CADTH

CADTH COMMON DRUG REVIEW

Clinical Review Report

Baricitinib (OLUMIANT)

(Eli Lilly Canada Inc.)

Indication: For use in combination with methotrexate (MTX) for the treatment of adult patients with moderate to severe rheumatoid arthritis who have responded inadequately to one or more disease-modifying antirheumatic drugs (DMARDs). Baricitinib may also be used as monotherapy in cases of intolerance to MTX.

Service Line:	CADTH Common Drug Review
Version:	Final (With Redactions)
Publication Date:	September 2019
Report Length:	85 Pages

CADTH

Disclaimer: The information in this document is intended to help Canadian health care decision-makers, health care professionals, health systems leaders, and policy-makers make well-informed decisions and thereby improve the quality of health care services. While patients and others may access this document, the document is made available for informational purposes only and no representations or warranties are made with respect to its fitness for any particular purpose. The information in this document should not be used as a substitute for professional medical advice or as a substitute for the application of clinical judgment in respect of the care of a particular patient or other professional judgment in any decision-making process. The Canadian Agency for Drugs and Technologies in Health (CADTH) does not endorse any information, drugs, therapies, treatments, products, processes, or services.

While care has been taken to ensure that the information prepared by CADTH in this document is accurate, complete, and up-to-date as at the applicable date the material was first published by CADTH, CADTH does not make any guarantees to that effect. CADTH does not guarantee and is not responsible for the quality, currency, propriety, accuracy, or reasonableness of any statements, information, or conclusions contained in any third-party materials used in preparing this document. The views and opinions of third parties published in this document do not necessarily state or reflect those of CADTH.

CADTH is not responsible for any errors, omissions, injury, loss, or damage arising from or relating to the use (or misuse) of any information, statements, or conclusions contained in or implied by the contents of this document or any of the source materials.

This document may contain links to third-party websites. CADTH does not have control over the content of such sites. Use of third-party sites is governed by the third-party website owners' own terms and conditions set out for such sites. CADTH does not make any guarantee with respect to any information contained on such third-party sites and CADTH is not responsible for any injury, loss, or damage suffered as a result of using such third-party sites. CADTH has no responsibility for the collection, use, and disclosure of personal information by third-party sites.

Subject to the aforementioned limitations, the views expressed herein are those of CADTH and do not necessarily represent the views of Canada's federal, provincial, or territorial governments or any third party supplier of information.

This document is prepared and intended for use in the context of the Canadian health care system. The use of this document outside of Canada is done so at the user's own risk.

This disclaimer and any questions or matters of any nature arising from or relating to the content or use (or misuse) of this document will be governed by and interpreted in accordance with the laws of the Province of Ontario and the laws of Canada applicable therein, and all proceedings shall be subject to the exclusive jurisdiction of the courts of the Province of Ontario, Canada.

The copyright and other intellectual property rights in this document are owned by CADTH and its licensors. These rights are protected by the Canadian *Copyright Act* and other national and international laws and agreements. Users are permitted to make copies of this document for non-commercial purposes only, provided it is not modified when reproduced and appropriate credit is given to CADTH and its licensors.

About CADTH: CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

CADTH

Table of Contents

Tables

Figures

Abbreviations

ACR	American College of Rheumatology
ANCOVA	analysis of covariance
bDMARDs	biologic disease-modifying antirheumatic drug
CDAI	Clinical Disease Activity Index
cMARDS	conventional disease-modifying antirheumatic drug
CDR	CADTH Common Drug Review
CI	confidence interval
CrI	credible interval
CRP	C-reactive protein
DAS28	Disease Activity Scale-28
DIC	deviance information criterion
DMARD	disease-modifying antirheumatic drug
ESR	erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
FACIT	Functional Assessment of Chronic Illness
HAQ-DI	Health Assessment Questionnaire–Disability Index
HRQoL	health-related quality of life
hs-CRP	high-sensitivity C-reactive protein
IDC	indirect comparison
IL	interleukin
JAK	Janus kinase
LDAS	low disease activity state
LSM	least squares mean
MCID	minimal clinically important difference
MCS	mental component summary
MRI	magnetic resonance imaging
mTSS	modified total Sharp score
MTX	methotrexate
NICE	The National Institute for Health and Care Excellence

NMA	network meta-analysis
NSAID	nonsteroidal anti-inflammatory drug
OR	odds ratio
PCS	physical component summary
RA	rheumatoid arthritis
RCT	randomized controlled trial
SD	standard deviation
SDAI	Simplified Disease Activity Index
SDC	smallest detectable change
SDD	smallest detectable difference
SF-36	Short Form (36) Health Survey
TNF	tumour necrosis factor
VAS	visual analogue scale

Drug	Baricitinib (Olumiant)
Indication	For use in combination with methotrexate (MTX) for the treatment of adult patients with moderate to severe rheumatoid arthritis who have responded inadequately to one or more disease-modifying antirheumatic drugs (DMARDs). Baricitinib may also be used as monotherapy in cases of intolerance to MTX.
Reimbursement Request	As per indication
Dosage Form(s)	2 mg tablet
NOC Date	August 17, 2018
Manufacturer	Eli Lilly Canada Inc.

Executive Summary

Introduction

Rheumatoid arthritis (RA) is chronic autoimmune disorder characterized by severe destructive inflammation of the distal joints, particularly of the hands. The inflammation breaks down cartilage and bone, resulting in severe pain, stiffness, deformities, and disability. The inflammation can affect other areas as well, including the eyes, lungs, heart, or skin. RA can strike at any age but is more commonly seen in adulthood. According to a 2011 report by the Arthritis Alliance of Canada, approximately 0.9% of the Canadian population suffers from RA.[1] Treatment of RA consists of both acute therapies, used to address intense flares of the disease, and more chronic therapies that are aimed at the underlying disease process itself, known as disease-modifying antirheumatic drugs (DMARDs). DMARDs consist of small molecules that address various pathways involved in inflammatory/immune processes and include a diverse array of drugs, such as the antimalarial drugs, sulfasalazine, leflunomide, and the most commonly used, methotrexate (MTX). As a group, these drugs are referred to as the conventional DMARDs (cDMARDs). Recently, these cDMARDs have been joined by biologic DMARDs (bDMARDs), a group of drugs with a shared design, being either monoclonal antibodies or fusion proteins. Common limitations of all approaches are increased risk of infection and, possibly, an increased risk, albeit rare, of certain cancers.

Baricitinib is a Janus kinase (JAK) inhibitor. JAK mediates the effects of cytokines and their production; thus, JAK inhibitors may have a global effect on various cytokine production compared with biologics, which tend to target specific cytokines. Baricitinib is the second JAK inhibitor approved in Canada, the first being tofacitinib, which was previously reviewed by the CADTH Common Drug Review (CDR). Although both are JAK inhibitors, baricitinib is highly selective for the JAK-1 and JAK-2 isoforms, which may confer a profile of cytokine inhibition different from that of tofacitinib, which inhibits JAK-1 and JAK-3. The clinical significance of these differences in pharmacodynamics has yet to be determined. Because JAK inhibitors target cytokines, they have much in common with biologics and are often lumped in with them. However, they are in fact small molecules and will likely be considered a third group of DMARD. Unlike bDMARDs, JAK inhibitors are administered orally. Baricitinib 2 mg orally, once daily, is approved by Health Canada, in combination with MTX, for the treatment of adult patients with moderate-to-severe RA who have responded inadequately to one or more DMARDs. Baricitinib is also approved for use as monotherapy in cases of intolerance to MTX.

The objective of the current review is to perform a systematic review of the beneficial and harmful effects of baricitinib 2 mg orally once daily in combination with MTX (or as monotherapy in cases of intolerance to MTX) for the treatment of adult patients with moderate-to-severe RA who have responded inadequately to one or more DMARDs.

Results and Interpretation

Included Studies

Two multinational, manufacturer-sponsored, double-blind randomized controlled trials met the inclusion criteria for this systematic review. BEACON (N = 527, three groups) and BUILD (N = 684, three groups) both enrolled patients with adult-onset RA, with insufficient response or intolerance to cDMARDs (BUILD), or with stable dosage on background cDMARDs but with insufficient response or intolerance to at least one bDMARD tumour necrosis factor (TNF) inhibitor (BEACON). Both studies were conducted between 2013 and 2014, in European, Asia, and the Americas (including sites in Canada) and had identical trial design: a 24-week, parallel, double-blind treatment period in which baricitinib 2 mg and baricitinib 4 mg were compared with placebo. The Health Canada–approved baricitinib 2 mg dose was the focus of this review. The primary outcome in each study was the proportion of patients achieving an American College of Rheumatology (ACR) improvement criteria of at least 20% (ACR20) at 12 weeks, and key secondary outcomes, all assessed at 12 weeks, included health-related quality of life (HRQoL) on the Health Assessment Questionnaire–Disability Index (HAQ-DI), the Disease Activity Scale and high-sensitivity C-reactive protein (DAS28-hs-CRP), and the Simplified Disease Activity Index (SDAI). The ACR criteria provide a composite measure of improvement in both swollen and tender joint counts and at least three of five additional disease criteria: 1) patient global assessment of disease activity; 2) physician global assessment of disease activity; 3) patient assessment of pain; 4) HAQ; 5) levels of either C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR). ACR20, ACR50, and ACR70 responses represent at least a 20%, 50%, and 70% improvement, respectively. The HAQ-DI is an instrument commonly used to assess HRQoL in RA. SDAI is a measure of disease activity consisting of physical examination, acute-phase response, patient self-assessment, and evaluator assessment. The DAS28-hs-CRP is a composite of the Disease Activity Scale and high-sensitivity C-reactive protein.

Major limitations included a lack of active comparators in the included studies and a relatively short duration of follow-up (24 weeks) for a drug with a relatively novel mechanism of action. There was insufficient evidence of long-term effectiveness and safety, particularly given that safety issues such as thrombosis and herpes zoster were noted at the higher 4 mg dose. There was a relatively higher proportion of premature withdrawals in the placebo group than in the baricitinib group (18% versus 10%, respectively). There was a large proportion of patients who opted for rescue therapy with baricitinib 4 mg after 16 weeks, particularly in the placebo group in both BEACON (22% of baricitinib patients versus 32% of placebo patients) and BUILD (9% of baricitinib patients versus 24% of placebo patients). There was no subgroup analysis performed for patients with prior MTX intolerance, which is a potential gap, given that baricitinib may be used as monotherapy in these patients.

Efficacy

The primary outcome in both the BUILD and BEACON trials was the percentage of patients achieving ACR20 at week 12. More participants in the baricitinib group than in placebo group achieved ACR20, in both BEACON (48.9% versus 27.3%) and BUILD (65.9% versus 39.5%), and these differences were statistically significant between groups in both BEACON (odds ratio, 2.7 [95% confidence interval (CI), 1.7 to 4.2], *P* = 0.001) and BUILD (odds ratio, 3.0 [95% CI, 2.0 to 4.4], *P* = 0.001). The clinical expert consulted by CADTH on this review pointed out that the 12-week primary end point is earlier than the usual 24-week primary end point in many previous trials of RA drugs and that there was a robust placebo response in both trials. Thus, the difference between drug and placebo arms was modest. ACR20 responses were also assessed at 24 weeks, although this was an exploratory outcome and not controlled for multiplicity. The proportion of patients achieving ACR20 at 24 weeks was higher with baricitinib than with placebo in BEACON (44.8% versus 27.3%) and BUILD (61.1% versus 42.1%). Other ACR outcomes were exploratory and not controlled for multiplicity; these included percentage of patients with ACR50 responses at 12 weeks, baricitinib versus placebo, in BEACON (20.1% versus 8.0%) and BUILD (33.6% vs. 12.7%) and at 24 weeks in BEACON (23.0% versus 13.1%) and BUILD (41.5% versus 21.5%). ACR70 responses in baricitinib versus placebo patients were also reported at 12 weeks in BEACON (12.6% versus 2.3%) and in BUILD (17.9% versus 3.1%) and at 24 weeks (BEACON: 13.2% versus 3.4%; BUILD: 25.3 versus 7.9%). Subgroup analyses of interest to this review were performed in BEACON based on inadequate response to prior bDMARDs (lack of efficacy, adverse event, other). In patients with lack of efficacy on prior bDMARDs, ACR20 at 12 weeks was 49.1% in the baricitinib group versus 27.1% in the placebo group, and in patients with previous adverse events on bDMARDs, ACR20 at 12 weeks was 45.5% in the baricitinib group versus 25.0% in the placebo group. Other subgroups of interest to this review — previous intolerance to MTX and patients with prior inadequate response to cDMARDs — were not investigated in either study. Note that BUILD included patients with inadequate response to cDMARDs; thus, a further subgroup analysis for this subpopulation was not necessary in this study.

The HAQ-DI was a secondary outcome of both studies, and statistical comparisons were controlled for multiplicity. HAQ-DI from baseline to week 12, when compared with placebo, was statistically significant in both BEACON (least squares mean difference between groups –0.20 [95% CI, –0.32 to –0.08], *P* = 0.001) and BUILD (–0.21 [95% CI, –0.30 to –0.11], *P* = 0.001). The results at week 24 were consistent with those at week 12 in both studies. The minimal clinically important difference (MCID) for the HAQ-DI is 0.22, according to the literature.[2,3] Therefore, these may not be clinically meaningful improvements over placebo. Other HRQoL outcomes that were assessed but not controlled for multiple comparisons included the Short Form (36) Health Survey (SF-36) and the EuroQol 5-Dimensions (EQ-5D) questionnaire. There were no statistically significant differences between baricitinib and placebo groups for the SF-36 mental component summary at week 12 or 24. The SF-36 physical component summary least squares mean difference between baricitinib and placebo at 24 weeks in BEACON was 4.3 (95% CI, 2.6 to 6.1) and in BUILD was 3.7 (95% CI, 2.0 to 5.4). On the EQ-5D Health State Index/Self-Perceived Health score (US algorithm), in BEACON the least squares mean difference between groups was 0.049 (95% CI, 0.018 to 0.081) and in BUILD, it was 0.013 (95% CI, 0.023 to 0.075). Similar results were reported when the UK algorithm was used.

The DAS28-hs-CRP is a composite of HAQ-DI and high-sensitivity C-reactive protein; it was a secondary outcome in both BEACON and BUILD. In each study, baricitinib reduced (improved) scores versus placebo, and these differences between groups at 12 weeks were statistically significant in BEACON (least squares mean difference between groups of –0.66 [95% CI, –0.96 to –0.35], *P* = 0.001) and in BUILD (least squares mean difference between groups of –0.75 [95% CI, –0.97 to –0.53], *P* = 0.001). No MCID was found for this outcome; therefore, it is unclear whether this represents a clinically meaningful improvement for baricitinib over placebo.

The proportion of patients achieving an MCID on the SDAI was assessed as a secondary outcome in both studies. In BUILD, the proportion of patients achieving a clinically significant improvement in SDAI was higher with baricitinib than with placebo (9.2% versus 0.9% of patients), and this difference was statistically significant (odds ratio not reported, *P* = 0.001). In BEACON, there was no statistically significant difference between groups. Other outcomes related to symptoms included duration and severity of joint stiffness, as well as fatigue; however, none of these outcomes were controlled for multiple comparisons.

The modified total Sharp score (mTSS) was not investigated in BEACON and was an exploratory outcome in BUILD. The mTSS was increased from baseline to 24 weeks in both baricitinib (least squares mean change from baseline of 0.33 [95% CI, 0.06 to 0.59]) and placebo (0.70 [95% CI, 0.42 to 0.98]) groups for a least squares mean difference between groups of –0.38 (95% CI, –0.74 to –0.01).

CDR reviewed numerous network meta-analyses involving baricitinib, including one performed by CADTH (see Appendix 7 for detailed review). The manufacturer submitted a network meta-analysis that found that patients with inadequate response to a prior TNF inhibitor who were treated with baricitinib did not respond (as measured by ACR and European League Against Rheumatism responses) as well as patients treated with tocilizumab or rituximab. In patients who failed on a cDMARD, ACR responses were similar between baricitinib and cDMARD and bDMARDs and cDMARD. There were no differences in ACR responses between baricitinib and tofacitinib in the manufacturer-submitted network meta-analysis. The two additional published network meta-analyses that were reviewed had results generally consistent with that of the manufacturer's submitted analysis.

Harms

In BEACON, 71% of baricitinib patients and 64% of placebo patients experienced an adverse event, while, in BUILD, the percentages were 67% in baricitinib and 71% in placebo groups. The most common adverse event was upper respiratory tract infection, occurring in 9% of baricitinib patients and 5% placebo patients in BEACON and in 6% of baricitinib and 8% of placebo patients in BUILD. Serious adverse events were reported in 7% of baricitinib patients and 9% of placebo patients in BEACON and in 7% of patients in each group in BUILD. Herpes zoster occurred in 1% of patients in each group in BEACON and 2% of baricitinib patients versus no placebo patients in BUILD. Withdrawal due to adverse event occurred in 4% of baricitinib and placebo patients in each study.

Notable harms identified for this review included infections, which occurred in 44% of baricitinib patients and 31% of placebo patients in BEACON and 31% of baricitinib versus 35% of placebo patients in BUILD. Serious infections occurred in 3% of patients in each group in BEACON and 3% of baricitinib and 2% of placebo patients in BUILD. Other notable harms included malignancies, thrombotic events, dyslipidemia, and elevations in hepatic enzymes; there were very few events and no clear differences between groups

within BEACON and BUILD. There was a numerical higher risk of elevated platelet counts with baricitinib treatment versus placebo in BUILD (19% versus 5%); however, there was a much smaller difference between groups in BEACON (18% versus 14%). Low neutrophil counts were seen in 6% of baricitinib patients and 2% of placebo patients in BEACON and 8% of baricitinib versus 4% of placebo patients in BUILD.

Potential Place in Therapy[1]

Baricitinib is the second JAK inhibitor licensed for use in RA. It is more selective than tofacitinib, a pan-JAK inhibitor, because it is an inhibitor of JAK-1 and JAK-2. Selectivity is expected to reduce adverse events such as infections and episodes of herpes zoster. However, the safety database does not suggest that baricitinib is safer, and "eyeball" comparison does not find it more effective than tofacitinib.

Baricitinib joins a crowded field of therapies for RA, competing with five "brand name" TNF inhibitors and a growing number of biosimilar versions, two interleukin (IL)-6 inhibitors, a blocker of T-cell activation (abatacept), a B-cell depletor (rituximab), and tofacitinib. Mixed treatment comparisons have not found that baricitinib is more effective than its competitors, so, aside from the convenience of once-daily oral therapy, baricitinib does not provide a striking reason for being selected as a first choice in RA treatment. There is no subset of RA patients in whom baricitinib might be the preferred choice.

Little has been learned about the durability of baricitinib therapy, and long-term safety is yet to be established. Until there is more information, it is unclear whether baricitinib will fulfill an unmet need in RA therapy. Currently, it will be part of a crowded therapeutic field, and because much more is known about the older drugs, baricitinib is not expected to develop a large market share in the short term.

Conclusions

The included studies showed that baricitinib at 2 mg, administered orally once daily, after 12 weeks of treatment, improved clinical responses using the ACR20 in a population of patients with RA who had either an inadequate response to cDMARDs or to TNF inhibitors. Between 70% and 80% of the study patients had concurrent treatment with MTX, either alone or in combination with one DMARD. Subgroup analysis demonstrated that ACR20 responses were unaffected by the reason for inadequate response to TNF inhibitor (i.e., lack of efficacy or adverse event). There was no subgroup analysis performed for patients with prior MTX intolerance, which represents a potential gap, given that these patients may receive baricitinib as monotherapy. There was a statistically significant improvement in HRQoL and in disease activity with baricitinib, yet the clinical relevance of the improvement remains unclear. There were numerous other outcomes that assessed symptoms and HRQoL; however, these were not adjusted for multiple statistical comparisons. The beneficial effects of baricitinib are likely no different from those of tofacitinib and bDMARDs, as consistently shown in network meta-analyses. The risk of notable harms, such as serious infections, malignancies, cardiovascular events, dyslipidemia, and elevated hepatic enzymes did not appear to differ between baricitinib and placebo, although the included studies were not powered to assess outcomes such as these. Long-term study of these potential adverse events and durability of treatment effect in a real-world setting is warranted.

[1] This information is based on information provided in draft form by the clinical expert consulted by CDR for the purpose of this review.

Table 1: Summary of Results

	BEACON		BUILD	
	BAR 2 mg N = 174	Placebo N = 176	BAR 2 mg N = 229	Placebo N = 228
Clinical Response (ACR)				
ACR20 response at week 12, n (%)	85 (48.9)	48 (27.3)	151 (65.9)	90 (39.5)
Difference in response rate [95% CI][a]	21.6 [11.7 to 31.5]		26.5 [17.6 to 35.3]	
Odds ratio [95% CI][a]	2.7 [1.7 to 4.2], *P* = 0.001		3.0 [2.0 to 4.4], *P* = 0.001	
Health-Related Quality Of Life: HAQ-DI				
Mean (SD) baseline	1.71 (0.55)	1.78 (0.57)	1.51 (0.62)	1.50 (0.60)
LSM (SE) change from baseline to week 12	–0.37 (0.04) N = 172	–0.17 (0.04) N = 176	–0.54 (0.036) N = 229	–0.34 (0.037) N = 224
Mean difference between groups [95% CI][b]	–0.20 [–0.32 to –0.08], *P* = 0.001		–0.21 [–0.30 to –0.11], *P* = 0.001	
Disease Activity: DAS28-hs-CRP				
Mean (SD) baseline	6.03 (0.89)	5.89 (0.94)	5.57 (0.96)	5.53 (0.91)
LSM (SE) change from baseline to week 12	–1.49 (0.11) N = 172	–0.83 (0.11) N = 174	–1.83 (1.22) N = 229	–1.05 (1.23) N = 228
Mean difference between groups [95% CI][b]	–0.66 [–0.96 to –0.35], *P* = 0.001		–0.75 [–0.97 to –0.53], *P* = 0.001	
Disease Activity: SDAI				
SDAI ≤ 3.3 response rate, n (%)	4 (2.3)	3 (1.7)	21 (9.2)	2 (0.9)
Difference in response rate [95% CI]	0.6 [–2.3 to 3.5]		8.3 [4.4 to 12.2]	
Odds ratio[a]	NR, *P* = 0.723		NR, *P* = 0.001	
Harms				
Treatment-emergent adverse event, n (%)	123 (70.7)	112 (63.6)	154 (67.2)	161 (70.6)
SAEs, n (%)	12 (6.9)	15 (8.5)	15 (6.6)	16 (7.0)
Herpes zoster	2 (1.1)	2 (1.1)	4 (1.7)	0
Death, n (%)	0	0	0	2 (1)
WDAE, n (%)	7 (4.0)	7 (4.0)	9 (3.9)	8 (3.5)
Notable Harms				
Infections	76 (43.7)	55 (31.3)	70 (30.6)	79 (34.6)
Patients with ≥ 1 MACE event[c]	0	0	0	2 (0.9)
Patients with ≥ 1 other cardiovascular event	1 (0.6)	0	1 (0.4)	2 (0.9)
Malignancies	1 (0.6)	0	0	0
Gastrointestinal perforation	0	0	0	0
Hepatic function abnormal	0	1 (0.6)	2 (0.9)	1 (0.4)

ACR = American College of Rheumatology; BAR = baricitinib; DAS = Disease Activity Scale; CI = confidence interval; HAQ-DI = Health Assessment Questionnaire–Disability Index; hs-CRP = high-sensitivity C-reactive protein; LSM = least squares mean; MACE = major adverse cardiovascular event; NR = not reported; SAE = serious adverse event; SD = standard deviation; SDAI = Simplified Disease Activity Index; SE = standard error; WDAE = withdrawal due to adverse event.

Note: All the analyses were based on ITT with missing data being treated with non-responder imputation: 1) patients who discontinued the study or the study treatment were counted as non-responders from the time they discontinued onward and 2) any patients who were missing post-baseline data at any visit were also counted as non-responders at all visits.

[a] 95% CI is calculated for the difference in response rates using the Newcombe-Wilson method without continuity correction. 95% CI and *P* value for odds ratio from logistic regression model adjusted for region + history of bDMARD use (< 3, ≥ 3) + treatment group. When logistic regression sample size requirements are not met, *P* value from Fisher's exact test is produced instead of odds ratio and 95% CI. In BUILD: 95% CI and *P* value for odds ratio from logistic regression model adjusted for region + baseline joint erosion status (yes/no) + treatment group.

[b] *P* value, LSM, mean difference, SE, and 95% CI from the analysis of covariance (ANCOVA) model adjusted for baseline value, study region, history of bDMARD (< 3, ≥ 3) and in BUILD was adjusted for baseline + region + baseline joint erosion status (yes/no) + treatment group.

[c] Cardiovascular death, myocardial infarction, or stroke.

Source: Clinical Study Reports for BEACON[4] and BUILD.[5]

Introduction

Disease Prevalence and Incidence

Rheumatoid arthritis (RA) is chronic autoimmune disorder characterized by the potential for severe destructive inflammation of the joints, particularly of the hands. The inflammation breaks down cartilage and bone, resulting in severe pain, stiffness, deformities, and disability. The inflammation can affect other areas as well, including the eyes, lungs, heart, or skin. RA can strike at any age but is more commonly seen in adulthood. According to a 2011 report by the Arthritis Alliance of Canada, approximately 0.9% of the Canadian population suffers from RA.[1]

Standards of Therapy

Treatment of RA consists of both acute therapies, used to address intense flares of the disease, and more chronic therapies, that are aimed at the underlying disease process itself, known as disease-modifying antirheumatic drugs (DMARDs). Conventional DMARDs (cDMARDs) consist of small molecules that address various pathways involved in inflammatory/immune processes and include a diverse array of drugs, such as antimalarials, sulfasalazine, leflunomide, and methotrexate. Gold injections are no longer used to treat arthritis, and immune suppressants, such as azathioprine and cyclosporine, are not highly effective yet are highly toxic, according to the clinical expert consulted by CADTH for this review. Recently, these drugs have been joined by biologic DMARDs (bDMARDs), a group of drugs with a shared design, being either monoclonal antibodies or fusion proteins. Tumour necrosis factor (TNF)-alpha inhibitors were the original biologics, but they have since been joined by drugs that target interleukins (IL-1, IL-6), as well as drugs that target stimulation of T cells, drugs that deplete B cells, and, now, drugs that target Janus kinases (JAKs), called JAK inhibitors. Typically, patients are started on one or more cDMARDs, most commonly methotrexate; if their disease progresses, they will work up to the biologics. Common limitations of all approaches are increased risk of infection and, possibly, an increased risk, albeit rare, of certain cancers.[6]

Drug

Baricitinib is a JAK inhibitor. JAK mediates the effects of cytokines and their production; thus, JAK inhibitors inhibit multiple cytokines rather than one specific cytokine, as seen with the bDMARDs. Because JAK inhibitors target cytokines, they have much in common with bDMARDs, and are often lumped in with them. However, they are in fact small molecules and will likely be considered a third group of DMARD. Unlike bDMARDs, JAK inhibitors are administered orally. Baricitinib 2 mg, once daily, is approved by Health Canada, in combination with methotrexate, for the treatment of adult patients with moderate-to-severe RA who have responded inadequately to one or more DMARDs. In the case of intolerance to methotrexate, baricitinib may be used as monotherapy.

Table 2: Key Characteristics of JAK Inhibitors, IL-6 Inhibitors, T-Cell Co-Stimulation Inhibitors, CD20 Inhibitors, IL-1 Inhibitors, and TNF Inhibitors

	Mechanism	Indication[a]	Monotherapy	Combination	Route
		Inadequate Response Required			
Baricitinib	JAK inhibitor	≥ 1 DMARD	Yes[b]	+ MTX	Oral
Tofacitinib	JAK inhibitor	MTX	Yes[b]	+ MTX	Oral
Tocilizumab	IL-6 inhibitor	≥ 1 DMARD or TNF inhibitor	Yes	+ MTX	SC or IV
Abatacept	T-cell co-stimulation inhibitor	≥ 1 DMARD or TNF inhibitor	Yes	+ DMARD[c]	SC or IV
Rituximab	CD20 inhibitor	≥ 1 TNF inhibitor	No	+ MTX	IV
Anakinra	IL-1 inhibitor	Not required	Yes	+ DMARD[d]	SC
Adalimumab	TNF inhibitor	Not required	Yes[b]	+ MTX[e]	SC
Etanercept			Yes	+ MTX	SC
Golimumab			No	+ MTX	SC or IV
Certolizumab pegol			Yes[b]	+ MTX	SC
Infliximab			No	+ MTX	IV

DMARD = disease-modifying antirheumatic drug; IL = interleukin; JAK = Janus kinase; MTX = methotrexate; SC = subcutaneous; TNF = tumour necrosis factor.

[a] Health Canada–approved indication (all approved for adults with moderately to severely active RA except anakinra, which is approved for active RA, severity not specified).

[b] If patient is intolerant to MTX.

[c] If first-line treatment, give with MTX.

[d] The DMARD used is usually MTX.

[e] Other DMARDs may also be used.

Source: Product Monographs from the e-CPS.[7]

Objectives and Methods

Objective

To perform a systematic review of the beneficial and harmful effects of baricitinib in combination with methotrexate for the treatment of adult patients with moderate-to-severe RA who have responded inadequately to one or more DMARDs.

Methods

All manufacturer-provided trials considered pivotal by Health Canada were included in the systematic review. Phase III studies were selected for inclusion based on the selection criteria presented in Table 3.

Table 3: Inclusion Criteria for the Systematic Review

Patient Population	Patients with moderate-to-severe RA who have responded inadequately to one or more disease-modifying antirheumatic drugs (DMARDs). Subgroups: Patients intolerant to methotrexate Patients with inadequate response to conventional DMARDs Patients with inadequate response to biologic DMARDs
Intervention	Baricitinib 2 mg orally once daily, in combination with methotrexate or as monotherapy in those patients who are unable to tolerate methotrexate
Comparators	• TNF-alpha inhibitors (infliximab, adalimumab, certolizumab pegol, golimumab, etanercept) • T-cell stimulation inhibitor (abatacept) • CD20 inhibitor (rituximab) • IL-6 inhibitors (tocilizumab, sarilumab) • JAK inhibitor (tofacitinib) • Nonbiologic DMARDs
Outcomes	**Efficacy outcomes** • Clinical response (ACR20, ACR50, ACR70)[a] • Radiographic response • Health-related quality of life[a] • Functional and disability outcomes[a] • Disease activity[a] • Health care resource utilization **Harms outcomes** • AEs,[a] SAEs,[a] WDAEs • Mortality • AEs of special interest (e.g., serious infection [including herpes zoster],[a] neutropenia, thrombocytopenia, malignancies, thrombosis [including increased platelets], major cardiovascular events, gastrointestinal perforations and other gastrointestinal SAEs,[a] liver toxicity, dyslipidemia)
Study Design	Published and unpublished phase III and IV RCTs

ACR = American College of Rheumatology; AE = adverse event; DMARD = disease-modifying antirheumatic drug; IL = interleukin; JAK = Janus kinase; RCT = randomized controlled trial; SAE = serious adverse event; TNF = tumour necrosis factor; WDAE = withdrawal due to adverse events.

[a] Outcomes identified as important to patients in input provided to the CADTH Common Drug Review.

The literature search was performed by an information specialist using a peer-reviewed search strategy.

Published literature was identified by searching the following bibliographic databases: MEDLINE All (1946–) via Ovid; Embase (1974–) via Ovid; and PubMed. The search strategy consisted of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concept was Olumiant (baricitinib).

No methodological filters were applied to limit retrieval to study type. Where possible, retrieval was limited to the human population. Retrieval was not limited by publication year or by language. Conference abstracts were excluded from the search results. See Appendix 2 for the detailed search strategies.

The initial search was completed on February 11, 2019. Regular alerts were established to update the search until the meeting of the CADTH Canadian Drug Expert Committee on June 19, 2019. Regular search updates were performed on databases that do not provide alert services.

Grey literature (literature that is not commercially published) was identified by searching relevant websites from the following sections of the *Grey Matters* checklist (https://www.cadth.ca/grey-matters): health technology assessment agencies**,** health economics, clinical practice guidelines, drug and device regulatory approvals, advisories and warnings**,** drug class reviews and databases. Google and other Internet search engines were used to search for additional Web-based materials. These searches were supplemented by reviewing the bibliographies of key papers and through contacts with appropriate experts. In addition, the manufacturer of the drug was contacted for information regarding unpublished studies.

Two CDR clinical reviewers independently selected studies for inclusion in the review based on titles and abstracts, according to the predetermined protocol. Full-text articles of all citations considered potentially relevant by at least one reviewer were acquired. Reviewers independently made the final selection of studies to be included in the review, and differences were resolved through discussion. Included studies are presented in Table 4, and excluded studies (with reasons) are presented in Appendix 3.

Results

Findings From the Literature

A total of two studies were identified from the literature for inclusion in the systematic review (Figure 1). The included studies are summarized in Table 4. A list of excluded studies is presented in Appendix 3.

Figure 1: Flow Diagram for Inclusion and Exclusion of Studies

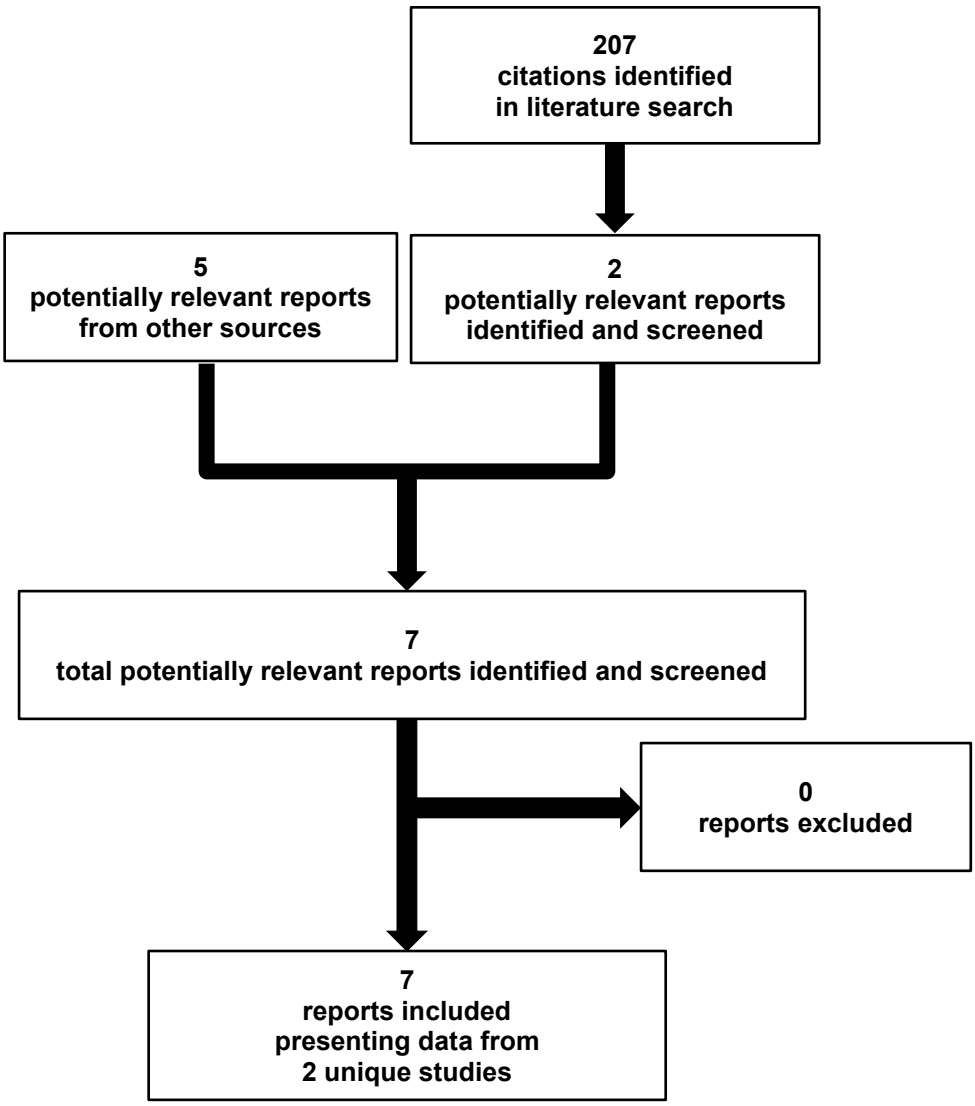


Table 4: Details of Included Studies

		BEACON	BUILD
Designs and Populations	**Study Design**	DB RCT	DB RCT
	Study Period	January 23, 2013, to September 2, 2014	January 10, 2013, to December 19, 2014
	Locations	140 sites: North America (including Canada), South America, Europe, Asia, Israel, Australia	182 sites: North America (including Canada), South America, Europe, Asia, India, Australia
	Randomized (N)	527	684
	Inclusion Criteria	• Adults diagnosed with adult-onset RA defined by the American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) 2010 Criteria for Classification of Rheumatoid Arthritis; • ≥ 6 tender joints (of 68 joints examined) and 6 swollen joints (of 66 joints examined); • high-sensitivity C-reactive protein (hs-CRP) measurement greater than the upper limit of normal (ULN); • receiving stable doses of background cDMARD; and • failed treatment at an approved dose with at least one biologic TNF-alpha inhibitor (experienced insufficient efficacy or were intolerant to treatment).	• Adults diagnosed with adult-onset RA defined by the ACR/EULAR 2010 Criteria for Classification of Rheumatoid Arthritis; • ≥ 6 tender joints (of 68 joints examined) and 6 swollen joints (of 66 joints examined); • hs-CRP measurement ≥ 1.2 times the ULN; • and had an insufficient response or were intolerant to cDMARDs.
	Exclusion Criteria	• Receiving prohibited RA therapies • Recent history of infection including active tuberculosis (TB) or untreated latent TB or other serious infections • Immunocompromized	• Receiving prohibited RA therapies • Recent history of infection including active TB or untreated latent TB or other serious infections • Immunocompromized
Drugs	**Intervention**	• Baricitinib 2 mg orally once daily • Baricitinib 4 mg orally once daily	• Baricitinib 2 mg orally once daily • Baricitinib 4 mg orally once daily
	Comparator(s)	Placebo	Placebo
Duration	**Phase**		
	Screening	3 to 42 days	3 to 42 days
	Double-blind	24 weeks	24 weeks
	Follow-up	4 weeks	4 weeks
Outcomes	**Primary End Point**	Proportion of patients achieving ACR20 at week 12	Proportion of patients achieving ACR20 at week 12
	Other End Points	**Major relevant secondary end points:** Baricitinib 2 mg q.d. versus placebo: • change from baseline to week 12 in HAQ-DI score • change from baseline to week 12 in DAS28-hs-CRP • proportion of patients who achieved an SDAI score ≤ 3.3 at week 12. **Other secondary objectives:** Baricitinib 4 mg and baricitinib 2 mg compared with placebo: • patients who achieved ACR20 at week 24 • patients who achieved ACR50 at week 12/24	**Major relevant secondary end points:** Baricitinib 2 mg q.d. versus placebo: • change from baseline to week 12 in HAQ-DI score • change from baseline to week 12 in DAS28-hs-CRP • proportion of patients achieving an SDAI score ≤ 3.3 at week 12. **Other secondary efficacy end points:** Baricitinib 4 mg and baricitinib 2 mg compared with placebo: • patients who achieved ACR20 at week 24 • patients who achieved ACR50 at week 12/24.

		BEACON	BUILD
		• patients who achieved ACR70 at week 12/24.	
NOTES	**Publications**	Genovese et al. 2016[8]	Dougados et al. 2017[9]

ACR = American College of Rheumatology; DAS28 = Disease Activity Scale-28; DB = double-blind; cDMARD = conventional disease-modifying antirheumatic drug; EULAR = European League Against Rheumatism; HAQ-DI = Health Assessment Questionnaire–Disability Index; hs-CRP = high-sensitivity C-reactive protein; q.d. = once daily; RA = rheumatoid arthritis; RCT = randomized controlled trial; SDAI = Simplified Disease Activity Index; TB = tuberculosis; TNF = tumour necrosis factor; ULN = upper limit of normal.

Source: Clinical Study Reports for BEACON[4] and BUILD.[5]

Included Studies

Description of Studies

Two pivotal, multinational, manufacturer-sponsored, double-blind randomized controlled trials met the inclusion criteria for this systematic review. BEACON and BUILD both enrolled adults with adult-onset RA. In the former, the patients' symptoms had been inadequately controlled on bDMARDs (TNF inhibitors) and, in the latter, they had been inadequately controlled on cDMARDs. Both studies had a 24-week double-blind treatment period in which baricitinib 2 mg and baricitinib 4 mg were compared with placebo. The Health Canada–approved baricitinib dose of 2 mg was the focus of this review. The primary outcome in each study was the proportion of patients achieving an American College of Rheumatology (ACR) improvement criteria of at least 20% (ACR20) response at 12 weeks, while key secondary outcomes that were accounted for multiplicity included health-related quality of life (HRQoL) on the Health Assessment Questionnaire–Disability Index (HAQ-DI), the Disease Activity Scale-28 and high-sensitivity C-reactive protein (DAS28-hs-CRP), and the Simplified Disease Activity Index (SDAI). There was a screening period in each study that ran from three to 42 days, and each study had follow-up of four weeks. Patients were invited to enter an extension study, called BEYOND, once they had completed these pivotal studies.

Randomization was carried out using an interactive voice response system in both studies, and in BEACON was stratified by region (US and Canada, Central/South America and Mexico, Europe, Asia, remainder of world) and by history of biologic DMARD use at screening (less than three different previous DMARDs, three or more different previous DMARDs). In BUILD, randomization was stratified by region and by baseline joint erosion status (yes or no). In both studies, patients with renal impairment could be assigned to either of the baricitinib 2 mg or 4 mg groups but were given only a 2 mg dose.

Subgroups analyses were planned in BUILD for gender, age, body mass index, weight, race, region, renal function, duration of RA, disease severity, joint erosion serology, background therapy, baseline modified total Sharp score (mTSS), and corticosteroid use. In BEACON, in addition to those subgroups in BUILD, number of previous bDMARDs and reason for inadequate response to TNF inhibitors were also explored, and both of these subgroups were of interest for this CDR review.

Populations

Inclusion and Exclusion Criteria

Both studies included adults diagnosed with adult-onset RA (ACR/EULAR 2010 definition) and hs-CRP greater than the upper limit of normal (in BUILD it had to be at least 1.2 times the upper limit of normal). Patients in BEACON had to have failed at least one TNF inhibitor (lack of efficacy or intolerance), while in BUILD patients had to have failed a previous conventional DMARD (lack of efficacy or intolerance). Patients with a recent history of infection, including active tuberculosis or other serious infection, were excluded, as were immunocompromized patients.

Baseline Characteristics

Patients in BEACON were approximately 55 years old, and in BUILD, were 52 years of age. The majority of patients in both studies were female (80% in BEACON and 82% in BUILD). As well, the majority of patients in both BEACON (83%) and BUILD (68%) were white. Patients in BEACON had had RA for longer (13.9 years since symptom onset) than patients in BUILD (7.4 years). In BEACON, 59% of patients were on a corticosteroid, while, in BUILD, 51% were on a corticosteroid currently. Demographic characteristics were reasonably similar between groups within studies. There was a higher proportion of patients in the placebo group in BEACON who were using a corticosteroid at baseline (53% in the baricitinib group versus 66% in the placebo group). Otherwise, there were no notable between-group differences in baseline characteristics between groups within studies.

Table 5: Summary of Baseline Characteristics

Characteristics	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
Age, years, mean (SD)	55.1 (11.1)	56.0 (10.7)	52.2 (12.3)	51.4 (12.5)
Female, n (%)	137 (79)	145 (82)	184 (80)	189 (83)
Race, n (%)				
White	144 (84)	147 (84)	156 (68)	153 (67)
American Indian or Alaska Native	12 (7)	9 (5)	2 (1)	3 (1)
Asian	9 (5)	11 (6)	61 (27)	60 (26)
Black or African American	9 (5)	8 (5)	9 (4)	10 (4)
Multiple	0	1 (1)	1 (< 1)	1 (< 1)
Native Hawaiian/Other Pacific Islander	0	0	0	1 (< 1)
Weight, kg, mean (SD)	83.0 (21.6)	82.0 (21.2)	74.9 (20.7)	75.6 (21.3)
Time from symptom onset of RA, years, mean (SD)	13.7 (8.0)	14.0 (9.6)	7.6 (7.6)	7.2 (7.5)
Time from symptom onset of RA category, n (%)				
< 1 year	0	1 (1)	21 (9)	18 (8)
≥ 1 year to < 5 years	20 (12)	28 (16)	89 (40)	111 (49)
≥ 5 years	154 (89)	147 (84)	115 (51)	99 (43)
Time from RA diagnosis, years, mean (SD)	12.3 (7.5)	12.8 (9.4)	6.5 (7.6)	5.9 (6.8)
Current use of corticosteroid, n (%)	92 (53)	116 (66)	117 (51)	114 (50)
Daily dose of corticosteroid, mg/day, mean (SD)	5.9 (2.7) N = 92 [a]	6.7 (2.6) N = 116 [a]	6.5 (2.5) N = 117 [a]	5.9 (2.6) N = 114 [a]

Characteristics	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
Methotrexate, average weekly dose, mg/week, mean (SD)	16.0 (4.8) N = 141[b]	15.9 (5.0) N = 143[b]	16.4 (4.7) N = 170[b]	16.0 (4.8) N = 167[b]
Number cDMARDs currently used, n (%)				
0	1 (1)	0	18 (8)	17 (8)
1	156 (90)	160 (91)	145 (63)	150 (66)
2	15 (9)	16 (9)	58 (25)	55 (24)
3 or more	2 (1)	0	8 (4)	6 (3)
Number cDMARDs previously used, n (%)				
0	0	0	3 (1)	1 (< 1)
1	66 (38)	75 (43)	104 (45)	96 (42)
2	54 (31)	47 (27)	61 (27)	81 (36)
3 or more	54 (31)	54 (31)	61 (27)	50 (22)
Type of cDMARDs currently used, n (%)				
Methotrexate only	126 (72)	131 (74)	111 (49)	109 (48)
Methotrexate + 1 other cDMARD	13 (8)	12 (7)	51 (22)	52 (23)
1 non-methotrexate cDMARD only	30 (17)	29 (17)	34 (15)	41 (18)
2 non-methotrexate cDMARDs only	2 (1)	4 (2)	7 (3)	3 (1)
3 or more cDMARDs	2 (1)	0	8 (4)	6 (3)
Number of previous bDMARDs used, n (%)				
0	0	1 (1)	NR	
1	69 (40)	81 (46)		
2	55 (32)	47 (27)		
3 or more	50 (29)	47 (27)		
Previous bDMARDs used, n (%)				
Abatacept	34 (20)	37 (21)	NR	
Adalimumab	85 (49)	78 (44)		
Anakinra	3 (2)	2 (1)		
Certolizumab	19 (11)	17 (10)		
Etanercept	90 (52)	107 (61)		
Golimumab	16 (9)	21 (12)		
Infliximab	51 (29)	44 (25)		
Rituximab	33 (19)	23 (13)		
Tocilizumab	36 (21)	36 (21)		
Mean (SD) joint counts				
Tender joint count (28 joints)	16.8 (6.8)	15.5 (6.9)	13.8 (7.2)	13.7 (7.0)
Swollen joint count (28 joints)	12.4 (6.1)	11.7 (5.8)	10.0 (5.5)	9.6 (4.7)
Tender joint count (68 joints)	31.0 (16.3)	28.3 (16.4)	23.5 (14.1)	24.3 (15.0)
Swollen joint count (66 joints)	18.6 (12.3)	17.2 (10.8)	13.6 (8.7)	13.1 (7.2)
Biomarkers				
hs-CRP, mg/L, mean (SD)	19.87 (22.48)	20.64(25.26)	18.2 (21.5)	17.7 (20.4)
DAS28-hs-CRP, mean (SD)	6.03 (0.89)	5.89 (0.94)	5.57 (0.96)	5.53 (0.91)
DAS28-ESR, mean (SD)	6.70 (0.98)	6.59 (0.93)	6.28 (0.99)	6.19 (1.00)
ESR, mm/h, mean (SD)	44.7 (23.5)	47.2 (24.0)	44.4 (22.7)	43.5 (25.1)

Characteristics	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
Rheumatoid factor status, n (%)				
Positive	128 (74)	130 (74)	177 (77)	171 (75)

BAR = baricitinib; bDMARD = biologic disease-modifying antirheumatic drug; cDMARD = conventional disease-modifying antirheumatic drug; DAS28 = Disease Activity Score-28; ESR = erythrocyte sedimentation rate; hs-CRP = high-sensitivity C-reactive protein; NR = not reported; SD = standard deviation.

[a] The numbers reflect only the subgroup of patients (outlined in the row above) who were taking corticosteroids at baseline.

[b] The numbers reflect only the subgroup of patients who were on methotrexate at baseline, described under the heading "type of cDMARDs currently used."

Source: Clinical Study Reports for BEACON[4] and BUILD.[5]

Interventions

Both included studies compared baricitinib 2 mg or 4 mg, once daily, with matched placebo. Patients continued on their background cDMARD therapy, at a stable dose. Patients could be "rescued" in either treatment arm, by being switched to the baricitinib 4 mg dose after 16 weeks of double-blind treatment in either study, if they were determined to be nonresponders. Patients with renal impairment were switched to the baricitinib 2 mg dose. A nonresponder was defined as lack of improvement of at least 20% in both tender joint count and swollen joint count at weeks 14 and 16 compared with baseline. Patients could be rescued only once; if they met the criteria a second time, they were withdrawn from the study.

In both trials, if patients were on methotrexate, they had to have been taking it for at least 12 weeks prior to entering the study and on a stable dose for at least eight weeks. Their dose did not change throughout the study, unless an adjustment was required for safety. Patients on concomitant hydroxychloroquine, sulfasalazine, or leflunomide (BUILD only) or azathioprine (BUILD only) had to be on a stable dose for at least eight weeks, and, again, a stable dose was required throughout the study. Nonsteroidal anti-inflammatory drugs (NSAIDs) were permitted; however, the patient had to be on a stable dose for at least six weeks prior to randomization, and dose increases or switching to another NSAID were not permitted unless the patient received rescue therapy. Other analgesics were also permitted during the study, but the dose could not be increased and a new analgesic could not be introduced unless the patient received rescue therapy. Prednisone, at a dose of up to 10 mg, was also permitted; however, the dose also had to be stable for at least six weeks prior to randomization, and, thus, patients could not begin therapy with a corticosteroid once the study had commenced. Parenteral corticosteroids and biologics (for any indication) were not permitted during the study. Intra-articular injections of corticosteroids, if required, were to be recorded as protocol deviations.

Outcomes

The primary outcome of both studies was the percentage of patients with ACR20 responses at 12 weeks. The ACR criteria provide a composite measure of improvement in both swollen and tender joint counts and at least three of five additional disease criteria: patient global assessment of disease activity; physician global assessment of disease activity; patient assessment of pain; HAQ; CRP or erythrocyte sedimentation rate (ESR). The ACR joint count for RA assesses 68 joints for tenderness and 66 joints for swelling. Patient and physician assessments are conducted using visual analogue scale (VAS) or Likert scale measurements. ACR20, ACR50, or ACR70 responses represent at least a 20%, 50%, or

70% improvement, respectively, in tender and swollen joint counts, as well as in three of the five additional disease criteria.

The mean change from baseline DAS28-hs-CRP at week 12 was a secondary outcome of both included studies. DAS28 is based on a 28-joint count that includes hands, wrists, elbows, shoulders, and knees. It omits the feet and ankle joints. In recent years, CRP has been used to calculate the DAS28 in place of ESR. The DAS28 is derived using the following formula:

$$DAS28\text{–}hs\text{–}CRP = 0.56 \times \sqrt{(t28)} + 0.28 \times \sqrt{(sw28)} + 0.014 \times GH + 0.36 \times ln(CRP + 1) + 0.96,$$

where CRP = C-reactive protein, DAS28 = Disease Activity Score-28, ln(CRP + 1) = natural logarithm of (CRP value +1), sw28 = swollen joint count of 28 joints, t28 = tender joint count of 28 joints, and GH = general health measured by Patient's Global Assessment of Disease Activity on a VAS of 100 mm.

A DAS28 score indicates an absolute level of disease activity, with a score of greater than 5.1 being considered high disease activity, while a DAS28 score lower than 3.2 indicates low disease activity state and a DAS28 score lower than 2.6 indicates remission.[10-12] To date, the minimal clinically important difference (MCID) for the DAS28 has not been determined.

The change from baseline to week 12 in HAQ-DI scores was a secondary outcome of both included studies. The full HAQ collects data on five generic, patient-centred health dimensions: 1) to avoid disability, 2) to be free of pain and discomfort, 3) to avoid adverse treatment effects, 4) to keep dollar costs of treatment low, and 5) to postpone death.12 The HAQ-DI is the disability assessment component of the HAQ. There are 20 questions in eight categories to assess a patient's physical functional status: dressing, arising, eating, walking, hygiene, reach, grip, and common activities.[2,3] For each of these categories, patients report the amount of difficulty they have in performing specific activities on a scale from zero (no difficulty) to three (unable to do). The eight category scores are averaged into an overall HAQ-DI score on a scale from zero (no disability) to three (completely disabled). A number of investigators have suggested that the MCID is 0.22; however, differences as small as 0.10 have been suggested as clinically important.[2]

The percentage of patients with a SDAI score of 3.3 or less (indicative of remission) at week 12 was a secondary outcome of both included studies. SDAI is a tool for measuring disease activity that integrates measures of physical examination, acute-phase response, patient self-assessment, and evaluator assessment.[4,5] SDAI is calculated by a simple numerical addition of the scores from the five following assessments: number of tender joints (0 to 28); number of swollen joints (0 to 28); C-reactive protein (mg/dL; 0.1 to 10.0); Patient Global Assessment of Disease Activity VAS (0 to 10.0 cm); and Physician Global Assessment of Disease Activity VAS (0 to 10.0 cm). The Clinical Disease Activity Index (CDAI) is similar to the SDAI, but it allows for immediate scoring because it does not include a laboratory result.[4,5] Both the SDAI and CDAI have been validated and show correlation with one another as well as with the DAS28.[13-15] Disease remission is defined as an SDAI score of 3.3 or less and as a CDAI score of 2.8 or less.[15,16]

The change from baseline to week 12 and week 24 in the Short Form (36) Health Survey (SF-36) was a secondary outcome of both studies and was not part of the statistical hierarchy. SF-36 is a generic health assessment questionnaire that has been used in

clinical trials to study the impact of chronic disease on HRQoL. The SF-36 consists of eight subdomains: physical functioning, pain, vitality, social functioning, psychological functioning, general health perceptions, and role limitations due to physical and emotional problems.[12] The SF-36 also provides two component summaries, the physical component summary (PCS) and the mental component summary (MCS). The eight subdomains are each measured on a scale of zero to 100, with an increase in score indicating improvement in health status. The MCID for either the PCS or MCS of the SF-36 is typically between 2.5 and 5 points.

Statistical Analysis

In BEACON, a target sample size of 175 patients per group was determined to have 97% power to detect a difference between baricitinib 4 mg and placebo for the ACR20 (assuming responses of 45% with baricitinib and 25% with placebo) and 80% power to detect a difference between the baricitinib 2 mg group and placebo for ACR20, assuming responses of 39% with baricitinib and 25% with placebo, all at week 12. In BUILD, a target sample size of 220 patients per group was identified as having greater than 95% power to detect a difference between baricitinib 4 mg and placebo (assuming ACR20 responses in 60% of baricitinib 4 mg patients and 35% of placebo patients) and greater than 90% power to detect a difference between baricitinib 2 mg and placebo (assuming ACR20 responses in 51% to 55% of patients in baricitinib 2 mg and 35% in placebo) at week 12. In both studies, these calculations were based on a two-sided chi-squared test and alpha of 0.05. It was not clear what the basis was for the assumed ACR20 responses in either intervention group or placebo.

In both trials, categorical outcomes (e.g., patients achieving ACR20) were analyzed using logistic regression adjusted for region and history of DMARD use at screening (less than three or three or more), with estimates of treatment effect by odds ratio (OR), the associated 95% confidence interval (CI), and *P* value. When the sample size requirements (more than five responders in any category for any factor) were not met, the *P* value from the Fisher's exact test was produced instead of the OR and 95% CI based on the logistic regression model. For continuous outcomes, analysis of covariance (ANCOVA) was applied with adjustment for region, history of bDMARD use at screening (less than three or three or more), and baseline value. A hierarchical testing procedure was employed to control for multiple statistical comparisons in both included studies. The baricitinib 4 mg doses were compared with placebo first, testing for superiority in terms of the proportion of patients achieving ACR20, then the mean change from baseline in HAQ-DI and in DAS28-hs-CRP at week 12 (Hochberg procedure), then the proportion of patients achieving SDAI of 3.3 or less (SDAI remission) at week 12. Comparisons for these three outcomes were then repeated for the baricitinib 2 mg dose, in the same order.

Missing data were accounted for using a variety of techniques. For dichotomous outcomes such as ACR responses (ACR20, ACR50, and ACR70), nonresponder imputation was used: patients who discontinued the study or the study treatment were counted as nonresponders from the time they discontinued onward. Patients receiving rescue therapy from week 16 were analyzed as nonresponders after the rescue visit. Any patients for whom post-baseline data at any visit were missing were also counted as nonresponders at all visits.

For analysis of key secondary continuous outcomes, a modified baseline observation carried forward method was used. For patients who discontinued the study or study treatment due to an adverse event, the baseline observation was carried forward to

subsequent time points. For those who discontinued for reasons other than an adverse event, the last non-missing data point was carried forward. For those receiving rescue therapy starting from week 16, the last non-missing observation at or before rescue was carried forward. If a patient had a missing baseline value or no change from baseline values after the previously mentioned imputation, then the change from baseline was set to zero (no improvement).

The modified last observation carried forward method was a general approach used to impute missing data for all continuous measures including safety analyses, unless otherwise specified. With respect to key secondary end points, the method was used as a sensitivity analysis to the modified baseline observation carried forward. For patients receiving rescue therapy starting at week 16, the last non-missing observation at or before rescue was carried forward, while, for all other patients who discontinued from the study or study treatment, the last non-missing post-baseline observation before discontinuation was carried forward.

The manufacturer used a hierarchical testing procedure to control for multiple statistical comparisons in both included studies. The baricitinib 4 mg doses were compared with placebo first, testing for superiority in terms of the proportion of patients achieving ACR20, then the mean change from baseline in HAQ-DI and in DAS28-hs-CRP at week 12 (Hochberg procedure), then the proportion of patients achieving SDAI of 3.3 or less (SDAI remission) at week 12. Comparisons for these three outcomes were then repeated for the baricitinib 2 mg dose, in the same order.

Analysis Populations

In each study, the modified intention-to-treat population included all randomized patients who had received at least one dose of study drug. Patients' data were analyzed according to the study drug to which they were assigned. The per-protocol population included all patients in the modified intention-to-treat population who complied with treatment, did not have significant protocol deviations, and whose study site did not have issues following the good clinical practice required by regulatory bodies. The safety population included all patients who received at least one dose of study drug and who did not discontinue for the reason "lost to follow-up" at the first post-baseline visit.

Patient Disposition

There were 10% of baricitinib patients and 18% of placebo patients who discontinued the study in BEACON and 9% versus 13% of baricitinib and placebo patients who discontinued the BUILD study. The most common reason for discontinuation of baricitinib was an adverse event, and the most common reason for discontinuation of placebo was lack of efficacy in BEACON and withdrawal by patient in BUILD.

Table 6: Patient Disposition

	BEACON		BUILD	
Characteristics	BAR 2 mg	PLACEBO	BAR 2 mg	Placebo
Screened	959		1,241	
Screen failure	432		557	
Entry criteria not met	392		500	
Withdrawal by patient	23		42	
Physician decision	11		8	
Lost to follow-up	3		6	
Sponsor decision	2		1	
Adverse event	1		0	
Randomized	527 (55)		684 (55)	
	174	176	229	228
Rescued by switching to baricitinib 4 mg	38 (22)	56 (32)	21 (9)	55 (24)
Discontinued	17 (10)	32 (18)	20 (9)	29 (13)
Adverse event	7 (4)	7 (4)	10 (4)	8 (4)
Lack of efficacy	4 (2)	16 (9)	4 (2)	7 (3)
Withdrawal by patient	6 (3)	7 (4)	5 (2)	11 (5)
Physician decision	0	1 (1)	1 (< 1)	0
Death	0	0	0	2 (1)
Lost to follow-up	0	0	0	1 (< 1)
Protocol violation	0	1 (1)	0	0
Entered long-term extension study (Study JADY)	151 (87)	139 (79)	197 (86)	189 (83)
Completed post-treatment follow-up	11 (6)	22 (13)	18 (8)	21 (9)
Ended study participation (did not complete post-treatment follow-up)	12 (7)	15 (9)	14 (6)	18 (8)

Source: Clinical Study Reports for BEACON[4] and BUILD.[5]

Exposure to Study Treatments

In BEACON, the mean duration of exposure was 146.8 (standard deviation [SD] 40.4) days with baricitinib and 136.5 (SD 44.9) days with placebo. In BUILD, the mean duration of exposure for baricitinib was 155.8 (SD 32.5) days and placebo was 143.8 (SD 41.9) days.

Critical Appraisal

Internal Validity

The baseline and disease characteristics were comparable across treatment arms in both trials; therefore, double blinding was likely maintained and there was no signal of loss of allocation concealment throughout both studies. Immune suppression and thus infection are known side effects of cytokine inhibitors for RA; however, infections are also very common events outside of clinical trials. Therefore, it is unlikely that an infection in a given patient would have been interpreted by that patient as a sign they were on baricitinib, unless it were a serious, uncommon infection such as herpes zoster, which occurred very infrequently in the included trials, with no obvious differences between groups.

There was a small number (two to four patients) of missing data in some groups for the key secondary outcomes in both BEACON and BUILD. However, the impact on the analysis is likely small. The amount of missing data was larger for some of the other secondary outcomes that were not part of the statistical hierarchy, at times up to eight patients missing per group.

There were a relatively large number of withdrawals in the placebo group in BEACON, and a numerical difference between groups (10% of patients withdrew in the baricitinib group and 18% in the placebo group). Additionally, a large percentage of patients in both BEACON and BUILD were rescued by switching to the baricitinib 4 mg dose from both the baricitinib 2 mg and placebo groups. The percentage of patients requiring rescue was lower in the baricitinib group than in the placebo group in BEACON (22% versus 32%) and in BUILD (9% versus 24%). However, this may not have impact on the analysis at 12 weeks, given that patients were allowed to seek rescue therapy after 16 weeks of the double-blind treatment period. Thus, the analyses at 24 weeks are likely subject to considerable bias to the high percentage of shift over to rescue therapy. Non-responder imputation was employed to account for patients with discontinuation of treatment, shift over to rescue therapy, or missing post-baseline data. This approach is likely to be conservative and to bias the difference in ACR response toward null. In fact, those patients who withdrew due to lack of efficacy could reasonably be considered treatment failures. Of note, there were numerically more patients in the placebo group versus the baricitinib group in BEACON who withdrew for this reason.

For continuous outcomes, the manufacturer employed a modified baseline observation carried forward approach, and, in this approach, the method of imputation differs depending on the reason (adverse event, other) for withdrawal. For patients who withdrew due to an adverse event, their baseline data were carried forward, while for patients withdrawing for other reasons, the last non-missing data point was carried forward. The number of patients who withdrew due to an adverse event was similar between groups in both of the included studies; therefore, this approach is unlikely to make much difference between groups in the number of patients who had their baseline data carried forward, and this method of imputation would likely bias toward null on the difference between baricitinib and placebo groups.

The clinical expert consulted for this review noted that there was a high placebo response rate for ACR20 responses in both studies. For example, 40% of placebo patients in BUILD had an ACR20 response at 12 weeks. Although baricitinib did demonstrate improvement over placebo in both included studies, the high placebo response might have biased results against the study drug. The clinical expert was concerned whether this high placebo response might have indicated an issue with screening or with clinical assessments required for the ACR20.

External Validity

There appears to have been a relatively large number of patients screened out of each study (45% in each). It is not completely clear why this was the case, although the most commonly cited reason for screen failure was “entry criteria not met.” An example would be the considerably high levels of hs-CRP in the study patients in both trials (mean range across trials from 18 mg/L to 21 mg/L at baseline; normal level of 10 mg/L or under), whereas patients with an hs-CRP lower than the upper normal level range were excluded. A high level of hs-CRP is indicative of severe inflammation, which would most likely signal highly active RA. The clinical expert consulted by CADTH on this review noted that, while a

high screen failure rate and focus on patients with high hs-CRP is not uncommon in trials of RA drugs, patients with RA who are failing or are inadequately responding to previous treatments may not present such high levels of hs-CRP in clinical practice. Therefore, it may be possible to assume that the treatment effect observed in both trials (e.g., a difference of 22% and 26% in ACR20 at week 12) may overestimate the actual treatment effect in those who have had inadequate response to previous DMARDs but who do not have a high level of hs-CRP. The ACR/EULAR 2010 criteria more appropriately classify these patients as likely to progress to persistent and potentially destructive disease. However, the 2010 criteria were not developed for use in clinical practice.

The included studies did not include an active comparator such as tofacitinib, and, thus, this limits the conclusions that can be drawn regarding the comparative efficacy and harms of baricitinib versus other available comparators for RA in Canada.

The length of follow-up in the double-blind comparison period was 24 weeks, and the primary efficacy analyses were conducted at 12 weeks. This is a relatively short duration for a drug that is intended to be administered on a long-term basis and that has a relatively novel mechanism of action. There was an extension to BEACON and BUILD; however, there was no longer a placebo control after patients entered the long-term extension study. The treatment effect assessment at week 24 was also compromised, as shift over to rescue therapy was allowed after week 16, which rendered it difficult to correctly estimate the treatment effect of the study drug. Therefore, the durability of the treatment effect over 12 weeks or even longer remains uncertain.

The higher dose of baricitinib (4 mg) was not approved by Health Canada, due in part to the increased risk of thrombosis seen at this dose. Therefore, it would have been useful to see whether a longer duration would have resulted in a higher risk of thrombotic events for the baricitinib 2 mg dose.

The clinical expert consulted by CADTH on this review believed the baseline characteristics reflected those of a typical population of patients in clinical trials of RA who would be candidates for baricitinib. Not surprisingly, the baseline characteristics related to disease activity differed between studies, as patients in BEACON, who had failed a previous biologic, were more advanced in their disease, and this was again considered appropriate by the clinical expert. Both studies included Canadian sites.

Baricitinib is indicated for use in combination with methotrexate; however, it may also be used as monotherapy in patients who are not able to tolerate methotrexate, based on the product monograph. One of the subgroups of interest for this CDR review was patients intolerant to methotrexate; however, this was not studied in either BEACON or BUILD. Thus, the efficacy of baricitinib in this subgroup is unclear.

Efficacy

Only those efficacy outcomes identified in the review protocol are reported in this section (Table 3). See Table 7 for detailed efficacy data.

Clinical Response

The primary outcome for both BUILD and BEACON was the proportion of patients achieving ACR20 at week 12. In both BEACON (48.9% of baricitinib patients and 27.3% of placebo patients) and BUILD (65.9% versus 39.5%), more participants in the baricitinib

group than in placebo group achieved ACR20, and these differences were statistically significant between groups in both BEACON (OR, 2.7 [95% CI, 1.7 to 4.2], *P* = 0.001) and BUILD (OR, 3.0 [95% CI, 2.0 to 4.4], *P* = 0.001) (Table 7). ACR20 responses were also assessed at 24 weeks, although this was an exploratory outcome and not controlled for multiplicity. The proportion of patients achieving ACR20 at 24 weeks was higher with baricitinib than with placebo in BEACON (44.8% versus 27.3%) and BUILD (61.1% versus 42.1%).

Although other ACR outcomes were exploratory and not controlled for multiplicity, a higher proportion of baricitinib patients versus placebo patients achieved ACR50 at 12 weeks in BEACON (20.1% versus 8.0%) and BUILD (33.6% versus 12.7%) and also at 24 weeks in BEACON (23.0% versus 13.1%) and BUILD (41.5% versus 21.5%). ACR70 responses were achieved by a higher proportion of baricitinib patients versus placebo patients at 12 weeks in BEACON (12.6% versus 2.3%) and in BUILD (17.9% versus 3.1%) and at 24 weeks (BEACON: 13.2% versus 3.4%; BUILD: 25.3 versus 7.9%). These results of ACR50 and ACR70 were generally consistent with the primary results on ACR20 at week 12, indicating a favourable treatment effect of baricitinib 2 mg once daily over placebo.

Subgroup analyses of interest to this review were performed on the primary outcome (ACR20 responses at week 12) based on prior reason for failure on bDMARDs (lack of efficacy, adverse event, other) and for number of previous bDMARDs in BEACON (Table 10). Results for the lack-of-efficacy subgroup were 49.1% for the baricitinib group versus 27.1% for the placebo group. In the “other” subgroup, there were only three patients across both groups; therefore, the fact that 50% responded in the baricitinib group and 100% in the placebo group is difficult to interpret.

Radiographic Response

The mTSS was not investigated in BEACON and was an exploratory outcome in BUILD. mTSS increased from baseline to 24 weeks in both baricitinib (least squares mean [LSM] change from baseline of 0.33 [95% CI, 0.06 to 0.59]) and placebo (LSM change 0.70 [95% CI, 0.42 to 0.98]) groups for an LSM difference between the baricitinib group and the placebo group of –0.38 (95% CI, –0.74 to –0.01) (Table 11).

Health-Related Quality of Life

The HAQ-DI was a secondary outcome of both studies, and statistical comparisons were controlled for multiplicity. In each of the studies, baricitinib reduced (improved) scores on the HAQ-DI from baseline to week 12 when compared with placebo. These differences were statistically significant in both BEACON (LSM difference between groups of –0.20 [95% CI, –0.32 to –0.08], *P* = 0.001) and in BUILD (LSM difference between groups of –0.21 [95% CI, –0.30 to –0.11], *P* = 0.001) (Table 7). In both studies, the week-24 responses on the HAQ-DI were consistent with those seen at week 12.

Other scales were used to assess HRQoL as exploratory outcomes: SF-36 and the EuroQol 5-Dimensions (EQ-5D) questionnaire (Table 11). There were no statistically significant differences between baricitinib and placebo groups for the SF-36 MCS at week 12 or 24. PCS increased (improved) from baseline in each of the groups in BEACON and BUILD. The LSM difference between baricitinib and placebo groups at 24 weeks in BEACON was 4.3 (95% CI, 2.6 to 6.1) and in BUILD was 3.7 (95% CI, 2.0 to 5.4).

On the EQ-5D Health State Index/Self-Perceived Health score, both baricitinib and placebo groups experienced an increase (improvement) in score from baseline. In BEACON, the LSM difference between groups was 0.049 (95% CI, 0.018 to 0.081), and, in BUILD, it was 0.013 (95% CI, 0.023 to 0.075). Similar results were reported when the UK algorithm was used.

Functional and Disability Outcomes

Duration of Joint Stiffness

Duration of morning joint stiffness was a secondary outcome of BUILD and BEACON and was not controlled for multiple comparisons. In BEACON, the median change from baseline to week 12 was –21.0 minutes (95% CI, –30.0 to –10.0 minutes) with baricitinib and –3.5 minutes (95% CI, –8.0 to 0.0 minutes) with placebo, for a median difference between groups of 15.0 minutes (95% CI, 2.0 to 30.0 minutes). In BUILD, the median difference between groups at week 12 was –15.7 minutes (95% CI, –27.9 to –5.6 minutes) (Table 11).

Severity of Morning Joint Stiffness

This outcome was measured using a numeric rating scale in which a reduction from baseline denotes improvement. This outcome was not adjusted for multiple comparisons. The LSM difference between the baricitinib group and the placebo group at week 12 was –0.6 (95% CI, –1.0 to –0.2) (Table 11).

Fatigue

The Worst Tiredness Numerical Rating Scale was used to assess fatigue in both studies, although only as an exploratory outcome in each. Scores decreased (improved) from baseline to week 12 in both the baricitinib and placebo groups in BEACON (LSM difference between groups of –0.7 [95% CI, –1.2 to –0.2]) and in BUILD (LSM difference between groups of –0.4 [95% CI, –0.8 to –0.0]) (Table 11).

Disease Activity

The DAS28-hs-CRP at week 12 was a secondary outcome in both BEACON and BUILD. In each study, baricitinib reduced (improved) scores versus placebo, and these differences between groups were statistically significant in BEACON (LSM difference between groups of –0.66 [95% CI, –0.96 to –0.35], $P = 0.001$) and in BUILD (LSM difference between groups of –0.75 [95% CI, –0.97 to –0.53], $P = 0.001$) (Table 7). In both studies, responses at week 24 were consistent with those seen at week 12.

The proportion of patients achieving an MCID on the SDAI at week 12 was assessed as a secondary outcome in both studies. In BUILD, the proportion of patients achieving a clinically significant improvement in SDAI was higher with baricitinib than with placebo (9.2% versus 0.9% of patients), and this difference was statistically significant (OR not reported, $P = 0.001$). In BEACON, there was no statistically significant difference between groups (Table 7). In both studies, the percentage of patients with SDAI response increased from week 12 to week 24, although at week 24 in BEACON there was still no statistically significant difference between baricitinib and placebo groups.

Health Care Resource Utilization

This outcome was not reported on in the Clinical Study Reports.

Table 7: Key Efficacy Outcomes

	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
Clinical Response (ACR)				
ACR20 response at week 12, n (%)	85 (48.9)	48 (27.3)	151 (65.9)	90 (39.5)
Difference in response rate [95% CI][a]	21.6 [11.7 to 31.5]		26.5 [17.6 to 35.3]	
Odds ratio [95% CI][a]	2.7 [1.7 to 4.2], *P* = 0.001		3.0 [2.0 to 4.4], *P* = 0.001	
ACR20 response at week 24, n (%)	78 (44.8)	48 (27.3)	140 (61.1)	96 (42.1)
Difference in response rate [95% CI]	17.6 [7.7 to 27.4]		19.0 [10.0 to 28.0]	
Odds ratio [95% CI]	2.3 [1.5 to 3.6], *P* = 0.001		2.2 [1.5 to 3.2], *P* = 0.001	
ACR50 response at week 12, n (%)	35 (20.1)	14 (8.0)	77 (33.6)	29 (12.7)
Difference in response rate [95% CI]	12.2 [5.0 to 19.3]		20.9 [13.4 to 28.4]	
Odds ratio [95% CI][b]	3.0 [1.6 to 5.9] *P* = 0.002		3.5 [2.2 to 5.6], *P* = 0.001	
ACR50 response at week 24, n (%)	40 (23.0)	23 (13.1)	95 (41.5)	49 (21.5)
Difference in response rate [95% CI]	9.9 [1.9 to 17.9]		20.0 [11.7 to 28.3]	
Odds ratio [95% CI]	2.0 [1.2 to 3.6], *P* = 0.015		2.6 [1.7 to 4.0], *P* = 0.001	
ACR70 response at week 12, n (%)	22 (12.6)	4 (2.3)	41 (17.9)	7 (3.1)
Difference in response rate [95% CI]	10.4 [5.0 to 15.8]		14.8 [9.4 to 20.3]	
Odds ratio [95% CI]	NR, *P* = 0.001		6.9 [3.0 to 15.9], *P* = 0.001	
ACR70 response at week 24, n (%)	23 (13.2)	6 (3.4)	58 (25.3)	18 (7.9)
Difference in response rate [95% CI]	9.8 [4.1 to 15.5]		17.4 [10.8 to 24.1]	
Odds ratio [95% CI]	NR, *P* = 0.001		4.0 [2.3 to 7.2], *P* = 0.001	
Health-Related Quality of Life: HAQ-DI				
HAQ-DI				
Mean (SD) baseline	1.71 (0.55)	1.78 (0.57)	1.51 (0.62)	1.50 (0.60)
LSM (SE) change from baseline to week 12	–0.37 (0.04) N = 172	–0.17 (0.04) N = 176	–0.54 (0.036) N = 229	–0.34 (0.037) N = 224
Mean difference between groups [95% CI][b]	–0.20 [–0.32 to –0.08], *P* = 0.001		–0.21 [–0.30 to –0.11], *P* = 0.001	
LSM (SE) change from baseline to week 24	–0.38 (0.05) N = 172	–0.15 (0.05) N = 172	–0.62 (0.039) N = 228	–0.38 (0.040) N = 220
Mean difference between groups [95% CI][b]	–0.23 [–0.35 to –0.12], *P* = 0.001		–0.24 [–0.35 to –0.14], *P* = 0.001	
Disease Activity: DAS28-hs-CRP				
Mean (SD) baseline	6.03 (0.89)	5.89 (0.94)	5.57 (0.96)	5.53 (0.91)
LSM (SE) change from baseline to week 12	–1.49 (0.11) N = 172	–0.83 (0.11) N = 174	–1.83 (1.22) N = 229	–1.05 (1.23) N = 228
Mean difference between groups [95% CI] [b]	–0.66 [–0.96 to –0.35], *P* = 0.001		–0.75 [–0.97 to –0.53], *P* = 0.001	
LSM (SE) change from baseline to week 24	–1.38 (0.13) N = 172	–0.81 (0.13) N = 170	–2.11 (0.090) N = 228	–1.30 (0.093) N = 220
Mean difference between groups [95% CI] [b]	–0.57 [–0.87 to –0.27], *P* = 0.001		–0.81 [–1.05 to –0.57], *P* = 0.001	
Disease Activity: SDAI				
SDAI ≤ 3.3 response rate at week 12, n (%)	4 (2.3)	3 (1.7)	21 (9.2)	2 (0.9)
Difference in response rate [95% CI]	0.6 [–2.3 to 3.5]		8.3 [4.4 to 12.2]	

	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
Odds ratio [95% CI][a]	NR, *P* = 0.723		NR, *P* = 0.001	
SDAI ≤ 3.3 response rate at week 24, n (%)	8 (4.6)	4 (2.3)	38 (16.6)	9 (3.9)
Difference in response rate [95% CI]	2.3 [–1.5 to 6.1]		12.6 [7.2 to 18.1]	
Odds ratio [95% CI][a]	NR, *P* = 0.257		4.9 [2.3 to 10.4], *P* = 0.001	

ACR = American College of Rheumatology; BAR = baricitinib; DAS28 = Disease Activity Scale-28; CI = confidence interval; HAQ-DI = Health Assessment Questionnaire–Disability Index; hs-CRP = high-sensitivity C-reactive protein; LSM = least squares mean; NR = not reported; SD = standard deviation; SDAI = Simplified Disease Activity Index; SE = standard error.

Note: All of the analysis was based on the intention-to-treat population with missing data being treated with non-responder imputation: 1) patients who discontinued the study or the study treatment were counted as non-responders from the time they discontinued onward, and 2) any patients who were missing post-baseline data at any visit were also counted as non-responders at all visits.

[a] 95% CI is calculated for the difference in response rates using the Newcombe-Wilson method without continuity correction. 95% CI and *P* value for OR from logistic regression model adjusted for region + history of bDMARD use (< 3, ≥ 3) + treatment group. When logistic regression sample size requirements are not met, *P* value from Fisher's exact test is produced instead of OR and 95% CI. In BUILD: 95% CI and *P* value for OR from logistic regression model adjusted for region + baseline joint erosion status (yes/no) + treatment group.

[b] *P* value, LSM, SE, and 95% CI from ANCOVA model adjusted for baseline value, study region, history of bDMARD (< 3, ≥ 3) and in BUILD adjusted for baseline + region + baseline joint erosion status (yes/no) + treatment group.

Harms

Only those harms identified in the review protocol are reported in this section. See Table 8 for detailed harms data.

Adverse Events

In BEACON, 71% of baricitinib patients and 64% of placebo patients experienced an adverse event, while, in BUILD, the percentages were 67% in baricitinib and 71% in placebo groups. The most common adverse event was upper respiratory tract infection, occurring in 9% of baricitinib patients and 5% placebo patients in BEACON and in 6% of baricitinib and 8% of placebo patients in BUILD (Table 8).

Serious Adverse Events

Serious adverse events were reported in 7% of baricitinib patients and 9% of placebo patients in BEACON and in 7% of patients in each group in BUILD. Herpes zoster as a serious adverse event occurred in 1% of patients in each group in BEACON and 2% of baricitinib patients versus no placebo patients in BUILD (Table 8).

Withdrawals Due to Adverse Events

There were 4% of patients who withdrew due to an adverse event in each of the baricitinib and placebo groups, in each of the included studies (Table 8).

Mortality

Across both studies, no patients died in the baricitinib group, while two patients died in the placebo group in BUILD (Table 8).

Notable Harms

Notable harms identified for this review included infections, which occurred in 44% of baricitinib patients and 31% of placebo patients in BEACON and 31% of baricitinib patients and 35% of placebo patients in BUILD. Serious infections occurred in 3% of patients in each group in BEACON and 3% of baricitinib patients and 2% of placebo patients in BUILD. Other notable harms included malignancies, thrombotic events, dyslipidemia, and elevations in hepatic enzymes, and there were very few events and no clear differences between groups within BEACON and BUILD. There was a numerical higher risk of elevated platelet counts with baricitinib treatment versus placebo in BUILD (19% versus 5%); however, there was a much smaller difference between groups in BEACON (18% versus 14%). Low neutrophil counts were seen with 6% of baricitinib patients and 2% of placebo patients in BEACON and 8% of baricitinib versus 4% of placebo in BUILD. No patients had a gastrointestinal perforation in either study (Table 8).

Table 8: Harms

	BEACON		BUILD	
	BAR 2 mg	PLACEBO	BAR 2 mg	PLACEBO
Adverse Events				
Treatment-emergent adverse event, week 24, n (%)	123 (70.7)	112 (63.6)	154 (67.2)	161 (70.6)
Most common (5% in either group), n (%)				
Diarrhea	10 (5.7)	12 (6.8)	7 (3.1)	8 (3.5)
Upper respiratory tract infection	16 (9.2)	8 (4.5)	14 (6.1)	18 (7.9)
Nasopharyngitis	12 (6.9)	7 (4.0)	10 (4.4)	18 (7.9)
Headache	17 (9.8)	11 (6.3)	15 (6.6)	8 (3.5)
Urinary tract infection	7 (4.0)	6 (3.4)	12 (5.2)	5 (2.2)
Bronchitis	6 (3.4)	6 (3.4)	6 (2.6)	12 (5.3)
Serious Adverse Events				
Total SAE, week 24, n (%)	12 (6.9)	15 (8.5)	15 (6.6)	16 (7.0)
Total SAE, ICH definition, week 24, n (%)	7 (4.0)	13 (7.4)	6 (2.6)	11 (4.8)
Herpes zoster	2 (1.1)	2 (1.1)	4 (1.7)	0
Anemia	1 (0.6)	0	1 (0.4)	1 (0.4)
Hemorrhagic anemia	0	1 (0.6)	0	0
Leukocytosis	0	1 (0.6)	0	0
Thrombocytopenia	0	1 (0.6)	0	0
Tachycardia	0	1 (0.6)	0	0
Atrial fibrillation	0	0	1 (0.4)	0
Myocardial infarction	0	0	0	1 (0.4)
Ventricular tachycardia	0	0	0	1 (0.4)
Diarrhea	0	1 (0.6)	0	0
Gastritis	1 (0.6)	0	0	0
Oral disorder	1 (0.6)	0	0	0
Cholecystitis	1 (0.6)	0	0	0
Diverticulum intestinal	0	0	0	1 (0.4)

	BEACON		BUILD	
	BAR 2 mg	PLACEBO	BAR 2 mg	PLACEBO
Gastrointestinal hemorrhage	0	0	0	1 (0.4)
Non-cardiac chest pain	0	0	1 (0.4)	0
Edema peripheral	0	0	1 (0.4)	0
Hepatic steatosis	0	0	1 (0.4)	0
Upper respiratory tract infection	1 (0.6)	0	0	0
Campylobacter gastroenteritis	1 (0.6)	0	0	0
Cellulitis	0	2 (1.1)	0	0
Intervertebral discitis	1 (0.6)	0	0	0
Osteomyelitis	1 (0.6)	0	0	0
Pneumonia	1 (0.6)	1 (0.6)	1 (0.4)	1 (0.4)
Tooth infection	0	1 (0.6)	0	0
Urinary tract infection	0	0	0	1 (0.4)
Bronchitis	0	0	0	1 (0.4)
Gastroenteritis	0	0	1 (0.4)	0
Wound infection, staphylococcal	0	0	0	1 (0.4)
Concussion	1 (0.6)	0	0	0
Fractures	0	7 (4.2)	0	2 (0.8)
Alcohol poisoning	0	1 (0.6)	0	0
Cardiac contusion	0	1 (0.6)	0	0
Laceration	0	1 (0.6)	0	0
Ligament sprain	1 (0.6)	0	0	0
Road traffic accident	1 (0.6)	1 (0.6)	0	0
Blood creatinine phosphokinase increased	0	1 (0.6)	0	0
Hepatic enzyme increased	0	1 (0.6)	0	0
Electrolyte imbalance	0	1 (0.6)	0	0
Hyperglycemia	0	1 (0.6)	0	0
Alanine aminotransferase increased	0	0	1 (0.4)	0
Glomerular filtration rate decreased	0	0	1 (0.4)	0
Malnutrition	0	1 (0.6)	0	0
Rheumatoid arthritis	0	3 (1.7)	0	0
Osteoarthritis	0	1 (0.6)	0	0
Back pain	0	0	0	1 (0.4)
Myopathy	0	0	0	1 (0.4)
Synovial cyst	0	0	0	1 (0.4)
Ovarian adenoma	1 (0.7)	0	0	0
Ovarian low-malignant-potential tumour	1 (0.7)	0	0	0
Carpal tunnel syndrome	1 (0.6)	0	0	0
Confusional state	0	1 (0.6)	0	0
Delirium	0	1 (0.6)	0	0
Migraine	0	0	1 (0.4)	0
Subarachnoid hemorrhage	0	0	0	1 (0.4)

	BEACON		BUILD	
	BAR 2 mg	PLACEBO	BAR 2 mg	PLACEBO
Depression	0	0	0	2 (0.9)
Irritability	0	0	0	1 (0.4)
Mood altered	0	0	0	1 (0.4)
Post-traumatic stress disorder	0	0	1 (0.4)	0
Suicidal ideation	0	0	0	1 (0.4)
Renal failure	0	1 (0.6)	0	1 (0.4)
Asthma	1 (0.6)	0	0	0
Pneumonia aspiration	1 (0.6)	0	0	0
Pulmonary mass	0	1 (0.6)	0	0
Acute respiratory distress syndrome	0	0	1 (0.4)	0
Acute respiratory failure	0	0	1 (0.4)	0
Lung cyst	0	0	0	1 (0.4)
Hypertension	0	2 (1.1)	0	0
Hypertensive crisis	0	1 (0.6)	0	0
Peripheral artery occlusive disease	1 (0.6)	0	0	0
Peripheral embolism	1 (0.6)	0	0	0
Psoriasis	0	0	1 (0.4)	0
Subcutaneous emphysema	0	0	0	1 (0.4)
Mortality				
Death, n (%)	0	0	0	2 (1)
Withdrawals Due to Adverse Event				
Discontinuation from study due to adverse event, week 24, n (%)	7 (4.0)	7 (4.0)	9 (3.9)	8 (3.5)
Notable Harms, n (%)				
Infections	76 (43.7)	55 (31.3)	70 (30.6)	79 (34.6)
Required treatment with antimicrobials	50 (28.7)	29 (16.5)	44 (19.2)	42 (18.4)
Cardiovascular/thrombosis				
Patients with ≥ 1 MACE[a] event	0	0	0	2 (0.9)
Cardiovascular death	0	0	0	1 (0.4)
Stroke (hemorrhagic)	0	0	0	1 (0.4)
Myocardial infarction	0	0	0	1 (0.4)
Patients with ≥ 1 other cardiovascular event	1 (0.6)	0	1 (0.4)	2 (0.9)
Serious arrhythmia	1 (0.6)	0	1 (0.4)	1 (0.4)
Coronary revascularization procedure	0	0	0	1 (0.4)
Platelets (billion cells/L) high	28 (17.9)	21 (13.5)	38 (19.4)	9 (4.7)
Low	0	2 (1.2)	2 (0.9)	4 (1.8)
Neutrophils (billion cells/L) low	11 (6.4)	175 3 (1.7)	18 (7.9)	8 (3.6)
High	28 (23.7)	43 (40.6)	36 (23.1)	45 (28.1)
Malignancies	1 (0.6)	0	0	0
Gastrointestinal perforation	0	0	0	0
Hepatic toxicity				

	BEACON		BUILD	
	BAR 2 mg	PLACEBO	BAR 2 mg	PLACEBO
Hepatic function abnormal	0	1 (0.6)	2 (0.9)	1 (0.4)
Aspartate aminotransferase increased	0	0	3 (1.3)	1 (0.4)
Alanine aminotransferase increased	0	1 (0.6)	5 (2.2)	2 (0.9)
Blood alkaline phosphatase increased	0	0	0	1 (0.4)
Blood bilirubin increased	NR	NR	0	0
Dyslipidemia				
Blood cholesterol increased	NR	NR	1 (0.4)	1 (0.4)
Hypercholesterolemia	1 (0.6)	2 (1.1)	5 (2.2)	2 (0.9)
Hyperlipidemia	3 (1.7)	1 (0.6)	2 (0.9)	2 (0.9)
Dyslipidemia	1 (0.6)	0	3 (1.3)	2 (0.9)

BAR = baricitinib; ICH = International Council on Harmonization; MACE = major adverse cardiovascular event; NR = not reported; SAE = serious adverse event.

[a] Cardiovascular death, myocardial infarction, or stroke.

Discussion

Summary of Available Evidence

Two multinational, manufacturer-sponsored, double-blind randomized controlled trials met the inclusion criteria for this systematic review, featuring patients who had an inadequate response to cDMARDs in one study and patients who had an inadequate response to TNF inhibitors in the other study. Both studies had a 24-week double-blind treatment period in which baricitinib 2 mg and baricitinib 4 mg were compared with placebo. The approved baricitinib 2 mg dose was the focus of this review. The primary outcome in each study was the proportion of patients achieving an ACR20 response at 12 weeks, while key secondary outcomes that were accounted for multiplicity included HRQoL measures: mean change from baseline on the HAQ-DI and the DAS28-hs-CRP, and patients with a clinically meaningful improvement on the SDAI.

Major limitations included a lack of active comparators in the included studies and a relatively short duration of follow-up (24 weeks) for a drug with a relatively novel mechanism of action, with safety issues noted at the higher, 4 mg dose. There were a relatively high proportion of withdrawals in the placebo group in one of the studies (BUILD) and a numerical difference between groups (10% withdrawals with baricitinib and 18% with placebo). A large proportion of patients opted for rescue therapy with baricitinib 4 mg after 16 weeks, particularly in the placebo group in both BEACON (22% versus 32% of patients) and BUILD (9% versus 24%). There was no subgroup analysis performed for patients with prior methotrexate intolerance, which is a potential gap, given that baricitinib may be used as monotherapy in these patients.

Efficacy

The included studies show that baricitinib 2 mg, once daily, demonstrated treatment efficacy over a 12-week period on the specific outcomes of disease activity, symptoms, and HRQoL important to patients with RA who have inadequate response or intolerance to one or more prior TNF inhibitors or cDMARDs. The clinical expert noted an unusually large placebo response in both studies, which may have reduced the treatment effect for ACR responses. Although the ACR20 is the more commonly assessed among primary and key secondary outcomes in clinical trials, the ACR50 and ACR70 criteria are considered more rigorous and more representative of an optimal clinical improvement.[17] As would be expected with more rigorous criteria for response, the number of patients achieving an ACR50 was lower than those achieving ACR20, and was lower still for those achieving ACR70 in both included studies. For example, in BUILD, ACR20 responses after 12 weeks of baricitinib were achieved in 66% of patients (40% in placebo), ACR50 responses in 34% baricitinib patients (13% in placebo), and ACR70 responses in 18% of baricitinib patients (3% in placebo). There were no clear changes in ACR20 responses from weeks 12 to 24 in either study; however, there were numerical increases in ACR50 and ACR70 responses from weeks 12 to 24 in both the baricitinib and placebo groups. The improvements on the HAQ-DI, a disease-specific HRQoL instrument, did not meet the MCID of 0.22 in either study. There was once again a large placebo response on the HAQ-DI, suggesting that a large placebo response may have made it difficult to find a clinically significant improvement for baricitinib over placebo. No MCID is available for the DAS28, and very few patients achieved a clinically significant improvement on the SDAI. Thus, the clinical significance of improvements in disease activity are not clear.

There were no randomized controlled trials available that directly compared baricitinib with any of the bDMARDs or the other JAK inhibitor, tofacitinib, which is currently available for the treatment of RA in Canada. CDR reviewed three available network meta-analyses involving baricitinib, including one performed by CADTH, and there was no clear evidence of baricitinib at the 2 mg once daily dosage being significantly better than other drugs used to treat RA (see Appendix 7 for detailed review). The manufacturer submitted a network meta-analysis that demonstrated that, in patients with inadequate response to a prior TNF inhibitor, patients treated with baricitinib did not respond as well for ACR and EULAR compared with patients treated with tocilizumab or rituximab. There were no differences in ACR responses between the baricitinib and tofacitinib groups in the manufacturer-submitted network meta-analysis. In patients who failed on a cDMARD, ACR responses were similar between baricitinib, cDMARDs, and bDMARDs. The two additional published network meta-analyses that were reviewed had results generally consistent with that of the manufacturer's submitted analysis.

Baricitinib and tofacitinib are both orally administered JAK inhibitors; however, they target JAKs in different ways. Baricitinib has approximately 100-fold selectivity for the JAK-1 and JAK-2 isoforms, and, thus, might be expected to have greater activity against the cytokines associated with these receptors, namely IL-2, IL-6, IL-12, and IL-23, as well as interferon gamma and transforming growth factor beta. Tofacitinib, however, is relatively selective for JAK-1 and JAK-3, and this may be associated with a different performance with respect to efficacy and harms.[18] However, to date, there are no direct comparisons of these two drugs in clinical trials; thus, all comparisons are limited to indirect comparisons such as those described in Appendix 7. As noted, there were no clear and consistent differences in efficacy between baricitinib and tofacitinib. Both drugs have significant immune-suppressing effects and, thus, by extension, serious infections are of concern, as are malignancies, and these are both highlighted in black box warnings.[19,20] The issue of thrombotic events is discussed in more detail in the next section; however, it should be noted that thrombotic events were seen primarily at the baricitinib 4 mg dose, which is not approved in Canada.

It would be prudent to study the efficacy and harms of JAK inhibitors, with their unique mechanism of action, beyond the 24-week double-blind period in the included studies. BEYOND was an open-label extension that included patients enrolled in both BEACON and BUILD, and this study is summarized in Appendix 6. The lack of a control group is a key limitation when trying to assess results, as is the fact that this is a relatively selected population, composed of patients who continued on the baricitinib 2 mg dose and did not require "rescue" to the 4 mg dose, suggesting that they were benefiting from baricitinib. Results from BEYOND suggest that the percentage of patients with ACR20, ACR50, and ACR70 responses increased at week 72 from the percentages seen in the core studies at week 12, in BEACON (ACR20: 49% to 62%) and BUILD (66% to 72%). Although this is not a large increase in either study, and it is a selected population, at the very least it suggests that there is no attenuation of response over time.

Baricitinib is indicated, in combination with methotrexate, for treatment of moderate or severe RA in patients who have failed one or more previous DMARDs. Failure was defined as inadequate response or intolerance to prior treatment with DMARDs. The clinical expert consulted on this review noted that there is no consistent, reliable definition of an "inadequate response" to a given therapy for RA. Both trials included a large patient population — of about 70% in BEACON and 50% in BUILD — that received baricitinib in combination with methotrexate alone and a smaller group — of about 8% in BEACON and

22% in BUILD — that received baricitinib in combination with methotrexate and one cDMARD.

BEACON enrolled patients who had failed at least one biologic TNF inhibitor, and BUILD featured patients who had failed cDMARDs. All of the included patients experienced a relatively high hs-CRP level, probably indicating severe inflammation due to poorly controlled RA. The clinical expert consulted by CADTH on this review noted that this enrichment of study populations with patients with high hs-CRP is not uncommon in clinical trials of RA drugs.

As would be expected, patients in BEACON had had RA longer than those in BUILD (time from RA symptom onset: 14 versus 7 years), suggesting that they had more advanced disease and, thus, were more likely to need to resort to therapy with biologics. Also as one would predict, there were fewer ACR20 responses in BEACON than in BUILD; however, the treatment effects were relatively consistent between the two studies, suggesting that baricitinib should achieve similar results regardless of whether patients had previously failed a cDMARD or a bDMARD. Subgroup analyses provided no suggestion that number of prior biologics or reason for failure on a biologic (adverse event, lack of efficacy, or "other") impacted response to baricitinib. Subgroups of interest to this CDR review included patients with inadequate response to bDMARDs or to cDMARDs, and these subgroups were addressed by the populations in BEACON (included patients with inadequate response to bDMARDs) and BUILD (included patients with inadequate response to cDMARDs). There was no subgroup analysis planned for patients intolerant to methotrexate, specifically; as a result, we do not know what the efficacy and harms would be in these patients, despite the fact that baricitinib may be used as monotherapy in patients who are intolerant to methotrexate.

Harms

Adverse events related to immune suppression are infections, including serious infections such as herpes zoster, and malignancies. Malignancies are a long-established adverse event associated with TNF inhibitors, for example. There was no clear evidence that baricitinib increased the risk of herpes zoster, although four patients developed this infection in BUILD and none in placebo. An increased risk of herpes zoster is one of the considerations that led Health Canada not to approve the baricitinib 4 mg dose.[21] There was no evidence of increased risk of malignancies in the included studies; however, it is unlikely that malignancies could develop over this short a time period. BEYOND is a long-term study extension with a follow-up of 48 months, summarized in Appendix 6; however, because it has no control group, there is no way to assess risk of malignancy from this study either.

A potential increased risk of thrombosis appears as a warning in the product monograph for baricitinib. Although there was a numerical increase in the proportion of patients with high platelet counts observed in BUILD (but not in BEACON), there were few thrombotic events and no indication of increased risk of major adverse coronary events with baricitinib. The FDA noted in its integrated review of safety that this elevation in platelets was a phenomenon observed across the various clinical trials of baricitinib, at both the 2 mg and the 4 mg doses.[21] As noted, the increase in thrombotic events was a key reason why the 4 mg dose was not approved for marketing by various regulatory bodies, including Health Canada. The FDA noted that the elevated platelet counts appear to peak at about two weeks but remain elevated above normal even after coming down from their peak.[21] The FDA also noted that the other two JAK inhibitors, tofacitinib and ruxolitinib, do not appear to have any issues with increased platelets; in fact, if anything, platelet counts tend to be reduced with these drugs.[21] Thus, it is unclear what the mechanism is behind the elevated platelet counts that have been observed at both doses of baricitinib, nor is it clear why thrombotic events were observed only at the higher 4 mg dose. It is also unclear whether the elevated platelet counts are related to the thrombotic events, or whether there is another cause. In March 2019, Health Canada issued a warning about an increased risk of thrombotic events with tofacitinib, suggesting that this may be a class effect.[22] Long-term safety surveillance, particularly for thrombotic events and other major adverse cardiovascular events, is warranted.

Potential Place in Therapy[2]

Baricitinib is the second JAK inhibitor licensed for use in RA. It is more selective than tofacitinib, being an inhibitor of JAK-1 and JAK-2 compared with tofacitinib, which is a pan-JAK inhibitor. Its selectivity is expected to reduce adverse events such as infections and episodes of herpes zoster. However, the safety database does not suggest that baricitinib is safer, and “eyeball” comparison does not find it more effective than tofacitinib.

Baricitinib joins a crowded field of therapies for RA and will compete with five “brand name” TNF inhibitors and a growing number of biosimilar versions, two IL-6 inhibitors, a blocker of T-cell activation (abatacept), a B-cell depletor (rituximab), and tofacitinib. Mixed treatment comparisons have not found that baricitinib is more effective than its competitors, so, aside from the convenience of once daily oral therapy, baricitinib does not provide a striking reason for being selected as a first choice in RA treatment. Furthermore, there is no subset of RA patients in whom baricitinib might be the preferred choice.

Little has been learned about the durability of baricitinib therapy, and long-term safety has yet to be established. Until there is more information, it is unclear whether baricitinib will fulfill an unmet need in RA therapy. Currently, it will be part of a crowded therapeutic field, and because so much more is known about the “older” drugs, baricitinib is not expected to develop a large market share in RA in the short term.

[2] This information is based on information provided in draft form by the clinical expert consulted by CDR reviewers for the purpose of this review.

Conclusions

The included studies showed that baricitinib at 2 mg, administered orally once daily, improved clinical responses using the ACR20 after 12 weeks of treatment in a population of patients with RA who had an inadequate response to either cDMARDs or TNF inhibitors. Between 70% and 80% of the study patients had concurrent treatment with methotrexate, either alone or in combination with one DMARD. Subgroup analysis demonstrated that ACR20 responses were unaffected by the reason for inadequate response to TNF inhibitor (i.e., lack of efficacy or adverse event). There was no subgroup analysis performed for patients with prior methotrexate intolerance, which represents a potential gap, given that these patients may receive baricitinib as monotherapy. There was a statistically significant improvement in HRQoL and in disease activity with baricitinib, yet the clinical relevance of the improvement remains unclear. There were numerous other outcomes that assessed symptoms and HRQoL; however, these were not adjusted for multiple statistical comparisons. It is likely that the beneficial effects of baricitinib are no different from those of tofacitinib and bDMARDs, as consistently shown in network meta-analyses. The risk of notable harms, such as serious infections, malignancies, cardiovascular events, dyslipidemia, and elevated hepatic enzymes did not appear to differ between baricitinib and placebo, although the included studies were not powered to assess outcomes such as these. Long-term study of these potential adverse events and durability of treatment effect in a real-world setting is warranted.

Appendix 1: Patient Input Summary

This section was prepared by CADTH staff based on the input provided by patient groups.

1. Brief Description of Patient Group(s) Supplying Input

Three patient groups — The Arthritis Society, the Canadian Arthritis Patient Alliance, and Arthritis Consumer Experts — provided input for this summary.

The Arthritis Society is Canada's principal health charity that provides education, programs, and support to Canadians living with arthritis. Founded in 1948, it is the largest non-government funder of arthritis research in Canada. The Arthritis Society provided combined patient input with the Canadian Arthritis Patient Alliance (CAPA). CAPA is a patient-driven, independent, national education and advocacy organization creating links among Canadians with arthritis. CAPA assists individuals living with arthritis in order for them to become more effective advocates, as well as to improve their quality of life. The patient input submission was prepared independently, without influence from any outside party; however, the manufacturer helped identify rheumatologists who had conducted clinical trials of baricitinib in Canada, so that these professionals could pass on the Society's contact information to patients for this submission. The Arthritis Society has received funding by Eli Lilly and other pharmaceutical companies. CAPA has received funding by pharmaceutical companies other than Eli Lilly.

Arthritis Consumer Experts (ACE) is Canada's largest not-for-profit patient-led organization that provides information, education, and support programs to people with arthritis, helping them take control of their disease and improve their quality of life. ACE also advocates on arthritis health policy issues. ACE has received funding by Eli Lilly Canada. The patient input submission was prepared independently without influence from any outside party.

2. Condition-Related Information

Data provided by The Arthritis Society was collected through a social media request for information and engaged 39 people living with rheumatoid arthritis (RA), including three patients who had experience with baricitinib. The data were gathered in February and March 2017, December 2018, and January 2019 in Canada. CAPA conducted a brief phone interview with a person who was diagnosed with RA 10 years ago, who participated in a clinical trial for baricitinib and has remained on the drug for approximately four years at the time the submission was prepared. Further information was obtained through personal experiences of the Board members living with RA, in addition to many years of interfacing with CAPA'S membership. ACE gathered information in Canada using its Survey Monkey platform from December 2018 to January 2019, and presented in its submission the feedback received from 10 patients, none of whom, however, had experience with baricitinib.

People living with RA may experience a variety of symptoms that significantly affect all aspects of their lives. The most frequently reported symptoms that have the most negative impact on quality of life include joint stiffness and swelling, joint pain, limitation of mobility, and ongoing fatigue. RA can significantly restrict the ability to perform daily activities; simple tasks most people take for granted can take some patients a long time and much effort or pain to complete. People with this condition must plan out or adjust their activities carefully and always keep in mind the state of their disease, what types of activities they can handle or need to give up, and how much help they may need. This adversely impacts employment, personal life and family responsibilities, recreational activities, as well as

physical and emotional health. Patients may also require the assistance of caregivers to help with many aspects of life at home.

There is no cure for RA, as treatments available only manage the signs and symptoms of the disease. In addition, structural joint damage caused by RA is typically irreversible. Disease activity is likely to fluctuate over time, without any way for patients to predict what tomorrow will look like.

3. Current Therapy-Related Information

Several patients reported trying multiple types of medications over time in order to control their symptoms, including disease-modifying antirheumatic drugs, both conventional and biologic, which is a common situation among people with RA. There is currently no way to predict which patients will respond to which medications. Patients need a variety of treatment options, providing alternatives in the event of treatment failure, waning efficacy over time, adverse events, or lack of full coverage. Specifically, some patients highlighted the need for treatments that confer better control of pain and fatigue. Other patients mentioned troubling adverse events with existing treatments, such as nausea, gastrointestinal upsets such as stomach pain and ulcers, infections, and injection-site reactions. The cost of treatment was reported as an issue for some patients, especially when the condition becomes disabling and employment is no longer possible. Other patients mentioned the difficulties of taking time off from work or away from their families in order to travel to see a specialist, receive treatment at an infusion clinic, or have regular exams and laboratory testing.

In light of these limitations of current therapy, people living with RA rely on caregivers in various ways, depending on a person's ability to cope with activities of daily living and their ability to still be employed. Living with a chronic condition that is potentially debilitating takes an emotional toll on patients, caregivers, and family members that cannot be underscored enough.

4. Expectations About the Drug Being Reviewed

Patients await the arrival of a new drug that would provide a greater reduction in the outcomes of inflammation, pain, and fatigue than the current treatment alternatives and, importantly, that would provide an option if other treatments are exhausted. Patients consider these outcomes the most debilitating symptoms of RA, yet several patients report that the currently available therapies show limited benefits, especially for pain and fatigue. If these outcomes could be improved, patients would expect to see a benefit in terms of their ability to work and be productive at work, as well as their ability to carry out activities of daily living, parenting tasks, and other important social roles. Patients expect that baricitinib may fill this unmet need. However, most of them are concerned about adverse events and serious adverse events. More specifically, patients would like to have treatment options associated with fewer gastrointestinal adverse events and infection rates than the current alternatives. There is also a concern among patients regarding potentially scary serious adverse events. However, one patient mentioned being willing to try baricitinib and take what they perceive as a small risk of a serious adverse event even if it only improves disease activity by 50%, For this patient, such improvement means being able to work longer and also decrease the need for prednisone. The ease of administration of baricitinib was also mentioned as an improvement over injections and infusions, which were reported to be painful and time-consuming. Finally, several patients expressed concerns over the cost of the medication for RA and expect new drugs to be more affordable.

The patient input submissions included four patients who have had experience with baricitinib. One patient had been on this medication for four years at the time she provided input for the submission. All four patients reported positive experiences with baricitinib and perceived the drug to be a real improvement over existing therapy. Patients reported significant relief of RA symptoms and levels of functioning that they described as close to normal, with fewer adverse events than they experienced with other drugs. Patients highlighted the ease of use of baricitinib compared with injections and, most importantly, infusions, which, as one patient mentioned, required taking time away from work.

Appendix 2: Literature Search Strategy

Clinical Literature Search

OVERVIEW	
Interface:	Ovid
Databases:	MEDLINE All (1946–present) Embase (1974–present) **Note:** Subject headings have been customized for each database. Duplicates between databases were removed in Ovid.
Date of Search:	Feb 11, 2019
Alerts:	Bi-weekly search updates until June 19, 2019
Study Types:	No search filters were applied
Limits:	No date or language limits were used Conference abstracts: excluded

SYNTAX GUIDE	
/	At the end of a phrase, searches the phrase as a subject heading
*	Before a word, indicates that the marked subject heading is a primary topic; or, after a word, a truncation symbol (wildcard) to retrieve plurals or varying endings
MeSH	Medical Subject Heading
exp	Explode a subject heading
.ti	Title
.ab	Abstract
.dq	Candidate term word (Embase)
.ot	Original title
adj#	Requires terms to be adjacent to each other within # number of words (in any order)
.hw	Heading word; usually includes subject headings and controlled vocabulary
.kf	Author keyword heading word (MEDLINE)
.kw	Author keyword (Embase)
.nm	Name of substance word
.pt	Publication type
.mp	Mapped term
.rn	Registry number
medall	Ovid database code: MEDLINE All, 1946 to present, updated daily
oemezd	Ovid database code; Embase, 1974 to present, updated daily

MULTI-DATABASE STRATEGY	
1	(baricitinib or olumiant* or "incb 028050" or incb 28050 or incb028050 or incb28050 or ly 3009104 or ly3009104 or ISP4442I3Y).ti,ab,ot,kf,hw,rn,nm.
2	1 use medal
3	*baricitinib/
4	(olumiant* or baricitinib or "incb 028050" or incb 28050 or incb028050 or incb28050 or ly 3009104 or ly3009104).ti,ab,kw,dq.
5	or/3-4
6	5 use oemezd
7	6 not (conference review or conference abstract).pt.
8	2 or 7
9	remove duplicates from 8

OTHER DATABASES		
PubMed	A limited PubMed search was performed to capture records not found in MEDLINE. Same MeSH, keywords, limits, and study types used as per MEDLINE search, with appropriate syntax used.	
Trial registries (ClinicalTrials.gov and others)	Same keywords, limits used as per MEDLINE search.	

Grey Literature

Dates for Search:	February 2019
Keywords:	Olumiant (baricitinib), rheumatoid arthritis
Limits:	No date or language limits used

Relevant websites from the following sections of the CADTH grey literature checklist *Grey Matters: a practical tool for searching health-related grey literature* (https://www.cadth.ca/grey-matters) were searched:

- health technology assessment agencies
- health economics
- clinical practice guidelines
- drug and device regulatory approvals
- advisories and warnings
- drug class reviews
- clinical trial registries
- databases (free)
- Internet search.

Appendix 3: Excluded Studies

Table 9: Excluded Studies

Reference	Reason for Exclusion
No excluded studies	

Appendix 4: Detailed Outcome Data

Table 10: Subgroup Data (Primary Outcome)

	BEACON	
	BAR 2 mg N = 174	PLACEBO N = 176
Clinical Response (ACR20) by		
Inadequate Response to TNF Inhibitor, n/N (%)		
Lack of efficacy	79/161 (49.1)	42/55 (27.1)
AE	5/11 (45.5)	4/16 (25.0)
Other	1/2 (50.0)	2/2 (100)
Number of Previous bDMARD Used, n/N (%)		
< 3	66/124 (53.2)	42/129 (32.6)
Odds ratio [95% CI]	2.36 [1.42 to 3.92]	
≥ 3	19/50 (38.0)	6/47 (12.8)
Odds ratio [95% CI]	4.19 [1.50 to 11.73]	
Interaction *P* value[a]	0.328	

BAR = baricitinib; bDMARD = biologic disease-modifying antirheumatic drug; CI = confidence interval; TNF = tumour necrosis factor.

[a] *P* value from the interaction of subgroup with treatment in the logistic regression model: treatment group + subgroup + treatment – by subgroup. When logistic regression sample size requirements (< 5 responders in any category for any subgroup) are not met, *P* value is not produced. OR and 95% CI within a subgroup obtained from the same model with subgroup and interaction terms removed. When logistic regression sample size requirements are not met, *P* value from Fisher's exact test is produced instead of OR and 95% CI.

Table 11: Other Efficacy Outcomes

	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
Functional/Disability				
Duration of Morning Joint Stiffness (ePRO Diary) Average Across 7 Days Preceding Week 12				
Mean (SD) baseline, minutes	149.1 (165.3) N = 174	131.7 (156.1) N = 176	144.4 (162.2) N = 223	142.4 (169.3) N = 221
Week 12 mean (SD), minutes	100.6 (149.5) N = 172	105.8 (123.0) N = 172	92.6 (132.1) N = 223	128.6 (167.6) N = 221
Median change, baseline to week 12 [95% CI], minutes	–21.0 [–30.0 to –10.0]	–3.5 [–8.0 to 0.0]		
Median week 12 [95% CI], minutes			44.4 [30.0 to 60.0]	60.0 [50.7 to 76.7]
Median difference between groups [95% CI], minutes[c]	15.0 [2.0 to 30.0], *P* = 0.003		–15.7 [–27.9 to –5.6], *P* = 0.002	
Severity of Morning Joint Stiffness NRS (ePRO Diary) Average Across 7 Days Preceding Week 12				
Mean (SD) baseline			5.5 (2.1)	5.5 (2.1)
Week 12 mean (SD)			3.5 (2.5) N = 223	4.2 (2.3) N = 221
LSM			3.5	4.1
LSMD between groups [95% CI][d]			–0.6 [–1.0 to –0.2], *P* = 0.002	

	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
Worst Tiredness NRS (ePRO Diary) Average Across 7 Days Preceding Week 12				
Mean (SD) baseline	7.2 (1.8)	6.9 (2.1)	5.7 (2.3)	5.8 (2.0)
Week 12 mean (SD)	5.4 (2.6) N = 172	6.0 (2.2) N = 172	4.0 (2.5) N = 223	4.5 (2.2) N = 221
LSM	–1.7	–1.0	4.1	4.5
LSMD between groups [95% CI] [d]	–0.7 [–1.2 to –0.2], *P* = 0.003		–0.4 [–0.8 to –0.0], *P* = 0.049	
Worst Joint Pain NRS (ePRO Diary) Average Across 7 Days Preceding Week 12				
Mean (SD) baseline	7.1 (1.7)	7.2 (1.8)	5.9 (2.2)	5.8 (2.0)
Week 12 mean (SD)	5.1 (2.6) N = 172	6.1 (2.3) N = 172	3.9 (2.5) N = 223	4.7 (2.2) N = 221
LSM	–1.9	–1.0	3.8	4.7
LSMD between groups [95% CI] [d]	–0.9 [–1.4 to –0.4], *P* = 0.001		–0.9 [–1.3 to –0.5], *P* = 0.001	
SF-36v2 Acute Physical and Mental Component Summaries				
Mental Component Summary Score				
Mean (SD) baseline	46.1 (13.1)	46.1 (13.7)	45.0 (11.5)	45.7 (11.5)
Week 12 mean (SD) change from baseline	3.4 (9.7) N = 168	1.6 (10.7) N = 168	3.6 (10.5) N = 229	3.3 (10.6) N = 218
LSM	3.0	1.2	3.1	3.2
LSMD [95% CI] [c]	1.9 [–0.1 to 3.8], *P* = 0.058		–0.1 [–2.0 to 1.8], *P* = 0.918	
Week 12, achieving MCID (improvement ≥ 5), n (%)	58 (33.3)	52 (29.5)	88 (38.4)	82 (36.0)
Odds ratio [95% CI] [a]	1.2 [0.8 to 1.9], *P* = 0.451		1.1 [0.8 to 1.7], *P* = 0.559	
Week 24, mean (SD) change from baseline	3.2 (11.5) N = 168	2.5 (10.8) N = 168	3.0 (10.4) N = 229	2.7 (11.5) N = 218
LSM	2.8	1.9	2.5	2.6
LSMD [95% CI] [c]	0.9 [–1.2 to 2.9], *P* = 0.401		–0.1 [–2.1 to 2.0], *P* = 0.955	
Week 24, achieving MCID (improvement ≥ 5), n (%)	45 (25.9)	38 (21.6)	72 (31.4)	64 (28.1)
Odds ratio [95% CI] [a]	1.3 [0.8 to 2.1], *P* = 0.339		1.2 [0.8 to 1.8], *P* = 0.391	
Physical Component Summary Score				
Mean (SD) baseline	28.7 (8.1)	28.2 (7.7)	32.5 (8.4)	32.2 (8.5)
Week 12, mean (SD) change from baseline	6.3 (8.8) N = 168	3.3 (8.0) N = 168	7.7 (8.5) N = 229	4.1 (7.3) N = 218
LSM	6.1	2.7	8.0	4.3
LSMD [95% CI] [c]	3.3 [1.6 to 5.1], *P* = 0.001		3.7 [2.1 to 5.3], *P* = 0.001	
Week 12, achieving MCID, (improvement ≥ 5), n (%)	86 (49.4)	56 (31.8)	130 (56.8)	92 (40.4)
Odds ratio [95% CI] [a]	2.2 [1.4 to 3.3], *P* = 0.001		1.9 [1.3 to 2.8], *P* = 0.001	
Week 24, mean (SD) change from baseline	6.4 (8.9) N = 168	2.4 (8.2) N = 168	8.5 (9.0) N = 229	4.9 (8.0) N = 218
LSM	6.2	1.9	9.0	5.3

	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
LSMD [95% CI] [c]	4.3 [2.6 to 6.1], *P* = 0.001		3.7 [2.0 to 5.4], *P* = 0.001	
Week 24, achieving MCID (improvement ≥ 5), n (%)	68 (39.1)	37 (21.0)	127 (55.5)	77 (33.8)
Odds ratio [95% CI] [a]	2.5 [1.5 to 4.0], *P* = 0.001		2.4 [1.7 to 3.6], *P* = 0.001	
SDAI Change from Baseline (mLOCF)				
Mean (SD) baseline	44.62 (13.58)	42.65 (13.75)	38.32 (13.42)	37.17 (11.94)
Week 12, mean (SD) change from baseline	–18.60 (15.62) N = 169	–11.76 (15.52) N = 170	–20.27 (13.24) N = 224	–13.22 (14.86) N = 218
LSM	–17.61	–11.29	–20.14	–13.73
LSMD [95% CI]	–6.32 [–9.40 to –3.24], *P* = 0.001		–6.41 [–8.68 to –4.15], *P* = 0.001	
Week 24, mean (SD) change from baseline	–17.80 (17.50) N = 169	–12.07 (17.50) N = 170	–21.87 (14.99) N = 224	–14.55 (16.37) N = 218
LSM	–15.89	–10.67	–21.80	–15.17
LSMD [95% CI]	–5.23 [–8.63 to –1.82], *P* = 0.003		–6.63 [–9.15 to –4.11], *P* = 0.001	
SDAI ≤ 3.3 response rate using NRI week 24 n (%)	8 (4.6)	4 (2.3)	38 (16.6)	9 (3.9)
Odds ratio [95% CI]	2.3 [–1.5 to 6.1], *P* = 0.257		12.6 [7.2 to 18.1], *P* = 0.001	
CDAI Change from Baseline (mLOCF)				
Mean (SD) baseline	42.62 (13.08)	40.62 (12.85)	36.50 (13.06)	35.45 (11.74)
Week 12 mean (SD) change from baseline	–18.03 (15.04) N = 169	–11.75 (15.25) N = 170	–19.40 (12.98) N = 224	–13.20 (14.53) N = 218
LSM	–17.13	–11.36	–19.40	–13.78
LSMD [95% CI]	–5.77 [–8.74 to –2.79], *P* = 0.001		–5.61 [–7.82 to –3.41], *P* = 0.001	
Week 24, mean (SD) change from baseline	–17.17 (16.96)	–12.19 (16.96)	–20.99 (14.48)	–14.29 (16.04)
LSM	–15.42	–10.96	–21.07	–15.01
LSMD [95% CI]	–4.46 [–7.75 to –1.17], *P* = 0.009		–6.05 [–8.48 to –3.62], *P* = 0.001	
CDAI ≤ 2.8 response rate, NRI, week 12, n (%)	5 (2.9)	3 (1.7)	23 (10.0)	4 (1.8)
Difference in response rate [95% CI]	1.2 [–2.0 to 4.3], *P* = 0.501		8.3 [4.0 to 12.5], *P* = 0.001	
Week 24 n (%)	8 (4.6)	6 (3.4)	35 (15.3)	9 (3.9)
Difference in response rate [95% CI]	1.2 [–2.9 to 5.3], *P* = 0.599		11.3 [6.0 to 16.6], *P* = 0.001	
ACR/EULAR Boolean Remission, Response Week 12, n (%)	4 (2.3)	4 (2.3)	16 (7.0)	1 (0.4)
Difference in response rate [95% CI]	0.0 [–3.1 to 3.2], *P* = 1.000		6.5 [3.1 to 10.0], *P* = 0.001	
Week 24, n (%)	7 (4.0)	2 (1.1)	29 (12.7)	8 (3.5)
Difference in response rate [95% CI]	2.9 [–0.4 to 6.2], *P* = 0.104		9.2 [4.2 to 14.1], *P* = 0.001	
EQ-5D-5L Health State Index/Self-Perceived Health Score				
Mean (SD) baseline (US algorithm)	0.606 (0.159)	0.595 (0.168)	0.637 (0.169)	0.659 (0.145)
Mean (SD) change from baseline to week 12	0.080 (0.152) N = 168	0.035 (0.167) N = 168	0.117 (0.151) N = 227	0.054 (0.155) N = 216
LSM	0.079	0.026	0.117	0.066

	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
LSMD [95% CI]	0.052 [0.022 to 0.083], *P* = 0.001		0.051 [0.027 to 0.075], *P* = 0.001	
Week 24, change from baseline	0.082 (0.170) N = 168	0.042 (0.166) N = 167	0.113 (0.172) N = 227	0.051 (0.149) N = 216
LSM	0.074	0.025	0.111	0.062
LSMD [95% CI]	0.049 [0.018 to 0.081], *P* = 0.003		0.049 [0.023 to 0.075], *P* = 0.001	
Health state index score (UK algorithm)				
Mean (SD) baseline	0.461 (0.233)	0.443 (0.250)	0.507 (0.249)	0.543 (0.214)
Mean (SD) change from baseline to week 12	0.116 (0.224) N = 168	0.052 (0.250) N = 168	0.167 (0.221) N = 227	0.074 (0.230) N = 216
LSM	0.114	0.036	0.165	0.092
LSMD [95% CI]	0.077 [0.033 to 0.121], *P* = 0.001		0.073 [0.038 to 0.108], *P* = 0.001	
Mean (SD) change from baseline to week 24	0.122 (0.250) N = 168	0.064 (0.245) N = 167	0.162 (0.254) N = 227	0.075 (0.218) N = 216
LSM	0.111	0.038	0.157	0.091
LSMD [95% CI]	0.073 [0.027 to 0.119], *P* = 0.002		0.066 [0.030 to 0.103], *P* = 0.001	
Self-Perceived Health Score (0 mm to 100 mm)				
Mean (SD) baseline	46.0 (20.8)	47.8 (22.4)	53.1 (20.5)	51.6 (19.7)
Mean (SD) change from baseline to week 12	14.1 (24.2) N = 168	4.1 (29.0) N = 168	13.4 (21.8) N = 227	5.7 (23.8) N = 216
LSM	11.3	2.9	13.5	4.5
LSMD [95% CI]	8.5 [3.6 to 13.3], *P* = 0.001		9.0 [5.2 to 12.7], *P* = 0.001	
Mean (SD) change from baseline to week 24	11.4 (26.5) N = 168	3.8 (27.8) N = 167	13.1 (25.8) N = 227	8.4 (25.1) N = 216
LSM	7.9	1.9	13.9	7.9
LSMD [95% CI]	6.0 [0.9 to 11.1], *P* = 0.022		6.0 [1.8 to 10.2], *P* = 0.005	
FACIT-F Change From Baseline				
Mean (SD) baseline	22.5 (10.0)	22.2 (10.6)	26.6 (11.5)	26.6 (11.1)
Week 12, mean (SD) change from baseline	8.8 (10.0) N = 170	5.9 (10.5) N = 170	8.7 (11.1) N = 227	7.6 (10.3) N = 216
LSM	8.3	5.2	8.5	7.5
LSMD [95% CI]	3.1 [1.0 to 5.1], *P* = 0.004	1.0 [–0.7 to 2.6], *P* = 0.247		
Week 12, achieving MCID (improvement ≥ 3.56)	111 (63.8)	85 (48.3)	145 (63.3)	134 (58.8)
Difference in MCID response rate [95% CI]	15.5 [5.2 to 25.8]	4.5 [–4.4 to 13.5]		
Odds ratio [95% CI]	1.9 [1.2 to 2.9], *P* = 0.004		1.21 [0.83 to 1.77], *P* = 0.323	
Week 24, change from baseline	8.8 (10.4)	6.6 (10.7)	9.2 (10.7)	7.8 (11.0)
LSM	8.1	5.7	9.2	7.9
LSMD [95% CI]	2.4 [0.3 to 4.6], *P* = 0.026		1.4 [–0.3 to 3.1], *P* = 0.117	
Week 24, achieving MCID (improvement ≥ 3.56)	87 (50.0)	66 (37.5)	135 (59.0)	97 (42.5)

	BEACON		BUILD	
	BAR 2 mg N = 174	PLACEBO N = 176	BAR 2 mg N = 229	PLACEBO N = 228
Difference in MCID response rate [95% CI]	12.5 [2.2 to 22.8]	16.4 [7.4 to 25.5]		
Odds ratio [95% CI]	1.7 [1.1 to 2.7], *P* = 0.015		1.95 [1.34 to 2.83], *P* = 0.001	
Radiology				
mTSS mean (SD) baseline using linear extrapolation	NR	NR	25.78 (40.26) N = 212	18.54 (31.47) N = 197
mTSS mean (SD) change from baseline, week 24	NR	NR	0.43 (1.19) N = 208	0.80 (2.86) N = 190
LSM [95% CI]	NR	NR	0.33 [0.06 to 0.59]	0.70 [0.42 to 0.98]
LSMD [95% CI]	NR	NR	–0.38 [–0.74 to –0.01], *P* = 0.043	

ACR = American College of Rheumatology; BAR = baricitinib; CDAI = Clinical Disease Activity Index; CI = confidence interval; EQ-5D-5L; European Quality of Life–5 Dimensions–5 Levels; EULAR = European League Against Rheumatism; FACIT-F = Functional Assessment of Chronic Illness Therapy–Fatigue; LSM = least squares mean; LSMD = least squares mean difference; MCID = minimal clinically important difference; mLOCF = modified last observation carried forward; mTSS = modified total Sharp score; NR = not reported; NRI = nonresponder imputation; NRS = numeric rating scale; SD = standard deviation; SDAI = Simplified Disease Activity Index; SF-36v2 = Short Form (36) Health Survey Version 2 Acute.

Appendix 5: Validity of Outcome Measures

Aim

To summarize the validity of the following outcome measures:

- American College of Rheumatology (ACR) response criteria ACR20, ACR50, and ACR70
- Health Assessment Questionnaire (HAQ) and Disability Index (HAQ-DI)
- Short Form (36) Health Survey (SF-36)
- Disease Activity Score 28 (DAS28)
- Simplified Disease Activity Index (SDAI)
- Clinical Disease Activity Index (CDAI)
- modified total Sharp score (mTSS)
- Functional Assessment of Chronic Illness Therapy (FACIT).

Findings

ACR criteria, HAQ-DI, SF-36, DAS28, SDAI, CDAI, mTSS, and FACIT-Fatigue are briefly summarized in Table 12.

Table 12: Validity and Minimal Clinically Important Difference of Outcome Measures

Instrument	Type	Evidence of Validity	MCID	References
ACR20 ACR50 ACR70	ACR20, ACR50, and ACR70 responses represent at least a 20%, 50%, and 70% improvement, respectively, in tender and swollen joint counts and in three of the five additional criteria: • Patient global assessment of disease activity • Physician global assessment of disease activity • Patient assessment of pain • Health Assessment Questionnaire • CRP or ESR.	Yes	ACR50	van Riel and van Gestel (2000)[23] Cohen et al. (2006)[24] Bansback et al. (2008)[25] ACR criteria (2007)[26] Chung et al. (2006)[17]
HAQ-DI	The HAQ–Disability Index (HAQ-DI) is the disability assessment component of the HAQ.	Yes	0.22 points	Bruce and Fries (2003)[2,3]
SF-36	The SF-36 consists of eight subdomains. The SF-36 provides two component summaries, PCS and MCS. The eight subdomains are each measured on a scale of zero to 100, with an increase in score indicating improvement in health status.	Yes	2.5 to 5.0	Gallagher et al. (2001)[27] Hays and Morales (2001)[28] Samsa et al. (1999)[29] Strand and Singh (2008)[30]
DAS28	DAS28 is an abbreviated version of the DAS, based on a 28-joint count that omits the feet and ankle joints.	Yes	Unspecified	Wells et al. (2009)[10] Crowson et al. (2009)[31]

Instrument	Type	Evidence of Validity	MCID	References
SDAI	The SDAI integrates measures of physical examination, acute-phase response, patient self-assessment, and evaluator assessment in order to simplify the assessment of disease activity in clinical practice.	Yes	Unspecified	Aletaha et al. (2005)[15] Fujiwara et al. (2013)[13] Gaujoux-Viala et al. (2012)[14] Smolen et al. (2014)[16]
CDAI	The CDAI is similar to the SDAI, but it allows for immediate scoring because it does not include a laboratory result.	Yes	Unspecified	Aletaha et al. (2005)[15] Fujiwara et al. (2013)[13] Gaujoux-Viala et al. (2012)[14] Smolen et al. (2014)[16]
mTSS	The mTSS is a composite measure of joint erosion and joint-space narrowing based on radiographic assessment.	Yes	4.6 units	Bruynesteyn et al. (2002)[32]
FACIT-Fatigue	The FACIT-Fatigue scale is a 13-item self-report measure of fatigue.	Yes	3 to 4 points	Cella et al. (2005)[33]

ACR = American College of Rheumatology; CDAI = Clinical Disease Activity Index; CRP = C-reactive protein; DAS = Disease Activity Score; ESR = erythrocyte sedimentation rate; FACIT = Functional Assessment of Chronic Illness Therapy; HAQ = Health Assessment Questionnaire; HAQ-DI = HAQ–Disability Index; MCID = minimal clinically important difference; MCS = mental component summary; mTSS = modified total Sharp score; PCS = physical component summary; SDAI = Simplified Disease Activity Index; SF-36 = Short Form (36) Health Survey.

American College of Rheumatology Response Criteria

The ACR criteria for assessing joint status was initially developed for patients with rheumatoid arthritis (RA).[23] ACR criteria provide a composite measure of improvement in both swollen and tender joint counts and at least three of five additional disease criteria:

- patient global assessment of disease activity
- physician global assessment of disease activity
- patient assessment of pain
- HAQ
- levels of either C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR).

The ACR joint count for RA assesses 68 joints for tenderness and 66 joints for swelling. Patient and physician assessments are conducted using visual analogue scale (VAS) or Likert scale measurements. ACR20, ACR50, or ACR70 responses represent at least a 20%, 50%, or 70% improvement, respectively, in tender and swollen joint counts as well as in three of the five additional disease criteria. This core set of measures included in the ACR response criteria was established through a consensus process of clinical experts. Individual criteria were selected based on their construct validity, face validity, content validity, criterion validity, and discriminant validity.[34] In the assessment of criterion validity, standards for comparison included death, physical disability, and radiologic evidence of joint damage. Physical functioning capacity, as measured by the HAQ, was considered a strong predictor of mortality, and many other risk factors for premature mortality were insignificant after adjusting for functional capacity. Predictors of radiographic progression included swollen joint counts and levels of acute-phase reactants such as ESR and CRP.[34] When considering the ability of an outcome measure to detect change, pain assessments, global assessments, tender joint counts, and HAQ scores all had strong discriminant validity.

The ACR20 is most commonly used as the primary end point in randomized controlled trials (RCTs) evaluating biologics in the treatment of RA. The FDA considers ACR20 a well-validated composite end point for assessing the signs and symptoms of RA, as noted in guidance provided to industry on the conduct of trials in patients with RA.[35] ACR50 and ACR70 are often reported in clinical trials and are considered more stringent outcome measures.

Chung et al.[17] conducted a meta-analysis of 21 RCTs of RA therapies published between 1997 and 2004 to compare the discriminant capabilities of the ACR50 and ACR20 responses and to determine whether ACR50 is as informative as ACR20 in distinguishing between active therapies and control groups. While both measures have the ability to distinguish an active therapy compared with a control therapy, the levels of improvement captured by ACR20 response do not generally represent an optimal clinical improvement. Furthermore, since the development of the ACR20 response criteria, much more aggressive therapies have been introduced in the treatment of RA, and larger clinical responses can be expected. This meta-analysis concluded that ACR20 and ACR50 are similar in distinguishing between active and control therapies but that ACR50 represents a more robust clinical response and may be a preferred end point in clinical trials.[17]

ACR70 is considered even more rigorous than ACR50. It is a component of the definitions established by the FDA in order to satisfy labelling requirements for RA drugs. Specifically, a “major clinical response,” as defined by the FDA, refers to a statistically significant increase in the proportion of patients achieving an ACR70 response, maintained over six months, with active therapy compared with a control group.[35]

With widespread use of the ACR criteria over the past 20 years, limitations associated with them have been identified. For example, while ACR response indicates the change from baseline, it does not indicate the final level of disease severity that the patient attains. This limitation also means that patients who are classified as ACR responders could have very different levels of disease.[25] Other criticisms of the ACR criteria are that most of its component measures are subjective, that dichotomous measures such as ACR lack sensitivity to change compared with continuous measures of response, and that the ACR20 response threshold is too low, relative to treatment goals applied in clinical practice.[26] In response to these criticisms, attempts have been made to develop improved outcome measures for RA, although none have widespread acceptance or are consistently used in clinical trials.[26,36]

Health Assessment Questionnaire and Disability Index

The HAQ was originally developed in 1978 at Stanford University.[37] It was one of the first self-reported functional status (disability) measures and has become the dominant instrument in many disease areas, including arthritis.[38] The HAQ has been widely validated in patients with RA.[38] The full HAQ collects data on five generic, patient-centred health dimensions: 1) to avoid disability, 2) to be free of pain and discomfort, 3) to avoid adverse treatment effects, 4) to keep dollar costs of treatment low, and 5) to postpone death.[12]

The HAQ–Disability Index (HAQ-DI) is the disability assessment component of the HAQ. It assesses a patient’s level of functional ability. There are 20 questions in eight categories to assess a patient’s physical functional status: dressing, arising, eating, walking, hygiene, reach, grip, and common activities.[2,3] For each of these categories, patients report the amount of difficulty they have in performing specific activities, and their responses are made on a scale from zero (no difficulty) to three (unable to do). The eight category scores

are averaged into an overall HAQ-DI score on a scale from zero (no disability) to three (completely disabled). Observational studies and RCTs have demonstrated that the HAQ-DI possesses face validity, content validity, construct validity, predictive validity, and discriminant validity. There is evidence suggesting that baseline HAQ scores are predictive of radiographic damage, work disability, and quality of life.[24,39] A number of investigators have suggested that the minimal clinically important difference (MCID) is 0.22; however, differences as small as 0.10 have been suggested as clinically important.[2]

Short Form (36) Health Survey

The SF-36 is a generic health assessment questionnaire that has been used in clinical trials to study the impact of chronic disease on health-related quality of life. The SF-36 consists of eight subdomains: physical functioning, pain, vitality, social functioning, psychological functioning, general health perceptions, and role limitations due to physical and emotional problems.[27] The SF-36 also provides two component summaries, the physical component summary (PCS) and the mental component summary (MCS). The eight subdomains are each measured on a scale of zero to 100, with an increase in score indicating improvement in health status. The MCID for either the PCS or MCS of the SF-36 is typically between 2.5 and 5 points.[28-30]

Disease Activity Score 28

The DAS is a measure of RA disease activity and includes the Ritchie Articular Index (0 to 78), which is performed on 53 joints; a 44-joint swollen joint count (0 to 44); ESR or CRP; and a general health item using a VAS (0 to 100).[40] DAS28 is an abbreviated version of the DAS, based on a 28-joint count that omits the feet and ankle joints. Thus, one obvious criticism of this scale is that a patient who had inflammation only of the feet and ankles would be counted as in remission.[41] The DAS components correlate well with each other and with the ACR criteria.[42-45] The DAS28 is a composite score derived using the following formula:

$$DAS28 = 0.56 \times \sqrt{(t28)} + 0.28 \times \sqrt{(sw28)} + 0.70 \times ln(ESR) + 0.014 \times GH$$

where DAS28 = Disease Activity Score–28, ESR = erythrocyte sedimentation rate, ln(ESR) = natural logarithm of ESR value, sw28 = swollen joint count of 28 joints, t28 = tender joint count of 28 joints, GH = general health measured by Patient's Global Assessment of Disease Activity on a VAS of 100 mm.

The formula was developed by comparing serial assessments of tender and swollen joint counts, ESR, and patient global assessment (global health) for a panel of patients with RA at times of poorly controlled RA and of well-controlled RA.[46] A DAS28 score indicates an absolute level of disease activity, with a score of greater than 5.1 being considered high disease activity, while a DAS28 score lower than 3.2 indicates low disease activity state (LDAS) and a DAS28 score lower than 2.6 indicates remission.[10-12]

In recent years, CRP has been used in place of ESR to calculate the DAS28. The trend of using CRP levels as opposed to ESR is mainly driven by greater availability, reduced cost, and increased sensitivity of CRP to short-term changes in disease activity.[10,31] The formula used to calculate the DAS28(CRP) is as follows:

$$DAS28(CRP) = 0.56 \times \sqrt{(t28)} + 0.28 \times \sqrt{(sw28)} + 0.014 \times GH + 0.36 \times in(CRP + 1) + 0.96$$

where CRP = C-reactive protein, DAS28 = Disease Activity Score–28, ln(CRP + 1) = natural logarithm of (CRP value +1), sw28 = swollen joint count of 28 joints, t28 = tender joint count of 28 joints, GH = general health measured by Patient's Global Assessment of Disease Activity on a VAS of 100 mm.

A DAS28 score indicates an absolute level of disease activity, with a score greater than 5.1 or greater being considered high disease activity, while a DAS28 score lower than 3.2 indicates LDAS and a DAS28 score lower than 2.6 indicates clinical remission.[10-12] Overall, the DAS28-CRP correlates well with DAS28-ESR, and both are validated measures for assessing disease activity in RA.[10,11,47-49] However, studies have shown that the DAS28-CRP score value is usually lower than the DAS28-ESR score.[11,47-53] The difference (DAS28-CRP minus DAS28-ESR) ranges from −0.2[47] to −0.8.[50] Because the definitions of remission (score lower than 2.6) are the same for both DAS28-CRP and DAS28-ESR, it was concluded that DAS28-CRP underestimates disease activity and overestimates the improvement in disease activity and the remission rate compared with DAS28-ESR. It was also suggested that DAS28-CRP should be evaluated using different criteria from those for DAS28-ESR.[49] Furthermore, the European League Against Rheumatism has recommended that the clinical implications of the DAS28 score (such as good response, moderate response, or no response) should be determined based on the baseline DAS28 scores (Table 13).[54] Finally, there is no MCID for change in DAS28 scores.

Table 13: European League Against Rheumatism Improvement Response Criteria (Disease Activity Score 28)

Baseline DAS28 Score	DAS28 Improvement Over Time Points		
	> 1.2	0.6 to 1.2	< 0.6
< 3.2	Good response	Moderate response	No response
3.2 to 5.1	Moderate response	Moderate response	No response
> 5.1	Moderate response	No response	No response

DAS28 = Disease Activity Score 28 items.

Source: Matsui et al. (2007).[49]

Simplified Disease Activity Index and the Clinical Disease Activity Index

The SDAI is a tool for measuring disease activity that integrates measures of physical examination, acute-phase response, patient self-assessment, and evaluator assessment. [4,5] It was originally developed to simplify the assessment of disease activity in clinical practice.[15]

The SDAI is calculated by a simple numerical addition of the scores from the five following assessments:

- number of tender joints (0 to 28)

- number of swollen joints (0 to 28)
- C-reactive protein in mg/dL (0.1 to 10.0)
- Patient Global Assessment of Disease Activity VAS (0 to 10.0 cm)
- Physician Global Assessment of Disease Activity VAS (0 to 10.0 cm).

The CDAI is similar to the SDAI, but it allows for immediate scoring because it does not include a laboratory result. [4,5] Therefore, CDAI is calculated by adding the scores from the four following assessments:

- number of tender joints (0 to 28)
- number of swollen joints (0 to 28)
- Patient Global Assessment of Disease Activity VAS (0 to 10.0 cm)
- Physician Global Assessment of Disease Activity VAS (0 to 10.0 cm).

Both the SDAI and CDAI have been validated and show correlation with one another as well as with the DAS28.[13-15] Disease remission is defined as an SDAI score of 3.3 or less and as a CDAI score of 2.8 or less.[15,16]

Modified Total Sharp Score

The Sharp scoring system, first developed in 1971, has undergone modifications over time and is now referred to as the modified total Sharp score. This method allows for the assessment of two different aspects of joint damage: articular erosions (representing direct invasion of cartilage and bone by the proliferating synovial pannus) and joint-space narrowing (representing destruction of surface cartilage). Data on the progression of joint structural damage are obtained by taking x-rays of specific joints (typically in the hands and feet) before treatment and at various points after treatment has been initiated. The most recent modification of the Sharp scoring system is that of van der Heijde, summarized in the following table.[55]

Sharp/van der Heijde Scoring System[56]	
Erosions	
0	Normal
1	Discrete erosions
2 to 3	Larger erosions according to surface area involved
4	Erosion extending over the middle of the bone
5	Complete collapse
Joint-Space Narrowing	
0	Intact bony outlines and normal joint space
1	Erosion < 1 mm in diameter or JSN
2	One or several small erosions (diameter > 1 mm)
3	Marked erosions
4	Severe erosions (usually no joint space left and the original bony outlines are only partly preserved)
5	Mutilating changes (the original bony outlines have been destroyed)

The van der Heijde erosion score includes 16 joints from the hands and wrists (graded from zero to five) and six joints from the feet (graded from zero to 10). The joint-space narrowing score includes 15 areas from the hands and wrists (graded from zero to four) and six areas

from the feet (also graded from zero to four). The maximum erosion score is 160 for hands and wrists and 120 for feet, while the maximum joint-space narrowing score is 120 for hands and 48 for feet.[57] Maximum total scores for both erosion and joint-space narrowing are shown below:

Erosion = (32 joints in hands and wrists × 5) + (12 joints in feet × 10) = 280

Joint-space narrowing = (30 joints in hands and wrists × 4) + (12 joints in feet × 4) = 168

The van der Heijde modification has become the most commonly used for a few reasons:

1. It includes both hands and feet.
2. It measures erosions and joint-space narrowing.
3. It covers a broad spectrum of joints, providing sensitivity to change.[58]

In the early stages of RA, inflammation, rather than actual damage to joints, appears to be the main contributor to increased disability.[59,60] The relationship between radiological and functional changes has been studied. A reanalysis of published data performed by Welsing et al. found that patients must reach a certain amount of radiological damage before an increase in damage will impact disability. The authors also found that changes in Sharp scores had a greater impact on disability with advancing age. A study by Sabin et al. found that radiologic damage assessed by the van der Heijde method was highly correlated with HAQ scores in a population with a mean disease duration of seven years. They also cited findings from another study, which found that Sharp scores became correlated with HAQ after six years' disease duration. At the other end of the spectrum, a study by Clarke et al. found that radiological scores assessed using the Genant method were positively correlated with HAQ in patients with 20 years' disease duration.[61] Therefore, radiological changes, assessed by Sharp scores, and functional changes, assessed by the HAQ, are not correlated with each other early in RA, but are correlated after several years of disease.

There are several limitations with using radiographs to assess clinical status in RA. Radiographs tend to change slowly in RA, requiring at least six months to a year to detect changes in a single patient. Inter-rater and intra-rater reliability is also a concern due to the subtle nature of changes and subjective interpretation. The images themselves can also vary between samples, due to positioning and quality. Radiographs should be read in random order to reduce the potential bias of interpretation at different time points.[62] Given these limitations, beginning in the early 1990s, the use of magnetic resonance imaging (MRI) was being examined as an alternative for assessing disease progression.[63] However, the use of MRI to assess clinical status of RA is limited by its cost and accessibility.

In a study by Bruynesteyn et al., the authors determined an MCID of 4.6 units for the Sharp/van der Heijde method, using a panel of experts.[32] They defined the MCID as a progression in radiologic joint damage that leads a rheumatologist to change the patient's therapy. This MCID was equal to, or slightly lower than, the smallest detectable difference (SDD) for this scoring system. The SDD represents the smallest change score that can be reliably discriminated from the measurement error of the scoring method.[64] The smallest detectable change (SDC) score is another method of measuring reliability. Similar to the MCID, the SDC score can provide guidance for interpreting whether there has been a real change in patient outcomes over time. A study by Navarro-Compan et al. assessed the level of agreement between two readers on radiographic images from patients with RA.[65] The authors found a SDC of 3.1 (range 2.3 to 4.3) using the 95% level of agreement

method and suggested that a score of 3.0 units is a reasonable cut-off for interpreting radiographic progression as clinically meaningful.[65]

Functional Assessment of Chronic Illness Therapy–Fatigue

The FACIT-Fatigue scale was originally developed for use in patients with cancer. It is one of a series of symptom subscales in the FACIT measurement system and has since been validated for use in patients with RA.[33]

FACIT-Fatigue is a patient self-report measure consisting of 13 statements. Patients are asked to indicate to what extent the statement applies to them over the course of the previous seven days. Each statement has five possible levels of response, scored on a scale of 0 to 4 (0 representing “not at all” and four representing “very much”), resulting in scores ranging from 0 to 52. Lower scores indicate higher levels of fatigue. A suggested MCID for the FACIT-Fatigue in patients with RA is between three and four points.[33] This MCID was found in a sample of 271 patients (77% female, 81% white, median age of 56 years [range 28 to 84]), with a median tender joint count of 26 (range 9 to 68) and a median swollen joint count of 15 (range 2 to 43).[33]

Conclusion

ACR response, HAQ-DI, SF-36, DAS28, SDAI, CDAI, mTSS, and FACIT-Fatigue were used as efficacy measures in RA trials. The ACR20, ACR50, and ACR70 indicate a percentage improvement from baseline (but not a final level of disease activity). ACR20 is most commonly reported in clinical trials; however, ACR50 or ACR70 are often cited as evidence of a more robust treatment effect. The HAQ is a comprehensive measure of the patient’s perception of functional status and has been widely validated in RA. The HAQ-DI (the disability component) is one of five components of the full HAQ. A suggested MCID in patients with RA is 0.22; however, differences as small as 0.10 have also been suggested. SF-36 is a generic health assessment questionnaire that consists of eight subdomains[27] but also provides two component summaries, the PCS and the MCS. The MCID is typically between 2.5 and 5 points.[28-30] The DAS28 measures an absolute rather than relative level of disease activity, and its components correlate well with each other and with the ACR components. However, it was reported that DAS28-CRP overestimates the improvement in disease activity and the remission rate compared with DAS28-ESR. The MCID for a change in DAS28 scores has not been specified. The SDAI and CDAI simplify the assessment of disease activity in clinical practice, as they are calculated by simple numerical addition. In addition, CDAI allows for immediate scoring because it does not include a laboratory result. Disease remission is defined as an SDAI score of 3.3 or less and as a CDAI score of 2.8 or less. The mTSS allows for the assessment of two different aspects of joint damage in the hands, wrists, and feet: articular erosions (representing direct invasion of cartilage and bone by the proliferating synovial pannus) and joint-space narrowing (representing destruction of surface cartilage). Some limitations of the mTSS include the time it takes for changes to appear on the radiographic image, inter-rater and intra-rater reliability, and the variability in images between samples, due to positioning and quality. An MCID of 4.6 units on the mTSS has been suggested. The FACIT-Fatigue is a self-report measure of fatigue that has been validated for use in patients with RA. The MCID for the FACIT-Fatigue scale has been cited as a three- to four-point change in score.

Appendix 6: Summary of a Long-Term Extension Study

The objective of this appendix is to summarize the clinical efficacy and harms of an extension study assessing the long-term efficacy and safety of baricitinib for the treatment of rheumatoid arthritis (RA). BEYOND (N = 3,073)[66] is an ongoing, manufacturer-sponsored, double-blind, partly randomized long-term extension study evaluating the long-term safety and tolerability of baricitinib 2 mg and 4 mg, once daily, in patients who had completed a previous baricitinib study, including BUILD and BEACON. Patients who received baricitinib 2 mg in BEACON or BUILD and who completed final active treatment continued to receive the 2 mg dose in BEYOND (N = 297). These patients were allowed to receive baricitinib 4 mg as rescue therapy if they had a Clinical Disease Activity Index (CDAI) score greater than 10 at or after three months following enrolment in BEYOND. Other patients received baricitinib 4 mg upon study entry, a dose that exceeds the Health Canada recommendation for baricitinib. Among these patients, those who met predefined criteria for sustained low disease activity or remission were randomized following a 1:1 ratio allocation to continue baricitinib 4 mg or to receive 2 mg in a blinded fashion. This appendix will present results for patients who continuously received the recommended dose of baricitinib 2 mg in the original studies and in the long-term extension study. BEYOND is currently ongoing; the data cut-off date for the interim analysis was August 10, 2015. At that point, no patient had yet completed the 48-month treatment period.

Details of the BEYOND long-term extension study are presented in Table 14.

One limitation of the BEYOND study was the fact that a majority of patients received baricitinib 4 mg. This represents a significant issue, as this dose exceeds the Health Canada recommendation for baricitinib. Only patients who received baricitinib 2 mg in BEACON and BUILD and who completed final active treatment continued to receive 2 mg in BEYOND (N = 297). [redacted]. Complete patient disposition is presented in Table 15. Although the level of discontinuations was expected, it limits the interpretation of the findings. In addition, the information presented in this appendix was obtained from an interim report, as the BEYOND study is currently ongoing. Therefore, the findings should be interpreted with caution. At the time of data cut-off, no patient had yet completed the 48-month treatment period.

Table 14: Details of BEYOND Long-Term Extension Study

		BEYOND
Designs and Populations	**Study Design**	Ongoing, multi-centre, DB, partly randomized, long-term extension study
	Total number of patients (N)	N = 3,073
	Patients receiving baricitinib 2 mg at study entry (N)	N = 297 Patients who received baricitinib 2 mg once daily in the original study continued to receive 2 mg once daily in BEYOND.
	Inclusion Criteria	Patients who completed final active treatment in studies BEAM, BEGIN, **BUILD, BEACON,** JADA, or JAGS
	Exclusion Criteria	Uncontrolled cerebrovascular, hepatic, renal, hematologic, or abnormal laboratory values, or permanently discontinued medication, during original study
Drugs	**Intervention**	Baricitinib 2 mg orally once daily through 48 months; or *Baricitinib 4 mg (exceeds HC-recommended dose), followed by step-down*[a] Background cDMARD, NSAID, corticosteroid, and/or other analgesics received at completion of original study
Duration	**Treatment Period**	Up to 48 months from enrolment in BEYOND *The BEYOND study is currently ongoing. The data cut-off date for the interim analysis was August 10, 2015. At that point, no patient had yet completed the 48-month treatment period.*
Outcomes	**Primary Objective**	Long-term safety and tolerability of baricitinib in patients who have completed a previous baricitinib rheumatoid arthritis study
	Outcomes	Efficacy: ACR20/50/70; SDAI and CDAI; HAQ-DI; DAS28; ACR/EULAR remission Safety: AEs and SAEs

ACR = American College of Rheumatology; AE = adverse event; CDAI = Clinical Disease Activity Index; cDMARD = conventional disease-modifying antirheumatic drug; DAS28 = Disease Activity Score 28; DB = double-blind; EULAR = European League Against Rheumatism; HAQ-DI = Health Assessment Questionnaire–Disability Index; HC = Health Canada; NSAID = nonsteroidal anti-inflammatory drug; SAE = serious adverse event; SDAI = Simplified Disease Activity Index.

[a] Patients who achieved a sustained low disease activity level (CDAI score ≤ 10) or sustained remission for patients who began in study JADZ (CDAI score ≤ 2.8), i.e., who maintain low disease activity or remission for at least three months in BEYOND, were randomized following a 1:1 ratio allocation to continue baricitinib 4 mg once daily or receive the 2 mg once-daily dose in a blinded fashion.

Source: BEYOND interim Clinical Study Report.[66]

Table 15: Patient Disposition

	BEYOND
Enrolled, N	[redacted]
Baricitinib 4 mg	[redacted]
Baricitinib 2 mg	[redacted]
From BUILD	[redacted]
From BEACON	[redacted]
Baricitinib 2 mg	
Enrolled	[redacted]
Rescued to baricitinib 4 mg	[redacted]
Not rescued	[redacted]
Discontinued	[redacted]
Most frequent reasons for discontinuation:	
Adverse event	[redacted]

	BEYOND
Lack of efficacy	[redacted]
Physician decision	[redacted]
Lost to follow-up	[redacted]
Death	[redacted]
Withdrawal by patient	[redacted]
Continuing in study	[redacted]

Source: BEYOND interim Clinical Study Report.[66]

Results

Baseline characteristics were not reported separately for patients previously in the BEACON or BUILD studies and receiving baricitinib 2 mg once daily. Therefore, no baseline characteristics were available to present in this appendix.

Efficacy outcomes were assessed using nonresponder imputation (American College of Rheumatology response rates) or last observation carried forward (change from baseline in Simplified Disease Activity Index [SDAI] and CDAI), without considering rescue status, from week 24 through the last available visit in the modified intention-to-treat population.

Because of the lack of control group in the BEYOND study, the results presented here are descriptive in nature and should be interpreted as such. [redacted]. Two of the most relevant health-related quality of life measures assessed in BEYOND were the SDAI and CDAI. [redacted]. An overview of efficacy is presented in Table 16.

Table 16: Summary of Results — Efficacy

Outcome	BEYOND	
	BAR 2 mg Patients Originating From BEACON (N = [redacted])	**BAR 2 mg Patients Originating From BUILD (N = [redacted])**
Clinical Response (ACR) **Response at Week 72, NRI**		
ACR20 n (%) [95% CI]	[redacted]	[redacted]
ACR50 n (%) [95% CI]	[redacted]	[redacted]
ACR70 n (%) [95% CI]	[redacted]	[redacted]
CDAI **Change From Baseline to Week 72, LOCF**		
Mean (SD) at baseline (week 24)	[redacted]	[redacted]

Outcome	BEYOND	
	BAR 2 mg Patients Originating From BEACON (N = [redacted])	BAR 2 mg Patients Originating From BUILD (N = [redacted])
Week 72, change from baseline, mean (SD)	[redacted]	[redacted]
[95% CI], *P* value	[redacted]	[redacted]
SDAI **Change From Baseline to Week 72, LOCF**		
Mean (SD) at baseline (week 24)	[redacted]	[redacted]
Week 72, change from baseline, mean (SD)	[redacted]	[redacted]
[95% CI], *P* value	[redacted]	[redacted]

ACR = American College of Rheumatology; BAR = baricitinib; CDAI = Clinical Disease Activity Index; CI = confidence interval; LOCF = last observation carried forward; NRI = nonresponder imputation; SD = standard deviation; SDAI = Simplified Disease Activity Index.

Source: BEYOND interim Clinical Study Report.[66]

Harms evaluations in the BEYOND study were performed on the evaluable safety data set, which was composed of all randomized patients, with data up to rescue. Harms outcomes were reported for patients receiving all doses of baricitinib combined, which limits the interpretation of the harms results. Nevertheless, an overview of the most relevant harms findings in patients originating from BEACON and BUILD is presented in Table 17.

The inclusion of patients receiving a higher dose of baricitinib is likely to bias the results against baricitinib.

Overall, harms results were consistent with the findings from BEACON and BUILD. There were a greater number of cases of herpes zoster in the BEYOND study than in patients receiving baricitinib 2 mg in BEACON and BUILD. The inclusion of patients receiving baricitinib 4 mg is likely to contribute to these findings.

Table 17: Summary of Results — Harms

Outcome	BEYOND	
	All Patients Originating From BEACON (N = [redacted])	All Patients Originating From BUILD (N = [redacted])
Key Harms Outcomes **Harms outcomes were reported for the overall population (populations receiving baricitinib 2 mg and 4 mg combined)**		
Patient-years of exposure	[redacted]	[redacted]
SAEs n (%)	[redacted]	[redacted]
WDAEs n (%)	[redacted]	[redacted]
Notable harms, n (%)		
Infection SAEs	[redacted]	[redacted]
Herpes zoster	[redacted]	[redacted]
Cardiac disorders SAEs	[redacted]	[redacted]
GI SAEs	[redacted]	[redacted]
NMSC	[redacted]	[redacted]
Malignancy other than NMSC	[redacted]	[redacted]
Hepatobiliary SAEs	[redacted]	[redacted]
ALT ≥ 3 × ULN	[redacted]	[redacted]
Neutrophils < 1,000 cells/mm^3	[redacted]	[redacted]

ALT = alanine aminotransferase; GI = gastrointestinal; NMSC = non-melanoma skin cancer; SAE = serious adverse event; ULN = upper limit of normal; WDAE = withdrawal due to adverse events.

Source: BEYOND interim Clinical Study Report[66] and BEYOND Clinical Study Report Addendum.[67]

Conclusion

The long-term efficacy and safety of baricitinib 2 mg once daily for the treatment of RA are being assessed in the ongoing long-term extension study BEYOND. Because of the lack of control group, findings from BEYOND are descriptive in nature and should be interpreted as such. [redacted]. Harms outcomes reported did not raise any new concerns regarding the safety of baricitinib. [redacted]. Therefore, at this point, uncertainty remains regarding the sustainability of beneficial treatment effects and long-term safety of baricitinib 2 mg once daily beyond the original trials' duration of 24 weeks.

Appendix 7: Summary of Indirect Comparisons

Background

Because of a lack of direct evidence comparing baricitinib with other biologic disease-modifying antirheumatic drugs (bDMARDs), the manufacturer performed a network meta-analysis (NMA) to estimate the efficacy of baricitinib in patients with rheumatoid arthritis relative to other conventional DMARDs (cDMARDs) and bDMARDs. The objective of this appendix is to summarize and critically review the unpublished NMA performed by the manufacturer and other available published indirect evidence that examines the relative efficacy and harms of baricitinib relative to other treatments for rheumatoid arthritis.

Methods

One NMA submitted by the manufacturer was reviewed in this section.[68] In addition, an information specialist performed a literature search to identify indirect treatment comparisons that included relevant interventions (e.g., baricitinib 2 mg), populations, and outcomes. Two relevant publications were identified and are reviewed in this section.[69,70]

Description of Indirect Treatment Comparisons Identified

An overview of the patients, interventions, outcomes, and study designs included in the three reports is provided in Table 18. In general, the scope of these reviews was similar with respect to patient populations, but there were some differences in the number of comparators selected and the number of outcomes analyzed. Bae et al. focused on the comparison of the two available Janus kinase (JAK) inhibitors (baricitinib versus tofacitinib) and limited their efficacy analysis to American College of Rheumatology (ACR)20. Choy et al. attempted to broaden their analyses to include data on serious infections and serious adverse events.

Table 18: Description of Indirect Comparisons Reviewed

	Manufacturer-Submitted Network Meta-Analysis[68]	Bae et al.[69]	Choy et al.[70]
Population	• Population 1: cDMARD-naive (limited or no prior treatment with cDMARD) • Population 2: TNF–inadequate response • Population 3: adults with moderate-to-severe active RA who were previously treated with cDMARDs but not bDMARDs (cDMARD–inadequate response) • Population 4: patients with inadequate response to MTX only	• RA patients with inadequate response to DMARDs or biologics	• RA patients with inadequate response to cDMARDs or a TNF inhibitor
Interventions	• Baricitinib 2 mg or 4 mg + cDMARD	• Baricitinib 2 mg or 4 mg + cDMARD	• Sarilumab
Comparisons	The following agents as monotherapy or with methotrexate • Abatacept • Adalimumab	• Tofacitinib 5 mg or 10 mg + MTX • Adalimumab + MTX • Placebo	• Baricitinib 2 mg • Abatacept • Adalimumab • Certolizumab pegol

	Manufacturer-Submitted Network Meta-Analysis[68]	Bae et al.[69]	Choy et al.[70]
	• Certolizumab pegol • Etanercept • Golimumab • Infliximab • Rituximab • Tocilizumab • Tofacitinib • Sarilumab • Methotrexate • cDMARDs • Placebo		• Etanercept • Golimumab • Infliximab • Rituximab • Etanercept • Tocilizumab • Tofacitinib • Methotrexate • cDMARDs • Placebo • Other investigational agents
Outcomes	• ACR20/ACR50/ACR70 • EULAR • HAQ-DI	• ACR20 • SAEs	• ACR20/ACR50/ACR70 • DAS28 response • HAQ-DI • mTSS • SAEs and serious infections
Study design	Randomized controlled trials	Randomized controlled trials	Randomized controlled trials
Funding	Eli Lilly	Korean Ministry of Health	Sanofi; Regeneron

ACR = American College of Rheumatology; bDMARD = biologic or targeted synthetic disease-modifying antirheumatic drug; cDMARD = conventional disease-modifying antirheumatic drugs; DAS28 = Disease Activity Scale-28; EULAR = European League Against Rheumatism; HAQ-DI = Health Assessment Questionnaire–Disability Index; mTSS = modified total Sharp score; MTX = methotrexate; RA = rheumatoid arthritis; SAE = serious adverse event; TNF = tumour necrosis factor.

Review and Appraisal of Indirect Comparisons

Review of Indirect Comparison Submitted by the Manufacturer

Objectives and Rationale

The objective of the indirect comparison (IDC) was to estimate between-treatment differences in efficacy between baricitinib and cDMARDs and between baricitinib and other bDMARDs. Comparators are listed in Table 18. Populations of interest were:

1. cDMARD-naive (limited or no prior treatment with cDMARDs)
2. Tumour necrosis factor (TNF)-inadequate response
3. cDMARD–inadequate response (previously treated with cDMARDs but not bDMARDs)
4. Inadequate response to methotrexate (MTX) only.

The two populations reviewed in this appendix are the TNF–inadequate response population and the cDMARD–inadequate response population (included patients with inadequate response to MTX). These are the populations that match the inclusion criteria for this CADTH Common Drug Review (CDR) clinical review and also contain data using the Health Canada–approved 2 mg dose of baricitinib.

Methods for Indirect Comparison Submitted by the Manufacturer

Study Eligibility and Selection Process

The authors performed a systematic review of literature published between 1999 and 2016 for the TNF–inadequate response population and between 1999 and 2017 for the cDMARD–inadequate response population. This excluded more recent trials. Since the number of trials was low in the TNF inhibitor–inadequate response network (N = 8), addition

of more trials from years 2017 to 2019 could have had an impact on the results. The authors did not describe which databases were searched. Inclusion and exclusion criteria for the study selection were not stated, although it was evident from the included studies that the authors selected randomized controlled trials (RCTs) that were performed in adults with rheumatoid arthritis. Titles and abstracts were reviewed independently by two researchers; if consensus was not reached, a third researcher was consulted.

Data Extraction

The authors included eight RCTs for the TNF–inadequate response population (including BEACON) and 40 RCTs for the cDMARD–inadequate response population (including BUILD and BEAM). The BEACON and BUILD studies contained baricitinib 2 mg dose groups, and these were the two studies that met the inclusion criteria for this CDR report. The BEAM study did not contain a baricitinib 2 mg dose group but was also included to populate the network for the cDMARD–inadequate response population. Information regarding the individual study characteristics is summarized in Table 19 and Table 20.

Table 19: Tumour Necrosis Factor–Inadequate Response — Study Characteristics

Study	Year of Publication	Interventions (N)	24-Week ACR Available?	Any Restrictions on Prior Biologics?
BEACON	**2016**	• **PL + cDMARD (176)** • **Baricitinib 2 mg + cDMARD (174)** • **Baricitinib 4 mg + cDMARD (177)**	**Yes**	**No**
ATTAIN	2005	• PL + cDMARD (133) • Abatacept 10 mg/kg + cDMARD (258)	Yes	No
BREVACTA	2014	• PL + cDMARD (219) • Tocilizumab 162 mg + cDMARD (437)	Yes	No
GO-AFTER	2009	• PL + cDMARD (155) • Golimumab 50 mg + cDMARD (153) • Golimumab 100 mg + cDMARD (153)	Yes	Natalizumab, rituximab at any time, anakinra in last 4 weeks; alefacept, efalizumab in last 3 months
ORAL STEP	2013	• PL + MTX (132) • Tofacitinib 5 mg + MTX (133) • Tofacitinib 10 mg + MTX (134)	No	No
RADIATE	2008	• PL + MTX (160) • Tocilizumab 4 mg/kg + MTX (163) • Tocilizumab 8 mg/kg + MTX (175)	Yes	No
REALISTIC	2012	• PL + cDMARD (212) • Certolizumab 400 mg + cDMARD (851)	No	Excluded if used ≥ 2 TNF inhibitors, rituximab, or abatacept
REFLEX	2006	• PL + MTX (209) • Rituximab 1,000 mg + MTX (311)	Yes	No

cDMARD = conventional disease-modifying antirheumatic drug; MTX = methotrexate; PL = placebo; TNF = tumour necrosis factor.

Source: Manufacturer's submission.[68,71]

The TNF–inadequate response studies were published between 2005 and 2016. Most of the TNF–inadequate response studies had no restrictions on prior biologic drug usage. Across the TNF–inadequate response study arms, mean ages ranged from 52 to 56 years, and the mean duration of rheumatoid arthritis ranged from 8.6 years to 14 years (data are not shown here but were provided by the manufacturer). Proportion of males across the

studies ranged from 14% to 26%, and mean Health Assessment Questionnaire–Disability Index (HAQ-DI) scores at baseline ranged from 1.5 to 1.9.

The cDMARD–inadequate response studies were published between 1999 and 2017. Most studies included patients who were either cDMARD-naive and/or MTX-naive. Across the cDMARD–inadequate responder study arms, mean ages ranged from 46 to 58 years. Proportion of males across the studies ranged from 4% to 26%, and mean Disease Activity Scale-28 using erythrocyte sedimentation rate (DAS28-ESR) baseline scores ranged from 5.0 to 6.9 (data are not shown here but were provided by the manufacturer).

Table 20: Conventional Disease-Modifying Antirheumatic Drug– (Including Methotrexate–) Inadequate Responders — Study Characteristics

Study Year	Interventions (N)	Response Criteria to Prior Treatment	cDMARD/MTX-Naive?	bDMARD-Naive?
Abe et al. 2006	• Infliximab + MTX (49) • MTX (47)	Active disease despite > 3 months MTX	No	Yes
ACT-RAY 2014	• Tocilizumab + MTX (279) • Tocilizumab + PL (277)	Active disease despite ≥ 12 weeks MTX	No	Yes
ADACTA 2013	• Adalimumab (163) • Tocilizumab (163)	Intolerance or inappropriate for MTX	No	Yes
AIM 2006	• Abatacept + MTX (433) • MTX (219)	Active RA despite ≥ 3 months MTX	No	Yes
AMPLE 2014	• Abatacept + MTX (318) • Adalimumab + MTX (328)	Inadequate response to MTX	No	Yes
APPEAL 2012	• Etanercept + MTX (197) • DMARD + MTX (103)	≥ 3 months of MTX at stable dose	No	Yes
ARMADA 2003	• Adalimumab + MTX (67) • MTX (62)	≥ 6 months of MTX; failed 1 to 4 DMARDs (besides MTX)	No	Yes
ATTEST 2008	• Abatacept + MTX (156) • Infliximab + MTX (165) • MTX (110)	Inadequate response to ≥ 3 months MTX	No	Yes
ATTRACT 1999	• Infliximab • PL + MTX • MTX	Active disease despite > 3 months MTX	Mixed (MTX-only group was MTX-naive)	Mixed (MTX-only group was mixed)
BEAM 2017	• **Baricitinib 4 mg +MTX (487)** • **Adalimumab + MTX (330)** • **PL + MTX (488)**	**Inadequate response to MTX**	**No**	**Yes**
BUILD 2017	• **Baricitinib 2 mg + MTX (229)** • **PL (228)**	**Inadequate response to ≥ 1 cDMARD**	**No**	**Yes**
BREVACTA 2014	• Tocilizumab (437) • PL (219)	Inadequate response to ≥ 8 weeks DMARD	No	Mixed (~20% failed ≥ 1 TNF inhibitor)
CHANGE 2008 24 weeks	• Adalimumab (91) • PL (87)	Failed ≥ 1 DMARD	No	Yes
CNTO 2008	• Golimumab + MTX (35) • MTX (35)	Active disease despite > 3 months MTX	No	Yes

Study Year	Interventions (N)	Response Criteria to Prior Treatment	cDMARD/MTX-Naive?	bDMARD-Naive?
52 weeks				
Coombe et al. 2006 104 weeks	• Etanercept (103) • Etanercept + SSZ (101) • SSZ (50)	Inadequate response to ≥ 4 months SSZ	No	Yes
De Filippis et al. 2006 54 weeks	• Etanercept +MTX (16) • Infliximab + MTX (16)	No response to > 6 months of DMARDs	No	Yes
Edwards et al. 2004 48 weeks	• Rituximab (40) • Rituximab + MTX (40)	Active disease despite ≥ 10 mg MTX/week	No	Yes
EXXELERATE 2016 104 weeks	• Certolizumab pegol +MTX • Adalimumab + MTX (N not reported)	Active disease despite 12 weeks of MTX	No	Yes
GO-FORTH 2012 156 weeks	• Golimumab + MTX (89) • MTX (90)	Active disease despite ≥ 3 months of MTX	No	Yes
GO-FORWARD 2009 52 weeks	• Golimumab + MTX (89) • MTX (133)	Active disease despite > 3 months MTX	No	Yes
JESMR 2010 104 weeks	• Etanercept + MTX (77) • Etanercept (74)	Received MTX ≥ 3 months	No	Yes
Keystone et al. 2004 52 weeks	• Adalimumab +MTX (207) • MTX (200)	Active disease despite > 3 months MTX	No	Yes
Kim et al. 2007 24 weeks	• PL (63) • Adalimumab (65)	Inadequate response to ≥ 6 months MTX	No	Unclear
Lan et al. 2004 12 weeks	• Etanercept (29) • MTX (29)	Received MTX at stable dose	No	Unclear
Li et al. 2013 56 weeks	• PL + MTX (132) • Golimumab + MTX	Active disease despite ≥ 4 weeks MTX	No	Yes
Machado et al. 2014 24 weeks	• Etanercept + MTX (284) • DMARD+MTX (145)	Active disease despite > 3 months MTX	No	Yes
MOBILITY 2015 62 weeks	• Sarilumab 150 mg + MTX (430) • Sarilumab 200 mg + MTX (428) • MTX (428)	Active disease despite > 3 months MTX	No	Mixed
MONARCH 2017 48 weeks	• Sarilumab • Adalimumab	Intolerant to MTX or active disease despite > 3 months MTX	No	Yes
Moreland et al. 1999 26 weeks	• PL (80) • Etanercept (78)	Inadequate response to 1 to 4 DMARDs	No	Unclear

Study Year	Interventions (N)	Response Criteria to Prior Treatment	cDMARD/MTX-Naive?	bDMARD-Naive?
Nishimoto et al. 2004 12 weeks	• Tocilizumab (55) • PL (54)	Unresponsive to MTX or immunosuppressant	No	Unclear
RAPID-C 2017 24 weeks	• PL + MTX (113) • Certolizumab + MTX (312)	Inadequate response to MTX	No	Unclear
RA-SCORE 2016 52 weeks	• Rituximab 500 mg + MTX (62) • Rituximab 1 g + MTX (60)	Inadequate response to MTX	No	Yes
REALISTIC 2012 12 weeks	• Certolizumab (851) • MTX (212)	Inadequate response to DMARD	No	Some previous TNF inhibitor usage
SATORI 2009 24 weeks	• MTX (66) • Tocilizumab + MTX (61)	Inadequate response to weeks MTX	No	Unclear
SERENE 2010 48 weeks	• MTX (172) • Rituximab 1 g + MTX (68) • Rituximab 2 g + MTX (170)	Active disease despite > 12 weeks MTX	No	Yes
STAR 2003 24 weeks	• Adalimumab + DMARD (318) • DMARD (318)	Inadequate response to DMARD	No	Yes
START 2006 54 weeks	• MTX (363) • Infliximab + MTX (360)	Active disease despite > 12 weeks MTX	No	Unclear
TEMPO 2004 52 weeks	• Etanercept (223) • MTX (228) • Etanercept + MTX (231)	Inadequate response to ≥ 1 DMARD	No	Unclear
van de Putte et al. 2004 26 weeks	• Adalimumab (113) • PL (110)	Failed ≥ 1 DMARD	No	Unclear
Weinblatt et al. 1999 24 weeks	• MTX (30) • Etanercept (59)	Active disease despite ≥ 6 months MTX	No	Unclear

bDMARD = biologic disease-modifying antirheumatic drug; cDMARD = conventional disease-modifying antirheumatic drug; DMARD = disease-modifying antirheumatic drug; MTX = methotrexate; PL = placebo; RA = rheumatoid arthritis; SSZ = sulfasalazine; TNF = tumour necrosis factor.

Source: Manufacturer's submission.[68]

Comparators

Population With Inadequate Response to Tumour Necrosis Factor Inhibitors

In the eight included trials, there were seven different bDMARDs used across 20 treatment arms. All trials shared placebo + MTX/cDMARD as a common comparator. No trials compared more than one bDMARD. The regimens are described in Table 19. According to the clinical expert consulted for this review, the dosages used in the included trials are similar to the dosages used in Canada to treat similar patients, with the exception of the baricitinib 4 mg once daily and tofacitinib 10 mg twice daily dosages that were used in some trials. These higher doses of JAK inhibitors were not approved by Health Canada.

Population With Inadequate Response to Conventional Disease-Modifying Antirheumatic Drugs

In the 40 included trials, there were 11 different bDMARDs across 87 treatment arms. Most trials shared cDMARD as a common comparator. Seven trials compared more than one bDMARD. The regimens are described in Table 20. According to the clinical expert consulted for this review, the dosages used in the included trials are similar to the dosages used in Canada to treat similar patients, with the exception of the baricitinib 4 mg once daily dosage that was used in the BEAM trial. This higher dosage of baricitinib is not approved by Health Canada.

Outcomes

Outcomes of interest were ACR20, ACR50, and ACR70 response (median odds ratios), European League Against Rheumatism (EULAR) response (median odds ratios), and HAQ-DI response (median difference). Most of the data were analyzed at the 24-week time point. No harms outcomes were assessed.

The inclusion criteria used in this CDR report included other outcomes that were not analyzed by the manufacturer, such as health care utilization, serious adverse events, serious infections, and withdrawals due to adverse events. The manufacturer did not assess radiographic outcomes.

Quality Assessment of Included Studies

The authors reported a brief summary of their quality assessment of the included trials.

Population With Inadequate Response to Tumour Necrosis Factor Inhibitors

Eight trials were included, of which seven were assessed for quality. The authors assessed approximately half the studies as not clearly describing methods for allocation concealment and randomization. They reported that approximately half of the studies lacked clarity regarding who was blinded in the study (e.g., outcome assessors, patients, treating physicians). Approximately 75% of the studies had an acceptable balance of prognostic factors between groups, according to the NMA authors. There were no sensitivity analyses performed based on study quality.

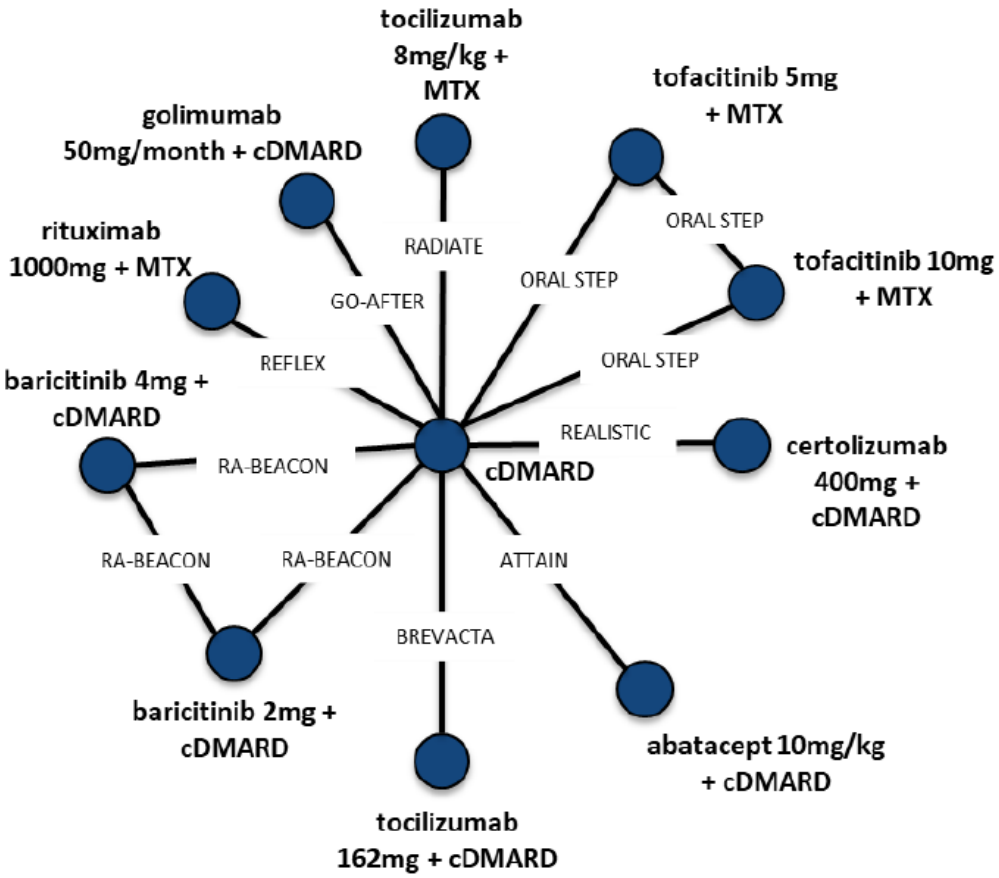
Population With Inadequate Response to Conventional Disease-Modifying Antirheumatic Drugs

Forty trials were included, and 38 were assessed for quality. The authors assessed approximately 65% of the studies as not clearly describing methods for allocation concealment and randomization. They reported that approximately half of the studies lacked clarity regarding who was blinded in the study (e.g., outcome assessors, patients, treating physicians). Approximately 85% of the studies had an acceptable balance of prognostic factors between groups, according to the NMA authors. In this population, the authors performed one sensitivity analyses related to study design and study quality, in which open-label studies were removed. This sensitivity analysis did not result in significant changes to the results, relative to the base-case analysis.

Evidence Network

Figures 2 and 3 depict the networks of evidence using the included studies. Figure 2 depicts the network for the TNF inhibitor–inadequate responder population in all included studies. Figure 3 depicts the network for the cDMARD–inadequate responder population and represents the network used to assess ACR20.

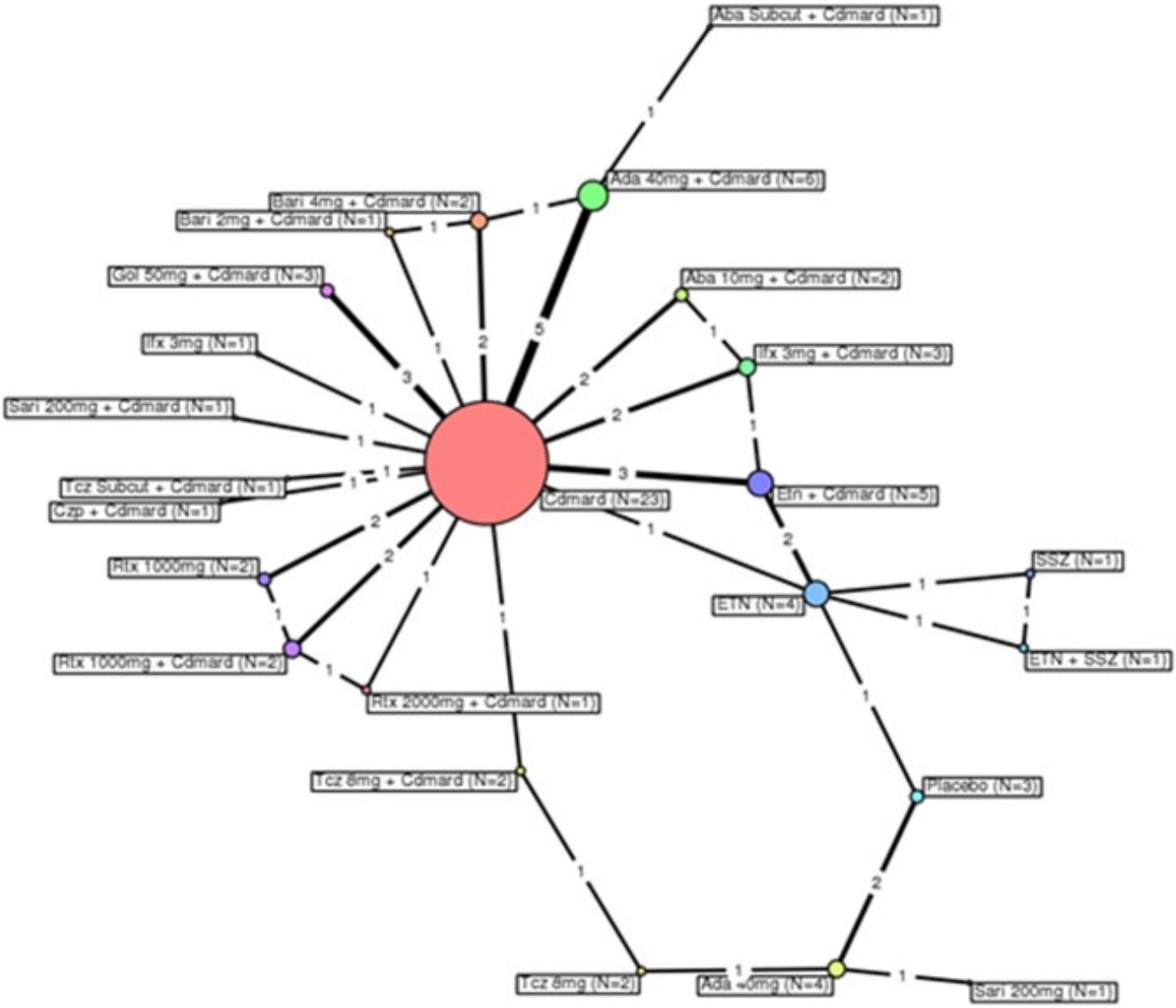
Figure 2: Tumour Necrosis Factor Inhibitor–Inadequate Responders — Network of Evidence



cDMARD = conventional disease-modifying antirheumatic drug; MTX = methotrexate TNF = tumour necrosis factor.

Source: Manufacturer's submission.[71]

Figure 3: Conventional Disease-Modifying Antirheumatic Drug– (Including Methotrexate–) Inadequate Responders — Network of Evidence for ACR20 Comparisons



Aba = abatacept; Ada = adalimumab; Bari = baricitinib; Cdmard = conventional disease-modifying antirheumatic drug; Czp = certolizumab pegol; ETN = etanercept; Ifx = infliximab; MTX = methotrexate; Rtx = rituximab; Tcz = tocilizumab; Sari = sarilumab; SSZ = sulfasalazine.

Source: Manufacturer's submission.[68]

Indirect Comparison Methods

Statistical Methods

The NMAs were performed using Bayesian methods, and both random-effects and fixed-effects models were employed. The choice of a fixed- or random-effects model was evaluated on the basis of model fit as measured by deviance information criterion (DIC), sensitivity of results, and assessment of residual deviance. The authors used both binary

and probit modelling approaches. The main analyses were conducted at the 24-week time point.

Bayesian analyses used vague priors — N(0,10,000) — for treatment effect coefficients. In some cases, the random-effects models were sensitive to the choice of vague priors, resulting in wide credible intervals. This was due to the combination of a low number of included studies, studies with small sample size, and a high degree of heterogeneity. In models in which unstable credible intervals were observed, informative priors were used.

The model used two chains. The first 530,000 simulations were discarded to allow for model convergence, and an additional 1,060,000 simulations (thinning of 53) were used to estimate the posterior probabilities from a sample of 40,000. Model fit was assessed with the DIC and the posterior mean of the total residual deviance. A good model fit was indicated by a total residual deviance approximately equal to the number of data points available. When comparing two DIC values, a difference of five or more was regarded as a meaningful difference. Convergence was verified by trace plots, monitoring the Monte Carlo error, and with Gelman-Rubin diagnostics.

Population With Inadequate Response to Tumour Necrosis Factor Inhibitors

The authors selected fixed-effects models as the primary approach for most comparisons for the TNF inhibitor–inadequate response population, based on model fit. The authors mentioned that sensitivity analyses showed consistent results for ACR and EULAR responses in this population, but no details were provided. No sensitivity analyses were provided by the manufacturer, and it was unclear whether any had been performed in this population because of the low number of studies.

Population with Inadequate Response to Conventional Disease-Modifying Antirheumatic Drugs

The authors selected the random-effects model as the primary approach for the cDMARD–inadequate response population, based on model fit. Sensitivity analyses were conducted, including the removal of specific studies (e.g., for high heterogeneity, inconsistency, open-label studies, Asian-Pacific studies, and/or previous low/unknown MTX dose) or addition of studies (e.g., up to 20% of background bDMARDs). Meta-regression was performed using factors such as year of study, MTX dose (low versus normal), mean duration of disease, and early versus established rheumatoid arthritis. Models were fit using baseline and treatment effect separately. Frequentist models using random and fixed effects were also fitted as sensitivity analyses.

There was no formal assessment of consistency. However, the BEAM study was included in the analysis of cDMARD–inadequate responders, and this study had an adalimumab group and a baricitinib 4 mg group. In the BEAM publication, the authors claimed that baricitinib 4 mg was superior to adalimumab at 12 weeks for the ACR20 response. In contrast, the NMA submitted by the manufacturer did not demonstrate any differences in response for baricitinib 2 mg versus adalimumab.

Results

Population with Inadequate Response to Tumour Necrosis Factor Inhibitors

Table 21 summarizes selected statistically significant treatment effects for baricitinib 2 mg + cDMARD relative to other comparators in the manufacturer's NMA in the population with inadequate response to TNF inhibitors.

American College of Rheumatology Response at Week 24 (Binary Model Analysis, Fixed-Effects Model)

The network included the comparators cDMARD alone and cDMARD plus baricitinib 4 mg, abatacept 10 mg, tocilizumab 162 mg, and golimumab 50 mg. The following is a summary of the results for which there was a higher or lower response between baricitinib 2 mg + cDMARD and another comparator in the network. There was no difference in response for any other comparison based on the reported 95% credible intervals.

[redacted]

[redacted]

American College of Rheumatology Response at Week 24 (Probit Model Analysis, Fixed-Effects Model)

The network included the comparators cDMARD alone and cDMARD plus baricitinib 4 mg, abatacept 10 mg, tocilizumab 162 mg, golimumab 50 mg, tocilizumab 8 mg + MTX, and rituximab 1,000 mg + MTX. The following is a summary of the results for which there was a higher or lower response between baricitinib 2 mg + cDMARD and another comparator in the network. There was no difference in response for any other comparison based on the reported 95% credible intervals.

[redacted]

[redacted]

European League Against Rheumatism Response at Week 24 (Fixed-Effects Binary Model Analysis)

[redacted]

European League Against Rheumatism Response at Week 24 (Fixed-Effects Probit Model Analysis)

[redacted]

Health Assessment Questionnaire–Disability Index Response at Week 24 (Fixed-Effects Binary Model Analysis)

[redacted]

Table 21: Population With Inadequate Response to Tumour Necrosis Factor Inhibitors: Statistically Significant Relative Treatment Effects of Baricitinib 2 mg Versus Comparators

	Week 24, Median Odds Ratio (95% CrI) Fixed-Effects Model		Week 24, Median Odds Ratio (95% CrI) Fixed-Effects Probit Model	
Comparator	Result Favours BAR 2 mg + cDMARD	Result Favours Comparator	Result Favours BAR 2 mg + cDMARD	Result Favours Comparator
cDMARD	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

ABA = abatacept; BAR = baricitinib; cDMARD = conventional disease-modifying antirheumatic drug; CrI = credible interval; EGR = EULAR Good Response; MTX = methotrexate; RTX = rituximab; TCZ = tocilizumab; TNF = tumour necrosis factor.

Note: This table summarizes selected statistically significant treatment effects for baricitinib 2 mg relative to other comparators in the manufacturer's NMA. Comparisons to baricitinib that were not statistically significant are not reported here. The fixed-effects model results are presented here because this model was selected as the primary approach based on model fit.

[a] HAQ-DI data are median difference.

Source: Manufacturer's submission.[68]

Population With Inadequate Response to Conventional Disease-Modifying Antirheumatic Drugs (Including Methotrexate)

Table 22 summarizes selected statistically significant treatment effects for baricitinib 2 mg + cDMARD relative to other comparators in the manufacturer's NMA in the population with inadequate response to cDMARDs.

American College of Rheumatology Response at Week 24 (Binary Model Analysis, Random-Effects Model)

The network included the comparators placebo, tocilizumab 8 mg alone, adalimumab 40 mg alone, infliximab alone, etanercept + sulfasalazine, etanercept alone, sulfasalazine alone, cDMARD alone, sarilumab alone, as well as cDMARD in combination with baricitinib 2 mg, tocilizumab 8 mg, abatacept 10 mg, abatacept subcutaneous, adalimumab, infliximab, etanercept, rituximab 1 g, golimumab 50 mg, sarilumab 200 mg, or certolizumab pegol.

[redacted]

American College of Rheumatology Response at Week 24 (Probit Model Analysis, Fixed-Effects Model)

[redacted]

Table 22: Population With Inadequate Response to cDMARDs: Statistically Significant Relative Treatment Effects of Baricitinib 2 mg Versus Comparators

	Week 24, Median Odds Ratio (95%CrI) Random-Effects Model		Week 24, Median Odds Ratio (95%CrI) Random-Effects Probit Model	
Comparator	**Result Favours BAR 2 mg + cDMARD**	**Result Favours Comparator**	**Result Favours BAR 2 mg + cDMARD**	**Result Favours Comparator**
cDMARD	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

BAR = baricitinib; cDMARD = conventional disease-modifying antirheumatic drug; CrI = credible interval; MTX = methotrexate.

Note: This table summarizes selected statistically significant treatment effects for baricitinib 2 mg relative to other comparators in the manufacturer's NMA. Comparisons to baricitinib that were not statistically significant are not reported here. The random-effects model results are presented here because this model was selected as the primary approach based on model fit.

Source: Manufacturer's submission.[68]

Sensitivity Analyses

The ability to perform sensitivity analysis in the TNF inhibitor–inadequate response population was limited by the low number of studies.

In the cDMARD–inadequate response population, there were an adequate number of studies to allow several analyses. The sensitivity analyses were largely consistent with the results of the base-case analyses. One unique finding of the sensitivity analyses is that there were several analyses showing that baricitinib 2 mg + cDMARD was superior to etanercept monotherapy, whereas this was rarely observed in the base-case analyses. This is an unsurprising finding, given that etanercept was used without MTX or other cDMARDs in that comparator group.

Critical Appraisal

Manufacturer-Submitted Network Meta-Analysis

The validity of statistical models, particularly the appropriateness of probit model versus binary model, is not completely clear, e.g., it is unclear how to handle ordinal nature of the outcome, as stated, using an extension of the binary logistic regression model. The binary model seems more straightforward to interpret than the probit model, especially when the two sets of results differ. Clinically, the ordinal nature of the ACR response seems to make sense, as different levels of change have different importance to patients.

Due to the limitations of odds ratios as a measure of the strength of relative effect, when the assumption of a rare event is not satisfied (i.e., which is obviously the case in ACR response), there may be a much wider 95% credible interval (CrI) of the odds ratio than that of relative risk. Therefore, converting an odds ratio to relative risk and interpreting accordingly would help ensure that the conclusion, based on a wide CrI of the odds ratio with regard to statistical non-significance between the comparisons, is appropriate.

The high placebo effect and considerable variations across trials may signal a significant heterogeneity, which could have compromised the validity of findings.

The NMA reporting did not harmonize in all respects with published guidance on NMA reporting such as PRISMA.[72] Some methodological elements were not described clearly (e.g., inclusion/exclusion criteria of studies, names of databases searched), and publication bias was not assessed. While there was some variation in the population demographics, patients' stage of disease, and disease severity in the trials, the studies included in the reviews were generally thought to include populations that are reflective of the Canadian population with rheumatoid arthritis expected to use baricitinib, according to the clinical expert consulted by CADTH.

The authors included relevant comparators, and the included trials formed connected networks, although not all networks contained all of the included trials because data were limited in some trials. For example, some trials did not have data at 24 weeks. The number of trials in the network was low for some of the TNF inhibitor–inadequate response population. For example, many of the ACR networks included only six trials, and the HAQ-DI network included only three trials.

The TNF inhibitor–inadequate response population analysis included outcomes of ACR, EULAR, and HAQ-DI responses. In the cDMARD–inadequate response population, ACR responses were reported. The omission of any harms analysis is a significant limitation of this NMA; there was no assessment of the relative risk of adverse events or withdrawals due to adverse events.

The authors included sensitivity analyses based on several prognostic factors and factors related to quality of the studies (e.g., blinding) and to generalizability (e.g., removing Asian studies). A sensitivity analysis based on dropout rates would also have been informative, but this was not provided. These rates vary across studies and over the 20-year time study publication period covered in the cDMARD–inadequate response population.

Review of Indirect Comparison by Bae et al.[69]

Objectives and Rationale for Bae et al.

The objective was to assess the relative efficacy and safety of tofacitinib and baricitinib in patients with RA with an inadequate response to cDMARDs or bDMARDs.

Methods for Bae et al.

Literature databases were searched for relevant tofacitinib and baricitinib studies published prior to April 2018. Authors selected RCTs performed in patients who responded inadequately to either cDMARDs or bDMARDs. Data were extracted by two independent reviewers. The Jadad scoring system was used to assess study quality. Authors stated that they relied on PRISMA guidelines to structure their report.[72] A Bayesian random-effects NMA was performed on the included trials. The outcomes selected for analysis were

ACR20 and serious adverse events. Authors reported funding from the Korean Ministry of Health.

Results of Bae et al.

Twelve RCTs, including 5,883 patients, met the inclusion criteria. The authors assessed the quality of all included studies as "high." The authors included six baricitinib trials, including BEACON,[8] BEAM,[73] BUILD,[9] and three small phase II studies. Treatment comparators included tofacitinib 5 mg/10 mg + MTX, adalimumab + MTX, baricitinib 4 mg/2 mg + MTX, and placebo + MTX. Results demonstrated no difference in ACR20 response between baricitinib 2 mg + MTX versus tofacitinib + MTX or versus adalimumab + MTX. There was no difference in the odds of serious adverse events for baricitinib 2 mg + MTX relative to any comparator.

The authors concluded that "tofacitinib 10 mg + MTX and baricitinib 4 mg + MTX were the most efficacious interventions for RA patients with an inadequate response to DMARD or biologics therapy."

Critical Appraisal of Bae et al.

A strength of this NMA was that it included phase II and phase III trials of two JAK inhibitors, as well as a literature search covering the period up to April 2018. Weaknesses of the analysis include combining data from heterogeneous trial populations; as well, populations of TNF inhibitor–inadequate responders were analyzed in the same network as cDMARD–inadequate responders. Another drawback of this NMA is the lack of data on other outcomes (e.g., ACR50, ACR70, HAQ-DI).

Review of Indirect Comparison by Choy et al.[70]

Objectives and Rationale for Choy et al.

To compare the efficacy and safety of subcutaneous sarilumab 200 mg and 150 mg plus cDMARDs versus other cDMARDs and bDMARDs in inadequate responders to cDMARDs or TNF inhibitors.

Methods for Choy et al.

A systematic literature search was performed to identify phase II, III, or IV RCTs published before December 2016. Literature was screened by two reviewers. Authors stated that they relied on PRISMA guidelines to structure their report.[72] Bayesian random-effects and fixed-effects models were used to analyze the outcomes in the two networks. The outcomes selected for analysis were ACR20, ACR50, ACR70, DAS28, HAQ-DI, mTSS, SAEs, and serious infections. All efficacy outcomes were evaluated at 24 weeks.

Results for Choy et al.

A total of 46 RCTs were included for the cDMARD–inadequate response population, and nine RCTs were included for the TNF inhibitor–inadequate response population. The authors included the baricitinib trial BEACON[8] in the TNF inhibitor–inadequate response network and BEAM[73] and BUILD[9] in the cDMARD–inadequate response analysis.

In the population with inadequate response to TNF inhibitors, there were no statistically significant differences observed between baricitinib 2 mg + cDMARD versus sarilumab 200 mg + cDMARD for ACR20, ACR70, or HAQ-DI. The authors reported that sarilumab had a higher ACR50 response (median estimate of effect 0.126, [95% CrI, 0.006 to 0.248]),

DAS28 response (median estimate of effect 0.169, [95% CrI, 0.074 to 0.265]). The authors also reported that rates of serious infections and serious adverse events were similar between sarilumab (150 mg or 200 mg) and baricitinib 2 mg.

In the population with inadequate response to cDMARDs, there were no statistically significant differences observed between baricitinib 2 mg + cDMARD and sarilumab 200 mg + cDMARD for ACR20, ACR50, ACR70, DAS28, or HAQ-DI. There were no data available for serious infections or serious adverse events. The authors reported that sarilumab showed a greater improvement in mTSS at week 24 (change from baseline –0.721 [95% CrI, –1.353 to –0.087]).

The authors concluded that "sarilumab + cDMARD had superior efficacy and similar safety versus placebo + cDMARDs and at least similar efficacy and safety versus bDMARDs" in inadequate responders to cDMARDs or TNF inhibitors.

Critical Appraisal of Choy et al.

The authors described significant heterogeneity, such as geographical location and variations in placebo response rate, among the included trials. Sparse data in the TNF inhibitor–inadequate responder network precluded meta-regression analyses to adjust for confounding factors. The authors stated that the rates of TNF inhibitor failure at baseline were not 100% in all included trials. A disadvantage of this indirect comparison report is that the authors focused on sarilumab as the main comparator and did not present the effects of baricitinib relative to other bDMARDs.

Discussion

In the absence of direct treatment comparisons, an IDC may serve to illuminate relative treatment effects. However, heterogeneity of study design features and study populations was a weakness common to all three IDCs reviewed. Studies within each NMA differed in terms of factors such as previous exposure to bDMARDs (number and type), placebo response rates, duration of disease, prior MTX usage, and disease severity at baseline. Authors attempted to overcome some of these differences through sensitivity analyses and meta-regression analyses, but this was not feasible for the smaller networks (e.g., Bae et al. and the manufacturer's TNF inhibitor–inadequate response network).

Baricitinib was included in recent analyses by CADTH and The National Institute for Health and Care Excellence (NICE).[74] CADTH has published a NMA of cDMARDs and bDMARDs in rheumatoid arthritis.[75] The CADTH report included data on baricitinib 4 mg but not 2 mg. Results indicated, for example, that baricitinib 4 mg + cDMARD had a higher ACR50 response relative to MTX monotherapy, cDMARD + MTX, etanercept monotherapy, and tocilizumab monotherapy; all other comparisons with baricitinib 4 mg for ACR50 were not statistically significant. Baricitinib was also recently reviewed by NICE, and its summary of the results is similar to what was described in the NMAs in this appendix.[74,76] The NICE appraisal found that baricitinib (2 mg or 4 mg) + cDMARD showed similar EULAR response rates as cDMARDs alone in both the TNF inhibitor–inadequate response population and the cDMARD–inadequate response population. The exceptions mentioned by NICE were that rituximab + cDMARD showed better EULAR response rates than baricitinib + cDMARD in the TNF inhibitor–inadequate response population, and that tocilizumab + cDMARD may have a better EULAR response than baricitinib + cDMARD in the cDMARD–inadequate response population.

Conclusion

In the absence of sufficient head-to-head trials, the manufacturer conducted an IDC based on a systematic review of RCTs, comparing the efficacy of baricitinib 2 mg with abatacept, adalimumab, certolizumab pegol, etanercept, golimumab, infliximab, rituximab, tocilizumab, tofacitinib, sarilumab, MTX, cDMARDs, and placebo. The literature search was not up-to-date, so the NMAs may have excluded relevant trials. Analyses in populations with inadequate response to TNF inhibitors suggest that patients did not respond as well to baricitinib 2 mg relative to tocilizumab and rituximab for some ACR and EULAR response outcomes. Analyses in populations with inadequate response to cDMARDs showed similar ACR response rates for baricitinib 2 mg relative to other bDMARDs, with the exception of etanercept monotherapy. There was no evidence that the impact on quality of life (e.g., HAQ-DI) differs for baricitinib 2 mg relative to other bDMARDs. There were no differences in ACR response rates for comparisons of baricitinib versus tofacitinib in the manufacturer's NMA analyses. No safety outcomes were evaluated in the manufacturer's NMAs.

Two published NMAs were reviewed, and the results were generally congruent with the findings of the manufacturer's IDC. The results from one NMA suggest that there is no difference in ACR20 response or serious adverse events for baricitinib relative to the other JAK inhibitor, tofacitinib. This NMA was limited by the low number of RCTs and a small network. The results from another NMA suggest that sarilumab had a higher response for some efficacy outcomes, but this was not consistently observed across all outcomes.

References

1. Arthritis Alliance of Canada. The impact of arthritis in Canada:today and over the next 30 years. 2011; http://www.arthritisalliance.ca/en/impactquick. Accessed 2019 May 28.
2. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. *Health Qual Life Outcomes.* 2003;1:20.
3. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. *J Rheumatol.* 2003;30(1):167-178.
4. Clinical Study Report: JADW. A randomized, double-blind, placebo-controlled, phase 3 study evaluating the efficacy and safety of baricitinib (LY3009104) in patients with moderately to severely active rheumatoid arthritis who have had an inadequate response to tumor necrosis factor inhibitors [CONFIDENTIAL internal manufacturer's report]. 2015 Aug 9.
5. Clinical Study Report: JADX. A randomized, double-blind, placebo-controlled, phase 3 study to evaluate the efficacy and safety of baricitinib (LY3009104) in patients with inadequate response to conventional disease-modifying antirheumatic drugs with moderately to severely active rheumatoid arthritis [CONFIDENTIAL internal manufacturer's report]. Indianapolis (IN): Eli Lilly and Company; 2015 Nov 5.
6. Singh JA, Saag KG, Bridges SL, Jr., et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis rheumatol.* 2016;68(1):1-26.
7. e-CPS. Ottawa (ON): Canadian Pharmacists Association; 2018: https://www.e-therapeutics.ca. Accessed 2019 April 23.
8. Genovese MC, Kremer J, Zamani O, et al. Baricitinib in patients with refractory rheumatoid arthritis. *N Engl J Med.* 2016;374(13):1243-1252.
9. Dougados M, van der Heijde D, Chen YC, et al. Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. *Ann Rheum Dis.* 2017;76(1):88-95.
10. Wells G, Becker JC, Teng J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. *Ann Rheum Dis.* 2009;68(6):954-960.
11. Inoue E, Yamanaka H, Hara M, Tomatsu T, Kamatani N. Comparison of Disease Activity Score (DAS)28- erythrocyte sedimentation rate and DAS28- C-reactive protein threshold values. *Ann Rheum Dis.* 2007;66(3):407-409.
12. Anderson J, Caplan L, Yazdany J, et al. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. *Arthritis Care Res (Hoboken).* 2012;64(5):640-647.
13. Fujiwara M, Kita Y. Reexamination of the assessment criteria for rheumatoid arthritis disease activity based on comparison of the Disease Activity Score 28 with other simpler assessment methods. *Mod Rheumatol.* 2013;23(2):260-268.
14. Gaujoux-Viala C, Mouterde G, Baillet A, et al. Evaluating disease activity in rheumatoid arthritis: which composite index is best? A systematic literature analysis of studies comparing the psychometric properties of the DAS, DAS28, SDAI and CDAI. *Joint Bone Spine.* 2012;79(2):149-155.
15. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. *Clin Exp Rheumatol.* 2005;23(5 Suppl 39):S100-108.
16. Smolen JS, Aletaha D. Scores for all seasons: SDAI and CDAI. *Clin Exp Rheumatol.* 2014;32(5 Suppl 85):S-75-79.
17. Chung CP, Thompson JL, Koch GG, Amara I, Strand V, Pincus T. Are American College of Rheumatology 50% response criteria superior to 20% criteria in distinguishing active aggressive treatment in rheumatoid arthritis clinical trials reported since 1997? A meta-analysis of discriminant capacities. *Ann Rheum Dis.* 2006;65(12):1602-1607.
18. Chaplin S. Janus kinase inhibitors for autoimmune disorders. *Prescriber.* 2017;28(12):33-37.
19. Olumiant (baricitinib): 2 mg tablets [product monograph]. Toronto (ON): Eli Lilly Canada Inc.; 2018 Aug 14.
20. Xeljanz (tofacitinib) 5mg and 10mg tablets [product monograph]. Kirkland (QC): Pfizer Canada ULC; 2019 Feb 4: https://pdf.hres.ca/dpd_pm/00049539.PDF. Accessed 2019 May 28.
21. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. Medical review(s). *Olumiant (baricitinib) tablets. Company: Eli Lilly and Company. Application No. 207924. Approval date: 05/31/2018 (FDA approval package)*. Rockville (MD): U. S. Food and Drug Administration (FDA); 2017: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/207924Orig1s000TOC.cfm. Accessed 2019 Jan 22.
22. Health Canada. Clinical trial finds an increased risk of blood clots in the lungs and of death in rheumatoid arthritis patients taking high dose of tofacitinib (sold as Xeljanz or Xeljanz XR). *(Recalls and safety alerts)* 2019; www.healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2019/69336a-eng.php. Accessed 2019 Apr 23.
23. van Riel PL, van Gestel AM. Clinical outcome measures in rheumatoid arthritis. *Ann Rheum Dis.* 2000;59 Suppl 1:i28-i31.
24. Cohen JD, Dougados M, Goupille P, et al. Health assessment questionnaire score is the best predictor of 5-year quality of life in early rheumatoid arthritis. *J Rheumatol.* 2006;33(10):1936-1941.
25. Bansback N, Ara R, Karnon J, Anis A. Economic evaluations in rheumatoid arthritis: a critical review of measures used to define health states. *Pharmacoeconomics.* 2008;26(5):395-408.

26. American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response. *Arthritis Rheum.* 2007;57(2):193-202.

27. Gallagher EJ, Liebman M, Bijur PE. Prospective validation of clinically important changes in pain severity measured on a visual analog scale. *Ann Emerg Med.* 2001;38(6):633-638.

28. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. *Ann Med.* 2001;33(5):350-357.

29. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics.* 1999;15(2):141-155.

30. Strand V, Singh JA. Improved health-related quality of life with effective disease-modifying antirheumatic drugs: evidence from randomized controlled trials. *Am J Manag Care.* 2008;14(4):239-253.

31. Crowson CS, Rahman MU, Matteson EL. Which measure of inflammation to use? A comparison of erythrocyte sedimentation rate and C-reactive protein measurements from randomized clinical trials of golimumab in rheumatoid arthritis. *J Rheumatol.* 2009;36(8):1606-1610.

32. Bruynesteyn K, van der Heijde D, Boers M, et al. Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference. *Arthritis Rheum.* 2002;46(4):913-920.

33. Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the functional assessment of chronic illness therapy fatigue scale relative to other instrumentation in patients with rheumatoid arthritis. *J Rheumatol.* 2005;32(5):811-819.

34. Cohen S, Hurd E, Cush J, et al. Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor antagonist, in combination with methotrexate: results of a twenty-four-week, multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum.* 2002;46(3):614-624.

35. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices, and biological products for the treatment of rheumatoid arthritis (RA). Rockville (MD): Food and Drug Administration; 1999: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071579.pdf. Accessed 2019 May 28.

36. Felson DT, Furst DE, Boers M. Rationale and strategies for reevaluating the ACR20. *J Rheumatol.* 2007;34(5):1184-1187.

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

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

39. Linde L, Sorensen J, Ostergaard M, Horslev-Petersen K, Hetland ML. Health-related quality of life: validity, reliability, and responsiveness of SF-36, 15D, EQ-5D RAQoL, and HAQ in patients with rheumatoid arthritis. *J Rheumatol.* 2008;35(8):1528-1537.

40. Fransen J, Antoni C, Mease PJ, et al. Performance of response criteria for assessing peripheral arthritis in patients with psoriatic arthritis: analysis of data from randomized, controlled trials of two TNF inhibitors. *Ann Rheum Dis.* 2006;65(10):1373-1378.

41. Landewe R, van der Heijde D. The validity of a rheumatoid arthritis medical records-based index of severity compared with the DAS28. *Arthritis Res Ther.* 2006;8(3):107.

42. van der Heijde DM, van 't Hof MA, van Riel PL, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis.* 1990;49(11):916-920.

43. van der Heijde DM, van 't Hof MA, van Riel PL, van Leeuwen MA, van Rijswijk MH, van de Putte LB. Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. *Ann Rheum Dis.* 1992;51(2):177-181.

44. Prevoo ML, van't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum.* 1995;38(1):44-48.

45. van Gestel AM, Anderson JJ, van Riel PL, et al. ACR and EULAR improvement criteria have comparable validity in rheumatoid arthritis trials. *J Rheumatol.* 1999;26(3):705-711.

46. Ranganath VK, Khanna D, Paulus HE. ACR remission criteria and response criteria. *Clin Exp Rheumatol.* 2006;24(6 Suppl 43):S14-S21.

47. Siemons L, Vonkeman HE, Ten Klooster PM, van Riel PL, van de Laar MA. Interchangeability of 28-joint disease activity scores using the erythrocyte sedimentation rate or the C-reactive protein as inflammatory marker. *Clin Rheumatol.* 2014;33(6):783-789.

48. Fernandez-Montero JV, Barreiro P, Soriano V. HIV protease inhibitors: recent clinical trials and recommendations on use. *Expert Opin Pharmacother.* 2009;10(10):1615-1629.

49. Matsui T, Kuga Y, Kaneko A, et al. Disease Activity Score 28 (DAS28) using C-reactive protein underestimates disease activity and overestimates EULAR response criteria compared with DAS28 using erythrocyte sedimentation rate in a large observational cohort of rheumatoid arthritis patients in Japan. *Ann Rheum Dis.* 2007;66(9):1221-1226.

50. Tamhane A, Redden DT, McGwin G, Jr., et al. Comparison of the disease activity score using erythrocyte sedimentation rate and C-reactive protein in African Americans with rheumatoid arthritis. *J Rheumatol.* 2013;40(11):1812-1822.

51. Clinical and economic overview: biological response modifier agents for adults with rheumatoid arthritis. *(CADTH Therapeutic review)*. Ottawa (ON): CADTH; 2010: http://www.cadth.ca/media/pdf/TR_RA_Clinical_and_Economic_Overview_e.pdf. Accessed 2019 May 28.

52. McKesson Pharmaclick. 2014; https://www.mckesson.ca/. Accessed 2014.

53. Ontario Ministry of Health and Long-Term Care. Exceptional Access Program (EAP). Toronto (ON)2014.

54. Disease activity score in 28 joints (DAS28). Philadelphia (PA): The Institute for Continuing Healthcare Education; 2009.

55. Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. *Ann Rheum Dis.* 2001;60(9):817-827.

56. Guillemin F, Billot L, Boini S, Gerard N, Odegaard S, Kvien TK. Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis. *J Rheumatol.* 2005;32(5):778-786.

57. Sokka T. Radiographic scoring in rheumatoid arthritis: a short introduction to the methods. *Bull NYU Hosp Jt Dis.* 2008;66(2):166-168.

58. Landewe R, van der Heijde D. Radiographic progression in rheumatoid arthritis. *Clin Exp Rheumatol.* 2005;23(5 Suppl 39):S63-S68.

59. Voskuyl AE, Dijkmans BA. Remission and radiographic progression in rheumatoid arthritis. *Clin Exp Rheumatol.* 2006;24(6 Suppl 43):S37-S40.

60. Drossaers-Bakker KW, de Buck M, van Zeben D, Zwinderman AH, Breedveld FC, Hazes JM. Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time. *Arthritis Rheum.* 1999;42(9):1854-1860.

61. Sahin F, Kotevoglu N, Taspinar S, Yilmaz F, Kuran B. Comparison of functional disability scales and their relevance to radiological progression in patients with rheumatoid arthritis in remission. *Clin Exp Rheumatol.* 2006;24(5):540-545.

62. Swinkels HL, Laan RF, van 't Hof MA, van der Heijde DM, de Vries N, van Riel PL. Modified sharp method: factors influencing reproducibility and variability. *Semin Arthritis Rheum.* 2001;31(3):176-190.

63. Foley-Nolan D, Stack JP, Ryan M, et al. Magnetic resonance imaging in the assessment of rheumatoid arthritis--a comparison with plain film radiographs. *Br J Rheumatol.* 1991;30(2):101-106.

64. Landewe R, van der Heijde D. Is radiographic progression a realistic outcome measure in clinical trials with early inflammatory arthritis? *Clin Exp Rheumatol.* 2003;21(5 Suppl 31):S37-S41.

65. Navarro-Compan V, van der HD, Ahmad HA, Miller CG, Wolterbeek R, Landewe R. Measurement error in the assessment of radiographic progression in rheumatoid arthritis (RA) clinical trials: the smallest detectable change (SDC) revisited. *Ann Rheum Dis.* 2014;73(6):1067-1070.

66. Clinical study report: BEYOND. JADY abbreviated clinical study report. A phase 3, multicenter study to evaluate the long-term safety and efficacy of baricitinib in patients with rheumatoid arthritis [CONFIDENTIAL manufacture's submission]. Indianapolis (IN): Eli Lilly and Company; 2015.

67. Clinical study report: BEYOND addendum. JADY abbreviated clinical study report addendum. A phase 3, multicenter study to evaluate the long-term safety and efficacy of baricitinib in patients with rheumatoid arthritis [CONFIDENTIAL manufacturer's submission]. Indianapolis (IN): Eli Lilly and Company; 2016 Feb 1.

68. Network meta-analysis to evaluate the clinical efficacy of rheumatoid arthritis treatments. In: CDR submission: Olumiant (baricitinib), 2 mg tablets [CONFIDENTIAL manufacturer's submission] Surrey (GB): Eli Lilly and Company; 2018 Sep.

69. Bae SC, Lee YH. Comparison of the efficacy and safety of tofacitinib and baricitinib in patients with active rheumatoid arthritis: a Bayesian network meta-analysis of randomized controlled trials. *Z Rheumatol.* 2018;06:06.

70. Choy E, Freemantle N, Proudfoot C, et al. Evaluation of the efficacy and safety of sarilumab combination therapy in patients with rheumatoid arthritis with inadequate response to conventional disease-modifying antirheumatic drugs or tumour necrosis factor alpha inhibitors: systematic literature review and network meta-analyses. *RMD Open.* 2019;5(1):e000798.

71. CDR submission: Olumiant (baricitinib), 2 mg tablets [CONFIDENTIAL manufacturer's submission]. Toronto (ON): Eli Lilly Canada Inc.; 2018 Dec 4.

72. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med.* 2019;162(11):777-784.

73. Taylor PC, Keystone EC, van der Heijde D, et al. Baricitinib versus placebo or adalimumab in rheumatoid arthritis. *N Engl J Med.* 2017;376(7):652-662.

74. National Institute for Health and Care Excellence. Baricitinib for moderate to severe rheumatoid arthritis. *(Technology appraisal guideance TA466)* 2017; https://www.nice.org.uk/guidance/ta466/chapter/3-Committee-discussion#indirect-comparison. Accessed 2019 May 28.

75. Drugs for the management of rheumatoid arthritis *(CADTH Health technology assessment)*. Ottawa (ON): CADTH; 2018: https://www.cadth.ca/drugs-management-rheumatoid-arthritis. Accessed 2019 May 28.

76. Ren S, Bermejo I, Simpson E, et al. Baricitinib for previously treated moderate or severe rheumatoid arthritis: an evidence review group perspective of a NICE single technology appraisal. *Pharmacoeconomics.* 2019;36(7):769-778.