg for ACR20 and continuous secondary endpoints such as DAS28 and HAQ-DI, with trends more pronounced at early time points (Table 3). These integrated analysis results were supportive of the results based on the individual phase 3 studies, and did not change the review team's conclusions about the dose comparison. In particular, there was some suggestion of greater benefit for 4 mg compared to 2 mg, but the evidence was not very strong and the estimated differences were relatively small. Such conclusions should be combined with conclusions about comparative safety between the doses to inform dosing recommendations.

Table 1. Results from Analyses of Studies JADX and JADW to Compare Baricitinib 4 mg versus 2 mg with Respect to ACR20 Response and Key Continuous ACR Components at Week 12

		Placebo	BARI 2 mg	BARI 4 mg	BARI 4 mg vs. BARI 2 mg Comparison[1] (95% CI)
JADW (N=525)	ACR20	27%	49%	55%	1.3 (0.8, 2.0)
	HAQ-DI	-0.17	-0.37	-0.41	-0.03 (-0.14, 0.07)
	TJC	-9	-12	-14	-2 (-5, 0)
	SJC	-5	-7	-9	-2 (-3, 0)
	Pain	-9	-17	-22	-5 (-10, 0)
	PaGA	-9	-20	-23	-3 (-7, 2)
	PhGA	-17	-31	-35	-5 (-9, 0)
	CRP	1	-5	-9	-4 (-8, 0)
JADX (N=660)	ACR20	39%	66%	62%	0.8 (0.6, 1.2)
	HAQ-DI	-0.36	-0.57	-0.56	0.01 (-0.08, 0.11)
	TJC	-10	-13	-13	0 (-1, 1)
	SJC	-6	-9	-9	0 (-2, 2)
	Pain	-16	-25	-23	2 (-2, 6)
	PaGA	-17	-25	-26	-1 (-5, 4)
	PhGA	-22	-32	-34	-3 (-6, 1)
	CRP	0	-9	-9	0 (-3, 2)

Source: Dr. Abugov's Primary Statistical Review and Applicant's CSRs

Abbreviations: BARI = baricitinib; CI = confidence interval; ACR20 = American College of Rheumatology 20% response; HAQ-DI = health assessment questionnaire disability index; TJC = tender joint count; SJC = swollen joint count; Pain = patient pain score; PaGA = patient global assessment; PhyGA = physician global assessment; CRP = C-reactive protein

[1] Metric for comparison is odds ratio for ACR20 response and difference in mean change for all other endpoints

Table 2. Results from Analyses of Study JADY to Compare Patients Remaining on Baricitinib 4 mg versus Patients Stepping Down to Baricitinib 2 mg with Respect to Mean Changes in Key Efficacy Endpoints at Week 12

	Remain on BARI 4 mg (N=146)	Step Down to BARI 2 mg (N=147)	P-value for Difference
CDAI	0.6	2.0	0.02
TJC	0.3	1.2	0.02
SJC	0.1	0.6	0.02
HAQ-DI	0.04	0.06	0.6
Pain	0.7	2.5	0.3
PaGA	1.2	2.5	0.5
PhyGA	0.0	2.7	0.05
DAS28	0.14	0.36	0.01
SDAI	0.69	2.19	0.02

Source: Dr. Abugov's Primary Statistical Review

Cell contents in first two columns are mean changes from time of step-down randomization to 12 weeks later

Abbreviations: BARI = baricitinib; CDAI = clinical disease activity index; TJC = tender joint count; SJC = swollen joint count; HAQ-DI = health assessment questionnaire disability index; Pain = patient pain score; PaGA = patient global assessment; PhyGA = physician global assessment; DAS28 = disease activity score based on 28 joints; SDAI = simplified disease activity index

Table 3. Results from Applicant's Integrated Analyses of Studies JADA, JADN, JADX, and JADW to Compare Baricitinib 4 mg versus 2 mg with Respect to Key Efficacy Endpoints over Time

		Placebo (N=551)	BARI 2 mg (N=479)	BARI 4 mg (N=480)	BARI 4 mg vs. BARI 2 mg Difference (95% CI)
ACR20 Response	Week 2	19%	31%	40%	9% (3%, 15%)
	Week 4	28%	45%	54%	8% (2%, 15%)
	Week 8	34%	53%	59%	7% (0%, 13%)
	Week 12	35%	59%	61%	2% (-4%, 8%)
DAS28 Mean Change	Week 2	-0.60	-0.99	-1.26	-0.27 (-0.39, -0.15)
	Week 4	-0.75	-1.33	-1.59	-0.27 (-0.40, -0.13)
	Week 8	-0.95	-1.59	-1.86	-0.26 (-0.41, -0.12)
	Week 12	-1.02	-1.73	-1.97	-0.24 (-0.40, -0.09)
HAQ-DI Mean Change	Week 2	-0.17	-0.23	-0.30	-0.06 (-0.11, -0.01)
	Week 4	-0.20	-0.30	-0.37	-0.06 (-0.12, -0.01)
	Week 8	-0.23	-0.39	-0.44	-0.05 (-0.11, 0.01)
	Week 12	-0.24	-0.43	-0.47	-0.03 (-0.10, 0.03)

Source: Tables 4.5–4.7, Applicant's July 21, 2016 response to FDA information request

Abbreviations: BARI = baricitinib; CI = confidence interval; ACR20 = American College of Rheumatology 20% response; DAS28 = disease activity score based on 28 joints; HAQ-DI = health assessment questionnaire disability index

---

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

GREGORY P LEVIN
01/06/2017

ROBERT ABUGOV
01/09/2017

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

Statistical Review

CLINICAL STUDIES

NDA / Sequence Number:	NDA 207924 / Seq 0000
Drug Name:	Baricitinib
Proposed Indication:	Rheumatoid arthritis (RA)
Applicant:	Eli Lilly and Company
Date(s):	Received: January 15, 2016 PDUFA Due Date: January 15, 2017
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Robert Abugov, Ph.D.
Concurring Reviewer:	Gregory Levin, Ph.D.
Statistics Supervisor:	Thomas Permutt, Ph.D. (Division Director)
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Raj Nair M.D. (Medical Officer) Janet Maynard, M.D. (Medical Team Leader) Sarah Yim, M.D. (Associate Division Director) Badrul A Chowdhury, M.D., Ph.D. (Division Director)
Project Manager:	Jessica Lee

Keywords: NDA review, Clinical Studies

Contents

Tables

Figures

1 EXECUTIVE SUMMARY

Four randomized, double-blind, trials demonstrate effectiveness of baricitinib for the treatment of rheumatoid arthritis. Studies JADX and JADW, which enrolled patients refractory to treatment with conventional disease modifying antirheumatic drugs (cDMARDS) and tumor necrosis factor (TNF) inhibitors respectively, randomized enrollees to baricitinib 4 mg oral tablet administered QD (B4), baricitinib 2 mg oral tablet administered QD (B2), or placebo (Pbo) as add-on therapies to cDMARDS, showed statistically significant effects of B4 and B2 compared to Pbo for the proportion of patients exhibiting a positive American College of Rheumatology 20 (ACR20) response. Study JADV, which randomized bDMARD naive patients refractory to treatment with methotrexate (MTX) to B4, adalimumab 40 mg SC Q2W (A), or Pbo as add-on therapies to MTX, showed statistically significant effects of B4 for ACR20 response compared to both Pbo and A. Study JADZ, which randomized biologic disease modifying antirheumatic drugs (bDMARD) and cDMARD naive patients to B4, B4 administered concurrently with MTX (B4MTX), or MTX, showed improved ACR20 response rates for either B4 or B4MTX compared to MTX monotherapy.

There were no significant differences between B4 and B2 for ACR20 response rate. However, exploratory analyses of the continuous components of ACR20 indicate nominal superiority of B4 to B2 in study JADV, and numeric, but not nominally significant trends, in study JADW which were consistent with superiority of B4 to B2.

Studies JADV and JADZ demonstrated that, compared to placebo, B4 inhibited radiographic progression. One study, JADX, evaluated the impact of B2 on radiographic progression. Using an exploratory analysis of all recorded data, regardless of patient adherence to randomized treatment, there was no statistically significant difference between B2 and Pbo. Because that analysis was only exploratory, and because convincing statistical significance was lacking for B2 in a single study, without additional radiographic evaluations of B2 in any additional phase 3 studies, inhibition of radiographic progression by B2 was not established.

In study JADZ, B4 was superior to MTX for the primary endpoint, ACR20 at week 24 ($p<.001$), with a p-value sufficiently low to argue in favor of substantial evidence of superiority in a single trial. However, superiority of B4 to MTX for inhibition of radiographic progression was not established ($p=.1$). In study JADV, superiority of B4 to A for the primary endpoint, ACR20 response at week 12, was less convincing for a single study ($p=.01$), and superiority of B4 to A for inhibition of radiographic progression was not established ($p=.6$).

2 INTRODUCTION

2.1 Overview

2.1.1 Drug Class and Indication

Baricitinib, a small molecule inhibitor of the Janus kinase family, is proposed for the treatment of rheumatoid arthritis.

2.1.2 History of Drug Development

IND 102204 to evaluate baricitinib for the treatment of RA was initiated on May 6, 2008, with concurrence for the first clinical trial granted on July 13, 2009.

In end-of-phase-2 meeting minutes on July 22, 2012, the Division acknowledged that evaluating radiographic progression in RA clinical trials has become difficult. In particular, because therapies are now available which minimize radiographic progression, inclusion of a placebo arm for an extended period has become ethically problematic. Nevertheless, long term radiographic outcomes cannot be reliably extrapolated from early trial data. The Division invited the applicant to propose alternative ideas regarding radiographic outcomes, but recommended that the applicant proceed first with clinical trials that evaluate core clinical outcomes, such as proportion of patients experiencing remission, ACR responses, and HAQ-DI.

Dosing regimens were also discussed. The applicant stated that 2 mg baricitinib daily (QD) was likely to prove ineffective, and that carrying only a 4 mg dose QD into phase 3 testing was being considered. The Division stated that carrying multiple doses into phase 3 would be necessary to determine a dose providing an acceptable balance between safety and efficacy. The applicant asked whether examining continued remission among patients stepping down from a 4 mg dose to a 2 mg dose would adequately characterize effects of the 2 mg dose. The Division stated that it would be preferable to include the 2 mg dose during the controlled period of the trials to directly determine its benefit – risk profile.

The applicant provided a preliminary protocol for study JADV, to be conducted in patients with inadequate response to methotrexate. The Division agreed with the proposed analysis of ACR20 in which missing data was to be imputed as non-response. For HAQ-DI, the Division stated that baseline observation carried forward (BOCF) imputation of data for patients discontinuing treatment due to adverse event would be acceptable, but asked that additional sensitivity analyses be designed, including use of BOCF for any patients who discontinue treatment or study prior to week 12.

The applicant asked whether results from an active adalimumab control arm could be included on the product label. The Division responded that, although that the content of the label would be a review issue, if a given study is considered adequate and important for inclusion in the prescribing information, and that study included active controls, then information from those control groups would likely be included.

The applicant asked whether inclusion of major clinical response (ACR 70 response for 6 continuous months) would be an acceptable endpoint. The placebo control would be discontinued at week 24, but patients would be evaluated at week 52 based on their original randomization. The Division stated that major clinical response would be impractical to pursue given the need to limit placebo-controlled periods in RA, and that other milestones of a high degree of disease control may be more suitable, such as various remission criteria.

The applicant also provided a preliminary protocol for study JADX, to be conducted in patients with inadequate responses to conventional disease modifying anti-rheumatic drugs, JADW, to be conducted in patients with inadequate responses to tumor necrosis factor inhibitors, and JADZ, to be conducted in patients with early RA who are naive to disease modifying anti-rheumatic drugs. The Division agreed with proposed gatekeeping strategies for studies JADW and JADZ, but disagreed with the gatekeeping strategy for study JADX. The Division further stated that it was unclear whether patient reported outcomes in studies JADV and JADX regarding morning joint stiffness, tiredness, and pain represented benefits distinct from control of disease activity in RA as represented by ACR response (b) (4)

For study JADZ, the proposed endpoint would compare baricitinib to methotrexate using a noninferiority margin of 12%. The Division responded that the proposed noninferiority margin was reasonable. The Division noted that protocol violations (e.g. non-adherence, patient switching treatment, misclassification of primary endpoint, or measurement errors) or patient discontinuation (i.e. missing data imputation) could affect the efficacy analysis both in the modified intent to treat (mITT) population and the per-protocol (PP) population. The assessment of efficacy would therefore need to evaluate results from analyses using both of these populations (i.e. mITT and PP), and any differences in results would need close examination. Furthermore, to minimize bias due to patient discontinuation or protocol violation, the Division recommended that the applicant minimize or avoid missing data due to patient discontinuation and protocol violations (e.g. non-adherence, and misclassification of the primary endpoint) by carefully planning the study design and by continually monitoring trial conduct.

The applicant proposed a matched observational comparator cohort of RA patients to be followed in parallel with the phase 3 safety extension trials to further character characterize risks associated with use of baricitinib. The Division responded that, while a matched, prospective observational cohort may provide better contextual information than the published literature, it

was still unlikely that comparison with an external group of patients would be adequate to allay concerns arising from comparisons to treatment groups within the study.

Regarding proposed long term extension study JADY, the division noted that inclusion of an active comparator arm would make the study more informative.

In all studies, the applicant proposed to provide baricitinib 4 mg as the rescue medication. The Division responded that use of an unproven experimental drug as rescue medication posed ethical challenges. The applicant agreed to provide further evidence that use of baricitinib 4 mg as rescue medication would not pose such challenges.

In a review of statistical analysis plans for the proposed studies communicated to the applicant on December 30, 2013, the Division disagreed with the applicant's proposal to impute missing radiographic data using linear extrapolation. Even at week 24, the Division noted that linear extrapolation could be associated with a disproportionate number of outliers. The Division also disagreed with the applicant's proposed multiple imputation sensitivity analysis because it was unclear whether adequate covariates exist to support use of multiple imputation. Further, even if such covariates existed, they were not prespecified in the statistical analysis plan, and any analyses would in consequence be suitable only for design of future trials. The applicant was instead advised to provide an analysis of treatment effect which included data collected after discontinuation of initially randomized treatment (retrieved data) if such data was available.

The applicant was further advised to provide an analysis of the proportion of patients without radiographic progression (change from baseline modified Sharp score ≤ 0). The Division further noted that discontinuations of initially randomized treatment at or beyond week 16 could pose a review issue for claims of continued effectiveness beyond week 16 in studies JADV, JADX, and JADW.

On December 1, 2014, the Division again communicated that the applicant's proposed imputation methods were unacceptable, and stated that tipping point multiple imputation sensitivity analyses should be conducted which would vary assumptions about the mean changes from baseline among treated and placebo patients who had missing data at study endpoint. These varying assumptions should include cases where dropouts on treatment have worse future outcomes than dropouts on control.

Further, the applicant was advised to ensure that analysis datasets include a column or columns which clearly indicate whether each observation was imputed, missing, observed while the patient was on randomized treatment, or observed after the patient discontinued randomized treatment.

In an evaluation of statistical analysis plans communicated to the applicant on April 10, 2014, the Division noted long term efficacy claims from study JADY should include analyses which incorporate adequate controls as well as control of type 1 error in the face of multiple endpoints. Further, the Division recommended continued collection of radiographic data from all randomized patients, regardless of adherence to study medication. The Division also noted that single study JADY may not be sufficient to demonstrate label claims for the proposed endpoints.

In a type C written response on October 10, 2014, the Division accepted the applicant's proposal to assess duration of morning stiffness in the clinical development program due to prior precedent in labeling. Of note, the prior precedent was for duration of morning stiffness, rather than severity of morning stiffness. Duration of morning stiffness should be included in the statistical hierarchy and evaluated in two studies if there were plans to propose labeling related to morning stiffness. Adequacy of data to support labeling claims would be a review issue. The Division further agreed to the applicant's proposal to evaluate treatment effects on median duration of morning joint stiffness using the Hodges-Lehmann point estimate and confidence interval.

Additionally, with respect to the primary and key secondary efficacy analyses and as recommended in the 2010 National Research Council report *The Prevention and Treatment of Missing Data in Clinical Trials*, the Division advised the applicant should explicitly define the causal estimand of interest that is being targeted, and justify that the estimand is meaningful and can be estimated with minimal and reasonable assumptions. The Division stated that the de facto treatment effect (i.e., the treatment effect in all patients randomized regardless of adherence) would be an appropriate estimand. For this reason, we believe that analyses incorporating all data post discontinuation of randomized treatment, regardless of any subsequent switch to other therapy are appropriate for primary or key secondary efficacy analyses. The Division recognized that this would require efficacy data to have been collected in subjects who prematurely discontinued study therapy.

In Reponses communicated to the applicant on June 5, 2014, the Division reiterated its position regarding the need for estimation of the de facto estimand, and requested that the applicant propose appropriate sensitivity analyses to evaluate the impact of missing data on treatment outcomes. The Division stated that use of LOCF, linear extrapolation, and MMRM on observed data only would not provide suitable sensitivity analyses. Further, the Division reiterated that use of linear extrapolation would not be suitable for a primary analysis of radiographic data, and also suggested that, to account for lack of normality in the distribution for radiographic outcome, the applicant may wish to consider use of van der Waerden normal scores in the analysis of covariance for change from baseline modified total Sharp score (mTSS).In a type C written response on January 16, 2015, the Division reiterated concerns regarding the proposed use of imputed data. For example, BOCF relies on strong and unverifiable assumptions about patient outcomes after withdrawal. In addition, it is not necessarily conservative with respect to the de

facto estimand, as patients who drop out from experimental therapy may have worse prognoses, and therefore go on to have worse future outcomes, than patients who drop out on the control arm. Therefore, the Division stated that additional supportive analyses would be critical.

The Division reiterated that the de facto estimand (i.e., the treatment effect in all patients randomized regardless of adherence) was an important measure of effectiveness. It appeared from the phase 3 protocols that patients who permanently discontinued study drug but maintained consent would continue to be followed for efficacy assessments. Therefore, for any endpoints proposed for inclusion on the product label or in the analysis hierarchy before endpoints proposed for inclusion on the product label, the applicant was advised to provide analyses which incorporate data collected from all randomized patients, regardless of whether they discontinued initially assigned randomized treatment, and regardless of whether they failed to actively maintain contact with their investigational site. In addition, regardless of the results of other supportive analyses, the applicant was advised to ensure that the NDA submission provides tipping point sensitivity analyses which include the possibility that dropouts on baricitinib have worse outcomes than dropouts on placebo.

Further, the applicant was advised to ensure that analysis datasets include a column or columns which clearly indicate whether each observation was imputed, missing, observed while the patient was on randomized treatment, or observed after the patient discontinued randomized treatment.

In pre-NDA meeting minutes communicated to the applicant on October 1, 2015, the Division noted that conclusions from study JADY regarding 'durability of effect' would be limited because it lacked a control or comparator group and because of a potential 'survivor bias' in which patients who remained on study would have been those who experience the most benefit from the drug.

The Division reiterated concerns regarding statistical analyses of primary and gated secondary endpoints regarding imputation of missing data, again noting that the de facto estimand would be considered an important measure of effectiveness. In addition, the Division repeated its advice to ensure, for all key efficacy endpoints, that the NDA submission provides tipping point sensitivity analyses which include the possibility that dropouts on baricitinib have worse outcomes than dropouts on placebo, and that analysis datasets include a column or columns which clearly indicate whether each observation was imputed, missing, observed while the patient was on randomized treatment, or observed after the patient discontinued randomized treatment.

The Division acknowledged the applicant's plan to collect data after discontinuation of randomized treatment. The applicant's response indicated that tipping point analyses would be conducted by adding a delta score to imputed scores from the baricitinib arm only. The Agency reiterated that tipping point analyses would require a two-dimensional array of delta scores,

constructed by using all observed data (including data collected after treatment discontinuation and use of ancillary therapies) and varying the delta score for missing data from the baricitinib and control arms independently. The Agency provided rationale for this request, noting that inference about the treatment effect (and therefore the tipping point) depends not only on the difference between assumed mean outcomes among dropouts on the two treatment arms, but also on the assumed mean outcome among dropouts on the placebo arm. An example was discussed, with different dropout rates on the two arms, to illustrate. The Agency recommended that a grid of deltas from initial imputed values be provided with, for example, deltas for baricitinib on the horizontal axis, and deltas for control on the vertical axis, with associated mean, 95% confidence interval, and p-value for the difference between treatment and control provided using multiple imputation for each pair of deltas.

Based on the discussion during the pre-NDA meeting, in a post-meeting response, the applicant agreed to include the following sensitivity analyses utilizing tipping point methodology in the Integrated Summary of Efficacy (ISE) document of the NDA submission:

> "Two-dimensional tipping point analyses where the sensitivity parameter (delta) for imputation of data from patients with missing data would be allowed to vary independently for the placebo group (or the active control group for studies without placebo) and for the experimental treatment (baricitinib) group.
>
> The tipping point results would be included in a 2-dimensional table summarizing (for each pair of delta values) the estimated treatment mean difference with 95% confidence interval and associated p-value from the primary analysis method. Alternatively, a graphical representation of the tipping point results could be provided in a manner similar to that shown in references provided by FDA.
>
> Efficacy variables to be included in the tipping point sensitivity analyses would include ACR20 and the set of continuous endpoints that are included in each study's statistical gatekeeping evaluation. All variables would be assessed at the primary analysis time point for the specific endpoint.
>
> To conduct the ACR20 tipping point analysis, sensitivity parameters would be converted back to the natural scale (i.e., probability of response) for inclusion in the results table."

2.1.3 Data Sources

Data sources for the current review are located at

\\cdsesub1\evsprod\NDA207924\0000\m5\datasets.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Review of this submission was made challenging by the applicant's failure to heed presubmission advice from the Division. Data, programs, and documentation were often inadequate to evaluate key endpoints. In particular, prior to completion of the statistical analysis plan, the applicant was repeatedly warned about FDA concerns with the use of linear extrapolation for imputation of missing radiographic data and the use of LOCF for imputation of missing data for other continuous endpoints (Table 1). In addition, the applicant was advised to analyze all data regardless of adherence to randomized treatment.

In spite of such cautions from FDA, summarized in Table 1, linear extrapolation was used for analysis of radiographic endpoints and LOCF was used for analysis of most other continuous endpoints, without submission of any of the requested analyses using all available data regardless of adherence to treatment. Programs used for analyses, and even the statistical analysis plans, were also omitted from the original submission, but were provided upon information request during the review.

Table 1. FDA Communications to Sponsor Prior to Submission

Date	Comment to Sponsor
July 2012	Avoid missing data for JADZ
	Do not use linear extrapolation for radiographic endpoints
	BOCF for AE discontinuations JADV HAQ-DI acceptable
December 2013	Linear extrapolation not usable
	Multiple imputation sensitivity analysis not acceptable
April 2014	Control of type 1 error over multiple endpoints for JADY required
December 2014	Multiple imputation tipping point sensitivity analyses required
	Datasets should delineate observation methods: imputation vs observed
October 2014	De facto estimand – include all patients regardless of adherence
June 2014	Reiterate need for evaluation of de facto estimand
	MMRM, LOCF, linear extrapolation not suitable sensitivity analyses
January 2015	Reiterated concerns around inappropriate imputation, including BOCF
	Reiterated need for de facto estimand based estimates
PreNDA Oct 2015	Reiterated need for de facto estimand based estimates
	Reiterated need for imputation information in datasets
	Reiterated need for tipping point analyses, in two-dimensions

Because of persistent issues with the appropriate handling of missing and post-rescue data and with the adequacy of documentation and program code to replicate and understand analyses conducted, a number of information requests were sent to the applicant during the review cycle:

Table 2. Information Requests to Applicant after Submission of NDA

Date	Comment to Sponsor
January 2016	Provide analysis programs for all endpoints proposed on label
February 2016	Provide study JAD-V, -W, -Z disposition tables at time of primary endpoint
March 2016	Describe and provide results from tipping point analyses
	Clarify whether analyzed data collected regardless of adherence
April 2016	Provide analysis programs for endpoints proposed on label which were not included in response to January 2016 IR
	Provide statistical analysis plans for studies in submission
	Explain flags used in statistical analyses
June 2016	Provide tipping point analyses based on de-facto estimand
August 2016	Clarify use of further undocumented flags in tipping point analyses
August 2016	Provide analysis programs for all endpoints proposed on label
September 2016	Provide de-facto analyses of ACR20 components

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

The present submission provides results from four randomized, double blind, parallel arm trials, JADV, JADW, JADX, and JADZ to evaluate the effects of baricitinib for the treatment of RA (Table 3, Table 4, and Table 5). Study JADV was conducted in patients with inadequate response to MTX, JADX in patients with inadequate responses to cDMARDs, JADW in patients with inadequate response to TNF inhibitors, and JADZ in patients completely naive to DMARDs. An additional trial, JADY, was an extension study examining, in this submission, the long term safety of baricitinib and the effects of reducing the baricitinib dose among patients successfully responding to treatment with baricitinib (Table 5).

Study JADV (Table 3) was a parallel group, double blind, double dummy, placebo controlled trial randomizing 1260 adult patients with moderately to severely active RA who were bDMARD naive, had inadequate response to MTX, and at least some erosive joint damage, to B4, A, or Pbo. Randomization was in a 3:2:3 ratio stratified by region and number of eroded

joints. All treatments were added to therapy with MTX. The Pbo treatment group continued to W24, with the possibility of rescue to B4 offered at W16 to patients in the Pbo and A trial arms. The primary endpoint was W12 ACR20[1] on randomized treatment.

Study JADX (Table 4) was a parallel group, double blind, double dummy, placebo controlled trial randomizing 660 adult patients with moderately to severely active RA who were bDMARD naive with inadequate response to cDMARDs, to B4, B2, or Pbo. Randomization was in a 1:1:1 ratio stratified by region and presence of joint erosion. All treatments were added to therapy with cDMARDs. The Pbo group was continued to W24, with the possibility of rescue to B4 offered at W16 to patients in the Pbo and B2 trial arms. The primary endpoint was W12 ACR20 on randomized treatment without rescue.

Study JADW (Table 4) was a parallel group, double blind, double dummy, placebo controlled trial randomizing 525 adult patients with moderately to severely active RA who were anti-TNF inadequate responders and who were on stable doses of cDMARDs, to B4, B2, or Pbo. Randomization was in a 1:1:1 ratio stratified by region and presence of joint erosion. All treatments were added to therapy with cDMARDs. The Pbo group was continued to W24, with rescue to B4 offered at W16 to patients in the Pbo and B2 trial arms. The primary endpoint was W12 ACR20 on randomized treatment without rescue.

Study JADZ (Table 5) was a parallel group, double blind, double dummy active controlled trial randomizing 550 adult patients with moderately to severely active RA who had limited experience with MTX, who were naive to other cDMARDS, and naive to bDMARDs, to B4, B4MTX, or MTX. Randomization was in a 3:4:4 ratio stratified by region and presence of joint erosion. All treatment groups continued to W52, with rescue to B4MTX offered to B4 and Pbo patients at W24. The primary endpoint evaluated noninferiority of B4 to MTX alone for W24 ACR 20 on randomized treatment without rescue.

The portion of extension study JADY (Table 5) evaluated in this review enrolled patients who completed studies JAD-V, -W, -X, and -Z randomized to B4 who achieved low disease activity (CDAI $\leq$ 10 for studies JADV, JADX, and JADW) or remission (CDAI < 2.8 for study JADZ). All patients remained on the add-on medications from their respective studies, with half rerandomized to receive a reduction in dose from B4 to B2. The study was parallel group, double

---

[1] ACR20 response is defined as a reduction of at least 20% in the number of swollen joints, a reduction of at least 20% in the number of tender joints, and a reduction of 20% in three of the following five parameters: physician global assessment, patient global assessment, patient assessment of pain, C-reactive protein or erythrocyte sedimentation rate, or health assessment questionnaire disability index (HAQ-DI).

blind, and double dummy. Rescue was allowed after step down in this extension study. Patients rerandomized to B2 who originated from studies JADV, JADW, and JADX were eligible for rescue to B4 at or after 12 weeks following enrollment into JADY. For patients enrolled from study JADZ, rescue via increases in MTX or other cDMARDS was provided at any time during the study.

In studies JADV, JADW, JADX, and JADZ, rescued patients without satisfactory improvement following at least four weeks of rescue treatment were discontinued from treatment. In study JADV, rescued patients who continued to meet nonresponder criteria for 4 weeks after rescue or at any time point thereafter were discontinued from treatment. In studies JADX and JADW patient rescued at week 16 were discontinued from treatment if they met nonresponder criteria at Week 20. In study JADZ, rescued patients not experiencing improvements in signs and symptoms following at least 4 weeks of rescue treatment were discontinued from treatment, with the understanding that the benefits of MTX as titrated (10 mg Q1W increasing in 5 mg increments Q4W to 20 mg) may not be evident within 4 weeks of rescue.

In study JADY, discontinuation of treatment was recommended if, after three months of rescue therapy and adjustment of concomitant therapies, CDAI was ≤ 22 for patients originating from studies JAD-V, -W and X or if CDAI was ≤ 10 for patients originating from study JADZ. However, such rescue will not be of concern here, (b) (4)

Baseline radiographs were obtained for each patient at screening, with additional radiographs recorded for study JADV at weeks 16, 24, and W52, for study JADX at week 24, and for study JADZ at weeks 12, 24, and 52. For study JADX, radiographs were recorded at week 16 for patients receiving rescue therapy. In both Studies JADV and JADX, radiographs were recorded at Week 24 for patients who received rescue at Week 16 but remained in the study. Among patients discontinuing prematurely from the study, radiographs were performed at the early termination visit only if the immediately prior radiographs were recorded more than 12 weeks earlier. Radiographs were not recorded in study JADW.

Table 3. Design for Study JADV[1]

Study	Design[a]	Population	Endpoints
JADV	B4 + MTX	Mod to sev active RA	Primary:
	A + MTX	Age ≥ 18	ACR20 on treatment W12
(II)[b]	Pbo + MTX	MTX inadequate resp	
		Erosive damage	Major Secondary:
	DB, DD, PG	bDMARD naive	ΔHAQ-DI W12
		TJC and SJC ≥ 6	mTSS W24
	Rescue[c]:	≥ 1 erosion	DAS28 W12
	Pbo, A from W16	RF or ACPA + if < 3	SDAI ≤ 3.3 W12
		CRP ≥ 1.2 ULN	ACR20 W12 B4 vs A
	Pbo + MTX to W24		Morning stiffness dur W12
	B4 + MTX to W52	exclusions:	DAS28 B4 vs A W12
	A + MTX to W52	> 10 mg/day predn	Morning stiffness sev W12
		MTX, HCQ and SSZ	Worst tiredness W12
			Worst pain W12
		N=480/320/480	
		Strat: region, joint erosion (<3, ≥3 previous)	

[1.] Abbreviations: B4 and B2 baricitinib tablets 2 mg or 4 mg QD, MTX methotrexate, A adalimumab 40 mg SC Q2W, DB double blind, DD double dummy, PG parallel group, W12 W16 W24 weeks 12, 16, and 24 respectively, b biologic, DMARD disease modifying antirheumatic drug, TJC tender joint count, SJC swollen joint count, erosion in hand, wrist, or foot joints based on radiographs, RF rheumatoid factor, ACPA anti-citrullinated protein antibody, CRP c-reactive protein, predn prednisone, HCQ hydroxychloroquine, SSZ sulfasalazine, NSAID nonsteroidal anti-inflammatory, ULN upper limit of normal, ACR20 20% improvement in American College of Rheumatology criteria Δ = change from baseline, HAQ-DI health assessment questionnaire disability index, mTSS modified total Sharp score, DAS28 Disease Activity Score 28 using high-sensitivity C-reactive protein, SDAI simplified disease activity index

[a] Dose for renally impaired patients randomized or rescued to baricitinib is B2 for all studies in this submission

[b] (b) (4)

[c] Rescue to B4 if lack of improvement of at least 20% in both tender joint count and swollen joint count at both Week 14 and Week 16 compared to baseline. After Week 16, rescue therapy will be offered to patients at the discretion of the investigator based on TJCs and SJCs.

Table 4. Designs for Studies JADX and JADW[1.]

Study	Design	Population	Endpoints
JADX	B4 + cDMARD	Mod to severe active RA	Primary (W12):
	B2 + cDMARD	Age ≥ 18	ACR20 on treatment
(III)	Pbo + cDMARD	cDMARD inadequate resp	
		On cDMARD (incl MTX)	Major Secondary (W12):
	DB, DD, PG	bDMARD naive	B4 vs P
		TJC and SJC ≥ 6	ΔHAQ-DI
	Rescue[a]:	CRP ≥ 1.2 ULN	DAS28
	Pbo, B2 from W16		SDAI ≤ 3.3
		exclusions:	Morning stiffness dur
	End W24	> 10 mg/day predn	B2 vs P
		MTX and HCQ and SSZ	ACR 20
			Worst Tiredness
		N = 220/220/220	B4 vs P
		Strat: region,	Morning stiffness sev
		joint erosion (Y/N)	Worst tiredness
			Worst pain
			B2 vs P
			HAQ-DI
			DAS28
			SDAI ≤ 3.3
JADW	B4 + cDMARD	Mod to severe active RA	Primary (W12):
	B2 + cDMARD	Age ≥ 18	ACR20 on trt
(IV)	Pbo + cDMARD	anti-TNF inadequate resp	
		On cDMARD	Major Secondary (W12):
		TJC and SJC ≥ 6	Δ HAQ-DI
	DB, DD, PG	CRP or hsCRP ≥ 1.0 ULN	DAS28-hsCRP
			SDAI ≤ 3.3
	Rescue[a]:	exclusions:	
	Pbo, B2 from W16	> 10 mg/day predn	
		MTX and HCQ and SZZ	
	End W24		
		N = 175/175/175	
		Strat: region, bDMARD	
		(<3, ≥3 previous)	

[1.] Abbreviations: cDMARD conventional DMARD including MTX, TNF tumor necrosis factor

[a] Rescue to B4 if lack of improvement of at least 20% in both TJC and SJC at both week 14 and week 16 compared to baseline. At week 20, rescue therapy offered to patients at the discretion of the investigator based on TJCs and SJCs.

Table 5. Designs for Studies JADZ and JADY

Study	Design	Population	Endpoints
JADZ	B4	Mod to severe active RA	Primary (W24):
	B4MTX	Age ≥ 18	ACR20 on trt
(I)	MTX	b and c DMARD naive	B4 to MTX
		≤ 3 weeks MTX	
		+ RF or anti CCP	Major Secondary (W24):
	DB, DD, PG, AC	TJC and SJC ≥ 6	B4 + MTX vs MTX
		CRP ≥ 1.2 ULN	ACR20 on trt
	Rescue[a]:		HAQ-DI
	MTX, B4 from W24	exclusions:	DAS28
		> 10 mg/day predn	B4 vs MTX
	End W52		ACR20 on trt
		N = 150/200/200	HAQ-DI
		Strat: region	DAS28-hsCRP
		joint erosion (Y/N)	B4 + MTX vs MTX
			mTSS
			SDAI ≤ 3.3
			B4 vs MTX
			SDAI ≤ 3.3
JADY	B4 +	B4 from JADV, W, X, Z	Exploratory Only:
	B2 +	B4 ≥ 15 months	ACR20, 50, 70 on trt
(V)		Not rescued original study	DAS28-hsCRP
	Extension Study	JADV, W, X CDAI ≤ 10 for	DAS28-ESR
		≥ 12 weeks in study JADY	LDA and remissn
	DB, DD, PG	JADZ CDAI ≤ 2.8 for ≥ 12	EULAR28 response
	Rescue W12 JADW, X, V	weeks in study JADY	ACR/EULAR remissn
	Rescue anytime JADZ		HAQ-DI ≥ 0.22
	Escape W24 JADW,X, V	N = 175/175/175	HAQ-DI ≥ 0.3
		Strat: region, bDMARD	CDAI and SDAI
		(<3, ≥3 previous)	LDA and remissn
			mTSS, JSN, JSE
			Morning joint stiffness

[1.] Abbreviations: B4MTX B4 administered concomitantly with MTX, RF rheumatoid factor, anti CCP cyclic citrullinated peptide antibody, EULAR European League Against Rheumatism, CDAI clinical disease activity index, LDA low disease activity CDAI ≤ 10, remissn remission

[a.] Rescue to B4MTX if lack of improvement of at least 20% in both TJC and SJC at Week 24 compared to baseline.

3.2.2 Statistical Methodologies

3.2.2.1 Studies JADW, JADX, JADY, and JADZ

Categorical efficacy variables for studies JAD -W, -X, -Y, and -Z were analyzed using logistic regression, with independent factors treatment and any stratification variables included in the randomization (see Table 3 and Table 4). Continuous efficacy variables were analyzed using ANCOVA, with independent factors treatment, baseline, and any stratification variables included in the randomization (see Table 3 and Table 4).

Efficacy analyses were generally conducted on the modified-intent-to-treat (mITT) population, defined as patients receiving at least one dose of the study drug. An exception was the analysis for radiographic progression, in which analyzed patients not only had to receive one dose of the study drug, but also were required to have non-missing baseline measurement as well as at least one non-missing post baseline measurement. Type 1 error rates in the face of multiple endpoints and doses, was controlled at the .05 level of significance using analysis hierarchies defined graphically as in Bretz et al,[2]. Analysis hierarchies particular to each study are provided in Appendix 6.2.

Non-response was recorded for binary response data missing or collected after permanent treatment discontinuation or escape. Therefore, we consider these variables to be composite endpoints defined by remaining on randomized treatment through the time point of interest and meeting the binary response criteria at the time point of interest. For key secondary endpoints, missing continuous data was imputed using modified baseline observation carried forward (mBOCF), with BOCF used after patients discontinued the study or study treatment due to an adverse event, and last observation carried forward (LOCF) used after patients who discontinued the study or study treatment due to other reasons. For most other secondary endpoints, missing continuous data was imputed using modified last observation carried forward (mLOCF), in which data for patients who received rescue therapy was imputed using LOCF from the time at which rescue occurred and in which, for all other patients who discontinued the study for any reason, LOCF was conducted by carrying forward the last post-baseline observation.

Radiographic data missing or collected after treatment discontinuation or escape was imputed using linear extrapolation, with analysis via ANCOVA for the timepoint of interest. For patient

[2] Bretz, F., Maurer, W., Brannath, W. and Posch, M. (2009) A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine 28(4), 586-604

reported outcomes, such as duration and severity of morning joint stiffness, worst tiredness, worst joint pain, and recurrence of morning stiffness, endpoints were defined as the mean over seven days prior to study endpoint. If less than four days of data was available, the seven day window was shifted one day back in time until at least four days of data was available in the window. LOCF for the last seven days of treatment was used for patients discontinuing the study, with the window similarly shifted for patients with fewer than four days of data. If there were no windows with at least four days of data, the first post-baseline observation was carried forward to study endpoint.

Two-dimensional tipping point sensitivity analyses were conducted to evaluate the potential effect of missing data on the reliability of efficacy results. Analyses were conducted for the primary endpoint, ACR20 at week 12, for DAS-28 hsCRP, HAQ-DI, and mTSS in studies JADV and JADZ, as well as other secondary endpoints such as duration of morning stiffness. For binary endpoints, the tipping point analyses were conducted using single imputation and, for continuous endpoints, the tipping point analyses were conducted using multiple imputation.

3.2.2.2 Study JADY

Although the analyses provided by the applicant for study JADY were similar to those for studies JAD-W, -X, -Y, and -Z, all endpoints in study JADY were examined without control of type 1 error, i.e., without any prespecified analysis hierarchy. As discussed later, this design flaw has obscured interpretation of treatment effects.

3.2.2.3 Further Notes on Tipping Point Analyses

The applicant conducted tipping point analyses to evaluate the sensitivity of results obtained from imputation of missing data. The tipping point analyses first imputed missing data using Markov chain Monte Carlo (MCMC) simulations to establish monotonicity of missingness, with simulations based on regression for the timepoint of interest. Data at other timepoints and treatment were used as covariates in these imputations based on the missing-at random assumption. The potential impact of missing data on the de facto estimand was then examined by adding or subtracting 'shifts' to the imputed data. In general, shifts applied by the applicant to the treatment arm were extensively explored, while shifts applied to control arms were minimal. However, because the impact of a shift on mean treatment effect is proportional to the amount of missing data in the trial arm to which the shift is applied, application of only minimal shifts to control arms, or more generally, to trial arms with more trial discontinuations, could omit shifts which lead to reversal of statistically significant results, and which are therefore important to include in the analyses. Therefore, I conducted additional tipping point analysis, with methods described in the following paragraphs.

The tipping point analyses provided in this review include four shifts in each trial arm:

1. missing data imputed but not shifted, designated as the 'zero shift imputation;' this corresponds to a missing-at-random assumption conditional on treatment and data at other time points,

2. imputed missing data shifted so that the adjusted mean[3] of imputed data for placebo equals the adjusted mean of imputed data for treatment (cells will be shaded in blue); this will be designated as the 'null imputation,'

3. imputed missing data shifted so that the adjusted mean benefit of imputed data for control exceeds that of the treatment by one half the overall treatment effect seen in 1 above; this will be designated as the 'null-plus-one-half imputation,' and

4. imputed missing data shifted so that the adjusted mean benefit of imputed data for control exceeds that of the treatment by the overall treatment effect seen in 1 above; this will be designated as the 'null-plus-one imputation.'

Note that, shifts applied to means in 1 to 4 above do not necessarily extend far enough to identify tipping points. Instead, the shifts are applied to evaluate estimates and evidence of treatment effects under a specific set of plausible and implausible assumed differences between treatments among patients whose outcomes are unknown. The first non-zero shift in the tipping point analyses, the null imputation, corresponds to the null hypothesis in these superiority trials, i.e. the assumption that, among patients with missing data, patients discontinuing from treatment have the same outcomes on average as those from the control.

The null imputation may be used to estimate the treatment effect. It provides several advantages over other methods. First, the null imputation does not require the strong and unverifiable assumption that data is missing at random (MAR). The MAR assumption is likely implausible for the evaluation of the de facto estimand given that patients who drop out are no longer taking study treatment, and any expected effects of treatment would usually not be expected to persist after treatment discontinuation. In addition, the null imputation arguably provides a middle ground between imputation scenarios; it could be plausibly argued that patients who withdraw from the experimental treatment arm are a sicker subset of patients, and therefore potentially more resistant to treatment, than dropouts on the control arm, and that if their outcomes had been recorded after withdrawal they would have been worse than those with missing data on placebo.

[3] calculated using the statistical model used to test efficacy, often called the 'least square mean.'

On the other hand, it could also be plausibly argued that patients on the treatment arm who withdrew had superior later outcomes to those who withdrew from placebo, because treatment had some lasting effects. In the absence of real data, in many settings it may be reasonable to choose a middle ground between these scenarios and to estimate treatment effects, the null imputation, which provides neither an advantage nor disadvantage of treatment relative to the control arm for patients with missing data. Third, with the use of such an approach that does not assume persistence of treatment effects among patients who dropped out, one may hope that missing data will quickly become a rarity.

On the other hand, the null imputation is only one of a possible range of reasonable scenarios and, like any other missing data approach, it relies on an unverifiable assumption about the missing outcomes.

Given the above advantages and disadvantages, this reviewer advocates using the null imputation to estimate treatment effects unless convincing evidence is provided that outcomes from patients who adhere to randomized treatment and remain in the trial are predictive of outcomes in patients who discontinue randomized treatment or who discontinue participation in the trial.

The other designated shifts evaluate the extent to which, compared to control, poor de-facto outcomes among patients who withdraw from the experimental treatment may impact estimates and evidence of treatment effects. In this particular submission, if the null plus one imputation still provides statistically significant effects, superiority over placebo will be considered robust, and if statistical significance is lacking at the null imputation, superiority over placebo will be seen as not convincingly demonstrated. However, note carefully that use of the null plus one imputation is only one of a number of plausible metrics for robustness — other shifts, e.g., by number of standard deviations from the mean, could also be employed.

To achieve the shift for the null imputation in 2 above, the difference between treatment and control in the imputed data was estimated using the prespecified statistical model for efficacy, and that estimate was subtracted or added to imputed patient outcomes at zero shift. To achieve the shift for the null-plus one-half imputation in 3 above, the estimated difference between treatment and control among all patients at the zero shift, was halved and subtracted or added to patient outcomes imputed according to the null imputation. Similarly, for the null plus one shift, the full treatment effect at the zero shift was subtracted or added to patient outcomes imputed according to the null imputation.

In the applicant's tipping point analyses for change from baseline, patients were not considered missing, i.e. were completely excluded from the tipping point analyses, when their baseline values were not recorded. From the limited viewpoint of assessing the impact of treatment on outcome, exclusion of such patients from the tipping point analyses could be justified because data collected before randomization is expected to be missing completely at random. However, if

the failure of patients to appear for collection of baseline data was associated with patient outcome, it is possible that the population targeted in the label would not match that evaluated in the trial. For tipping point analyses provided in this review, both situations were evaluated, i.e., the case in which patients without baseline scores were eliminated from the tipping point analysis, and the case in which patients without baseline scores were considered missing in the tipping point analysis.

Operationally, each tipping point analysis was conducted with 20 imputations. Monotonicity of missingness was first ensured by simulating imputed data using random draws from the posterior predictive distribution of regression parameters. Then, with monotonicity established, remaining timepoints were imputed using Markov chain Monte Carlo (MCMC) simulations based on the normal distribution.

The applicant's tipping point analyses had the potential to underestimate between imputation variance. In particular, only a single imputation was conducted to ensure monotonicity, and then, based on values from that single imputation, multiple imputations were conducted for remaining timepoints. Thus, for each patient, the multiple imputations were correlated, all stemming from the same initial imputation for monotonicity. The tipping point analyses presented in this review rectify this potential problem by ensuring that each imputation is generated independently, with no overlap among imputations.

Each of 20 initial imputations conducted without shifts were analyzed using the preplanned analysis and compiled into a single result using Rubin's method. The compiled result was considered the treatment effect at shift (0,0) for treatment and control respectively. For the tipping point analyses, shifts were then added to these 20 imputed datasets and reanalyzed.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

There were no obvious differences between treatments for baseline characteristics in the submitted studies (Appendix A; Table 67, Table 68, Table 69, and Table 70). The mean age of patients in the treatment arms in these studies ranged from 49 to 56 years, and a majority were female and white or Asian. Patterns of patient disposition did not contradict efficacy of B4 or B2 (Table 4, Table 6, Table 8, and Table 12). Compared to control, baricitinib treated patients tended to have lower early discontinuation rates, higher adverse event rates, and lower discontinuation rates due to lack of efficacy.

Table 6. Patient Disposition. Week 12, mITT Population, Study JADV

	Pbo N (%)	**B4 N (%)**	**A N (%)**
Randomized	488	487	330
Completed W12	454 (93%)	472 (97%)	312 (95%)
Early discontinuation	34 (7%)	15 (3%)	18 (5%)
Adverse event	11 (2.3%)	11 (2.3%)	6 (1.8%)
Death	0	0	0
Lack of efficacy	11 (2%)	1 (0.2%)	2 (0.6%)
Entry criteria not met	0 (0%)	1 (0.2%)	2 (0.6%)
Patient withdrawal	12 (2.5%)	2 (0.4%)	9 (2.7%)

Source: Sequence 0006 Regulatory Response Table 1

Table 7. Patient Disposition. Week 24, mITT Population, Study JADV

	Pbo N (%)	**B4 N (%)**	**A N (%)**
Randomized	488	487	330
Rescued	128 (26%)	35 (7%)	40 (12%)
Completed W24	435 (89%)	458 (94%)	306 (93%)
After rescue	125 (26%)	34 (7%)	39 (12%)
Early discontinuation (w/out rescue)	53 (11%)	29 (6%)	24 (7%)
Adverse event	16 (3.3%)	18 (3.7%)	7 (2.1%)
Death	0	2 (0.4%)	0
Lack of efficacy	16 (3.3%)	1 (0.2%)	3 (0.9%)
Entry criteria not met	0	1 (0.2%)	1 (0.2%)
Patient withdrawal	12 (2.5%)	2 (0.4%)	9 (2.7%)

Source: Sequence CSR Table JADV 10.1

Table 8. Patient Disposition. Week 12, mITT Population, Study JADW

	Pbo N (%)	**B2 N (%)**	**B4 N (%)**
Randomized	176	174	177
Completed W12	154 (88%)	160 (92%)	166 (94%)
Early discontinuation	22 (12%)	14 (8%)	11 (6%)
Adverse event	4 (2.3%)	6 (3.4%)	7 (4.0%)
Lack of efficacy	12 (6.8%)	3 (1.7%)	2 (1.1%)
Entry criteria not met	0 (0%)	0 (0%)	0 (0%)
Patient withdrawal	5 (2.8%)	5 (2.9%)	1 (0.6%)

Source: Sequence 0006 Regulatory Response Table 2

Table 9. Patient Disposition. Week 24, mITT Population, Study JADW

	Pbo N (%)	**B2 N (%)**	**B4 N (%)**
Randomized	176	174	177
Rescued	56 (32%)	38 (22%)	33 (19%)
Completed W24	89 (50%)	119 (68%)	126 (71%)
Post-rescue	55 (31%)	38 (22%)	32 (18%)
Early discontinuation (w/out rescue)	31 (18%)	17 (10%)	18 (10%)
Adverse event	7 (4.0%)	7 (4.0%)	10 (5.6%)
Death	0	0	1 (0.6%)
Lack of efficacy	15 (8.5%)	4 (2.3%)	3 (2.8%)
Entry criteria not met	0	0	0
Patient withdrawal	7 (4.0%)	6 (3.4%)	1 (0.6%)

Source: Sequence CSR Table JADW 10.1

Table 10. Patient Disposition. Week 12, mITT Population, Study JADX

	Pbo N (%)	**B2 N (%)**	**B4 N (%)**
Randomized	228	229	227
Completed W12	206 (90%)	216 (94%)	212 (93%)
Early discontinuation	22 (10%)	13 (6%)	15 (7%)
Adverse event	5 (2.2%)	7 (3.1%)	6 (2.6%)
Lack of efficacy	4 (1.8%)	3 (1.3%)	0 (0%)
Entry criteria not met	0 (0%)	0 (0%)	0 (0%)
Patient withdrawal	10 (4.4%)	3 (1.3%)	6 (2.6%)

Source: Sequence 0006 Regulatory Response Table 3

Table 11. Patient Disposition. Week 24, mITT Population, Study JADX

	Pbo N (%)	**B2 N (%)**	**B4 N (%)**
Randomized	228	229	227
Rescued	55 (24%)	21 (9%)	15 (7%)
Completed W24	145 (64%)	189 (83%)	189 (83%)
Post-rescue	54 (24%)	20 (9%)	14 (6%)
Early discontinuation (w/out rescue)	28 (12%)	19 (8%)	23 (11%)
Adverse event	2 (0.9%)	0	0
Death	2 (0.9%)	0	0
Lack of efficacy	7 (3.1%)	3 (1.3%)	0
Patient withdrawal	11 (4.8%)	5 (2.2%)	8 (3.5%)

Source: Sequence CSR Table JADX 10.1

Table 12. Patient Disposition. Week 24, mITT Population, Study JADZ

	MTX N (%)	**B4 N (%)**	**B4 + MTX N (%)**
Randomized	210	159	215
Completed W24	182 (87%)	145 (91%)	192 (89%)
Early discontinuation	28 (13%)	14 (9%)	23 (11%)
Adverse event	2 (1.0%)	6 (3.8%)	14 (6.5%)
Lack of efficacy	10 (4.8%)	1 (0.6%)	0 (0%)
Entry criteria not met	1 (0.5%)	0 (0%)	0 (0%)
Patient withdrawal	10 (4.8%)	4 (2.5%)	6 (2.8%)

Source: CSR Table JADZ.10.1

Table 13. Patient Disposition. Week 52, mITT Population, Study JADZ

	MTX N (%)	**B4 N (%)**	**B4 + MTX N (%)**
Randomized	210	159	215
Rescued	26 (12%)	7 (4%)	6(3%)
Completed W52	140 (76%)	131 (86%)	169 (81%)
Early discontinuation	44 (24%)	21 (14%)	40 (19%)
Adverse event	7 (3.8%)	9 (5.9%)	23 (11.0%)
Lack of efficacy	12 (6.5%)	2 (1.3%)	1 (0.5%)
Entry criteria not met	1 (0.5%)	1 (0.7%)	1 (0.5%)
Patient withdrawal	15 (8.2%)	6 (3.9%)	13 (6.2%)

Source: CSR Table JADZ.10.1

3.2.4 Results and Conclusions

3.2.4.1 Primary Endpoint: ACR20

Statistically significant differences between baricitinib doses and placebo were seen for B4 and B2 at week 12 in studies JAD-V, -W, and -X (Table 14). Differences at week 12 between B4 and B2 in studies JAD-W and –X were not statistically significant and did not trend in consistent directions (Table 14). Results at week 24 were similar to those at week 12 (Table 14).

Table 14. ACR20 and Remain on Randomized Treatment, Studies JADV, W, X

Wk	Study (backgrd)	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
		B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
12	JADV (MTX)	70 (339/487)		40 (196/488)	3.6 (<.001) (2.7, 4.7)		
	JADW (cDMARD)	55 (98/177)	49 (85/174)	27 (48/176)	3.4 (<.001) (2.2, 5.4)	2.7 (<.001) (1.7, 4.2)	1.3 (0.3) (0.8, 2)
	JADX (cDMARD)	62 (140/227)	66 (151/229)	39 (90/228)	2.5 (<.001) (1.7, 3.7)	3.0 (<.001) (2.0, 4.4)	0.8 (.4) (0.6, 1.2)
24[a]	JADV (MTX)	74 (360/487)		37 (179/488)	5 (<.001) (3.8, 6.7)		
	JADW (cDMARD)	46 (82/177)	45 (78/174)	27 (48/176)	2.4 (<.001) (1.5, 3.7)	2.3 (<.001) (1.5, 3.6)	1.0 (.9) (0.7, 1.6)
	JADX (cDMARD)	65 (148/227)	61 (140/229)	42 (96/228)	2.6 (<.001) (1.8, 3.9)	2.2 (<.001) (1.5, 3.2)	1.2 (0.3) (0.8, 1.8)

source: reviewer programs ACR JADV 2016 03 14.sas, ACR JADW W12 &24 2016 03 14.sas, ACR JADX W12 &24 2016 03 14.sas, CSR Tables JADV.11.9, .11.21, JADW.11.9, 11.14, JADX 11.9, 11.18

[a] Exploratory endpoints, p-values nominal only

Results for ACR50 and ACR70 were similar to those for ACR20, with significant differences between placebo and both baricitinib doses in studies JAD-V, -W, and –X (Table 15 and Table 16), and with no significant differences between the B4 and B2 doses.

Table 15. ACR50 and Remain on Randomized Treatment, Studies JADV, W, X

Wk	Study (backgrd)	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
		B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
12	JADV (MTX)	45 (219/487)		17 (82/488)	4.2 (<.001) (3.1, 5.7)		
	JADW (cDMARD)	28 (50/177)	20 (35/174)	8 (14/176)	4.7 (<.001) (2.5, 8.9)	3 (.001) (1.6, 5.9)	1.5 (.09) (0.9, 2.5)
	JADX (cDMARD)	33 (76/227)	34 (77/229)	13 (29/228)	3.5 (<.001) (2.2, 5.7)	3.5 (<.001) (2.2, 5.6)	1 (.9) (0.7, 1.5)
24[a]	JADV (MTX)	51 (246/487)		19 (94/488)	4.4 (<.001) (3.3, 5.9)		
	JADW (cDMARD)	29 (52/177)	23 (40/174)	13 (23/176)	2.8 (<.001) (1.6, 4.9)	2 (.02) (1.2, 3.6)	1.4 (.2) (0.8, 2.3)
	JADX (cDMARD)	44 (100/227)	41 (95/229)	21 (49/228)	2.9 (<.001) (1.9, 4.5)	2.6 (<.001) (1.7, 4)	1.1 (.6) (0.8, 1.6)

source: reviewer programs ACR JADV 2016 03 14.sas, ACR JADW W12 &24 2016 03 14.sas, ACR JADX W12 &24 2016 03 14.sas, CSR Tables JADV 11.23, JADW 11.15, JADX 11.19

[a] Exploratory endpoints, p-values nominal only

Table 16. ACR70 and Remain on Randomized Treatment, Studies JADV, W, X

Wk	Study (backgrd)	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
		B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
12	JADV (MTX)	19 (92/487)		5 (23/488)	4.9 (<.001) (3, 7.9)		
	JADW[a] (cDMARD)	11 (20/177)	13 (22/174)	2 (4/176)	9.0 (.002) (3.9, 14.2)	10.4 (.001) (5.0, 15.8)	0.9 (.6) (0.4, 1.6)
	JADX (cDMARD)	18 (41/227)	18 (41/229)	3 (7/228)	7.3 (<.001) (3.2, 16.6)	6.9 (<.001) (3, 15.9)	1 (.9) (0.7, 1.7)
24[b]	JADV (MTX)	30 (145/487)		8 (39/488)	4.9 (<.001) (3.4, 7.2)		
	JADW[a] (cDMARD)	17 (30/177)	13 (23/174)	3 (6/176)	13.5 (.001) (7.4, 19.7)	9.8 (.001) (4.1, 15.5)	1.3 (.3) (0.7, 2.4)
	JADX (cDMARD)	24 (55/227)	25 (58/229)	8 (18/228)	3.9 (<.001) (2.2, 6.9)	4 (<.001) (2.3, 7.2)	1.0 (.8) (0.6, 1.5)

source: reviewer programs ACR JADV 2016 03 14.sas, ACR JADW W12 &24 2016 03 14.sas, ACR JADX W12 &24 2016 03 14.sas SCR Tables JADV 11.24 , JADW 11.16, JADX 11.20

[a] Statistical significance evaluated using Fisher's Exact Test, odds ratio and 95% CI via logistic transformation

[b] Exploratory endpoints, p-values nominal only

In study JADZ, conducted in DMARD naive patients, B4 was superior to MTX at week 24 for ACR20 (Table 17). Further evaluation showed B4MTX treatment significantly superior to MTX and B4MTX not significantly superior to B4 and with no numerical advantage, suggesting that the addition of B4 to MTX will improve outcomes in patients naive to MTX, but that there is no evidence that the addition of MTX to B4 will improve outcomes in this patient population. Results for differences between treatments at weeks 12 and 52 were similar to those at week 24 (Table 17). Similar results were seen for ACR50 (Table 18) and ACR70 (Table 19).

Table 17. ACR20 and Remain on Randomized Treatment, Study JADZ

Wk	**% Responders (Responders/Total)**			**Odds Ratio (p-value) (95% CI)**		
	B4MTX	**B4**	**MTX**	**B4MTX: MTX**	**B4: MTX**	**B4MTX: B4**
12[a]	77 (165/215)	79 (125/159)	59 (123/210)	2.4 (<.001) (1.6, 3.6)	2.6 (<.001) (1.6, 4.2)	0.9 (.7) (0.6, 1.5)
24	78 (168/215)	77 (122/159)	62 (130/210)	2.2 (.001) (1.4, 3.4)	2.0 (.003) (1.3, 3.2)	1.1 (.7) (0.7, 1.8)
52[a]	73 (156/215)	73 (116/159)	56 (117/210)	2.1 (<.001) (1.4, 3.2)	2.2 (<.001) (1.4, 3.4)	1 (.9) (0.6, 1.6)

source: ACR JADZ W12 &24&52 2016 03 14.sas, JADZ t-acr20-nri-wk12.sas, CSR Table JADZ 11.17, 11.18

[a] Exploratory endpoints, p-values nominal only

Table 18. ACR50 and Remain on Randomized Treatment, Study JADZ

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4MTX	B4	MTX	B4MTX: MTX	B4: MTX	B4MTX: B4
12[a]	60 (129/215)	55 (87/159)	33 (69/210)	3.1 (<.001) (2.1, 4.7)	2.5 (<.001) (1.6, 3.9)	1.2 (.3) (0.8, 1.9)
24	63 (136/215)	60 (95/159)	43 (91/210)	2.3 (<.001) (1.5, 3.4)	2.0 (.002) (1.3, 3)	1.2 (.5) (0.8, 1.8)
52[a]	62 (133/215)	57 (91/159)	38 (79/210)	2.8 (<.001) (1.9, 4.2)	2.3 (<.001) (1.5, 3.5)	1.2 (.4) (0.8, 1.9)

source: ACR JADZ W12 &24&52 2016 03 14.sas, CSR Table JADZ.11.19

[a] Exploratory endpoints, p-values nominal only

Table 19. ACR70 and Remain on Randomized Treatment, Study JADZ

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4MTX	B4	MTX	B4MTX: MTX	B4: MTX	B4MTX: B4
12[a]	33 (72/215)	31 (49/159)	16 (33/210)	2.8 (<.001) (1.7, 4.5)	2.5 (<.001) (1.5, 4.2)	1.1 (.6) (0.7, 1.8)
24	40 (85/215)	42 (67/159)	21 (45/210)	2.5 (<.001) (1.6, 3.9)	2.8 (<.001) (1.8, 4.5)	.9 (.6) (0.6, 1.4)
52[a]	46 (99/215)	42 (67/159)	25 (53/210)	2.6 (<.001) (1.7, 4)	2.2 (<.001) (1.4, 3.5)	1.2 (.4) (0.8, 1.8)

source: ACR JADZ W12 &24&52 2016 03 14.sas, CSR Table 11.20

[a] Exploratory endpoints, p-values nominal only

In study JADV, for ACR-20, -50, and -70 at weeks 12 and 24, both A and B4 were superior to Pbo, and B4 was superior to A (Table 20, Table 21, and Table 22) when evaluated before patient rescue.

Table 20. ACR20 and Remain on Randomized Treatment, Study JADV, B4 vs Adalimumab

Wk	**% Responders (Responders/Total)**			**Odds Ratio (p-value) (95% CI)**		
	B4	**A**	**Pbo**	**B4:Pbo**	**A:Pbo**	**B4:A**
12	70 (339/487)	61 (202/330)	40 (196/488)	3.6 (<.001) (2.7, 4.7)	2.4 (<.001) (1.8, 3.3)	1.5 (.01) (1.1, 2)
24[a]	74 (360/487)	66 (219/330)	37 (179/488)	5 (<.001) (3.8, 6.7)	3.5 (<.001) (2.6, 4.7)	1.4 (.02) (1.1, 2)

source: ACR JADZ W12 &24&52 2016 03 14.sas, CSR Table JADV 11.9, 11.15 (week 12),
[a] Exploratory endpoints, p-values nominal only

Table 21. ACR50 and Remain on Randomized Treatment, Study JADV, B4 vs Adalimumab

Wk	**% Responders (Responders/Total)**			**Odds Ratio (p-value) (95% CI)**		
	B4	**A**	**Pbo**	**B4:Pbo**	**A:Pbo**	**B4:A**
12	45 (219/487)	35 (115/330)	17 (82/488)	4.2 (<.001) (3.1, 5.7)	2.7 (<.001) (2, 3.8)	1.5 (.005) (1.1, 2.1)
24[a]	51 (246/487)	45 (150/330)	19 (94/488)	4.4 (<.001) (3.3, 5.9)	3.6 (<.001) (2.6, 4.9)	1.2 (.2) (0.9, 1.6)

source: ACR JADZ W12 &24&52 2016 03 14.sas, CSR Table JADV 11.23
[a] Exploratory endpoints, p-values nominal only

Table 22. ACR70 and Remain on Randomized Treatment, Study JADV, B4 vs Adalimumab

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4	A	Pbo	B4:Pbo	A:Pbo	B4:A
12	19 (92/487)	13 (42/330)	5 (23/488)	4.9 (<.001) (3, 7.9)	3.0 (<.001) (1.8, 5.1)	1.6 (.02) (1.1, 2.4)
24[a]	30 (145/487)	22 (72/330)	8 (39/488)	4.9 (<.001) (3.4, 7.2)	3.3 (<.001) (2.1, 5)	1.5 (.02) (1.1, 2.1)

source: ACR JADZ W12 &24&52 2016 03 14.sas, CSR Table JADV 11.24

[a] Exploratory endpoints, p-values nominal only

In summary, for ACR response rates, studies JAD-W, -X, and -V demonstrate superiority of B4 and B2 to placebo when provided as an add-on therapy to cDMARDs (JADW and JADX) or MTX (JADV). There was no clear difference between B4 and B2.

Study JADV indicates superiority of B4 over MTX, and B4+MTX over MTX, in patients naive to DMARDS. There was no evidence that administration of MTX in addition to B4 in such patients improves patient outcomes over use of B4 alone.

Further, study JADV suggests superiority of B4 to adalimumab in inadequate responders to MTX.

Tipping point analyses for these results were not provided in the sponsor's submission. However, the primary ACR20 endpoint comparisons were measured at or before rescue, and therefore large impacts of withdrawal from treatment were considered unlikely to be an issue. Nevertheless, de facto estimates of treatment effect were requested ; such estimates confirmed statistical significance of differences between treatment and control for the primary endpoint comparisons.

3.2.4.2 Components of ACR20, Change from Baseline

There were statistically significant differences between placebo and both doses of baricitinib in change from baseline for nearly all ACR components at weeks 12 and 24 in studies JAD-V, -W, and -X (Table 23, Table 24, and Table 25). Numerical differences comparing B4 vs A, B4 vs MTX, and B4 vs B2 generally favored B4. Given that there was considerable patient escape prior to Week 24, and that the modified LOCF approach utilized by the applicant relies on strong and unverifiable assumptions, results at Week 12 (prior to escape and at which there was minimal missing data) are considered more reliable.

Table 23. ACR20 Components. Study JADV, mLOCF

Param	Wk	Δ Baseln (Baseln) (N)			Difference (p-value) (95% CI)		
		B4	A	Pbo	B4-Pbo	A - Pbo	B4 - A
hsCRP	12	-14.82	-12.31	-2.25	-12.57	-10.05	-2.52
		(22.31)	(21.76)	(19.76)	(<.001)	(<.001)	(0.029)
		(483)	(328)	(485)	(-14.6, -10.54)	(-12.31, -7.79)	(-4.78, -0.26)
	24	-15.13	-11.69	-2.94	-12.19	-8.75	-3.44
		(22.31)	(21.76)	(19.76)	(<.001)	(<.001)	(0.004)
		(483)	(328)	(485)	(-14.31, -10.07)	(-11.11, -6.39)	(-5.8, -1.08)
HAQ-DI	12	-0.66	-0.56	-0.34	-0.32	-0.22	-0.1
		(1.57)	(1.59)	(1.55)	(<.001)	(<.001)	(0.004)
		(482)	(327)	(484)	(-0.38, -0.26)	(-0.29, -0.14)	(-0.17, -0.03)
	24	-0.75	-0.63	-0.35	-0.4	-0.28	-0.12
		(1.57)	(1.59)	(1.55)	(<.001)	(<.001)	(0.002)
		(482)	(327)	(484)	(-0.47, -0.33)	(-0.35, -0.2)	(-0.2, -0.04)
Pain	12	-31.49	-26.35	-17.14	-14.35	-9.21	-5.14
		(61.86)	(61.01)	(59.63)	(<.001)	(<.001)	(0.001)
		(482)	(327)	(484)	(-17.17, -11.53)	(-12.36, -6.07)	(-8.28, -2.00)
	24	-33.61	-28.84	-17.49	-16.13	-11.35	-4.77
		(61.86)	(61.01)	(59.63)	(<.001)	(<.001)	(0.003)
		(482)	(327)	(484)	(-18.99, -13.26)	(-14.54, -8.16)	(-7.96, -1.58)
PaGA	12	-31.21	-26.57	-16.74	-14.47	-9.83	-4.64
		(63.13)	(63.75)	(60.92)	(<.001)	(<.001)	(0.003)
		(482)	(327)	(484)	(-17.18, -11.77)	(-12.85, -6.82)	(-7.65, -1.63)
	24	-33.12	-29.15	-17.01	-16.12	-12.14	-3.98
		(63.13)	(63.75)	(60.92)	(<.001)	(<.001)	(0.013)
		(482)	(327)	(484)	(-18.94, -13.29)	(-15.28, -8.99)	(-7.12, -0.84)
PhGA	12	-38.13	-32.58	-24.33	-13.8	-8.25	-5.55
		(65.67)	(65.34)	(64.15)	(<.001)	(<.001)	(<.001)
		(479)	(325)	(482)	(-16.34, -11.26)	(-11.08, -5.42)	(-8.38, -2.72)
	24	-41.83	-36.56	-25.92	-15.91	-10.64	-5.27
		(65.67)	(65.34)	(64.15)	(<.001)	(<.001)	(<.001)
		(479)	(325)	(482)	(-18.64, -13.18)	(-13.68, -7.59)	(-8.32, -2.23)

PaGA patient global assessment, PhGA physician global assessment

Table 23 (continued)

Param	Wk	Δ Baseln (Baseln) (N)			Difference (p-value) (95% CI)		
		B4	A	Pbo	B4-Pbo	A - Pbo	B4-A
SJC	12	-13.2	-11.83	-7.74	-5.46	-4.1	-1.36
		(23.40)	(23.46)	(23.3)	(<.001)	(<.001)	(0.056)
		(483)	(328)	(485)	(-6.71, -4.2)	(-5.49, -2.7)	(-2.76, 0.04)
	24	-14.91	-13.92	-7.93	-6.98	-5.99	-0.99
		(23.40)	(23.46)	(23.30)	(<.001)	(<.001)	(0.185)
		(483)	(328)	(485)	(-8.3, -5.66)	(-7.46, -4.52)	(-2.46, 0.48)
TJC	12	-9.46	-8.72	-5.66	-3.8	-3.06	-0.74
		(15.03)	(15.37)	(15.51)	(<.001)	(<.001)	(0.115)
		(483)	(328)	(485)	(-4.62, -2.97)	(-3.98, -2.14)	(-1.66, 0.18)
	24	-10.22	-9.72	-6.02	-4.21	-3.71	-0.50
		(15.03)	(15.37)	(15.51)	(<.001)	(<.001)	(0.327)
		(483)	(328)	(485)	(-5.11, -3.31)	(-4.71, -2.7)	(-1.5, 0.5)

Source: JADV t-crp-chg-wk0to52.sas, JADV t-haq-chg-wk0to52.sas, JADV t-patga-chg-wk0to52.sas, JADV t-tjc26-chg-wk0to52.sas, JADV, t-sjc66-chg-wk0to52.sas, CSR JADV Tables 14 21, 14.23, 14.17, 14 16, 14.18, 14.20, 14.19

SJC swollen joint count, TJC tender joint count

Table 24. ACR20 Components. Study JADW, mLOCF

Param	Wk	Δ Baseln (Baseln) (N)			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
hsCRP	12	-9	-5.24	1.07	-10.07	-6.31	-3.76
		(19.94)	(19.77)	(20)	(<.001)	(0.001)	(0.050)
		(175)	(173)	(172)	(-13.83, -6.3)	(-10.08, -2.53)	(-7.52, 0)
	24	-7.7	-4.64	3.62	-11.32	-8.26	-3.06
		(19.94)	(19.77)	(20)	(<.001)	(<.001)	(0.149)
		(175)	(173)	(172)	(-15.49, -7.15)	(-12.44, -4.08)	(-7.23, 1.1)
HAQ-DI	12	-0.41	-0.37	-0.17	-0.24	-0.2	-0.03
		(1.74)	(1.7)	(1.79)	(<.001)	(<.001)	(0.513)
		(175)	(172)	(171)	(-0.34, -0.13)	(-0.31, -0.1)	(-0.14, 0.07)
	24	-0.43	-0.38	-0.15	-0.28	-0.23	-0.05
		(1.74)	(1.7)	(1.79)	(<.001)	(<.001)	(0.387)
		(175)	(172)	(172)	(-0.39, -0.17)	(-0.35, -0.12)	(-0.16, 0.06)
PAIN	12	-22.28	-17.11	-8.78	-13.5	-8.33	-5.17
		(65.88)	(62.7)	(65.45)	(<.001)	(<.001)	(0.039)
		(175)	(172)	(172)	(-18.39, -8.61)	(-13.25, -3.41)	(-10.08, -0.27)
	24	-24.04	-18.84	-8.77	-15.27	-10.08	-5.2
		(63.92)	(62.14)	(65.26)	(<.001)	(<.001)	(0.056)
		(175)	(172)	(172)	(-20.58, -9.97)	(-15.41, -4.74)	(-10.52, 0.12)
PaGA	12	-23	-20.38	-8.94	-14.06	-11.44	-2.61
		(66.37)	(67.69)	(66.67)	(<.001)	(<.001)	(0.291)
		(175)	(172)	(172)	(-18.91, -9.2)	(-16.32, -6.56)	(-7.47, 2.25)
	24	-24.81	-20.33	-8.85	-15.96	-11.49	-4.47
		(64.82)	(66.61)	(66.72)	(<.001)	(<.001)	(0.087)
		(175)	(172)	(172)	(-21.08, -10.84)	(-16.63, -6.35)	(-9.59, 0.65)
PhGA	12	-35.13	-30.64	-17.2	-17.93	-13.43	-4.5
		(67.16)	(67.05)	(66.62)	(<.001)	(<.001)	(0.070)
		(171)	(170)	(171)	(-22.78, -13.08)	(-18.29, -8.58)	(-9.35, 0.36)
	24	-33.98	-29.12	-19.81	-14.17	-9.31	-4.86
		(64.1)	(65.77)	(66.54)	(<.001)	(<.001)	(0.076)
		(172)	(170)	(171)	(-19.52, -8.81)	(-14.67, -3.94)	(-10.22, 0.51)

Table 24 (continued)

Param	Wk	Δ Baseln (Baseln) (N)			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
TJC	12	-13.98	-11.57	-8.66	-5.31	-2.9	-2.41
		(28.01)	(31.23)	(28.4)	(<.001)	(0.035)	(0.078)
		(175)	(173)	(170)	(-8, -2.63)	(-5.6, -0.21)	(-5.09, 0.27)
	24	-13.02	-10.02	-8.43	-4.59	-1.59	-3
		(27.54)	(31.04)	(28.12)	(0.002)	(0.282)	(0.042)
		(175)	(173)	(170)	(-7.49, -1.7)	(-4.5, 1.32)	(-5.89, -0.11)
SJC	12	-8.91	-7.3	-5.13	-3.78	-2.17	-1.61
		(16.35)	(18.53)	(17.38)	(<.001)	(0.013)	(0.064)
		(175)	(173)	(170)	(-5.49, -2.07)	(-3.88, -0.45)	(-3.32, 0.1)
	24	-9.14	-6.16	-4.81	-4.34	-1.35	-2.98
		(15.56)	(18.28)	(17.02)	(<.001)	(0.167)	(0.002)
		(175)	(173)	(170)	(-6.25, -2.42)	(-3.27, 0.57)	(-4.89, -1.07)

Source: jadw t-crp-chg-mlocf-wk0to24.sas, jadw t-haq-chg-mlocf-wk0to24.sas, jadw t-patga-chg-mlocf-wk0to24.sas, jadw t-patpain-chg-mlocf-wk0to24.sas, jadw t-phyga-chg-mlocf-wk0to24.sas, jadw t-sjc66-chg-mlocf-wk0to24.sas, t-tjc68-chg-mlocf-wk0to24.sas, csr Tables 14.28 for HAQ-DI,

Table 25. ACR20 Components. Study JADX, mLOCF

Param	Wk	Δ Baseln (Baseln) (N))			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
hsCRP	12	-8.83	-8.51	0	-8.83	-8.51	-0.32
		(14.14)	(18.18)	(17.85)	(<.001)	(<.001)	(0.803)
		(223)	(228)	(223)	(-11.37, -6.29)	(-11.02, -6)	(-2.85, 2.21)
	24	-7.24	-8.37	-1.57	-5.67	-6.81	1.14
		(14.14)	(18.18)	(17.85)	(<.001)	(<.001)	(0.486)
		(223)	(228)	(223)	(-8.89, -2.45)	(-9.99, -3.62)	(-2.06, 4.34)
HAQDI	12	-0.56	-0.57	-0.36	-0.19	-0.21	0.01
		(1.55)	(1.51)	(1.5)	(<.001)	(<.001)	(0.760)
		(222)	(228)	(220)	(-0.29, -0.1)	(-0.31, -0.11)	(-0.08, 0.11)
	24	-0.62	-0.62	-0.38	-0.24	-0.24	0
		(1.55)	(1.51)	(1.5)	(<.001)	(<.001)	(0.941)
		(222)	(228)	(220)	(-0.34, -0.14)	(-0.35, -0.14)	(-0.1, 0.11)
Pain	12	-23.38	-25.4	-15.59	-7.79	-9.81	2.02
		(57.36)	(59.43)	(57.37)	(<.001)	(<.001)	(0.358)
		(222)	(228)	(220)	(-12.14, -3.45)	(-14.12, -5.5)	(-2.29, 6.33)
	24	-27.91	-27.38	-19.56	-8.35	-7.82	-0.54
		(57.36)	(59.43)	(57.37)	(<.001)	(<.001)	(0.807)
		(222)	(228)	(220)	(-12.72, -3.99)	(-12.14, -3.49)	(-4.86, 3.79)
Pat GA	12	-25.85	-25.32	-16.8	-9.05	-8.52	-0.53
		(60.23)	(61.54)	(60.49)	(<.001)	(<.001)	(0.808)
		(222)	(228)	(220)	(-13.33, -4.76)	(-12.77, -4.27)	(-4.77, 3.72)
	24	-29.15	-27.64	-18.8	-10.35	-8.84	-1.51
		(60.23)	(61.54)	(60.49)	(<.001)	(<.001)	(0.482)
		(222)	(228)	(220)	(-14.61, -6.09)	(-13.06, -4.62)	(-5.73, 2.71)
Phy GA	12	-34.2	-31.64	-22.06	-12.15	-9.58	-2.57
		(64.06)	(64.31)	(62.12)	(<.001)	(<.001)	(0.193)
		(216)	(223)	(221)	(-16.03, -8.26)	(-13.43, -5.74)	(-6.43, 1.3)
	24	-40.23	-36.39	-26.4	-13.83	-9.99	-3.83
		(64.09)	(64.31)	(62.12)	(<.001)	(<.001)	(0.071)
		(219)	(224)	(221)	(-18.01, -9.64)	(-14.14, -5.84)	(-8.00, 0.33)

Table 25 (continued)

Param	Wk	Δ Baseln (Baseln) (N)			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
SJC	12	-12.94	-12.88	-9.77	-3.17	-3.11	-0.06
		(24.25)	(23.55)	(24.35)	(0.002)	(0.002)	(0.953)
		(222)	(228)	(223)	(-5.19, -1.14)	(-5.11, -1.1)	(-2.08, 1.96)
	24	-15.66	-14.08	-10.24	-5.42	-3.84	-1.58
		(24.24)	(23.55)	(24.35)	(<.001)	(<.001)	(0.152)
		(223)	(228)	(223)	(-7.59, -3.25)	(-5.99, -1.69)	(-3.74, 0.58)
TJC	12	-8.61	-8.61	-6.12	-2.48	-2.49	0.01
		(13.49)	(13.67)	(13.11)	(<.001)	(<.001)	(0.993)
		(222)	(228)	(223)	(-3.63, -1.34)	(-3.63, -1.35)	(-1.14, 1.15)
	24	-9.33	-8.95	-6	-3.32	-2.95	-0.37
		(13.49)	(13.67)	(13.11)	(<.001)	(<.001)	(0.558)
		(223)	(228)	(223)	(-4.59, -2.06)	(-4.2, -1.7)	(-1.63, 0.88)

Source: Reviewer programs JADX t-haq-chg-mlocf-wk0to24.sas, JADX t-crp-chg-wk0to24.sas, JADX t-patga-chg-wk0to24.sas, JADX t-tjc68-chg-wk0to24.sas, CSR Tables14 30, 14.29, 14.26 14.27, 14.28, 14.25, 14.24

As in the analysis of ACR20, ACR components in study JADZ indicated superiority of B4 to MTX monotherapy in DMARD naive patients. There was no evidence that addition of MTX to B4 monotherapy improved outcomes, while there was evidence that addition of B4 to MTX monotherapy improved outcomes. The superiority of B4 to MTX, and suggested lack of additional effect of MTX when added to B4 monotherapy suggests potential use of B4 as a primary treatment for RA should its safety profile prove benign relative to MTX.

Table 26. ACR20 Components, Study JADZ, mLOCF

Param	Wk	Δ Baseln (Baseln) (N)			Difference (p-value) (95% CI)		
		B4MTX	B4	MTX	B4MTX -B4	B4MTX -MTX	B4-MTX
hsCRP	12	-16.05	-16.91	-10.11	0.87	-5.94	-6.81
		(24.29)	(23.75)	(21.89)	(0.647)	(<.001)	(<.001)
		(210)	(159)	(205)	(2.85, -4.58)	(-9.41, -2.47)	(-10.54, -3.07)
	24	-18.35	-16.43	-11.68	-1.93	-6.67	-4.74
		(24.29)	(23.75)	(21.89)	(0.189)	(<.001)	(0.001)
		(210)	(159)	(205)	(4.81, -0.96)	(-9.37, -3.98)	(-7.64, -1.84)
	52	-18.01	-16.1	-10.71	-1.91	-7.3	-5.38
		(24.29)	(23.75)	(21.89)	(0.231)	(<.001)	(<.001)
		(210)	(159)	(205)	(5.05, -1.22)	(-10.23, -4.37)	(-8.53, -2.23)
HAQDI	12	-0.98	-0.92	-0.61	-0.06	-0.37	-0.31
		(1.59)	(1.64)	(1.67)	(0.349)	(<.001)	(<.001)
		(209)	(159)	(204)	(0.17, -0.06)	(-0.48, -0.26)	(-0.43, -0.19)
	24	-1.03	-1.04	-0.74	0.01	-0.29	-0.3
		(1.59)	(1.64)	(1.67)	(0.832)	(<.001)	(<.001)
		(209)	(159)	(204)	(0.11, -0.13)	(-0.4, -0.18)	(-0.42, -0.18)
	52	-1.06	-0.99	-0.71	-0.07	-0.35	-0.28
		(1.59)	(1.64)	(1.67)	(0.271)	(<.001)	(<.001)
		(209)	(159)	(204)	(0.19, -0.05)	(-0.46, -0.23)	(-0.40, -0.16)
Pain	12	-38.44	-38.47	-26.66	0.03	-11.78	-11.81
		(62.89)	(64.14)	(64.98)	(0.989)	(<.001)	(<.001)
		(209)	(159)	(204)	(4.63, -4.7)	(-16.14, -7.41)	(-16.5, -7.12)
	24	-41.28	-40.86	-29.92	-0.42	-11.36	-10.94
		(62.89)	(64.14)	(65.14)	(0.861)	(<.001)	(<.001)
		(209)	(159)	(203)	(5.17, -4.32)	(-15.81, -6.92)	(-15.72, -6.17)
	52	-43.44	-40.33	-30.72	-3.1	-12.72	-9.61
		(62.89)	(64.14)	(64.98)	(0.208)	(<.001)	(<.001)
		(209)	(159)	(204)	(7.94, -1.73)	(-17.24, -8.19)	(-14.47, -4.75)

Table 26 (continued)

Param	Wk	Δ Baseln (Baseln) (N)			Difference (p-value) (95% CI)		
		B4MTX	B4	MTX	B4MTX -B4	B4MTX -MTX	B4-MTX
Pat GA	12	-38.01	-38	-26.93	-0.01	-11.08	-11.07
		(63.39)	(65.04)	(65.43)	(0.996)	(<.001)	(<.001)
		(209)	(159)	(204)	(4.56, -4.54)	(-15.33, -6.82)	(-15.64, -6.5)
	24	-39.95	-41.15	-30.4	1.21	-9.54	-10.75
		(63.39)	(65.04)	(65.52)	(0.609)	(<.001)	(<.001)
		(209)	(159)	(203)	(3.43, -5.84)	(-13.88, -5.2)	(-15.41, -6.09)
	52	-42.6	-40.39	-29	-2.2	-13.59	-11.39
		(63.39)	(65.04)	(65.43)	(0.365)	(<.001)	(<.001)
		(209)	(159)	(204)	(6.98, -2.57)	(-18.06, -9.13)	(-16.19, -6.59)
Phy GA	12	-45.76	-44.63	-34.98	-1.13	-10.78	-9.65
		(66.38)	(68.08)	(66.5)	(0.582)	(<.001)	(<.001)
		(207)	(157)	(201)	(5.16, -2.9)	(-14.55, -7.01)	(-13.71, -5.6)
	24	-49.68	-49.21	-38.29	-0.48	-11.39	-10.92
		(66.39)	(68.13)	(66.49)	(0.818)	(<.001)	(<.001)
		(208)	(156)	(200)	(4.55, -3.59)	(-15.2, -7.59)	(-15.02, -6.81)
	52	-50.43	-49.45	-40.17	-0.98	-10.27	-9.28
		(66.39)	(68.08)	(66.5)	(0.660)	(<.001)	(<.001)
		(208)	(157)	(201)	(5.38, -3.41)	(-14.37, -6.16)	(-13.71, -4.86)

Table 26 (continued)

Param	Wk	Δ Baseln (Baseln) (N)			Difference (p-value) (95% CI)		
		B4MTX	**B4**	**MTX**	**B4MTX -MTX**	**B4-MTX**	**B4MTX -B4**
SJC	12	-19.05	-16.8	-13.81	-2.25	-5.24	-2.99
		(27.7)	(26.44)	(26.12)	(0.039)	(<.001)	(0.007)
		(208)	(159)	(205)	(4.4, 0.11)	(-7.25, -3.24)	(-5.14, -0.84)
	24	-20.15	-18.54	-16.13	-1.61	-4.02	-2.41
		(27.61)	(26.44)	(26.1)	(0.126)	(<.001)	(0.023)
		(210)	(159)	(204)	(3.68, -0.46)	(-5.96, -2.09)	(-4.49, -0.33)
	52	-20.87	-19.12	-16.83	-1.75	-4.04	-2.3
		(27.61)	(26.44)	(26.12)	(0.129)	(<.001)	(0.048)
		(210)	(159)	(205)	(4.01, -0.51)	(-6.16, -1.93)	(-4.57, -0.02)
TJC	12	-12.49	-11.05	-8.82	-1.44	-3.67	-2.23
		(16.16)	(16.12)	(16.2)	(0.034)	(<.001)	(0.001)
		(208)	(159)	(205)	(2.77, 0.11)	(-4.92, -2.43)	(-3.57, -0.9)
	24	-13.07	-12.16	-10.39	-0.91	-2.68	-1.77
		(16.11)	(16.12)	(16.21)	(0.169)	(<.001)	(0.008)
		(210)	(159)	(204)	(2.21, -0.39)	(-3.9, -1.47)	(-3.08, -0.47)
	52	-13.21	-12.05	-10.22	-1.16	-2.99	-1.83
		(16.11)	(16.12)	(16.2)	(0.099)	(<.001)	(0.010)
		(210)	(159)	(205)	(2.54, -0.22)	(-4.28, -1.7)	(-3.22, -0.44)

source: reviewer programs jadz t-haq-ch-wk0to52.sas, jadz t-crp-chg-wk0to52.sas, jadz t-patga-chg-wk0to52.sas, jadz t-tjc68-chg-wk0to52.sas, CSR Tables jadz.14.23, 14.26, 14.20, 14.21, 14.22, 14.19, 14.18

3.2.4.3 Secondary Endpoint: DAS28-hsCRP

Percentages of patients with DAS28-hsCRP ≤ 3.2 or ≤ 2.6 showed results similar to those from analyses of ACR20, with significant differences between placebo and B4 or B2 (Table 27, Table 28). In all three studies, B4 was numerically superior to B2, with statistical significance in study JADW but not in studies JADV or JADX.

Table 27. DAS28-hsCRP ≤ 3.2 and Remain on Randomized Treatment, Studies JADV, W, X

Wk	Study (backgrd)	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
		B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
12	JADV (MTX)	44 (214/487)		14 (67/488)	5.2 (<.001) (3.8, 7.2)		
	JADW (MTX)	32 (56/177)	24 (42/174)	9 (16/176)	4.8 (<.001) (2.6, 8.9)	3.3 (<.001) (1.8, 6.2)	1.4 (.13) (0.9, 2.3)
	JADX (cDMARD)	39 (89/227)	36 (82/229)	17 (39/228)	3.3 (<.001) (2.1, 5.1)	2.8 (<.001) (1.8, 4.3)	1.2 (.4) (0.8, 1.8)
24[a]	JADV (MTX)	52 (255/487)		19 (93/488)	4.8 (<.001) (3.6, 6.4)		
	JADW (cDMARD)	33 (59/177)	20 (35/174)	11 (20/176)	3.9 (<.001) (2.2, 6.9)	2.0 (.02) (1.1, 3.7)	2.0 (.007) (1.3, 3.2)
	JADX (cDMARD)	52 (117/227)	46 (106/229)	24 (54/228)	3.5 (<.001) (2.4, 5.3)	2.9 (<.001) (1.9, 4.3)	1.2 (.3) (0.8, 1.8)

source: reviewer programs ACR JADV 2016 03 14.sas, ACR JADW W12 &24 2016 03 14.sas, ACR JADX W12 &24 2016 03 14.sas, CSR Tables JADV 11.34, JADW 11.25, JADX 11.29

[a] Exploratory endpoints, p-values nominal only

Table 28. DAS28-hsCRP ≤ 2.6 and Remain on Randomized Treatment, Studies JADV, W, X

Wk	Study (backgrd)	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
		B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
12	JADV (MTX)	24 (119/487)		4 (21/488)	7.6 (<.001) (4.7, 12.4)		
	JADW (MTX)	16 (29/177)	11 (19/174)	4 (7/176)	4.8[b] (.001) (2, 11.3)	3.0[b] (.02) (1.2, 7.4)	1.6 (.2) (0.8, 3)
	JADX (cDMARD)	26 (58/227)	26 (59/229)	9 (20/228)	3.7 (<.001) (2.1, 6.5)	3.7 (<.001) (2.1, 6.4)	1 (.97) (0.7, 1.6)
24[a]	JADV (MTX)	34 (168/487)		8 (38/488)	6.4 (<.001) (4.4, 9.4)		
	JADW (cDMARD)	21 (38/177)	11 (19/174)	6 (11/176)	4.2 (<.001) (2.1, 8.6)	1.9 (0.11) (0.9, 4.1)	2.2 (0.01) (1.3, 4)
	JADX (cDMARD)	33 (75/227)	31 (70/229)	11 (24/228)	4.2 (<.001) (2.5, 7.1)	3.8 (<.001) (2.3, 6.3)	1.1 (0.6) (0.8, 1.7)

source: reviewer programs ACR JADV 2016 03 14.sas, ACR JADW W12 &24 2016 03 14.sas, ACR JADX W12 &24 2016 03 14.sas, CSR Tables JADV 11.34, JADW 11.25, JADX 11.29

[a] Exploratory endpoints, p-values nominal only

[b] p-values from exact test

For DAS28-hsCRP ≤ 3.2 and ≤ 2.6, in study JADZ conducted in MTX naive patients, B4 was significantly superior to MTX, B4MTX was superior to MTX alone, and B4MTX was not significantly superior to B4 (Table 29, Table 30).

Table 29. DAS28-hsCRP ≤ 3.2 and Remain on Randomized Treatment, Study JADZ

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4MTX	B4	MTX	B4MTX: MTX	B4: MTX	B4MTX: B4
12[a]	56 (121/215)	47 (74/159)	30 (62/210)	3.1 (<.001) (2.1, 4.7)	2.1 (<.001) (1.4, 3.2)	1.5 (.059) (1, 2.3)
24	60 (129/215)	57 (91/159)	38 (80/210)	2.5 (<.001) (1.7, 3.7)	2.2 (<.001) (1.4, 3.4)	1.1 (.6) (0.7, 1.7)
52[a]	63 (136/215)	57 (91/159)	38 (79/210)	2.9 (<.001) (2.0, 4.4)	2.3 (<.001) (1.5, 3.5)	1.3 (.2) (0.8, 2)

source: ACR JADZ W12 &24&52 2016 03 14.sas, CSR Table 11.29

[a] Exploratory endpoints, p-values nominal only

Table 30. DAS28-hsCRP ≤ 2.6 and Remain on Randomized Treatment, Study JADZ

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4MTX	B4	MTX	B4MTX: MTX	B4: MTX	B4MTX: B4
12[a]	36 (77/215)	28 (44/159)	16 (33/210)	3 (<.001) (1.9, 4.8)	2.1 (.006) (1.2, 3.4)	1.5 (.10) (0.9, 2.3)
24	40 (87/215)	40 (64/159)	24 (50/210)	2.2 (<.001) (1.5, 3.4)	2.2 (<.001) (1.4, 3.5)	1 (.97) (0.7, 1.5)
52[a]	49 (105/215)	44 (70/159)	24 (50/210)	3.2 (<.001) (2.1, 4.8)	2.6 (<.001) (1.6, 4)	1.2 (.3) (0.8, 1.9)

source: ACR JADZ W12 &24&52 2016 03 14.sas, CSR Table 11.29

3.2.4.4 Secondary Endpoint: SDAI ≤ 3.3

Statistically significant differences between baricitinib treatments and placebo were seen for SDAI ≤ 3.3 at weeks 12 and 24 in studies JADV and JADX (Table 31 and Table 32), but not in study JADW (Table 33). One possible explanation for lack of significance in study JADW is the lower sample size of study JADW compared to the other studies.

In study JADV, B4 was not significantly superior to A.

Table 31. SDAI < 3.3 and Remain on Randomized Treatment, Study JADV

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4	A	Pbo	B4:Pbo	A:Pbo	B4:A
12	8 (41/487)	7 (24/330)	2 (9/488)	4.9 (<.001) (2.4, 10.3)	4.2 (<.001) (1.9, 9.3)	1.2 (0.576) (0.7, 2)
24[a]	16 (78/487)	14 (45/330)	3 (15/488)	6.1 (<.001) (3.4, 10.7)	4.8 (<.001) (2.6, 8.9)	1.3 (0.277) (0.8, 1.9)

source: acr jadv 2016 03 14.sas, CSR Table JADV 14.25 (exact test)

Table 32. SDAI < 3.3 and Remain on Randomized Treatment, Study JADX

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
12	9 (20/227)	9 (21/229)	1 (2/228)	11.7 (0.001) (2.7, 51.1)	11.5 (0.001) (2.6, 49.8)	1 (0.949) (0.5, 2)
24[a]	15 (34/227)	17 (38/229)	4 (9/228)	4.3 (<.001) (2, 9.3)	4.9 (<.001) (2.3, 10.4)	0.9 (0.646) (0.5, 1.5)

source: ACR JADX W12&24 2016 03 08, CSR Table JADX 11.12 (exact test),

[a] Exploratory endpoints, p-values nominal only

Table 33. SDAI < 3.3 and Remain on Randomized Treatment, Study JADW

Wk	% Responders (Responders/Total)			Odds Ratio (p-value) (95% CI)		
	B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
12	5 (9/177)	2 (4/174)	2 (3/176)	3.1 (0.053) (0.8, 11.9)	1.4 (0.7) (0.3, 6.5)	2.2 (0.14) (0.7, 7.4)
24[a]	9 (16/177)	5 (8/174)	2 (4/176)	4.4 (0.010) (1.4, 13.5)	2.2 (0.218) (0.6, 7.4)	2 (0.116) (0.8, 4.9)

source: acr jadw.sas, CSR Table JADW 11.12 (exact test)
[a] Exploratory endpoints, p-values nominal only

In study JADZ, differences between B4 and MTX and between B4MTX and MTX for SDAI ≤ 3.3 were statistically significant, while the difference between B4MTX and B4 was not statistically significant (Table 34).

Table 34. SDAI ≤ 3.3 and Remain on Randomized Treatment, Study JADZ

Wk	% Responders (Responders/Total)			Odds Ratio[a.] (p-value) (95% CI)		
	B4MTX	B4	MTX	B4MTX: MTX	B4: MTX	B4MTX: B4
12[a]	20 (42/215)	14 (22/159)	6 (13/210)	3.8 (<.0001) (2, 7.4)	2.5 (0.013) (1.2, 5.1)	1.5 (0.150) (0.9, 2.7)
24	23 (49/215)	22 (35/159)	10 (22/210)	2.6 (<.001) (1.4, 4.4)	2.5 (<.001) (1.6, 3.6)	1 (0.885) (0.6, 1.7)
52[a]	30 (65/215)	25 (39/159)	13 (28/210)	2.9 (<.001) (1.8, 4.8)	2.2 (0.005) (1.3, 3.7)	1.3 (0.214) (0.8, 2.2)

source: ACR JADZ W12 &24&52 2016 03 14.sas, CSR Table 11.14
[a.] Results differed between use of proc glimmix and proc logistic on same dataset
[a]Exploratory endpoints, p-values nominal only

3.2.4.5 *Secondary Endpoint: Radiographic Progression*

In study JADX, compared to placebo, B4 reduced mTSS at week 24 in an analysis using linear extrapolation of missing and post-escape data (Table 35), an analysis including only observed data while on randomized treatment (Table 36), and an analysis including all observed data, including data collected after treatment discontinuation and escape to B4 (Table 37). Results were not as consistent for B2, with lack of evidence of an effect in the analyses based on only observed data. In addition, tipping point analyses comparing B2 and B4 to placebo indicated lack of statistical significance, with p-values exceeding .05 for the null imputation (Table 38 and Table 39). The results correspond (with allowance for differences between simulations) to those from the applicant (sequence 0020 pp 17 to 19) which found that small differences from zero shift led to lack of statistical significance of effects.

That the effects of B4 and B2 compared to placebo were similar for the primary analysis suggests an exploratory analysis in which additional statistical power is gained by merging data from B4 and B2 into a single baricitinib treatment, 'B.' This analysis again did not indicate statistically significant differences between B and Pbo for the null imputation (Table 40).

Table 35. Radiographic Progression, with Linear Extrapolation, Study JADX

Wk	Param	Δ mTSS (N)			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
24[a]	mTSS	0.15	0.33	0.7	-0.55	-0.38	-0.18
		(198)	(208)	(190)	(.003)	(.04)	(.3)
					(-0.92, -0.19)	(-0.74, -0.01)	(-0.54, 0.18)
	JSN	0.04	0.03	0.23	-0.20	-0.21	0.01
		(198)	(208)	(190)	(.02)	(.01)	(.9)
					(-0.37, -0.03)	(-0.37, -0.04)	(-0.15, 0.17)
	JSE	0.11	0.3	0.47	-0.36	-0.17	-0.19
		(198)	(208)	(190)	(.005)	(.2)	(.13)
					(-0.61, -0.11)	(-0.42, 0.07)	(-0.43, 0.06)

source: JADX mtss 2016 05 25.sas, CSR Table JADX 11.45

[a] Exploratory endpoint, p-values nominal only

Table 36. Radiographic Progression, Observed and Remain on Randomized Treatment, Study JADX

Wk	Param	Change from Baseline mTSS (N)			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
24[a]	mTSS	0.2 (174)	0.27 (178)	0.53 (137)	-0.32 (.03) (-0.62, -0.03)	-0.25 (.09) (-0.55, 0.04)	-0.07 (.6) (-0.34, 0.2)
	JSN	0.06 (174)	0.05 (178)	0.18 (137)	-0.12 (.12) (-0.28, 0.03)	-0.13 (.09) (-0.29, 0.02)	0.01 (.9) (-0.13, 0.15)
	JSE	0.14 (174)	0.22 (178)	0.34 (137)	-0.2 (.055) (-0.4, 0)	-0.12 (.2) (-0.33, 0.08)	-0.08 (.4) (-0.27, 0.11)

source: JADX mtss 2016 05 25.sas

[a] Exploratory endpoint, p-values nominal only

Table 37. Radiographic Progression, All Observed Data, Study JADX

Wk	Param	Change from Baseline mTSS (N)			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
24[a]	mTSS	0.19 (184)	0.34 (188)	0.49 (167)	-0.3 (.03) (-0.58, -0.03)	-0.15 (.3) (-0.42, 0.13)	-0.15 (.3) (-0.42, 0.11)
	JSN	0.06 (184)	0.06 (188)	0.14 (167)	-0.08 (.2) (-0.22, 0.06)	-0.08 (.3) (-0.22, 0.06)	0 (.9) (-0.14, 0.13)
	JSE	0.13 (184)	0.28 (188)	0.35 (167)	-0.22 (.03) (-0.42, -0.02)	-0.07 (.5) (-0.27, 0.13)	-0.15 (.13) (-0.34, 0.04)

source: xray jadx all data 2016 10 13.sas

[a] Exploratory endpoint, p-values nominal only

Table 38. Tipping Point Radiographic Progression, B4 vs Pbo, Study JADX, Week 24

	Treatment Effect (p-value)[a]			
Shift	**B4**			
Pbo	0	0.144	0.280	0.415
0	-0.27 (.058)	-0.26 (.07)	-0.25 (.08)	-0.24 (.1)
-0.144	-0.25 (.08)	-0.24 (.09)	-0.23 (.11)	-0.22 (.13)
-0.280	-0.23 (.11)	-0.22 (.12)	-0.21 (.14)	-0.2 (.2)
-0.415	-0.21 (.14)	-0.2 (.2)	-0.19 (.2)	-0.18 (.2)

	Treatment Effect (p-value)[b]			
Shift	**B4**			
Pbo	0	0.264	0.400	0.536
0	-0.27 (.04)	-0.23 (.09)	-0.21 (.12)	-0.19 (.2)
-0.264	-0.21 (.12)	-0.16 (.2)	-0.14 (.3)	-0.12 (.4)
-0.400	-0.17 (.2)	-0.13 (.3)	-0.11 (.4)	-0.09 (.5)
-0.536	-0.14 (.3)	-0.1 (.5)	-0.08 (.6)	-0.05 (.7)

a. Patients missing baseline excluded from tipping point analysis
b. Patients missing baseline included in tipping point analysis

Table 39. Tipping Point Radiographic Progression, B2 vs Pbo, Study JADX, Week 24

	Treatment Effect (p-value)[a]			
Shift	**B2**			
Pbo	0	0.010	0.076	0.141
0	-0.27 (.058)	-0.27 (.059)	-0.26 (.064)	-0.26 (.069)
-0.010	-0.27 (.06)	-0.27 (.06)	-0.26 (.065)	-0.26 (.07)
-0.076	-0.26 (.069)	-0.26 (.07)	-0.25 (.08)	-0.25 (.08)
-0.141	-0.25 (.08)	-0.25 (.08)	-0.24 (.09)	-0.24 (.09)

	Treatment Effect (p-value)[b]			
Shift	**B2**			
Pbo	0	0.136	0.208	0.279
0	-0.27 (.04)	-0.25 (.06)	-0.24 (.07)	-0.23 (.09)
-0.136	-0.24 (.07)	-0.22 (.1)	-0.21 (.12)	-0.19 (.15)
-0.208	-0.22 (.1)	-0.2 (.13)	-0.19 (.2)	-0.18 (.2)
-0.279	-0.2 (.12)	-0.18 (.2)	-0.17 (.2)	-0.16 (.2)

a. Patients missing baseline excluded from tipping point analysis
b. Patients missing baseline included in tipping point analysis

Table 40. Tipping Point Radiographic Progression, B vs Pbo, Study JADX, Week 24

	Treatment Effect (p-value)[a]			
Shift	**B**			
Pbo	0	0.080	0.180	0.281
0	-0.20 (.11)	-0.19 (.12)	-0.18 (.14)	-0.17 (.2)
-0.080	-0.19 (.13)	-0.18 (.14)	-0.17 (.2)	-0.16 (.2)
-0.180	-0.18 (.2)	-0.17 (.2)	-0.16 (.2)	-0.15 (.2)
-0.281	-0.16 (.2)	-0.15 (.2)	-0.14 (.2)	-0.13 (.3)

	Treatment Effect (p-value)[b]			
Shift	**B**			
Pbo	0	0.156	0.254	0.353
0	-0.20 (.08)	-0.17 (.13)	-0.15 (.2)	-0.14 (.2)
-0.156	-0.16 (.2)	-0.13 (.2)	-0.12 (.3)	-0.1 (.4)
-0.254	-0.13 (.2)	-0.11 (.3)	-0.09 (.4)	-0.07 (.5)
-0.353	-0.11 (.3)	-0.08 (.4)	-0.07 (.5)	-0.05 (.6)

source: xray jadx tipping point 2016 08 02.sas

a. Patients missing baseline excluded from tipping point analysis

b. Patients missing baseline included in tipping point analysis

Study JADV showed statistically significant differences between B4 and Pbo in mTSS at weeks 24 and 52 in analyses using linear extrapolation of missing data and data post-escape (Table 41), observed data while on randomized treatment (Table 42), and all observed data, including data collected after treatment discontinuation and escape to B4 (Table 43). Tipping point analyses confirmed the statistical differences between B4 and Pbo at week 24 (Table 44) but not at week 52 (Table 45). Differences between B4 and A in this study were not statistically significant (Table 41). Results from tipping point analyses comparing B4 to Pbo at W24 correspond to analyses by the applicant (sequence 0020 pp 20-26, B4 vs Pbo), the tipping point analyses at W52 are only provided here for the present review.

Table 41. Radiographic Progression, with Linear Extrapolation, Study JADV

Wk	Param	Δ mTSS (N)			Difference (p-value) (95% CI)		
		B4	A	Pbo	B4-Pbo	A-Pbo	B4-A
16[a]	mTSS	0.35	0.28	0.69	-0.33	-0.4	0.07
		(454)	(300)	(444)	(<.001)	(<.001)	(.5)
					(-0.53, -0.14)	(-0.63, -0.18)	(-0.15, 0.29)
	JSN	0.11	0.08	0.20	-0.09	-0.12	0.03
		(454)	(300)	(444)	(.074)	(.03)	(.6)
					(-0.18, 0.01)	(-0.22, -0.01)	(-0.08, 0.14)
	JSE	0.25	0.21	0.5	-0.25	-0.29	0.04
		(454)	(300)	(444)	(<.001)	(<.001)	(.6)
					(-0.38, -0.11)	(-0.44, -0.14)	(-0.11, 0.19)
24	mTSS	0.41	0.33	0.90	-0.49	-0.56	0.07
		(470)	(312)	(452)	(<.001)	(<.001)	(.6)
					(-0.73, -0.25)	(-0.83, -0.29)	(-0.19, 0.34)
	JSN	0.12	0.10	0.29	-0.17	-0.19	0.01
		(470)	(312)	(452)	(.003)	(.004)	(.8)
					(-0.29, -0.06)	(-0.31, -0.06)	(-0.11, 0.14)
	JSE	0.29	0.24	0.61	-0.32	-0.38	0.06
		(470)	(312)	(452)	(<.001)	(<.001)	(.5)
					(-0.48, -0.15)	(-0.56, -0.19)	(-0.13, 0.24)
52[a]	mTSS	0.71	0.6	1.8	-1.1	-1.2	0.1
		(473)	(312)	(452)	(<.001)	(<.001)	(.7)
					(-1.55, -0.64)	(-1.71, -0.69)	(-0.4, 0.61)
	JSN	0.21	0.19	0.58	-0.37	-0.39	0.02
		(473)	(312)	(452)	(<.001)	(.001)	(.9)
					(-0.58, -0.16)	(-0.63, -0.15)	(-0.22, 0.25)
	JSE	0.51	0.42	1.23	-0.73	-0.82	0.09
		(473)	(312)	(452)	(<.001)	(<.001)	(.6)
					(-1.05, -0.41)	(-1.18, -0.46)	(-0.27, 0.44)

source: JADV mtss 2016 06 02.sas, CSR JADV Table 11.10

[a] Exploratory endpoints, p-values nominal only

Table 42. Radiographic Progression, Observed and on Randomized Treatment, Study JADV

Wk	Param	Change from Baseline mTSS (N)			Difference (p-value) (95% CI)		
		B4	A	Pbo	B4-Pbo	A-Pbo	B4-A
16[a]	mTSS	0.36 (450)	0.28 (298)	0.70 (434)	-0.34 (<.001) (-0.54, -0.14)	-0.42 (<.001) (-0.64, -0.2)	0.08 (.5) (-0.15, 0.3)
	JSN	0.11 (450)	0.08 (298)	0.20 (434)	-0.09 (.063) (-0.19, 0.01)	-0.12 (.026) (-0.23, -0.01)	0.03 (.6) (-0.08, 0.14)
	JSE	0.25 (450)	0.21 (298)	0.51 (434)	-0.25 (<.001) (-0.39, -0.11)	-0.3 (<.001) (-0.45, -0.14)	0.05 (.6) (-0.11, 0.2)
24	mTSS	0.36 (424)	0.29 (280)	0.79 (363)	-0.44 (<.001) (-0.65, -0.22)	-0.51 (<.001) (-0.75, -0.26)	0.07 (.6) (-0.17, 0.31)
	JSN	0.11 (424)	0.09 (280)	0.26 (363)	-0.15 (.01) (-0.26, -0.04)	-0.17 (.01) (-0.29, -0.04)	0.02 (.8) (-0.1, 0.14)
	JSE	0.25 (424)	0.2 (280)	0.54 (363)	-0.29 (<.001) (-0.43, -0.15)	-0.34 (<.001) (-0.5, -0.18)	0.05 (.5) (-0.1, 0.21)
52[a]	mTSS	0.44 (389)	0.29 (238)				0.15 (.2) (-0.1, 0.41)
	JSN	0.18 (389)	0.08 (238)				0.10 (.13) (-0.03, 0.22)
	JSE	0.26 (389)	0.21 (238)				0.05 (.5) (-0.11, 0.22)

source: JADV mtss 2016 05 20.sas

[a] Exploratory endpoints, p-values nominal only

Table 43. Radiographic Progression, All Observed Data, Study JADV

Wk	Param	Change from Baseline mTSS (N)			Difference (p-value) (95% CI)		
		B4	A	Pbo	B4-Pbo	A-Pbo	B4-A
16[a]	mTSS	0.36	0.28	0.70	-0.34	-0.42	0.08
		(450)	(298)	(434)	(<.001)	(<.001)	(.5)
					(-0.54, -0.14)	(-0.64, -0.2)	(-0.15, 0.3)
	JSN	0.11	0.08	0.20	-0.09	-0.12	0.03
		(450)	(298)	(434)	(.063)	(.03)	(.6)
					(-0.19, 0.01)	(-0.23, -0.01)	(-0.08, 0.14)
	JSE	0.25	0.21	0.51	-0.25	-0.3	0.05
		(450)	(298)	(434)	(<.001)	(<.001)	(.6)
					(-0.39, -0.11)	(-0.45, -0.14)	(-0.11, 0.2)
24	mTSS	0.36	0.30	0.8	-0.43	-0.5	0.07
		(444)	(299)	(426)	(<.001)	(<.0001)	(.6)
					(-0.66, -0.21)	(-0.75, -0.25)	(-0.18, 0.31)
	JSN	0.1	0.09	0.25	-0.15	-0.15	0
		(444)	(299)	(426)	(.006)	(.01)	(.9)
					(-0.26, -0.04)	(-0.27, -0.04)	(-0.11, 0.12)
	JSE	0.27	0.21	0.55	-0.28	-0.35	0.06
		(444)	(299)	(426)	(<.001)	(<.0001)	(.5)
					(-0.43, -0.13)	(-0.51, -0.18)	(-0.1, 0.23)
52[a]	mTSS	0.55	0.48	0.98	-0.43	-0.49	0.06
		(422)	(282)	(408)	(.002)	(.002)	(.7)
					(-0.7, -0.16)	(-0.8, -0.19)	(-0.24, 0.37)
	JSN	0.19	0.2	0.32	-0.12	-0.12	0
		(422)	(282)	(408)	(.1)	(.14)	(.98)
					(-0.26, 0.02)	(-0.28, 0.04)	(-0.16, 0.16)
	JSE	0.36	0.29	0.66	-0.31	-0.37	0.07
		(422)	(282)	(408)	(<.001)	(<.001)	(.5)
					(-0.48, -0.13)	(-0.57, -0.18)	(-0.13, 0.26)

source: xray jadv all obs 2016 10 28.sas

[a] Exploratory endpoints, p-values nominal only

Table 44. Tipping Point Radiographic Progression, B4 vs Placebo, Study JADV, Week 24

	Treatment Effect (p-value)[a]			
Shift	**B4**			
Pbo	0	0.498	0.730	0.962
0	-0.46 (<.001)	-0.43 (<.001)	-0.42 (<.001)	-0.4 (<.001)
-0.498	-0.43 (<.001)	-0.4 (<.001)	-0.38 (<.001)	-0.37 (.001)
-0.730	-0.41 (<.001)	-0.38 (<.001)	-0.37 (.001)	-0.35 (.002)
-0.962	-0.4 (<.001)	-0.37 (.001)	-0.35 (.002)	-0.34 (.003)

[a.] Patients missing baseline excluded from tipping point analysis

	Treatment Effect (p-value)[b]			
Shift	**B4**			
Pbo	0	0.544	0.776	1.008
0	-0.46 (<.000)	-0.43 (<.001)	-0.42 (<.001)	-0.4 (<.001)
-0.544	-0.43 (<.001)	-0.39 (<.001)	-0.38 (<.001)	-0.36 (.001)
-0.776	-0.41 (<.001)	-0.38 (<.001)	-0.36 (.001)	-0.35 (.002)
-1.008	-0.39 (<.001)	-0.36 (.001)	-0.35 (.002)	-0.33 (.004)

source: xray jadv Tipping point 2016 08 05.sas

[b.] Patients missing baseline included in tipping point analysis

Table 45. Tipping Point Radiographic Progression, B4 vs Placebo, Study JADV, Week 52

	Treatment Effect (p-value)[a]			
Shift	**B4**			
Pbo	0	-0.184	-0.332	-0.481
0.000	-0.3 (.3)	-0.12 (.7)	0.03 (.9)	0.18 (.5)
-0.184	-0.11 (.7)	0.07 (.8)	0.21 (.4)	0.36 (.2)
-0.332	0.03 (.9)	0.21 (.4)	0.36 (.2)	0.51 (.08)
-0.481	0.18 (.5)	0.36 (.2)	0.51 (.07)	0.66 (.02)

[a.] Patients missing baseline excluded from tipping point analysis

	Treatment Effect (p-value)[a]			
Shift	**B4**			
Pbo	0	-0.778	-1.212	-1.647
0	-0.87 (.01)	-0.10 (.8)	0.33 (.3)	0.76 (.03)
-0.778	-0.09 (.8)	0.67 (.049)	1.10 (.002)	1.53 (<.0001)
-1.212	0.34 (.3)	1.11 (.002)	1.54 (<.0001)	1.97 (<.0001)
-1.647	0.77 (.03)	1.54 (<.0001)	1.97 (<.0001)	2.40 (<.0001)

source: xray jadv Tipping point W52 2016 08 23.sas

[a.] Patients missing baseline included in tipping point analysis

In analyses of study JADZ, B4 was not significantly superior to MTX alone at W24, but B4MTX was superior to MTX monotherapy for linear extrapolation of missing data (Table 46), for observed data while on randomized treatment (Table 47), and for all observed data regardless of adherence to treatment or initiation of rescue (Table 48).

Tipping point analyses weakly supported superiority of B4MTX to MTX, with imputed B4MTX superior to imputed MTX for the null plus one-half imputation, but not for the null plus one

imputation (Table 49).

Table 46. Radiographic Progression, with Linear Extrapolation, Study JADZ

Wk	Param	Δ mTSS (N)			Difference (p-value) (95% CI)		
		B4MTX	B4	MTX	B4MTX-MTX	B4-MTX	B4MTX-B4
12[a]	mTSS	0.27 (191)	0.32 (146)	0.47 (185)	-0.19 (.20) (-0.46, 0.08)	-0.14 (.3) (-0.43, 0.14)	-0.05 (.7) (-0.33, 0.24)
	JSN	0.03 (191)	0.03 (146)	0.12 (185)	-0.09 (.15) (-0.22, 0.03)	-0.10 (.2) (-0.23, 0.04)	0 (.96) (-0.13, 0.14)
	JSE	0.24 (191)	0.3 (146)	0.34 (185)	-0.10 (.3) (-0.28, 0.09)	-0.05 (.6) (-0.24, 0.15)	-0.05 (.6) (-0.25, 0.14)
24	mTSS	0.29 (198)	0.39 (152)	0.61 (191)	-0.32 (.03) (-0.6, -0.04)	-0.22 (.2) (-0.52, 0.08)	-0.1 (.5) (-0.4, 0.2)
	JSN	0.03 (198)	0.06 (152)	0.14 (191)	-0.10 (.13) (-0.23, 0.03)	-0.08 (.3) (-0.22, 0.07)	-0.03 (.7) (-0.17, 0.11)
	JSE	0.26 (198)	0.33 (152)	0.47 (191)	-0.22 (.03) (-0.42, -0.02)	-0.14 (.2) (-0.36, 0.07)	-0.08 (.5) (-0.29, 0.14)
52[a]	mTSS	0.40 (199)	0.80 (154)	1.02 (192)	-0.62 (.004) (-1.04, -0.2)	-0.23 (.3) (-0.67, 0.22)	-0.4 (.08) (-0.84, 0.05)
	JSN	0.06 (199)	0.25 (154)	0.21 (192)	-0.15 (.15) (-0.35, 0.05)	0.04 (.7) (-0.18, 0.25)	-0.18 (.09) (-0.39, 0.03)
	JSE	0.34 (199)	0.55 (154)	0.81 (192)	-0.47 (.002) (-0.77, -0.18)	-0.26 (.11) (-0.58, 0.06)	-0.21 (.2) (-0.53, 0.1)

source: JADZ mtss 2016 03 22.sas

[a] Exploratory endpoints, p-values nominal only

Table 47. Radiographic Progression, Observed and on Randomized Treatment, Study JADZ

Wk	Param	Δ mTSS (N)			Difference (p-value) (95% CI)		
		B4MTX	B4	MTX	B4MTX-MTX	B4-MTX	B4MTX-B4
12[a]	mTSS	0.28	0.32	0.47	-0.19	-0.15	-0.04
		(189)	(146)	(184)	(.2)	(.3)	(.8)
					(-0.46, 0.08)	(-0.43, 0.14)	(-0.33, 0.24)
	JSN	0.03	0.03	0.12	-0.09	-0.1	0.01
		(189)	(146)	(184)	(.2)	(.2)	(.9)
					(-0.22, 0.04)	(-0.24, 0.04)	(-0.13, 0.14)
	JSE	0.25	0.3	0.34	-0.10	-0.05	-0.05
		(189)	(146)	(184)	(.3)	(.6)	(.6)
					(-0.28, 0.09)	(-0.25, 0.15)	(-0.25, 0.15)
24	mTSS	0.31	0.38	0.65	-0.34	-0.27	-0.07
		(177)	(138)	(172)	(.03)	(.1)	(.7)
					(-0.65, -0.03)	(-0.6, 0.05)	(-0.39, 0.26)
	JSN	0.04	0.07	0.15	-0.11	-0.08	-0.03
		(177)	(138)	(172)	(.13)	(.3)	(.7)
					(-0.26, 0.03)	(-0.24, 0.07)	(-0.18, 0.13)
	JSE	0.27	0.31	0.5	-0.23	-0.19	-0.04
		(177)	(138)	(172)	(.04)	(.1)	(.7)
					(-0.44, -0.02)	(-0.42, 0.04)	(-0.27, 0.19)
52[a]	mTSS	0.32	0.6	1.17	-0.85	-0.57	-0.28
		(170)	(130)	(139)	(<.001)	(.02)	(.2)
					(-1.29, -0.4)	(-1.05, -0.09)	(-0.73, 0.18)
	JSN	0.01	0.21	0.27	-0.26	-0.06	-0.2
		(170)	(130)	(139)	(.02)	(.6)	(.079)
					(-0.47, -0.04)	(-0.29, 0.17)	(-0.42, 0.02)
	JSE	0.3	0.38	0.89	-0.59	-0.51	-0.08
		(170)	(130)	(139)	(<.001)	(.003)	(.6)
					(-0.9, -0.28)	(-0.84, -0.18)	(-0.4, 0.24)

source: JADZ mtss 2016 05 25.sas

[a] Exploratory endpoints, p-values nominal only

Table 48. Radiographic Progression, All Observed Data, Study JADZ

Wk	Param	Δ mTSS (N)			Difference (p-value) (95% CI)		
		B4MTX	B4	MTX	B4MTX-MTX	B4-MTX	B4MTX-B4
12[a]	mTSS	0.28	0.32	0.47	-0.19	-0.15	-0.04
		(189)	(146)	(184)	(.2)	(.3)	(.8)
					(-0.46, 0.08)	(-0.43, 0.14)	(-0.33, 0.24)
	JSN	0.03	0.03	0.12	-0.09	-0.1	0.01
		(189)	(146)	(184)	(.2)	(.2)	(.9)
					(-0.22, 0.04)	(-0.24, 0.04)	(-0.13, 0.14)
	JSE	0.25	0.3	0.34	-0.1	-0.05	-0.05
		(189)	(146)	(184)	(.3)	(.6)	(.6)
					(-0.28, 0.09)	(-0.25, 0.15)	(-0.25, 0.15)
24	mTSS	0.31	0.38	0.65	-0.34	-0.27	-0.07
		(177)	(138)	(172)	(.03)	(.1)	(.7)
					(-0.65, -0.03)	(-0.6, 0.05)	(-0.39, 0.26)
	JSN	0.04	0.07	0.15	-0.11	-0.08	-0.03
		(177)	(138)	(172)	(.13)	(.3)	(.7)
					(-0.26, 0.03)	(-0.24, 0.07)	(-0.18, 0.13)
	JSE	0.27	0.31	0.5	-0.23	-0.19	-0.04
		(177)	(138)	(172)	(.04)	(.1)	(.7)
					(-0.44, -0.02)	(-0.42, 0.04)	(-0.27, 0.19)
52[a]	mTSS	0.31	0.62	1.09	-0.78	-0.47	-0.31
		(174)	(135)	(160)	(<.001)	(.041)	(.2)
					(-1.2, -0.36)	(-0.92, -0.02)	(-0.75, 0.13)
	JSN	0.01	0.22	0.25	-0.24	-0.03	-0.21
		(174)	(135)	(160)	(.02)	(.8)	(.054)
					(-0.44, -0.04)	(-0.24, 0.18)	(-0.42, 0.00)
	JSE	0.3	0.41	0.85	-0.54	-0.44	-0.11
		(174)	(135)	(160)	(<.001)	(.007)	(.5)
					(-0.84, -0.25)	(-0.75, -0.12)	(-0.42, 0.2)

source: JADZ mtss all data 2016 11 08.sas

[a] Exploratory endpoints, p-values nominal only

Table 49. Tipping Point Radiographic Progression, B4MTX versus MTX, Study JADZ, Week 24

	Treatment Effect (p-value)[a]			
Shift	**B4MTX**			
MTX	0	0.269	0.439	0.609
0	-0.34 (.02)	-0.31 (.03)	-0.29 (.044)	-0.27 (.061)
-0.269	-0.31 (.03)	-0.28 (.05)	-0.26 (.068)	-0.24 (.09)
-0.439	-0.3 (.04)	-0.27 (.066)	-0.25 (.09)	-0.23 (.12)
-0.609	-0.28 (.053)	-0.25 (.09)	-0.23 (.11)	-0.21 (.15)

	Treatment Effect (p-value)[b]			
Shift	**B4MTX**			
MTX	0	0.389	0.564	0.739
0	-0.35 (.02)	-0.28 (.050)	-0.25 (.08)	-0.22 (.13)
-0.389	-0.28 (.049)	-0.22 (.13)	-0.19 (.2)	-0.16 (.3)
-0.564	-0.25 (.08)	-0.19 (.2)	-0.16 (.3)	-0.13 (.4)
-0.739	-0.23 (.11)	-0.16 (.3)	-0.13 (.4)	-0.1 (.5)

source: xray jadzTipping Point 2016 07 16.sas

a. Patients missing baseline excluded from tipping point analysis

b. Patients missing baseline included in tipping point analysis

Proportion of patients with no radiographic progression (Δ mTSS ≤ 0) was estimated using linear extrapolation for missing data. For patients showing progression at earlier timepoints, such extrapolation seems reasonable, since one can presume that once progression occurs, it does not reverse. However, linear extrapolation for patients who do not show progression at earlier timepoints will not necessarily reflect patient outcome, since progression may appear at later timepoints.

For proportion of patients with no radiographic progression, no statistically significant differences between Pbo and B4 or B2 were seen in study JADX (Table 50, Table 51). However, the difference between B4 and placebo was statistically significant in study JADV in an analysis based on linear extrapolation and an analysis including all observed data, including data collected after escape or treatment discontinuation (Table 52, Table 53). In study JADZ, B4MTX was superior to MTX at weeks 24 and 52, while differences between B4 and MTX and between B4MTX and B4 were not significant at the 24 week prespecified endpoint (Table 54, Table 55).

Table 50. Proportion Without Radiographic Progression , Study JADX, Linear Extrapolation

Wk	Proportion (N/Total)			Odds Ratio (p-value) (95% CI)		
	B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
24	.79 (137/174)	.72 (129/178)	.77 (105/137)	1.2 (.5) (0.7, 2.1)	0.8 (.4) (0.5, 1.4)	1.5 (.13) (0.9, 2.4)

source: Rad Prog 0 JADX.sas, CSR Table JADX 11.46

Table 51. Proportion Without Radiographic Progression , Study JADX, All Recorded Data

Wk	Proportion (N/Total)			Odds Ratio (p-value) (95% CI)		
	B4	B2	Pbo	B4:Pbo	B2:Pbo	B4:B2
24	.8 (144/181)	.71 (131/184)	.76 (118/156)	1.3 (.3) (0.8, 2.2)	0.8 (.4) (0.5, 1.3)	0.6 (0.050) (1, 2.7)

source: JADX Rad Prog 0 all data 2016 11 09.sas,

Table 52. Proportion Without Radiographic Progression, Study JADV, Linear Extrapolation

Wk	Proportion (N/Total)			Odds Ratio (p-value) (95% CI)		
	B4	A	Pbo	B4:Pbo	A:Pbo	B4:A
16[a]	0.81 (363/449)	0.83 (245/296)	0.72 (310/432)	1.7 (.002) (1.2, 2.3)	1.9 (<.001) (1.3, 2.7)	0.9 (.5) (0.6, 1.3)
24	0.82 (345/423)	0.83 (231/279)	0.71 (257/361)	1.8 (<.001) (1.3, 2.5)	2 (<.001) (1.3, 2.9)	0.9 (.7) (0.6, 1.4)
52[a]	0.8 (310/388)	0.83 (196/237)				0.8 (.4) (0.6, 1.3)

source: JADV Rad Prog 0 t-mtss-le-wk0to52 Obs.sas, *after* CSR Table JADV 11.49

[a] Exploratory endpoints, p-values nominal only

Table 53. Proportion Without Radiographic Progression, Study JADV, All Recorded Data

Wk	Proportion (N/Total)			Odds Ratio (p-value) (95% CI)		
	B4	A	Pbo	B4:Pbo	A:Pbo	B4:A
16[a]	.81 (364/450)	.83 (247/298)	.72 (311/434)	1.7 (0.002) (1.2, 2.3)	1.9 (<.001) (1.3, 2.7)	0.9 (.5) (0.6, 1.3)
24	.81 (361/444)	.82 (246/299)	.70 (300/426)	1.8 (<.001) (1.3, 2.5)	1.9 (<.001) (1.3, 2.8)	0.9 (.73) (0.6, 1.4)
52[a]	.78 (331/422)	.80 (227/282)	.66 (270/408)	1.8 (<.001) (1.4, 2.5)	2.1 (<.001) (1.5, 3)	0.9 (.51) (0.6, 1.3)

source: JADV Rad Prog 0 all data 2016 11 09.sas,

[a] Exploratory endpoints, p-values nominal only

Table 54. Proportion Without Radiographic Progression, Study JADZ, Linear Extrapolation

Wk	Proportion (N/Total)			Odds Ratio (p-value) (95% CI)		
	B4MTX	B4	MTX	B4MTX: MTX	B4MTX:B4	B4:MTX
12[a]	.82 (155/189)	.75 (109/146)	.76 (139/184)	1.5 (.13) (0.9, 2.5)	1.6 (.1) (0.9, 2.7)	0.9 (.8) (0.6, 1.6)
24	.8 (142/177)	.75 (104/138)	.66 (113/172)	2.2 (.002) (1.3, 3.6)	1.3 (.3) (0.8, 2.3)	1.6 (.06) (1, 2.7)
52[a]	.79 (135/170)	.70 (91/130)	.63 (88/139)	2.3 (.002) (1.4, 3.9)	1.7 (.063) (1, 2.9)	1.4 (.2) (0.8, 2.3)

source: JADZ Rad Prog 0 t-mtss-le-wk0to52 Obs.sas, *after* CSR Table JADZ 11.43

[a] Exploratory endpoints, p-values nominal only

Table 55. Proportion Without Radiographic Progression, Study JADZ, All Recorded Data

Wk	Proportion (N/Total)			Odds Ratio (p-value) (95% CI)		
	B4MTX	B4	MTX	B4MTX: MTX	B4MTX:B4	B4:MTX
12[a]	.82 (155/189)	.75 (109/146)	.76 (139/184)	1.5 (.13) (0.9, 2.5)	1.6 (.1) (0.9, 2.7)	0.9 (.8) (0.6, 1.6)
24	.8 (142/177)	.75 (104/138)	.66 (113/172)	2.2 (.002) (1.3, 3.6)	1.3 (.3) (0.8, 2.3)	1.6 (.061) (1, 2.7)
52[a]	.80 (139/174)	.70 (95/135)	.63 (100/160)	2.4 (<.001) (1.5, 4)	1.7 (.062) (1, 2.8)	1.5 (.13) (0.9, 2.4)

source: JADZ Rad Prog 0 all data.sas

[a] Exploratory endpoints, p-values nominal only

In summary, despite several cautionary comments issued by the Division prior to submission of the NDA, quantitative estimates of treatment effects submitted by the applicant used linear extrapolation to impute missing radiographic data and radiographic data after escape to baricitinib. As seen in the supportive analyses using all observed data and in the tipping point analyses, linear extrapolation may have exaggerated the effects of baricitinib. Furthermore, the appropriateness of linear extrapolation had patients continued on their randomized treatment is unknown and untestable and, even if appropriate, relies on a single-imputation imputation and does not appropriately account for the uncertainty in the imputation. Furthermore, the relevance of any de jure estimand such as that potentially targeted by linear extrapolation (i.e., the difference in radiographic progression had all patients remained on their randomized treatment) is questionable because it is hypothetical quantity - not all patients adhered to randomized treatment in the phase 3 studies, and not all patients are expected to adhere to treatment in clinical situations. Therefore, (b) (4) (b) (4) we believe that the de facto estimand (the difference in radiographic progression regardless of adherence or escape) is of interest here, and estimates of this estimand are more reliable (b) (4).

Using an approach intending to impute missing outcomes under the null hypothesis of no treatment effect, study JADX failed to provide statistically significant differences between the two baricitinib doses and Pbo at week 24, although numerical trends did point toward efficacy of baricitinib. With double the sample size, study JADV did show statistically significant differences favoring B4 over Pbo at week 24. Study JADZ, with sample sizes comparable to those of study JADX, weakly suggested superiority of B4MTX to MTX monotherapy among

patients naïve to MTX, with effects statistically significant for the null plus 1/2 imputation but not for the null plus 1 imputation. Results from study JADZ tended to numerically favor adalimumab over baricitinib for inhibition of radiographic progression; however none of those results were statistically significant.

In conclusion, studies JADV and JADZ provide evidence of effects of B4 on radiographic progression. One study, JADX, evaluated the impact of B2 on radiographic progression. Using an exploratory analysis on all recorded data, regardless of patient adherence to randomized treatment, there was no statistically significant difference between B2 and Pbo. Because that analysis was only exploratory, and because convincing statistical significance is lacking for B2 in a single study, without additional radiographic evaluation of B2 in any additional phase 3 studies, inhibition of radiographic progression by B2 has not been established.

3.2.4.6 Morning Joint Stiffness

Baricitinib significantly reduced median duration of morning joint stiffness at week 12 (Table 56 and Table 57), with median Hodges-Lehmann differences between baricitinib and placebo ranging from roughly 15 to 30 minutes (Table 56 and Table 57).

Table 56. Median Change from Baseline Duration of Morning Stiffness. Study JADV, Week 12

Δ Minutes (N)			**Median Difference (p-value) (95% CI)**		
B4	**A**	**Pbo**	**B4-Pbo**	**A-Pbo**	**B4-A**
-30	-13	-2	-30	-10	-10
(277)	(190)	(276)	(.001)	(.015)	(.03)
			(-45, -15)	(-25, 0)	(-29, 0)

source: CSR Table 11 52 reviewer program jadv t_mjsdur_chg_wk0to52.sas

Table 57. Median Change from Baseline Duration of Morning Stiffness. Study JADX, Week 12

Δ Minutes (N)			**Median Difference (p-value) (95% CI)**		
B4	**B2**	**Pbo**	**B4-Pbo**	**B2-Pbo**	**B4-B2**
-20	-30	-9	-14	-21	4
(222)	(223)	(221)	(.02)	(.004)	(.5)
			(-29, -2)	(-38, -7)	(-9, 19)

source: reviewer program stiffness duration jadx.sas, CSR Table 11 13 did not include change from baseline

Severity of morning stiffness was measured on an 11 point scale, with 0 representing no joint stiffness and 10 'as bad as you can imagine.' Baricitinib significantly reduced mean severity of morning joint stiffness at week 12, with differences between baricitinib and placebo ranging from 0.6 to 1.1 (Table 58 and Table 59).

Table 58. Change from Baseline Severity of Morning Stiffness. Study JADV, Week 12

Δ Severity (N)			Mean Difference (p-value) (95% CI)		
B4	**A**	**Pbo**	**B4-Pbo**	**A-Pbo**	**B4-A**
-2.46	-1.97	-1.38	-1.08	-0.60	-0.49
(478)	(320)	(476)	(<.0001)	(<.0001)	(.001)
			(-1.34, -0.82)	(-0.89, -0.31)	(-0.77, -0.20)

source: stiff sev jadv.sas,. CSR Table JADV 14.13 did not evaluate change from baseline.

Table 59. Change from Baseline Severity of Morning Stiffness. Study JADX, Week 12

Δ Severity (N)			Mean Difference (p-value) (95% CI)		
B4	**B2**	**Pbo**	**B4-Pbo**	**B2-Pbo**	**B4-B2**
-2.07	-1.95	-1.32	-0.75	-0.63	-0.12
(219)	(223)	(220)	(<.001)	(.002)	(.6)
			(-1.15, -0.35)	(-1.03, -0.24)	(-0.52, 0.28)

source: stiff sev tired pain jadx.sas, CSR Table JADX 14 14 did not evaluate change from baseline

In summary, for duration and severity of morning stiffness at week 12, B4 and B2 in study JADX and B4 in study JADV provided statistically significant improvements compared to placebo. Differences between B4 and B2 were not statistically significant in study JADX. In study JADV B4 was significantly superior to A.

3.2.4.7 Worst Tiredness

Worst tiredness was measured on an 11 point scale, with 0 representing no tiredness and 10 'as bad as you can imagine.' Baricitinib significantly reduced worst tiredness at week 12, with differences between baricitinib and placebo ranging from 0.4 to 0.8 (Table 60 and Table 61). Differences between B4 and B2 were not statistically significant, and differences favoring B4 over A were statistically significant.

Table 60. Change from Baseline Worst Tiredness. Study JADV, Week 12

Δ Worst Tiredness (N)			Mean Difference (p-value) (95% CI)		
B4	**A**	**Pbo**	**B4-Pbo**	**A-Pbo**	**B4-A**
-2.02	-1.7	-1.24	-0.78	-0.45	-0.32
(478)	(320)	(476)	(<.0001)	(.002)	(.03)
			(-1.04, -0.52)	(-0.74, -0.17)	(-0.61, -0.04)

source: stiff sev jadv.sas. CSR Table 14.14 did not address change from baseline

Table 61. Change from Baseline Worst Tiredness. Study JADX, Week 12

Δ Severity (N)			Mean Difference (p-value) (95% CI)		
B4	**B2**	**Pbo**	**B4-Pbo**	**B2-Pbo**	**B4-B2**
-1.68	-1.63	-1.23	-0.45	-0.40	-0.05
(219)	(223)	(220)	(.03)	(.049)	(.8)
			(-0.86, -0.05)	(-0.8, -0.00)	(-0.45, 0.35)

source: stiff sev tired pain jadx.sas CSR Table 11.15 did not address change from baseline

3.2.4.8 Worst Joint Pain

Worst joint pain was measured on an 11 point scale, with 0 representing no pain and 10 'joint pain as bad as you can imagine.' Baricitinib significantly reduced mean worst joint pain at week 12, with differences between baricitinib and placebo ranging from 0.9 to 1.2 (Table 62 and Table 63). B4 was significantly superior to A in study JADV.

Table 62. Change from Baseline Worst Joint Pain. Study JADV, Week 12

Δ Worst Joint Pain (N)			Mean Difference (p-value) (95% CI)		
B4	**A**	**Pbo**	**B4-Pbo**	**A-Pbo**	**B4-A**
-2.47	-1.83	-1.29	-1.19	-0.54	-0.65
(478)	(320)	(476)	(<.0001)	(<.001)	(<.0001)
			(-1.44, -0.94)	(-0.82, -0.26)	(-0.92, -0.37)

source: stiff sev jadv.sas, CSR table 14.15 did not address change from baseline

Table 63. Change from Baseline Worst Joint Pain. Study JADX, Week 12

Δ Worst Joint Pain (N)			Mean Difference (p-value) (95% CI)		
B4	**B2**	**Pbo**	**B4-Pbo**	**B2-Pbo**	**B4-B2**
-1.97	-1.99	-1.09	-0.89	-0.91	0.02
(219)	(223)	(220)	(<.0001)	(<.0001)	(.9)
			(-1.29, -0.49)	(-1.3, -0.51)	(-0.38, 0.42)

source: stiff sev tired pain jadx.sas, CSR table 14.16 did not address change from baseline

3.2.4.9 SF36

Baricitinib provided consistent, statistically significant improvements in physical component score, physical function, role physical, bodily pain, vitality, and general health at week 12 (Table 64), with no statistically significant effects for mental component score, and mixed results for role emotional, mental health, and social functioning. Differences between B4 and B2 were not statistically significant and did not follow any clear numeric patterns.

Table 64. SF36 Components and Domains at Week 12, Observed Data While on Randomized Treatment, Studies JADV, JADW and JADX

Param	Study	Change from Baseline			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
MCS	JADV	3.27		2.99	0.27		
		(474)		(458)	(.6)		
					(-0.86, 1.41)		
	JADW	1.84	2.81	1.24	0.6	1.57	-0.98
		(174)	(168)	(168)	(.5)	(.10)	(.3)
					(-1.28, 2.47)	(-0.32, 3.46)	(-2.86, 0.91)
	JADX	3.48	3.14	3.23	0.25	-0.09	0.34
		(219)	(229)	(218)	(.8)	(.9)	(.7)
					(-1.49, 1.99)	(-1.81, 1.63)	(-1.39, 2.06)
PCS	JADV	8.68		4.25	4.43		
		(474)		(458)	(<.001)		
					(3.52, 5.35)		
	JADW	6.37	6.03	2.64	3.73	3.38	0.34
		(174)	(168)	(168)	(<.001)	(<.001)	(.7)
					(2.04, 5.41)	(1.69, 5.08)	(-1.35, 2.03)
	JADX	7.24	7.96	4.29	2.95	3.67	-0.72
		(219)	(229)	(218)	(<.001)	(<.001)	(.3)
					(1.53, 4.37)	(2.27, 5.07)	(-2.12, 0.69)
PF	JADV	7.95		4.35	3.6		
		(474)		(458)	(<.001)		
					(2.57, 4.64)		
	JADW	4.69	5.08	1.34	3.35	3.74	-0.39
		(174)	(168)	(168)	(<.001)	(<.001)	(0.7)
					(1.57, 5.13)	(1.95, 5.53)	(-2.17, 1.39)
	JADX	6.67	7.22	3.97	2.7	3.25	-0.55
		(219)	(229)	(218)	(0.001)	(<.001)	(0.5)
					(1.09, 4.31)	(1.66, 4.84)	(-2.14, 1.04)

source: SF36 All.sas, *after* LOCF analyses by applicant

MCS mental component score, PCS physical component score, PF physical function

Table 64 (continued)

Param	Study	Change from Baseline			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
RP	JADV	7.7		4.35	3.35		
		(474)		(458)	(<.001)		
					(2.32, 4.39)		
	JADW	5.05	5.68	2.96	2.09	2.72	-0.63
		(174)	(168)	(168)	(.02)	(.003)	(.5)
					(0.29, 3.88)	(0.91, 4.52)	(-2.42, 1.17)
	JADX	5.63	6.45	4.26	1.38	2.19	-0.81
		(219)	(229)	(218)	(.083)	(.005)	(.300)
					(-0.18, 2.93)	(0.65, 3.73)	(-2.35, 0.73)
BP	JADV	9.16		4.74	4.42		
		(474)		(458)	(<.001)		
					(3.43, 5.42)		
	JADW	8.04	7.27	3.97	4.07	3.29	0.77
		(174)	(168)	(168)	(<.001)	(<.001)	(.4)
					(2.25, 5.88)	(1.47, 5.12)	(-1.05, 2.59)
	JADX	8.05	8.45	4.71	3.34	3.74	-0.40
		(219)	(229)	(218)	(<.001)	(<.001)	(.6)
					(1.78, 4.9)	(2.2, 5.29)	(-1.95, 1.14)
GH	JADV	5.37		3.12	2.25		
		(474)		(458)	(<.001)		
					(1.32, 3.18)		
	JADW	2.97	4.01	1.2	1.77	2.81	-1.04
		(174)	(168)	(168)	(.02)	(<.001)	(.2)
					(0.28, 3.26)	(1.31, 4.31)	(-2.54, 0.46)
	JADX	5.96	5.74	3.42	2.54	2.31	0.22
		(219)	(229)	(218)	(<.001)	(.002)	(.8)
					(1.06, 4.02)	(0.85, 3.77)	(-1.24, 1.69)
VT	JADV	6.31		3.94	2.36		
		(474)		(458)	(<.001)		
					(1.29, 3.44)		
	JADW	5.84	5.42	3.23	2.62	2.19	0.43
		(174)	(168)	(168)	(.006)	(.022)	(.7)
					(0.76, 4.48)	(0.32, 4.06)	(-1.45, 2.3)
	JADX	6.68	6.88	4.88	1.8	1.99	-0.2
		(219)	(229)	(218)	(.03)	(.02)	(.8)
					(0.14, 3.46)	(0.36, 3.63)	(-1.84, 1.44)

source: SF36 All.sas

RP role physical, BP bodily pain, GH general health, VT vitality

Table 64 (continued)

Param	Study	Change from Baseline			Difference (p-value) (95% CI)		
		B4	B2	Pbo	B4-Pbo	B2-Pbo	B4-B2
SF	JADV	5.46		2.99	2.47		
		(474)		(458)	(<.001)		
					(1.33, 3.61)		
	JADW	4.22	3.11	0.8	3.41	2.3	1.11
		(174)	(168)	(168)	(<.001)	(.02)	(.3)
					(1.43, 5.39)	(0.31, 4.3)	(-0.88, 3.09)
	JADX	4.2	4.58	3.57	0.62	1.01	-0.39
		(219)	(229)	(218)	(.5)	(.2)	(.650)
					(-1.07, 2.31)	(-0.65, 2.68)	(-2.06, 1.28)
RE	JADV	4.97		3.56	1.41		
		(474)		(458)	(.02)		
					(0.22, 2.59)		
	JADW	1.4	4.04	1.52	-0.12	2.53	-2.65
		(174)	(168)	(168)	(.9)	(.01)	(.01)
					(-2.13, 1.89)	(0.5, 4.55)	(-4.66, -0.63)
	JADX	4.24	3.89	3.43	0.81	0.46	0.35
		(219)	(229)	(218)	(.4)	(.6)	(.7)
					(-1.02, 2.63)	(-1.35, 2.26)	(-1.46, 2.15)
MH	JADV	3.94		3.66	0.29		
		(474)		(458)	(.6)		
					(-0.84, 1.41)		
	JADW	2.53	3.3	1.33	1.21	1.97	-0.76
		(174)	(168)	(168)	(.2)	(.04)	(.4)
					(-0.63, 3.04)	(0.12, 3.82)	(-2.61, 1.08)
	JADX	4.33	3.98	3.31	1.02	0.67	0.35
		(219)	(229)	(218)	(.3)	(.5)	(.7)
					(-0.78, 2.82)	(-1.11, 2.45)	(-1.43, 2.13)

source: SF36 All.sas

SF social functioning, RE role emotional, MH mental health

3.2.4.10 Maintenance of Low Disease Activity with Step-Down Dosing

(b) (4)

However, the endpoint proposed for inclusion, low disease activity, likely reflects 'cherry picking' of results from multiple exploratory endpoints. In particular, despite cautions from FDA, study JADY did not control type 1 error over multiple endpoints; therefore the selected results are likely biased and the actual 95% confidence interval is wider. For example, CDAI $\leq$ 2.8 was also evaluated, with no statistical significance between B4 and B2. Results for ACR20, the primary endpoint in the other studies in this submission, were omitted from the study report. For HAQ-DI, the endpoint near the top of the analysis hierarchy in the originating studies, the reported difference between B4 and B2 was not statistically significant.

Nevertheless, despite the applicant's misuse of 'statistical significance' and potential biased selection of results (b) (4), there is some evidence that step down dosing may negatively impact patient outcome. For example, numerical differences between B4 and B2 seem to consistently favor B4 for more customary metrics, such as the ACR components and other continuous endpoints, suggesting that step down of dose from B4 to B2 may have negative consequences (Table 65).

Table 65. Exploratory Analyses: Step Down Dosing, Study JADY, Week 12

Param	Mean Change from Baseline		P-value for Difference
	B4	B2	B4-B2
N =	(147)	(146)	
CDAI	0.60	2.04	.02
TJC	0.3	1.2	.02
SJC	0.1	0.6	.02
HAQ-DI	0.04	0.06	.6
Pat-Pain	0.7	2.5	.3
Pat-GA	1.2	2.5	.5
Phys-GA	0.0	2.7	.048
DAS28-hsCRP	0.14	0.36	.01
SDAI	0.69	2.19	.02

Source: CSR, Table 7.2

(b) (4)

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

Impacts of gender, age class (<45, 45 to 65, >65), race, ethnicity, and country (USA vs not USA) on the effect of baricitinib compared to placebo were evaluated in studies JADV, X, W, and Z by adding the subgroup to be evaluated and its interaction with treatment to the primary analysis model for ACR 20 at week 12 for studies JADV, JADX, and JADW and at week 24 for study JADZ. Since there were no discernable differences between B4 and B2, these treatments were merged, with the interaction term evaluated at the unadjusted .05 level of significance. As a final note, the term 'placebo' is used advisedly, because add-on therapies were provided in all of the studies evaluated in this submission.

With the exception of race in study JADV (Table 66), no statistically significant impacts of subgroups on treatment efficacy were seen. With regard to race in study JADV, all races but Native Americans had odds ratios for ACR20 greater than 1. For Native Americans, however, the odds ratio was 0.55 with unadjusted 95% confidence bounds of (.2, 2). In the analysis, there were 19 Native Americans in the placebo arm and 26 in the baricitinib arm. Given results in other studies, this difference may have been a chance finding.

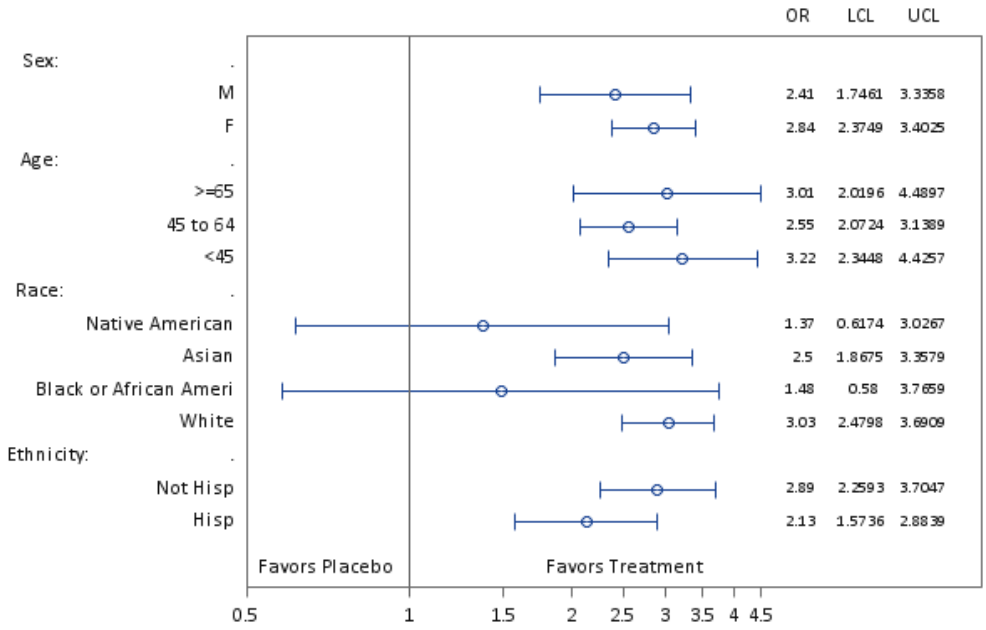
Inverse variance weighted meta-analysis across studies JADV, W, X, and Z showed numerically positive estimated effects of treatment in all subgroups (Figure 1).

Table 66. Treatment by Subgroup Interactions, ACR20

Study	P-value for Interaction of Treatment with				
	Sex	Age	Ethnicity	Country	Race
JADW	.6	.5	.8	.11	.6
JADZ	.2	.7	.3	.9	.3
JADX	.3	.2	.2	.8	.7
JADV	.3	.4	.096	.5	.025

source: subgroups acr jadv 2016 08 03.sas, subgroups acr jadw 2016 08 03.sas, subgroups acr jadx 2016 08 03.sas, subgroups acr jadz 2016 08 03.sas

Figure 1. Meta-analysis of Treatment Effect by Subgroup. ACR20, Studies JADV, W, X, and Z



source: snaps 2016 111 08.sas

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

The initial submission failed to address appropriate estimands for timepoints after rescue. In addition, analysis datasets were inadequately documented. After multiple information requests, this submission does provide adequate evidence of effectiveness to support approval of this drug for the treatment of rheumatoid arthritis, and demonstrates that the higher dose of this drug reduces radiographic progression.

Because rescued patients were considered nonresponders in the evaluation of binary response endpoints other than lack of radiographic progression, treatment effects for binary endpoints could have been driven by differences in the proportions of patients who failed to adhere to randomized treatment rather than by continued activity of baricitinib on the outcome of interest. For example, if, early in the trial, proportion rescued were greater among patients randomized to placebo than among patients randomized to baricitinib, imputation of nonresponse among rescued patients would suggest demonstration of baricitinib activity for any such composite endpoints, even, for example, aversion to chocolate.

For this reason, unless supported by appropriate de-facto analyses, estimates of post-rescue treatment effects for responder endpoints should be considered unreliable, potentially driven by withdrawal from randomized treatment. Similarly, to establish long-term drug activity for continuous endpoints, de facto estimates of treatment effects and associated tipping point analyses are needed.

5.2 Collective Evidence

The collective evidence from primary and secondary endpoints in these studies consistently supports effectiveness of baricitinib for the treatment of RA. In addition, the higher dose has been shown to inhibit radiographic progression.

5.3 Conclusions and Recommendations

Four randomized, double-blind trials demonstrate effectiveness of baricitinib for the treatment of rheumatoid arthritis. Studies JADX and JADW, which enrolled patients refractory to treatment with cDMARDS and TNF-inhibitors respectively, randomized enrollees to B4, B2, or Pbo as add-on therapies to cDMARDS, showed statistically significant effects of B4 and B2 compared to Pbo for the proportion of patients exhibiting a positive ACR20 response. Study JADV, which randomized bDMARD naive patients refractory to treatment with MTX to B4, A, or Pbo as add-on therapies to MTX, showed statistically significant effects of B4 for ACR20 response compared to both Pbo and A. Study JADZ, which randomized bDMARD and cDMARD naive patients to B4, B4MTX, or MTX, showed improved ACR20 response rates for either B4 or B4MTX compared to MTX monotherapy.

There were no significant differences between B4 and B2 for ACR20 response rate. However, exploratory analyses of the continuous components of ACR20 indicate nominal superiority of B4 to B2 in study JADV, and numeric, but not nominally significant trends, in study JADW which were consistent with superiority of B4 to B2.

Studies JADV and JADZ demonstrated that, compared to placebo, B4 inhibited radiographic progression. One study, JADX, evaluated the impact of B2 on radiographic progression using an exploratory analysis of all recorded data, regardless of patient adherence to randomized treatment. The analysis showed no statistically significant difference between B2 and Pbo. Because that analysis was only exploratory, and because statistical significance was lacking, inhibition of radiographic progression by B2 was not established.

In study JADZ, B4 was superior to MTX for the primary endpoint, ACR20 at week 24 ($p<.001$), with a p-value sufficiently low to argue in favor of substantial evidence of superiority in a single trial. However, superiority of B4 to MTX for inhibition of radiographic progression was not established ($p=.1$). In study JADV, superiority of B4 to A for the primary endpoint, ACR20 response at week 12, was less convincing for a single study (p-value=.01), and superiority of B4 to A for inhibition of radiographic progression not established ($p=.6$).

5.4 Labeling Recommendations

Consistent with conclusions from the current review, I have the following recommendations:

1. Responder endpoints on Tables 4, 5 and 6, and Figures 4 and 5 of the proposed label should exclude [redacted] If the appropriate data has been recorded, the applicant may, in a supplementary application, [redacted]
2. If comparative treatment information is determined to be appropriate, and if the risk-benefit is favorable, I recommend inclusion of comparisons from study JADZ, which demonstrates that B4 and B4MTX are superior to MTX for ACR20. However, it will be important to note, in the accompanying sentence, that there was insufficient evidence in this single trial to establish superiority of B4 and B4MTX to MTX for inhibition of radiographic progression, and that, for ACR20 superiority of B2 and B2MTX to MTX has not been established.
3. If substantial evidence of efficacy is required for inclusion of comparative drug efficacy, the comparisons from study [redacted] should be removed [redacted]
4. Discussion of patient reported outcomes should report differences between treatments and their associated confidence intervals calculated according to the statistical analysis plans.
5. Results from all domains and components should be included in the discussion of SF-36.
6. [redacted] Efficacy for [redacted] should not be claimed unless supported by tipping point analyses.
7. Discussions of physical function response in studies III and IV should explicitly state that the conclusions are supported at week 12.
8. The discussion [redacted] should therefore be removed from the product label. However, if included, the discussion additionally seems to imply [redacted]
9. Results from Table 6 at pre-rescue timepoints more appropriately address differences between the two proposed baricitinib doses.

10. To facilitate interpretation of multiple endpoints, any comparisons to approved products should be consolidated into a separate subsection.

6 APPENDICES

6.1 Baseline Demographic Characteristics

Table 67. Baseline Demographics, mITT Population, Study JADV

Category		Pbo	B4	A
Randomized		488	487	330
Age (mean)		53	54	53
	< 65 years	83%	79%	83%
	< 75 years	98%	97%	98%
Male (%)		22%	23%	24%
Race N (%)				
	White	302 (62.0%)	312 (64.1%)	204 (61.8%)
	Black	4 (0.8%)	2 (0.4%)	4 (1.2%)
	Asian	148 (30.4%)	143 (29.4%)	101 (30.6%)
	Other	7 (1.4%)	11 (2.3%)	3 (0.9%)
Weight (kg)		70	70	70

Source: CSR Table JADV.11.2

Table 68. Baseline Demographics, mITT Population, Study JADW

Category		Pbo	B2	B4
Randomized		176	174	177
Age (mean)		56	55	56
	< 65 years	77%	80%	77%
	< 75 years	98%	97%	97%
Male (%)		18%	21%	16%
Race N (%)				
	White	147 (83.5%)	144 (82.8%)	144 (82.8%)
	Black	8 (4.5%)	9 (5.2%)	7 (4.0%)
	Asian	11 (6.3%)	9 (5.2%)	12 (6.9%)
	Other	1 (0.6%)	0 (0%)	0 (0%)
Weight (kg)		82	83	81

Source: CSR Table JADW.11.1

Table 69. Baseline Demographics, mITT Population, Study JADX

Category		Pbo	B2	B4
Randomized		228	229	227
Age (mean)		51	52	52
	< 65 years	86%	86%	86%
	< 75 years	98%	98%	97%
Male (%)		17%	20%	18%
Race N (%)				
White		153 (67.1%)	156 (68.1%)	148 (65.5%)
	Black	10 (4.4%)	9 (3.9%)	9 (4.0%)
	Asian	60 (26.3%)	61 (26.6%)	59 (26.1%)
	Other	2 (0.8%)	1 (0.4%)	1 (0.4%)
Weight (kg)		76	75	78

Source: CSR Table JADX.11.1

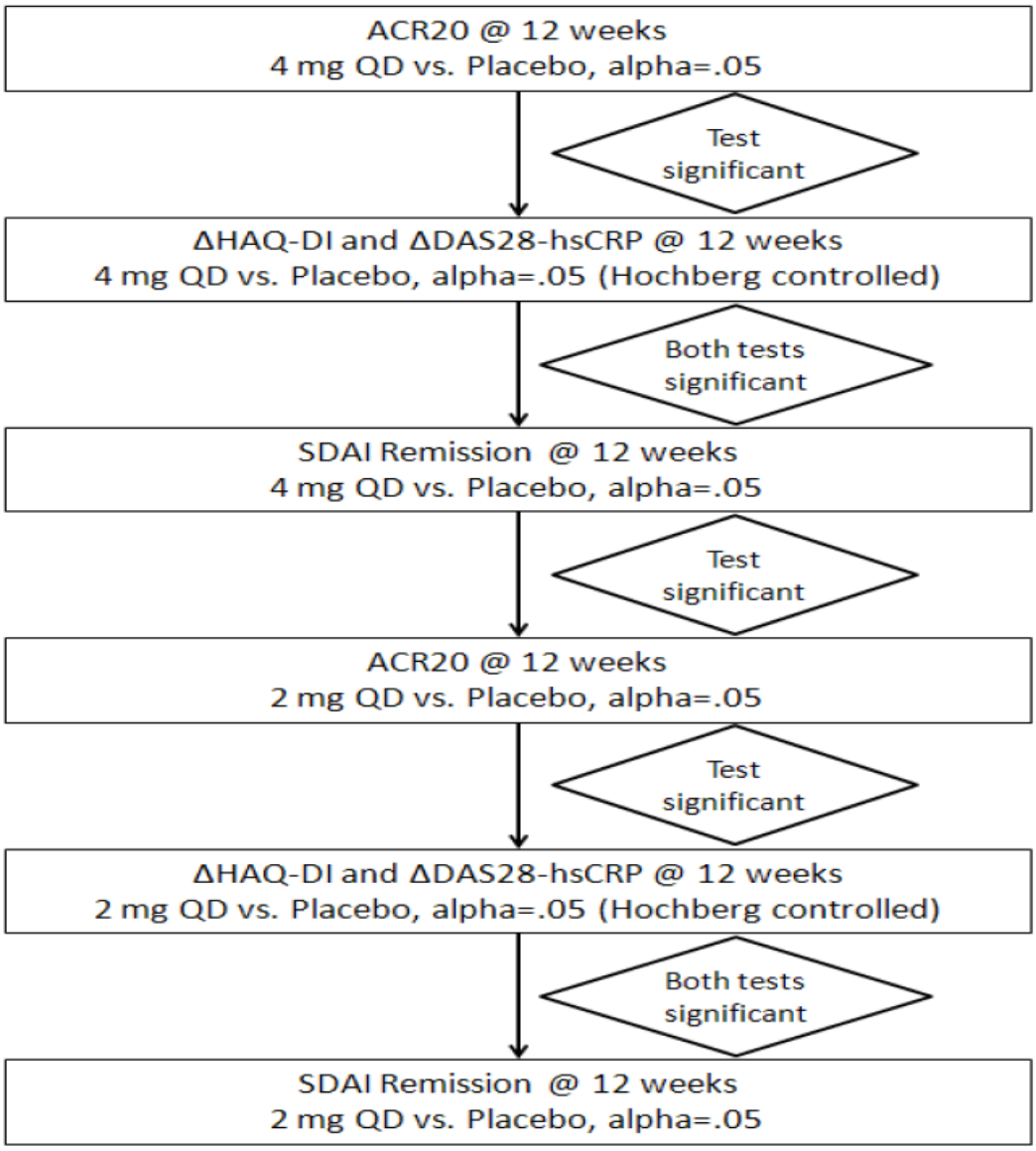
Table 70. Baseline Demographics, mITT Population, Study JADZ

Category		MTX	B4	B4 + MTX
Randomized		210	159	215
Age (mean)		51	51	49
	< 65 years	83%	84%	90%
	< 75 years	99%	98%	99%
Male (%)		30%	24%	27%
Race N (%)				
	White	128 (61.0%)	98 (61.6%)	123 (57.2%)
	Black	10 (4.8%)	5 (3.1%)	10 (4.7%)
	Asian	60 (28.6%)	44 (27.7%)	61 (28.4%)
	Other	1 (0.5%)	2 (1.3%)	1 (0.5%)
Weight (kg)		71	70	71

Source: CSR Table JADZ.11.1

6.2 Multiple Testing: Control of Type 1 Error

Figure 2. Multiple Test Procedure, Study JADW

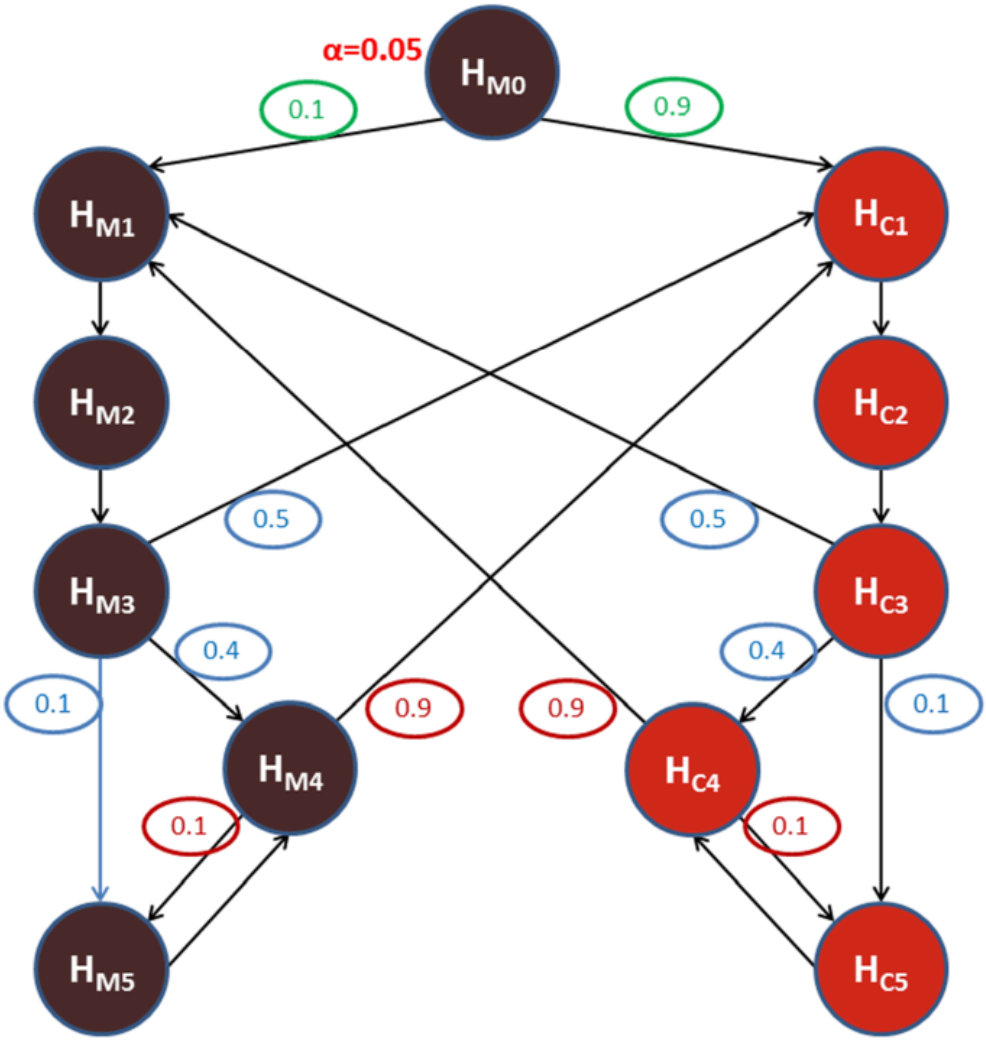


source: protocol Figure JADW.2

Key to Figure 2

Step 1: Test ACR20 at Week 12 for baricitinib 4 mg versus placebo at 1-sided alpha=0.025 (2-sided alpha=0.05). If the null hypothesis is rejected, proceed to Step 2. Otherwise, discontinue testing.
Step 2: Test the change from baseline in HAQ-DI at Week 12 and the change from baseline in DAS28-hsCRP at Week 12, both tests of baricitinib 4 mg versus placebo, at overall 1-sided alpha=0.025 (2-sided alpha=0.05) via a Hochberg procedure (Hochberg 1988). If both null hypotheses are rejected, proceed to Step 3. Otherwise, discontinue testing.
Details of the Hochberg procedure at this step:
Let p2 and p3 denote the unadjusted 1-sided p-values resulting from testing the hypotheses regarding HAQ-DI and DAS28-hsCRP, respectively.
If p2 and p3 $\leq$0.025, then Lilly will reject both null hypotheses (declaring statistical significance for both HAQ-DI and DAS28-hsCRP).
Otherwise, if p2 $\leq$0.0125 and p3 >0.025, then Lilly will reject only the null hypothesis (declaring statistical significance) related to HAQ-DI.
Otherwise if p3 $\leq$0.0125 and p2 >0.025, then Lilly will reject only the null hypothesis (declaring statistical significance) related to DAS28-hsCRP.
Otherwise Lilly will fail to reject both null hypotheses and discontinue testing.
Step 3: Test the proportion of patients achieving an SDAI score $\leq$3.3 (SDAI remission) at Week 12 for baricitinib 4 mg versus placebo at 1-sided alpha=0.025 (2-sided alpha=0.05). If the null hypothesis is rejected, proceed to Step 4. Otherwise, discontinue testing.
Step 4: Test ACR20 at Week 12 for baricitinib 2 mg versus placebo at 1-sided alpha=0.025 (2-sided alpha=0.05). If the null hypothesis is rejected, proceed to Step 5. Otherwise, discontinue testing.
Step 5: Test the change from baseline in HAQ-DI at Week 12 and the change from baseline in DAS28-hsCRP at Week 12, both tests of baricitinib 2 mg versus placebo, at overall 1-sided alpha=0.025 (2-sided alpha=0.05) via a Hochberg procedure and following similar details as outlined in Step 2. If both null hypotheses are rejected, proceed to Step 6. Otherwise, discontinue testing.
Step 6: Test the proportion of patients achieving an SDAI score $\leq$3.3 (SDAI remission) at Week 12 for baricitinib 2 mg versus placebo at 1-sided alpha=0.025 (2-sided alpha=0.05).

Figure 3. Multiple Test Procedure, Study JADZ



source: Figure JADZ.9.2 of CSR

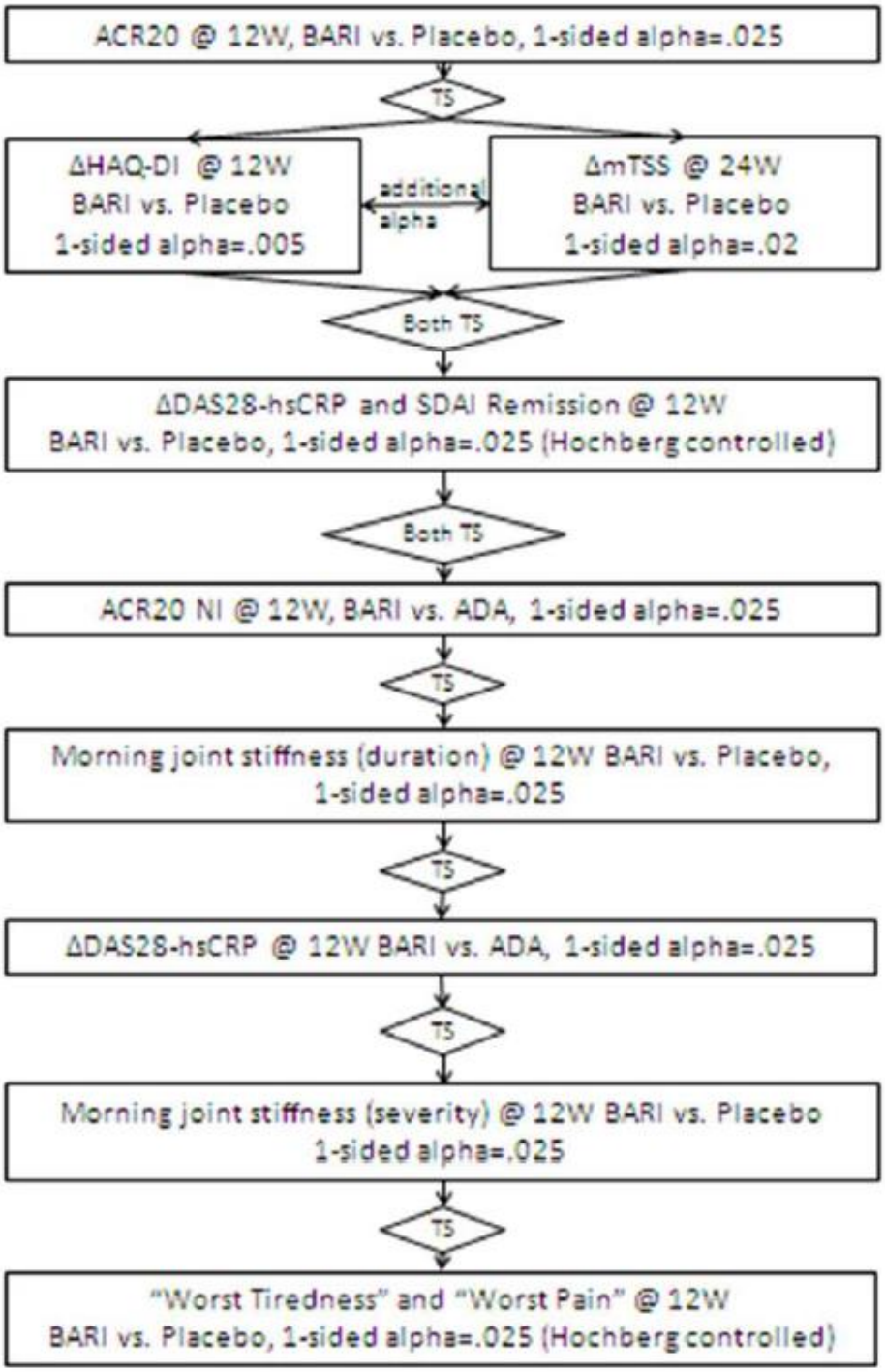
Key to Figure 3

- **HM0**: Proportion of patients who achieved ACR20 at Week 24, noninferiority of baricitinib monotherapy to MTX monotherapy
- **HM1**: Proportion of patients who achieved ACR20 at Week 24, superiority of baricitinib monotherapy to MTX monotherapy
- **HM2**: Mean change from baseline in DAS28-hsCRP at Week 24, baricitinib monotherapy vs MTX monotherapy

ortion of patients who achieved SDAI ≤3.3 at Week 24, baricitinib monotherapy vs MTX monotherapy

- **HM3**: Mean change from baseline in HAQ-DI at Week 24, baricitinib monotherapy vs MTX monotherapy
- **HM4**: Mean change from baseline in mTSS at Week 24, baricitinib monotherapy vs MTX monotherapy
- **HM5**: Proportion of patients who achieved SDAI ≤3.3 at Week 24, baricitinib monotherapy vs MTX monotherapy
- **HC1**: Proportion of patients who achieved ACR20 at Week 24, baricitinib plus MTX vs MTX monotherapy
- **HC2**: Mean change in DAS28-hsCRP at Week 24, baricitinib plus MTX vs MTX monotherapy
- **HC3**: Mean change in HAQ-DI at Week 24, baricitinib plus MTX vs MTX monotherapy
- **HC4**: Mean change in mTSS at Week 24, baricitinib plus MTX vs MTX monotherapy
- **HC5**: Proportion of patients who achieved SDAI ≤3.3 at Week 24, baricitinib plus MTX vs MTX monotherapy

Figure 4. Multiple Test Procedure, Study JADV



source: Figure JADV.2, Study Protocol

Key to Figure 4

Step 1: Test ACR20 at Week 12 for baricitinib versus placebo at 1-sided alpha = 0.025. If the null hypothesis is rejected, proceed to Step 2. Otherwise, discontinue testing.

Step 2: Test the change from baseline in HAQ-DI at Week 12 and the change from baseline in mTSS at Week 24, both tests of baricitinib versus placebo, at 1-sided alpha = 0.025 via a weighted Holm procedure (Holm 1979) with weights of 0.2 and 0.8, respectively. If both null hypotheses are rejected, proceed to Step 3. Otherwise, discontinue testing.

Details of the Holm procedure at this step:

Let p2 denote the unadjusted 1-sided p-value resulting from testing the hypothesis regarding HAQ-DI and let p3 denote the unadjusted 1-sided p-value resulting from testing the hypothesis regarding mTSS. Let p2* and p3* denote the corresponding weight-adjusted p-values (ie, p2* = p2/0.2 and p3*=p3/0.8). The following decision rules apply:
If min(p2*, p3*) >0.025, then Lilly will fail to reject both null hypotheses and discontinue testing.

Otherwise, if p2* ≤0.025, then Lilly will reject the null hypothesis for HAQ-DI (declaring statistical significance) and test mTSS using the unadjusted p-value. If p3 ≤0.025, Lilly will also reject the null hypothesis for mTSS (declaring statistical significance); however if p3 >0.025, then Lilly will fail to reject the null hypothesis for mTSS and discontinue testing.

Otherwise, if p3* ≤0.025, then Lilly will reject the null hypothesis for mTSS (declaring statistical significance) and test HAQ-DI using the unadjusted p-value. If p2 ≤0.025, Lilly will also reject the null hypothesis for HAQ-DI (declaring statistical significance); however if p2 >0.025, then Lilly will fail to reject the null hypothesis for HAQ-DI and discontinue testing.

Step 3: Test the change from baseline in DAS28-hsCRP at Week 12 and the proportion of patients achieving an SDAI score ≤3.3 (SDAI remission) at Week 12, both tests of baricitinib versus placebo, at 1-sided alpha=0.025 via a Hochberg procedure (Hochberg 1988). If both null hypotheses are rejected, proceed to Step 4. Otherwise, discontinue testing.

Details of the Hochberg procedure at this step:

Let p1 and p2 denote the unadjusted 1-sided p-values resulting from testing the hypotheses regarding DAS28-hsCRP and SDAI remission, respectively.

If p1 and p2 ≤0.025, then Lilly will reject both null hypotheses (declaring statistical significance for both DAS28-hsCRP and SDAI remission).

Otherwise, if p1 ≤0.0125 and p2 >0.025, then Lilly will reject only the null hypothesis (declaring statistical significance) related to DAS28-hsCRP.

Otherwise if p2 ≤0.0125 and p1 >0.025, then Lilly will reject only the null hypothesis (declaring statistical significance) related to SDAI remission.
Otherwise Lilly will fail to reject both null hypotheses and discontinue testing. Step 4: Test ACR20 at Week 12 for noninferiority of baricitinib to adalimumab at 1-sided alpha=0.025. Details concerning the noninferiority margin and statistical conclusions are found in Section 12.2.7. If the null hypothesis is rejected, proceed to Step 5. Otherwise, discontinue testing.

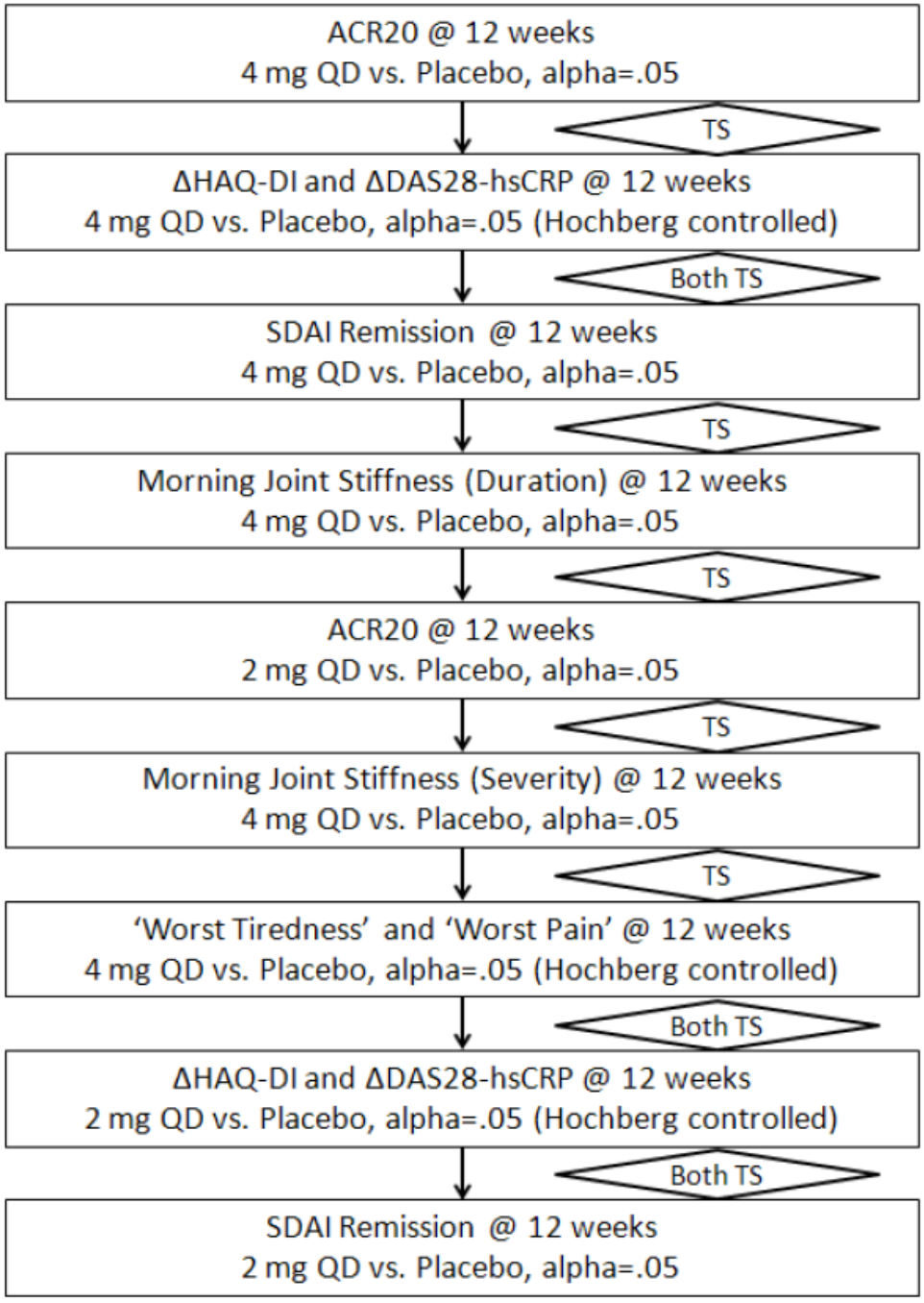
Step 5: Test mean duration of morning joint stiffness in the 7 days prior to Week 12 for baricitinib versus placebo at 1-sided alpha=0.025. If the null hypothesis is rejected, proceed to Step 6. Otherwise, discontinue testing.

Step 6: Test the change from baseline in DAS28-hsCRP at Week 12 for baricitinib versus adalimumab at 1-sided alpha=0.025. If the null hypothesis is rejected, proceed to Step 7. Otherwise, discontinue testing.

Step 7: Test mean severity of morning joint stiffness in the 7 days prior to Week 12 for baricitinib versus placebo at 1-sided alpha=0.025. If the null hypothesis is rejected, proceed to Step 8. Otherwise, discontinue testing.

Step 8: Test mean Worst Tiredness NRS in the 7 days prior to Week 12 and mean Worst Pain NRS in the 7 days prior to Week 12, both tests of baricitinib versus placebo, at 1-sided alpha=0.025 via a Hochberg procedure and following similar details as outlined in Step 3.

Figure 5. Multiple Test Procedure, Study JADX



Source: Figure JADX.2, study protocol

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

ROBERT ABUGOV
11/17/2016

GREGORY P LEVIN
11/17/2016

THOMAS J PERMUTT
11/17/2016
I concur.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDIES

NDA/BLA #:	NDA 207924 (IND 102204)
Drug Name:	Baricitinib (LY3009104)
Indication(s):	Oral Treatment for Rheumatoid Arthritis
Applicant:	Eli Lilly and Company Lilly Corporate Center, Indianapolis, Indiana 46285, USA (b) (4)
Date(s):	Received on 1/15/2016
Documents Reviewed:	Study 8291245 (transgenic mice) and Study 8253534 (rats) reports and electronic datasets submitted with the electronic submission on 3/12/2015 via IND 102204/S0258
Review Priority:	Regular
Biometrics Division:	Division of Biometrics VI
Statistical Reviewer:	Feng Zhou, M.S.
Concurring Reviewers:	Karl Lin, Ph.D., Team Leader
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products (DPARP)
Pharmacology Team:	Matthew Whittaker, Ph.D; Timothy W Robison, Ph.D
Project Manager:	Jessica K Lee, Project Manager
Keywords:	Carcinogenicity, Dose response

Table of Contents

1 Summary

This review evaluates statistically the tumorigenicity data of carcinogenicity studies of LY3009104 (NDA 207924). The studies included a 2-year study in the Sprague Dawley rats and a 26 week study in the Tg.rasH2 mice. The review analyzes the dose-response relationship of tumor incidence and mortality (including tumor-related mortality). The analyses of tumor data consisted of trend analyses for dose-response relationship in tumor incidence and pairwise comparisons in tumor incidence between individual treated groups and the vehicle control. For the mouse study, pairwise comparisons were also carried out between the positive control and the vehicle control groups.

From the statistical point of view, the review concludes that LY3009104 had a negative effect on survival. There were dose-dependent decreases in mortality in both sexes in rats. The tumor analysis showed no statistically significant dose-response relationship in tumor incidence for either sex in rats or mice.

Rat Study: Rats (60/sex/dose) were dosed by the oral gavage with LY3009104 daily for up to 104 weeks. The respective LY3009104 doses rats in the low (LD), mid (MD), and high-dose (HD) groups were 0, 1, 3, and 8 mg/kg in males and 0, 3, 8, and 25 mg/kg in females. The male rats and females were terminated at Weeks 95 and 91, respectively.

The survival analysis showed a statistically significant dose-response relationship in mortality in both males ($p= 0.0104$) and females ($p=0.0378$) in trend test. The pairwise comparisons showed only the HD group had statistically significant decreases in mortality when compared against the VC ($p=0.0077$ and 0.0274 in males and females, respectively). The respective survival rate in the VC, LD, MD, and HD groups at the termination (Week 95 or 91) were 32%, 38%, 47%, and 53% in males and 38%, 42%, 43%, and 55% in females.

There was no statistically significant dose-response relationship in incidence in any tumors in either sex.

Mouse Study: Mice (25/sex/dose) were dosed by the oral gavage with LY3009104 daily for up to 26 weeks. The respective LY3009104 dose in the low (LD), mid (MD), and high-dose (HD) groups was 0, 15, 40, and 300 mg/kg in males and 0, 10, 30, and 150 mg/kg in females. The study had two control groups: vehicle (VC), and positive control (PC). The PC mice (10/sex) were dosed with 75-mg/kg N-methyl-N-nitrosourea (MNU).

The survival analysis showed no statistically significant effects on mortality in either trend analysis or pairwise comparison in LY3009104 treatment groups in either sex. The pairwise comparisons showed a statistically significant increase in mortality between vehicle control and positive control ($p<0.0001$). The respective survival rates in the VC, LD, MD, HD, PC groups at the termination (Week 26) were 96%, 96%, 96%, 96%, and 30% in male mice; 88%, 92%, 88%, 92%, and 20% in female mice.

The tumor analysis did not show any statistically significant dose-response relationship in incidence in any tumors in males or females. The PC group showed statistically significant increases in the incidence of a number of tumors in both males and females ($p<0.05$), when compared against the vehicle control. Those tumor types included hemangiosarcoma (in males only) and lymphosarcoma in the entire body, papilloma and squamou in skin/subcutis, and squamous papilloma squamous in the stomach.

2 Background

The sponsor conducted two studies: a 26-week subcutaneous carcinogenicity study in transgenic TgRasH2 mice (8291245); a 24-month carcinogenicity study in the Sprague Dawley rats (8253534). This review analyzed the SAS data sets of these studies received from the sponsor on 3/12/2015 via submission IND 102204/S0258.

The pharmacologist Dr. Matthew Whittaker has a review issue about the early study termination at week 91 of all surviving females in the rat study. The sponsor terminated all groups due to decreased numbers of control females. However, there were still 23 surviving control females at the time of the study termination. The ECAC typically recommends study termination when the number of control animals decreases below 20.

The phrase "dose response relationship" refers to the linear component of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases. Results of this review have been discussed with the reviewing pharmacologist Dr. Matthew Whittaker.

3 Rat Study

Study Report: 8253534.pdf; SAS data: Tumor.xpt

This study assessed the carcinogenic potential of LY3009104 in male and female Sprague Dawley rats. The test material was administered daily by oral gavage at doses of 0, 1, 3, and 8 mg/kg for male rats and 0, 3, 8, and 25 mg/kg for female rats for at least 104 weeks. This review refers these dose groups as the vehicle control (VC), low (LD), mid (MD), and high (HD) dose groups, respectively. The dose volume was 10 mL/kg. There were 60 rats/sex/dose. Assessment of oncogenic potential was based on mortality, clinical observations, body weight, food consumption, and anatomic pathology. The sponsor claimed that due to low survival in the control animals, the study was terminated after Week 94 for the males and during Week 91 for the females. However, at least 28 animals per sex in each group survived to Week 80, which provided sufficient numbers of animals to adequately evaluate toxicity (FDA Draft Guidance for Industry, 2001).

3.1 Sponsor's Analyses

3.1.1 Survival Analysis

Adjusted survival data were analyzed using the National Cancer Institute Life Table Package (Thomas et al., 1977). Week 95 and 91 of the dosing phase was treated as the end of the study in the National Cancer Institute package for males and females, respectively. The methods consisted of the Cox-Tarone binary regression and the Gehan-Breslow nonparametric test on life tables. Those animals sacrificed at the scheduled interval and animals sacrificed for other reasons (gavage-related) were censored (i.e., considered not compound-related) in the analyses. Additionally, Kaplan-Meier product-limit estimate survival curves were produced for graphical

evaluation of survival. One-sided tail probabilities for trend and group comparisons were evaluated at $p < 0.05$ significance level.

Sponsor's concluded:
For males, a significant negative trend (p = 0.0061 by Cox-Tarone test and p = 0.0044 by Gehan-Breslow test) occurred for mortality along with a significant decrease in mortality in Group 4 compared with Group 1 (p = 0.0066 by Cox-Tarone test and p = 0.0033 by Gehan-Breslow test).

For females, a significant negative trend (p = 0.0054 by Cox-Tarone test and p = 0.0052 by Gehan-Breslow test) occurred for mortality along with a significant decrease in mortality in Group 4 compared with Group 1 (p = 0.0049 by Cox-Tarone test and p = 0.0033 by Gehan-Breslow test).

The Sponsor's Report of Total Unscheduled Deaths, n=60/sex/dose group)
Male Rats

Group	1	2	3	4
Dose (mg/kg)	0	1	3	8
Unadjusted Mortality (Rate)	40/60 (0.667)	37/60 (0.617)	32/60 (0.533)	28/60 (0.467)
Adjusted Mortality Kaplan-Meier Estimate (Final)	0.685	0.617	0.533	0.474
Standard Error	0.061	0.063	0.064	0.065

	Cox-Tarone Test	Gehan-Breslow Test
1 versus 2-4 p-value (one-sided)	0.0061-**	0.0044-**
1 versus 2 p-value (one-sided)	0.2874-	0.2736-
1 versus 3 p-value (one-sided)	0.0733-	0.0755-
1 versus 4 p-value (one-sided)	0.0066-**	0.0033-**

Female Rats

Group	1	2	3	4
Dose (mg/kg)	0	3	8	25
Unadjusted Mortality (Rate)	39/60 (0.650)	34/60 (0.567)	31/60 (0.517)	24/60 (0.400)
Adjusted Mortality Kaplan-Meier Estimate (Final)	0.660	0.576	0.543	0.429
Standard Error	0.062	0.064	0.066	0.066

	Cox-Tarone Test	Gehan-Breslow Test
1 versus 2-4 p-value (one-sided)	0.0054-**	0.0052-**
1 versus 2 p-value (one-sided)	0.1543-	0.1047-
1 versus 3 p-value (one-sided)	0.1483-	0.1471-
1 versus 4 p-value (one-sided)	0.0049-**	0.0033-**

[Source: page 2785 of study report of 8253534.pdf]
** = Significant at 1% level. - = Effect in the decreased direction.

3.1.2 Tumor Data Analysis

Neoplastic lesions were chosen for statistical analyses if the incidence in at least one of the dosed groups was increased or decreased by at least two occurrences over the vehicle control group. The occult (i.e., internal organ or nonobservable) tumors (incidental alone or incidental and fatal combined) were analyzed by the International Agency for Research on Cancer asymptotic fixed interval-based prevalence test (Peto et al., 1980). The cutoff points for the interval-based test were Study Weeks 0 to 52, 53 to 78, 79 to 92, 93 to before terminal necropsy, and terminal necropsy for males; and Study Weeks 0 to 52, 53 to 78, 79 to before terminal necropsy, and

terminal necropsy for females. The test was implemented using PROC MULTTEST in the SAS system (SAS, 2002-2008).
In the case of sparse tables (<10 total in all groups combined), the exact form of the test was used. Palpable (superficial as in mammary or skin) tumors were analyzed by the Cox-Tarone binary regression method (as under the Survival method section) using the tumor onset time or the first palpation time (as applicable) as a surrogate for the death time.

The one-sided trend and group comparisons were evaluated at the 5.0% significance level. Benign and malignant neoplastic incidences were evaluated separately and combined, where appropriate. The criterion for combination was based on the work of McConnell et al. (McConnell et al., 1986) and determined by the principal investigator for anatomic pathology. Incidences of multiple-organ and combined neoplastic findings, such as hemangioma, fibrosarcoma, and endometrial stromal polyp, were counted by animal, not by tissue type.

Nominal dose levels (i.e., the starting arithmetic dose levels) were used in the analyses. For all analyses, the two sexes were kept separate.

Adjustment for multiple testing: In order to control the overall false positive error, the sponsor tested the common and the rare tumors at 0.005 and 0.025 significance levels, respectively (Lin, 2000) for positive dose response relationships in individual tumor types, and at 0.01 and 0.05 for pairwise comparisons in individual tumor types. Tumors are considered by the sponsors as common with a background rate of ≥ 1% and as rare with a background incidence of < 1%.

Sponsor's concluded: All significant findings in males and females were found to be the negative trend or decreases when comparing the dosed groups to the control group. No positive findings were noted in neoplastic lesions for males and females.

3.2 Reviewer's Analyses

To verify the sponsor's analyses and to perform additional analyses suggested by the reviewing pharmacologist, this reviewer performed survival and tumor data analyses using data submitted electronically in IND 102204 on 3/12/2015.

3.2.1 Survival Analysis

The survival distributions of rats in all treatment groups were estimated using the Kaplan-Meier product limit method. For control, low, medium, mid-high, and high dose groups, the dose response relationship was tested using the likelihood ratio test and the homogeneity of survival distributions was tested using the log-rank test. The Kaplan-Meier curves for survival rates are given in Figures 1A and 1B in the appendix for male and female rats, respectively. The intercurrent mortality data are given in Tables 1A and 1B in the appendix for male and female rats, respectively. Results of the tests for dose response relationship and homogeneity of survivals, are given in Tables 1A and 1B in the appendix for male and female rats, respectively.

Reviewer's findings: The sponsor terminated the male rats at Week 95 and female rats at Week 91. This reviewer's analysis showed the numbers (percent) of death were 41 (68%), 37 (62%), 32 (53%), and 28 (47%) in male rats and 37 (62%), 35 (58%), 34 (57%), and 26 (45%) in female rats in the VC, LD, MD, and HD groups, respectively. The tests did show a statistically significant dose response relationship in mortality across VC and treated groups for male rats (p= 0.0104) and female rats

(p=0.0378). The pairwise comparisons did show a statistically significant mortality between VC and HD for male rats (p=0.0077) and female rats (p=0.0274).

Reviewer's comment: *The sponsor's analysis showed 40 (67%) death, while this reviewer's analysis showed 41 (68%) death in VC group for male rats. These differences are due to the facts that three animals (B34158 and B34155 in VC, B34376 in HD) that were dead accidentally before the week 74 were classified as deaths by this reviewer as deaths while as survivor by the sponsor; and two animals (B34210 in VC and B34385 in HD) that died naturally during their respective terminal sacrifice week (week 95) were classified as survivors by this reviewer, while as deaths by the sponsor.*

For females, the sponsor's analysis showed 39 (65%), 34 (58%), 31 (54%), and 24 (43%) death, while this reviewer's analysis showed 37 (62), 35 (58%), 34 (57%), and 26 (43%) death in VC, LD, MD, and HD group respectively. These differences are due to the fats that nine animals (B34451 in VC, B34525 in LD, B34577, B34599, B34609 in MD, B34661, B34663, B34667, B34708 in HD) that were dead accidentally before the week 60 were classified as deaths by this reviewer while as survivors by the sponsor and five animals (B34468, B34484, and B34474 in VC, B34648, B34650 in HD) that died naturally during their respective terminal sacrifice week and accidental death were classified as survivors by this reviewer while as deaths by the sponsor.

These discrepancies did not significantly affect the survival analysis results.

3.2.2 Tumor Data Analysis

The tumor data were analyzed for dose response relationships and pairwise comparisons of control group with each of the treated groups. Both the dose response relationship tests and pairwise comparisons were performed using the Poly-k method described in the paper of Bailer and Portier (1988) and Bieler and Williams (1993). In this method an animal that lives the full study period ($w_{\max}$) or dies before the terminal sacrifice but develops the tumor type being tested gets a score of s_h =1. An animal that dies at week w_h without developing the tumor before the end of the study gets a score of $s_h = \left(\frac{w_h}{w_{\max}} \right)^k$ <1. The adjusted group size is defined as Σs_h . As an interpretation, an animal with score s_h =1 can be considered as a whole animal while an animal with score $s_h < 1$ can be considered as a partial animal. The adjusted group size Σs_h is equal to N (the original group size) if all animals live up to the end of the study or if each animal that dies before the terminal sacrifice develops at least one tumor, otherwise the adjusted group size is less than N. These adjusted group sizes are then used for the dose response relationship (or the pairwise) tests using the Cochran-Armitage test. One critical point for Poly-k test is the choice of the appropriate value of k, which depends on the tumor incidence pattern with the increased dose. For long term 104 week standard rat and mouse studies, a value of k=3 is suggested in the literature. Hence, this reviewer used k=3 for the analysis of this data. For the calculation of p-values the exact permutation method was used. The tumor rates and the p-values of the tested tumor types are listed in Tables 5A and 5B in the appendix for male and female rats, respectively.

Multiple testing adjustment: For the adjustment of multiple testing of dose response relationship, the FDA guidance for the carcinogenicity study design and data analysis suggests the use of test levels α=0.005 for common tumors and α=0.025 for rare tumors for a submission with two species, and a significance level α=0.01 for common tumors and α=0.05 for rare tumors for a submission

with one species in order to keep the false-positive rate at the nominal level of approximately 10%. A rare tumor is defined as one in which the published spontaneous tumor rate is less than 1%. For multiple pairwise comparisons of treated group with control the FDA guidance the suggested the use of test levels α=0.01 for common tumors and α=0.05 for rare tumors, in order to keep the false-positive rate at the nominal level of approximately 10% for both submissions with two or one species.

It should be noted that the FDA guidance for multiple testing for dose response relationship is based on a publication by Lin and Rahman (1998). In this work the authors investigated the use of this rule for Peto analysis. However, in a later work Rahman and Lin (2008) showed that this rule for multiple testing for dose response relationship is also suitable for Poly-K tests.

The tumor rates and the p-values of the tested tumor types are listed in Tables 3A and 3B in the appendix for male and female rats, respectively.

Reviewer's findings: The tumor analysis did not show any statistically significant dose-response relationship in incidence in all tumors tested in male and female rats.

4 Mouse Study

Study Report: 8291245.pdf; SAS data: Tumor.xpt

This study assessed the carcinogenic potential of LY3009104 in male and female hemizygous Tg.rasH2 mice. The test material was administered daily by oral gavage at doses of 0, 15, 40, and 300 mg/kg for male mice and 0, 10, 30, and 150 mg/kg for female mice for approximately 26 weeks This review refers these dose groups as the vehicle control (VC), low (LD), mid (MD), and high (HD) dose groups, respectively. The dose volume was 10 mL/kg. There were 25 mice /sex/dose. There were 10 animals for each sex dosed with one intraperitoneal dose of N-methyl-N-nitrosourea (MNU) on Day 1 of the dosing phase at a dose level of 75 mg/kg IP at a dose volume of 10 mL/kg as the positive control (PC).

4.1 Sponsor's Analyses

4.1.1 Survival Analysis

Adjusted survival data were analyzed using the National Cancer Institute Life Table Package (Thomas et al., 1977). Week 27 of the dosing phase was treated as the end of the study in the National Cancer Institute package for males and females. The methods consisted of the Cox-Tarone binary regression and the Gehan-Breslow nonparametric test on life tables. Those animals sacrificed at the scheduled interval were censored (i.e., considered not compound-related) in analyses. Additionally, Kaplan-Meier productlimit estimate survival curves were produced for graphical evaluation of survival. One-sided tail probabilities for trend and group comparisons were evaluated at the $p < 0.05$ significance level.

Sponsor's findings: As depicted in the Kaplan-Meier curves and statistical analysis results in males, the positive control group had higher mortality than the vehicle control group (7/10 versus 1/25), with $p = 0.0001$ by Cox-Tarone test and $p = 0.0000$ by Gehan-Breslow tests. No significant increase in mortality was noted in males for any of the groups given the test article. As depicted in the Kaplan-Meier curves and statistical results in females, the positive control group had higher

mortality than the vehicle control group (8/10 versus 3/25), with p = 0.0001 by Cox-Tarone test and p = 0.0000 by Gehan-Breslow tests. No significant increase in mortality was noted in females for any of the groups given the test article.

The Sponsor's Report of Total Unscheduled Deaths, n=25/sex/dose group)

Male Rats

Group	1	2	3	4	5
Dose (mg/kg)	0 (Vehicle)	15	40	300	75 (Positive)
Unadjusted Mortality (Rate)	1/25 (0.040)	1/25 (0.040)	1/25 (0.040)	1/25 (0.040)	7/10 (0.700)
Adjusted Mortality Kaplan-Meier Estimate (Final)	0.040	0.040	0.040	0.040	0.700
Standard Error	0.039	0.039	0.039	0.039	0.145

Female Rats

Group	1	2	3	4	5
Dose (mg/kg)	0 (Vehicle)	10	30	150	75 (Positive)
Unadjusted Mortality (Rate)	3/25 (0.120)	2/25 (0.080)	3/25 (0.120)	2/25 (0.080)	8/10 (0.800)
Adjusted Mortality Kaplan-Meier Estimate (Final)	0.120	0.080	0.120	0.080	0.800
Standard Error	0.065	0.054	0.065	0.054	0.126

[Source: page 1418 of study report of 8291245.pdf]

4.1.2 Tumor Data Analysis

All non-palpable neoplastic lesions were analyzed by Cochran-Armitage test for trend (for Group 1 versus Groups 2 through 4) and by Fisher-Irwin exact test (Thakur et al., 1985) for group comparisons (including Group 5) versus the vehicle control group because no survival difference occurred among groups and the above survival-unadjusted tests are considered as appropriate. Neoplastic lesions were chosen for statistical analyses if the incidence in at least one of the dosed groups was increased or decreased by at least two occurrences over the control group. Analyses were performed using the SAS system (SAS, 2002-2008).

Palpable (superficial as in mammary or skin) tumors were analyzed by the Cox-Tarone binary regression method (as under the Survival method section) using the death time or the first palpation time (as applicable) as a surrogate for the tumor onset time.

One-sided positive trends in common (background incidence rate >1%) and rare (background incidence rate <1%) tumors (if applicable) defined by the principal investigator for anatomic pathology were evaluated at the 0.01 and 0.05 significance levels, respectively. High-dose group comparisons in common and rare tumors were evaluated at the 0.01 and 0.05 significance levels (FDA Draft Guidance for Industry, 2001). Other intermediate, pair-wise, one-sided group comparisons were evaluated at the 5.0% significance level. Benign and malignant neoplastic incidences were evaluated separately and combined, where appropriate. The criterion for combination was based on the work of McConnell et al. (McConnell et al., 1986) and determined by the principal investigator for anatomic pathology. Incidences of multiple-organ and combined neoplastic findings, such as hemangioma, fibrosarcoma, and endometrial stromal polyp, were counted by animal, not by tissue type. They were evaluated statistically if they met the selection criterion for the analysis.

Nominal dose levels (i.e., the starting arithmetic dose levels) were used in analyses. For all analyses, sexes were kept separate.

Sponsor's findings: No significant findings were noted for neoplastic lesions of animals in any of the groups given the test article versus the vehicle control group.

File Name: outputfile805671446.pdf

4.2 Reviewer's Analyses

To verify the sponsor's analyses and to perform additional analyses suggested by the reviewing pharmacologist, this reviewer performed survival and tumor data analyses using data submitted electronically in NDA 207924 on 8/24/2015.

4.2.1 Survival Analysis

The Kaplan-Meier curves for survival rates of all treatment groups are given in Figures 2A and 2B in the appendix for male and female mice, respectively. The intercurrent mortality data of all treatment groups are given in Tables 2A and 2B in the appendix for male and female mice, respectively. Results of the tests for dose response relationship and homogeneity of survivals for control, low, medium, and high dose groups are given in Tables 2A and 2B in the appendix for male and female mice, respectively.

Reviewer's findings: The studies were terminated on week 27. This reviewer's analysis showed 1 (4%) death for VC, LD, MD, and HD groups and 7 (70%) deaths for positive control group in male mice, and the numbers (percent) of deaths were 3 (12%), 2 (8%), 3 (12%), 2 (8%), and 8 (80%) for VC, LD, MD, HD, PC groups in female mice, respectively. The tests did not show any statistically significant dose response relationship in mortality across control and treated groups in either sex mice. The pairwise comparisons did not show any statistically significant increase in mortality in the treated groups compared to the vehicle control in either male or female mice. The pairwise comparisons did show a statistically significant mortality difference between vehicle control and positive control ($p<0.0001$).

4.2.2 Tumor Data Analysis

The tumor data were analyzed for dose response relationships and pairwise comparisons of control group with each of the treated groups using the same method that was used for the rats study. The tumor rates and the p-values of the tested tumor types are listed in Tables 4A, 4B, 5B, and 5B in the appendix for male and female mice, respectively.

Reviewer's findings: Because of the small group size and short study duration used in transgenic mouse studies, based on the statistical guideline for transgenic mouse studies, the significance level of 0.05 was used in the tests for dose response and pairwise comparisons in tumor incidences of both rare and common tumors. Based on this recommendation of adjustment for multiple testing discussed above, the tumor analysis did not show any statistically significant dose-response relationship in incidence in all tumor types tested in male and female mice. The PC group showed statistically significant increases in the incidence of a number of tumors in both males and females ($p<0.05$), when compared to the vehicle control. Those tumor types included M-Hemangiosarcoma (in males only) and M-Lymphosarcoma in body, B-Papilloma, squamou in skin/subcutis, and B-Papilloma, squamou in stomach.

Tumor Types with P-Values ≤ 0.05 for Pairwise Comparisons of VC and PC in Male Mice

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	75 mg/kg PC (N=10)	P-Value VC vs. PC
Body, Whole/Cav	M-Hemangiosarcoma	0	4	0 0019*
	M-Lymphosarcoma	0	6	<0.001*
Lung	B-Adenoma, bronchiol	2	3	0.0619

File Name: outputfile805671446.pdf

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	75 mg/kg PC (N=10)	P-Value VC vs. PC
Skin/Subcutis	B-Adenoma, sebaceous	1	3	0.0278
	B-Papilloma, squamou	0	5	<0.001*
Stomach, Nongla	B-Papilloma, squamou	0	5	<0.001*

Tumor Types with P-Values ≤ 0.05 for Pairwise Comparisons of VC and PC in Female Mice

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	75 mg/kg PC (N=10)	P-Value VC vs. PC
Body, Whole/Cav	M-Lymphosarcoma	0	8	<0.001*
Harderian Gland	B-Adenoma	1	3	0.0307
Skin/Subcutis	B-Papilloma, squamou	0	2	0 0369*
Stomach, Nongla	B-Papilloma, squamou	0	6	<0.001*

5 Conclusion

This review evaluates statistically the tumorigenicity data of carcinogenicity studies of LY3009104 (NDA 207924). The studies included a 2-year study in the Sprague Dawley rats and a 26 week study in the Tg.rasH2 mice. The review analyzes the dose-response relationship of tumor incidence and mortality (including tumor-related mortality). The analyses of tumor data consisted of trend analyses for dose-response relationship in tumor incidence and pairwise comparisons in tumor incidence between individual treated groups and the vehicle control. For the mouse study, pairwise comparisons were also carried out between the positive control and the vehicle control groups.

From the statistical point of view, the review concludes that LY3009104 had a negative effect on survival. There were dose-dependent decreases in mortality in both sexes in rats. The tumor analysis showed no statistically significant dose-response relationship in tumor incidence for either sex in rats or mice.

Feng Zhou
Mathematical Statistician

Secondary Reviewer: Atiar Mohammad Rahman, Ph.D., Acting Deputy Director, Biometrics-6
Concurring Reviewers: Karl Lin, Ph.D., Team Leader

cc:
Dr. Whittaker
Dr. Robison
Dr. Tsong
Ms. Zhou
Dr. Lin
Ms. Patrician

File Name: outputfile805671446.pdf

6 Appendix

Table 1A: Intercurrent Mortality Rate in Male Rats

	0 mg/kg/day VC (n=60)		1 mg/kg/day LD (n=60)		3 mg/kg/day MD (n=60)		8 mg/kg/day HD (n=60)	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	6	10.00	8	13.33	5	8.33	1	1.67
53 - 78	20	43.33	9	28.33	10	25.00	7	13.33
79 - 90	8	56.67	18	58.33	15	50.00	19	45.00
91 - 94	7	68.33	2	61.67	2	53.33	1	46.67
Terminal sacrifice	19	31.67	23	38.33	28	46.67	32	53.33

Test	VC, LD, MD, HD	VC vs. LD	VC vs. MD	VC vs. HD
Dose-Response (Likelihood Ratio)	0.0104	0.3688	0.0747	0 0077
Homogeneity (Log-Rank)	0.0489	0.3642	0.0711	0 0066

 # All Cum. %Cumulative Percentage except for Terminal sacrifice
****=Significant at 1% level**

Table 1B: Intercurrent Mortality Rate in Female Rats

	0 mg/kg/day VC (n=60)		3 mg/kg/day LD (n=60)		8 mg/kg/day MD (n=60)		25 mg/kg/day HD (n=60)	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	5	8.33	4	6 67	5	8.33	8	13.33
53 - 78	20	41.67	11	25.00	15	33.33	6	23.33
79 - 90	12	61.67	20	58.33	14	56.67	12	43.33
91 - 94	.	.	.	.	.	.	.	.
Terminal sacrifice	23	39.33	25	41.67	26	43.33	34	56.67

Test	VC, LD, MD, HD	VC vs. LD	VC vs. MD	VC vs. HD
Dose-Response (Likelihood Ratio)	0.0378	0.2680	0.3789	0 0274
Homogeneity (Log-Rank)	0.1642	0.2616	0.3725	0 0254

 # All Cum. %Cumulative Percentage except for Terminal sacrifice
****=Significant at 1% level**

Table 2A: Intercurrent Mortality Rate in Male Mice

	0 mg/kg/day VC (n=25)		15 mg/kg/day LD (n=25)		40 mg/kg/day MD (n=25)		300 mg/kg/day HD (n=25)		75 mg/kg of Urethane PC (n=10)	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 – 13	.	.	.	.	.	.	.	.	.	.
14 - 26	1	4.00	1	4.00	1	4.00	1	4 00	7	70.00
Terminal sacrifice	24	96 00	24	96 00	24	96 00	24	96.00	3	30.00

Test	VC, LD, MD, HD	VC vs. LD	VC vs. MD	VC vs. HD	VC vs. PC
Dose-Response (Likelihood Ratio)	0.9963	0 9885	0.9885	0.9885	<0 0001
Homogeneity (Log-Rank)	1.0000	0 9885	0.9885	0.9885	<0 0001

 # All Cum. %Cumulative Percentage except for Terminal sacrifice
****=Significant at 1% level**

File Name: outputfile805671446.pdf

Table 2B: Intercurrent Mortality Rate in Female Mice

	0 mg/kg/day VC (n=25)		10 mg/kg/day LD (n=25)		30 mg/kg/day MD (n=25)		150 mg/kg/day HD (n=25)		75 mg/kg of Urethane PCI (n=10)	
Week	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**
0 -13	.	.	.	.	.	.	.	.	2	20.00
14 - 26	3	12 00	2	8.00	3	12.00	2	8.00	6	80.00
Terminal sacrifice	22	88 00	23	92.00	22	88.00	23	92 00	2	20.00
Test	**VC, LD, MD, HD**		**VC vs. LD**		**VC vs. MD**		**VC vs. HD**		**VC vs. PC**	
Dose-Response (Likelihood Ratio)	0.7025		0 6473		0.9864		0.6403		0 0002	
Homogeneity (Log-Rank)	0.9299		0 6470		0.9863		0.6413		<0.0001	

 # All Cum. %Cumulative Percentage except for Terminal sacrifice

****=Significant at 1% level**

Table 3A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons with Vehicle Control – Male Rats

Kidney	M-Carcinoma, tubule	1	0	0	0	1.0000	1.0000	1.0000	1.0000
Liver	B-Adenoma, hepatocel	0	0	0	1	0.2775	.	.	0.5517
	M-Carcinoma, hepatoc	1	0	0	0	1.0000	1.0000	1.0000	1.0000
Lung	B-Adenoma, bronchiol	0	1	0	0	0.7746	0.5185	.	.
Muscle, Biceps	B-Granular cell tumo	0	1	0	0	0.7746	0.5185	.	.
Muscle, Other	M-Fibrosarcoma	0	1	0	0	0.7746	0.5185	.	.
Muscle, Skeleta	M-Rhabdomyosarcoma	0	0	1	0	0.5345	.	0.5357	.
Pancreas	B-Adenoma, acinar Ce	0	0	1	0	0.5318	.	0.5301	.
	B-Adenoma, islet cel	3	2	1	3	0.5033	0.8352	0.9526	0.7444
	M-Carcinoma, acinar	0	1	0	0	0.7746	0.5185	.	.
	M-Carcinoma, islet c	1	2	1	4	0.1166	0.5281	0.7874	0.2517
	C IsletCell M+B	**4**	**4**	**2**	**7**	**0.1866**	**0.6712**	**0.9235**	**0.3765**
Parathyroid	B-Adenoma	1	0	1	0	0.7822	1.0000	0.7823	1.0000
Pituitary	B-Adenoma, pars dist	22	27	28	29	0.3396	0.3073	0.3474	0.3148
	M-Carcinoma, pars di	1	0	2	0	0.7289	1.0000	0.5446	1.0000
Skin/Subcutis	B-Adenoma, sebaceous	0	0	2	2	0.0943	.	0.2899	0.3015
	B-Amelanotic melanom	0	1	1	0	0.6669	0.5185	0.5301	.

File Name: outputfile805671446.pdf

Organ Name	Tumor Name	0 mg VC N=60	1 mkd LD N=60	3 mkd MD N=60	8 mkd HD N=60	P-Value			
						Dos-Resp	VC vs. LD	VC vs. MD	VC vs. HD
	B-Basal cell tumor	0	2	1	1	0.4781	0.2657	0.5301	0.5517
	B-Fibroma	2	0	0	3	0.1414	1.0000	1.0000	0.5868
	B-Keratoacanthoma	3	0	2	3	0.3224	1.0000	0.8484	0.7444
	B-Leiomyoma	2	0	0	0	1.0000	1.0000	1.0000	1.0000
	B-Lipoma	2	1	0	1	0.7525	0.8929	1.0000	0.9138
	B-Papilloma, squamou	0	2	0	1	0.4907	0.2657	.	0.5517
	M-Carcinoma, Zymbal'	0	0	1	0	0.5345	.	0.5357	.
	M-Carcinoma, basal c	0	1	0	2	0.1348	0.5185	.	0.3015
	M-Fibrosarcoma	0	3	1	2	0.3712	0.1441	0.5357	0.3072
	M-Osteosarcoma	0	0	0	1	0.2775	.	.	0.5517
	M-Sarcoma	0	0	1	0	0.5318	.	0.5301	.
Spinal Cord	B-Astrocytoma	0	2	1	0	0.7771	0.2657	0.5357	.
Testis	B-Interstitial cell	0	0	2	2	0.0950	.	0.2780	0.3015
Thymus	B-Thymoma	1	1	1	1	0.5778	0.7713	0.7823	0.8019
Thyroid	B-Adenoma, C-cell	7	6	6	3	0.9507	0.7567	0.7705	0.9753
	B-Adenoma, follicula	0	1	1	3	0.0610	0.5185	0.5301	0.1632
	M-Carcinoma, C-cell	2	2	1	0	0.9635	0.7199	0.9041	1.0000
	M-Carcinoma, follicu	1	0	1	0	0.7794	1.0000	0.7762	1.0000
	C FollicularCell M+B	**1**	**1**	**2**	**3**	**0.1679**	**0.7651**	**0.5361**	**0.3801**

Table 3B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons with Vehicle Control – Female Rats

Organ Name	Tumor Name	0 mg VC N=60	3 mkd LD N=60	8 mkd MD N=60	25 mkd HD N=60	P-Value			
						Dos-Resp	VC vs. LD	VC vs. MD	VC vs. HD
Adrenal, Medull	B-Complex pheochromo	0	0	1	0	0 5056	.	0.5059	.
	B-Pheochromocytoma	0	1	2	1	0 3466	0 5227	0.2588	0.5281
Body, Whole/Cav	B-Hemangioma	0	1	0	0	0.7640	0 5227	.	.
	M-Leukemia, myelogen	0	0	0	1	0 2640	.	.	0.5281
	M-Lymphosarcoma	0	0	1	1	0.1977	.	0.5116	0.5281
Brain	B-Astrocytoma	1	0	1	1	0.4100	1 0000	0.7644	0.7801
	B-Granular cell tumo	0	0	0	1	0 2640	.	.	0.5281
Cecum	B-Adenoma	1	0	0	0	1 0000	1 0000	1.0000	1.0000
Cervix	M-Sarcoma, stromal	0	0	0	1	0 2640	.	.	0.5281
Femur	M-Osteosarcoma	0	0	1	0	0 5056	.	0.5059	.
Gingiva	M-Carcinoma, squamou	0	0	0	1	0 2682	.	.	0.5333
Jejunum	M-Leiomyosarcoma	0	0	0	1	0 2640	.	.	0.5281
Liver	B-Adenoma, hepatocel	3	0	2	0	0 9230	1 0000	0.8200	1.0000
Lung	M-Carcinoma, bronchi	1	0	0	0	1 0000	1 0000	1.0000	1.0000
Mammary Gland	B-Adenoma	0	1	1	0	0 6437	0 5227	0.5116	.
	B-Fibroadenoma	17	16	9	3	1 0000	0.7468	0.9849	1.0000
	B-Fibroma	0	2	0	0	0 8227	0 2704	.	.
	M-Carcinoma	13	12	8	5	0 9895	0.7008	0.9164	0.9927
Ovary	M-Pericytoma	0	1	0	0	0.7640	0 5227	.	.
Pancreas	B-Adenoma, islet cel	2	0	2	0	0 8418	1 0000	0.7009	1.0000
Parathyroid	B-Adenoma	0	0	1	0	0 5056	.	0.5059	.
Pituitary	B-Adenoma, pars dist	31	40	36	33	0 5113	0.1399	0.2103	0.3739
	B-Adenoma, pars inte	0	1	0	0	0.7640	0 5227	.	.
	M-Carcinoma, pars di	1	2	1	2	0 3810	0 5345	0.7644	0.5426
Skin/Subcutis	B-Amelanotic melanom	1	0	0	0	1 0000	1 0000	1.0000	1.0000

File Name: outputfile805671446.pdf

Organ Name	Tumor Name	0 mg VC N=60	3 mkd LD N=60	8 mkd MD N=60	25 mkd HD N=60	P-Value			
						Dos-Resp	VC vs. LD	VC vs. MD	VC vs. HD
	B-Keratoacanthoma	1	0	0	1	0.4595	1 0000	1.0000	0.7801
	B-Leiomyoma	0	1	0	0	0.7640	0 5227	.	.
	B-Lipoma	0	0	2	0	0 5167	.	0.2588	.
	M-Carcinoma, basal c	0	1	0	0	0.7640	0 5227	.	.
	M-Carcinoma, squamou	1	0	0	1	0.4573	1 0000	1.0000	0.7745
	M-Fibrosarcoma	1	0	1	0	0.7570	1 0000	0.7588	1.0000
	M-Sarcoma	1	0	0	0	1 0000	1 0000	1.0000	1.0000
Thymus	B-Thymoma	3	1	0	1	0.7965	0 9469	1.0000	0.9492
Thyroid	B-Adenoma, C-cell	2	2	1	2	0 5268	0.7255	0.8838	0.7332
	B-Adenoma, follicula	0	1	0	1	0 3342	0 5227	.	0.5281
	M-Carcinoma, C-cell	1	0	1	1	0.4091	1 0000	0.7588	0.7801
	M-Carcinoma, follicu	1	0	1	1	0.4100	1 0000	0.7644	0.7801
	C FollicularCell M+B	**1**	**1**	**1**	**2**	**0.2837**	**0.7751**	**0.7644**	**0.5426**
Tongue	M-Carcinoma, squamou	0	0	1	0	0 5084	.	0.5116	.
Uterus	B-Polyp, endometrial	1	1	0	0	0 9453	0.7751	1.0000	1.0000
Vagina	B-Granular cell tumo	3	0	0	0	1 0000	1 0000	1.0000	1.0000
	M-Sarcoma, stromal	0	0	2	0	0 5167	.	0.2588	.

Table 4A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons with Vehicle Control – Male Mice

Organ Name	Tumor Name	0 mkd VC N=25	10 mkd LD N=25	40 mkd MD N=25	300 mkd HD N=25	P-Value			
						Dos Response	C vs. LD	C vs. MD	C vs. HD
Body, Whole/Cav	B-Hemangioma	1	0	0	0	1 0000	1 0000	1.0000	1 0000
	M-Hemangiosarcoma	0	0	1	1	0.1830	.	0.5000	0 5000
Lung	B-Adenoma, bronchiol	2	1	4	0	0 8922	0 8901	0.3535	1 0000
Muscle, Skeleta	M-Rhabdomyosarcoma	0	1	0	0	0.7526	0 5102	.	.
Skin/Subcutis	B-Adenoma, sebaceous	1	0	0	0	1 0000	1 0000	1.0000	1 0000
	B-Keratoacanthoma	0	1	0	0	0.7526	0 5102	.	.
Stomach, Nongla	M-Carcinoma	1	0	0	0	1 0000	1 0000	1.0000	1 0000
Thymus	B-Thymoma	1	1	0	1	0 5332	0.7653	1.0000	0.7553
	M-Malignant thymoma	0	0	0	1	0 2551	.	.	0 5102
	C Thymoma B+M	**1**	**1**	**0**	**2**	**0.2453**	**0.7653**	**1.0000**	**0.5156**

Table 4B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons with Vehicle Control – Female Mice

Organ Name	Tumor Name	0 mkd VC N=25	10 mkd LD N=25	30 mkd MD N=25	150 mkd HD N=25	P-Value			
						Dos Response	C vs. LD	C vs. MD	C vs. HD
	C Thymoma B+M	2	0	1	0	0.8931	1.0000	0.8830	1.0000
Uterus	B-Polyp, endometrial	0	1	0	0	0.7553	0 5106	.	.

Table 5A: Tumor Rates and P-Values for Comparisons between VC and PC– Male Mice

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	75 mg/kg PC (N=10)	P-Value VC vs. PC
Body, Whole/Cav	B-Hemangioma	1	0	1.0000
	M-Hemangiosarcoma	0	4	0 0019*
	M-Lymphosarcoma	0	6	<0.001*
Lung	B-Adenoma, bronchiol	2	3	0.0619
Muscle, Skeleta	M-Rhabdomyosarcoma	0	0	.
Skin/Subcutis	B-Adenoma, sebaceous	1	3	0.0278
	B-Keratoacanthoma	0	0	.
	B-Papilloma, squamou	0	5	<0.001*
Stomach, Nongla	B-Papilloma, squamou	0	5	<0.001*
	M-Carcinoma	1	0	1.0000
Thymus	B-Thymoma	1	0	1.0000

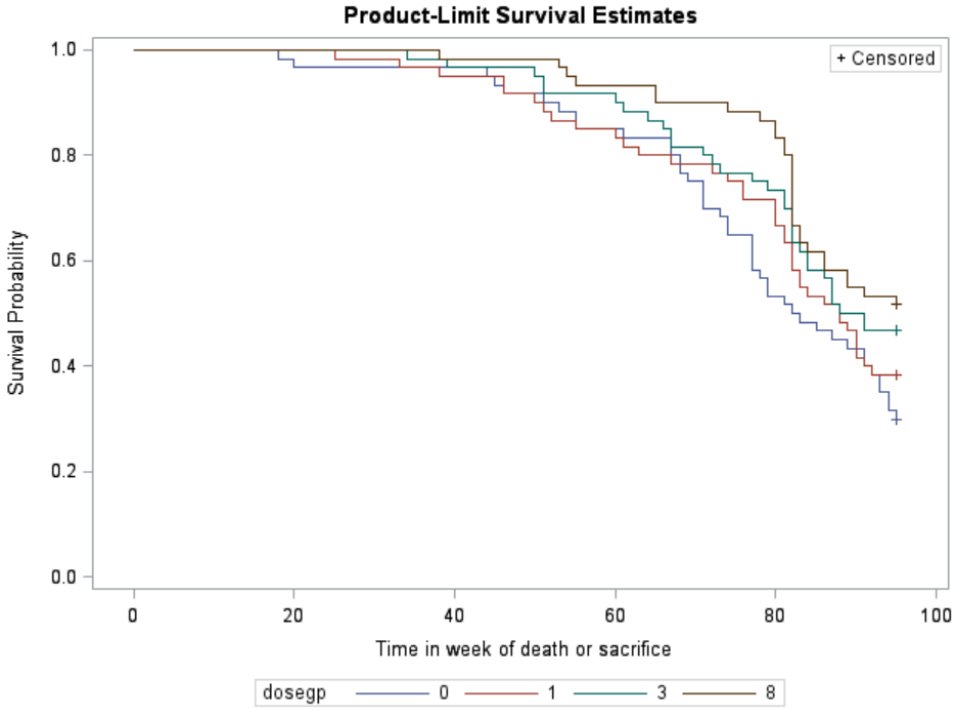
*Indicted the significant at 0.001 alpha levels. PC=75 mg/kg of N-methyl-N-nitrosourea.

Table 5B: Tumor Rates and P-Values for Comparisons between VC and PC– Female Mice

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	75 mg/kg PC (N=10)	P-Value VC vs. PC
Adrenal, Cortex	B-Adenoma, subcapsul	0	0	.
Body, Whole/Cav	B-Hemangioma	3	0	1.0000
	M-Hemangiosarcoma	0	1	0.2069
	M-Lymphosarcoma	0	8	<0.001*
Cervix	M-Sarcoma, stromal	0	0	.
Harderian Gland	B-Adenoma	1	3	0.0307
Lung	B-Adenoma, bronchiol	1	2	0.0999
Mammary Gland,	B-Adenoma	0	1	0.2333
Skin/Subcutis	B-Papilloma, squamou	0	2	0 0369*
Stomach, Nongla	B-Papilloma, squamou	0	6	<0.001*
Thymus	B-Thymoma	1	0	1.0000
	M-Malignant thymoma	1	0	1.0000
Uterus	B-Polyp, endometrial	0	1	0.2069

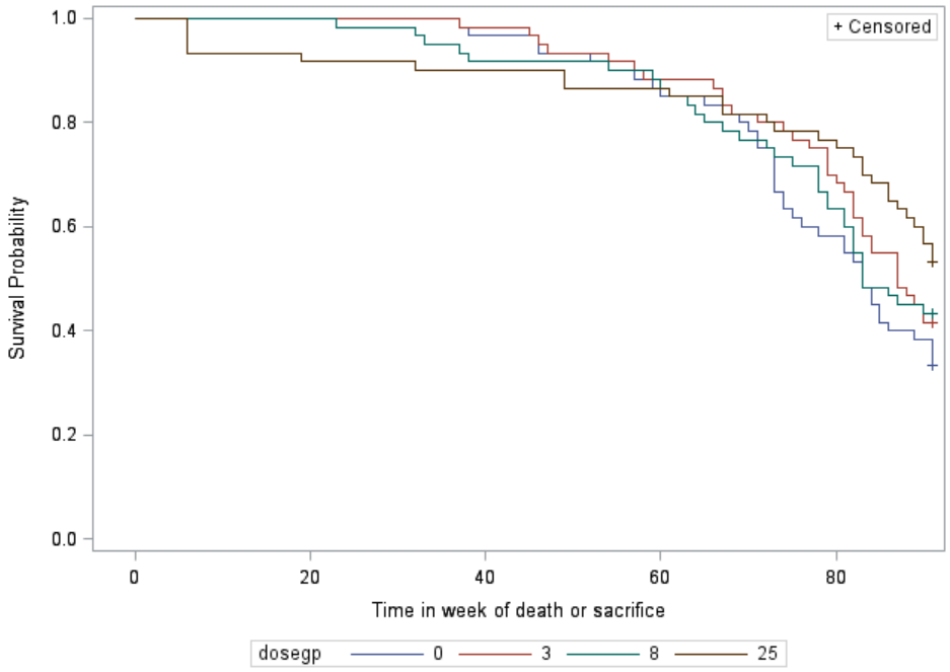
*Indicted the significant at 0.001 alpha levels. PC=75 mg/kg of N-methyl-N-nitrosourea.

Figure 1A: Kaplan-Meier Survival Functions for Male Rats



Note: dose group should be 0, 1, 3, and 8-mg/kg/day

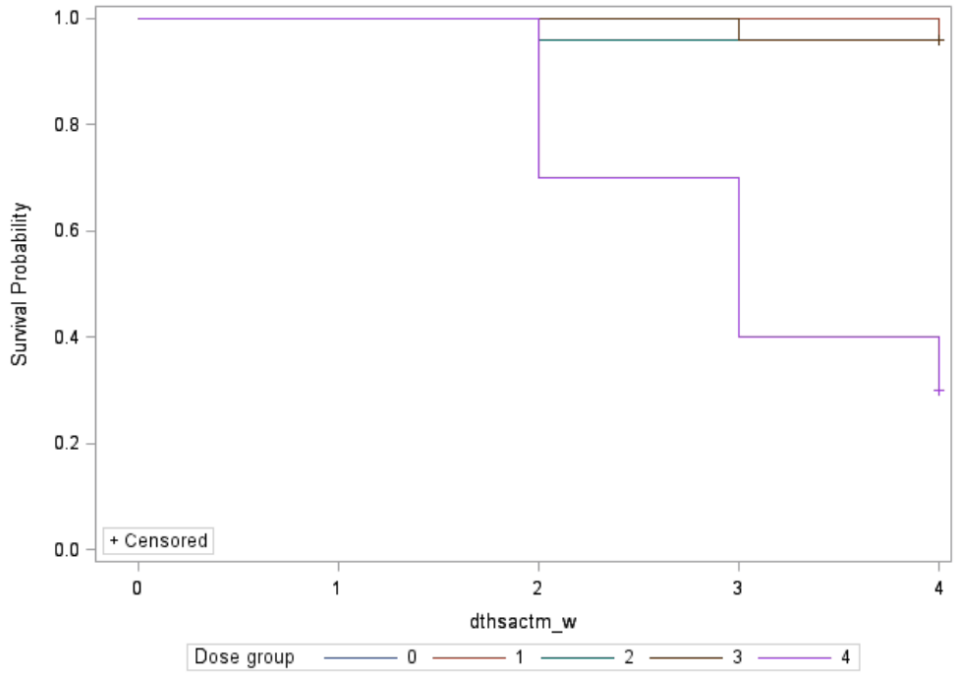
Figure 1B: Kaplan-Meier Survival Functions for Female Rats



Note: dose group should be 0, 3, 8, 25-mg/kg/day

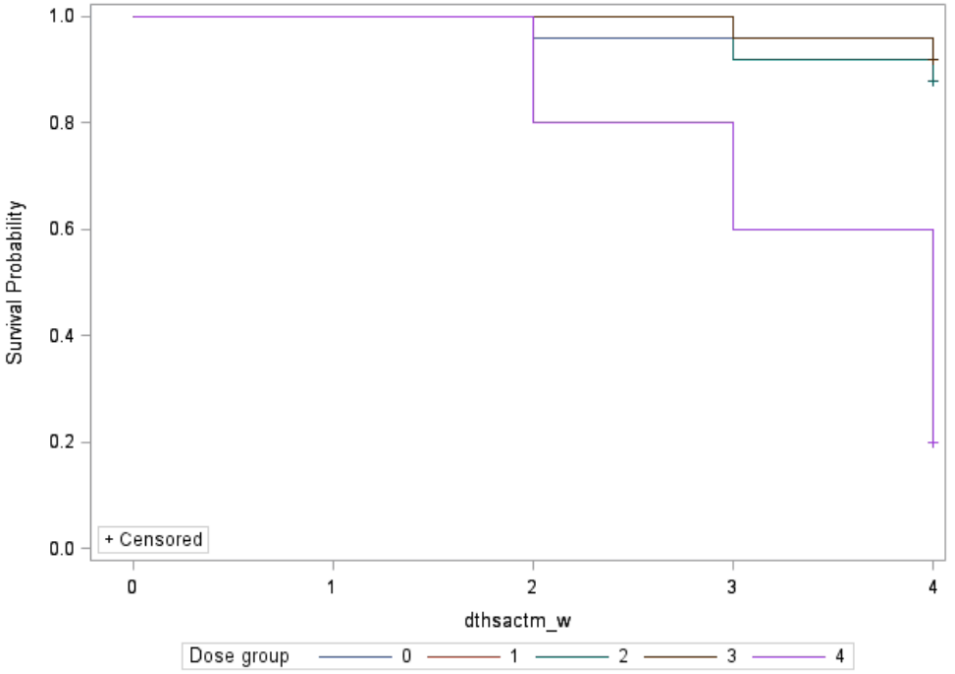
File Name: outputfile805671446.pdf

Figure 2A: Kaplan-Meier Survival Functions for Male Mice



Note: dose group should be 0, 15, 40, or 300-mg/kg/day and 4=75 mg/kg of N-methyl-N-nitrosourea, positive control

Figure 2B: Kaplan-Meier Survival Functions for Female Mice



Note: dose group should be 0, 10, 30, or 150-mg/kg/day and 4=75 mg/kg of N-methyl-N-nitrosourea, positive control

File Name: outputfile805671446.pdf

7 References

- McConnell, E. E., Solleveld, H. A., Swenberg, J. A., and Boorman, G. A., "Guidelines for Combining Neoplasms for Evaluation of Rodent Carcinogenicity Studies," Journal of National Cancer Institute, 76:283-289 (1986).
- Peto, R., Pike, M. C., Day, N. E., Gray, R. G., Lee, P. N., Parish, S., Peto, J., Richards, S., and Wahrendorf, J., "Guidelines for Simple, Sensitive Significance Tests for Carcinogenic Effects in Long-Term Animal Experiments," Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal, Annex to Supp. 2, pp. 311-426, International Agency for Cancer Research (IARC): Lyon (1980).
- SAS (Statistical Analysis System), Version 9.2, SAS Institute Inc.: Cary, North Carolina, USA (2002-2008).
- Thomas, D. G., Breslow, N., and Gart, J. J., "Trend and Homogeneity Analyses of Proportions and Life Table Data," Computers and Biomedical Research, 10:373-381 (1977).

--EOF--

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

FENG ZHOU
06/28/2016

KARL K LIN
06/28/2016
Concur with review