Janus Kinase Inhibitors and Biosimilars for Rheumatoid Arthritis: Effectiveness and Value

Final Evidence Report and Meeting Summary

January 9, 2020

Prepared for

ICER Staff and Consultants	
Jeffrey A. Tice, MD Professor of Medicine University of California, San Francisco	**Serina Herron-Smith** Research Assistant, Evidence Synthesis Institute for Clinical and Economic Review
Varun M. Kumar, MBBS, MPH, MSc (Former) Associate Director of Health Economics Institute for Clinical and Economic Review	**Laura Cianciolo** Program Manager Institute for Clinical and Economic Review
Rick Chapman, PhD, MS Director of Health Economics Institute for Clinical and Economic Review	**Pamela Bradt, MD, MPH** Chief Scientific Officer Institute for Clinical and Economic Review
Judith Walsh, MD, MPH Professor of Medicine University of California, San Francisco	**Steven D. Pearson, MD, MSc** President Institute for Clinical and Economic Review

None of the above authors disclosed any conflicts of interest.

DATE OF PUBLICATION: January 9, 2020

How to cite this document: Tice J, Kumar V, Chapman R, Walsh J, Herron-Smith S, Cianciolo L, Bradt P, Pearson S. Janus Kinase Inhibitors and Biosimilars for Rheumatoid Arthritis. Institute for Clinical and Economic Review, January 9, 2020. https://icer-review.org/material/ra-update-evidence-report/.

Jeffrey A. Tice served as the lead author for the report, led the systematic review and authorship of the comparative clinical effectiveness section, and wrote the background, other benefits, and contextual considerations sections of the report. Varun M. Kumar developed the cost-effectiveness model and potential budget impact analysis and authored the corresponding sections of the report. Laura Cianciolo authored the section on coverage policies, managed the timeline and public process, and performed quality controls. Eric Borrelli authored the section on clinical guidelines. Pamela Bradt and Steven D. Pearson provided methodologic guidance on the clinical and economic evaluations. We would like to thank Rick Chapman, Noemi Fluetsch, Foluso Agboola, David Rind, and Patty Synnott for their contributions to this report. We would also like to thank Jordan Amdahl, Rebecca Bornheimer, and Gerry Oster from Policy Analysis Inc. for their technical support with hēRo3 throughout the course of this review.

About ICER

The Institute for Clinical and Economic Review (ICER) is an independent non-profit research organization that evaluates medical evidence and convenes public deliberative bodies to help stakeholders interpret and apply evidence to improve patient outcomes and control costs. Through all its work, ICER seeks to help create a future in which collaborative efforts to move evidence into action provide the foundation for a more effective, efficient, and just health care system. More information about ICER is available at http://www.icer-review.org.

The funding for this report comes from government grants and non-profit foundations, with the largest single funder being the Laura and John Arnold Foundation. No funding for this work comes from health insurers, pharmacy benefit managers, or life science companies. ICER receives approximately 19% of its overall revenue from these health industry organizations to run a separate Policy Summit program, with funding approximately equally split between insurers/PBMs and life science companies. Life science companies relevant to this review who participate in this program include Genentech, Johnson & Johnson, and Sanofi. For a complete list of funders and for more information on ICER's support, please visit http://www.icer-review.org/about/support/.

About CTAF

The California Technology Assessment Forum (CTAF) – a core program of ICER – provides a public venue in which the evidence on the effectiveness and value of health care services can be discussed with the input of all stakeholders. CTAF seeks to help patients, clinicians, insurers, and policymakers interpret and use evidence to improve the quality and value of health care.

The CTAF Panel is an independent committee of medical evidence experts from across California, with a mix of practicing clinicians, methodologists, and leaders in patient engagement and advocacy. All Panel members meet strict conflict of interest guidelines and are convened to discuss the evidence summarized in ICER reports and vote on the comparative clinical effectiveness and value of medical interventions. More information about CTAF is available at https://icer-review.org/programs/ctaf/.

The findings contained within this report are current as of the date of publication. Readers should be aware that new evidence may emerge following the publication of this report that could potentially influence the results. ICER may revisit its analyses in a formal update to this report in the future.

The economic models used in ICER reports are intended to compare the clinical outcomes, expected costs, and cost effectiveness of different care pathways for broad groups of patients. Model results therefore represent average findings across patients and should not be presumed to represent the clinical or cost outcomes for any specific patient. In addition, data inputs to ICER models often come from clinical trials; patients in these trials and provider prescribing patterns may differ in real-world practice settings.

In the development of this report, ICER's researchers consulted with several clinical experts, patients, manufacturers, and other stakeholders. The following clinical experts provided input that helped guide the ICER team as we shaped our scope and report. None of these individuals is responsible for the final contents of this report or should be assumed to support any part of this report, which is solely the work of the ICER team and its affiliated researchers.

For a complete list of stakeholders from whom we requested input, please visit: https://icer-review.org/material/ra-update-stakeholder-list/.

Expert Reviewers

Andrew L. Concoff, MD, FACR, CAQSM
Executive Vice President, Chief Value Medical Officer, United Rheumatology

Dr. Andrew L. Concoff has served as a consultant and/or speaker to Flexion Pharmaceuticals, Exagen, Inc., and UCB.

Christopher Phillips, MD
Chair, American College of Rheumatology Insurance Subcommittee
Rheumatologist, Paducah Rheumatology

Dr. Christopher Phillips has served as a principal investigator on several Rinvoq™ trials conducted through an independent multi-specialty clinical research center.

Angus B. Worthing, MD, FACP, FACR
Chair, American College of Rheumatology Committee on Government Affairs

No relevant conflicts of interest to disclose, defined as more than $10,000 in health care company stock or more than $5,000 in honoraria or consultancies during the previous year from health care manufacturers or insurers.

We would also like to thank Andrew Laster, MD, FACR, for the guidance and feedback he provided throughout the duration of the review.

Expert Health Economics Reviewer

Matthew Stevenson, PhD, BSc
Professor of Health Technology Assessment, School of Health and Related Research (ScHARR), University of Sheffield

No relevant conflicts of interest to disclose, defined as more than $10,000 in health care company stock or more than $5,000 in honoraria or consultancies during the previous year from health care manufacturers or insurers.

Table of Contents

List of Acronyms Used in this Report

ACPA	Anticitrullinated protein antibody
ACR	American College of Rheumatology
AHRQ	Agency for Healthcare Research and Quality
AIC	Akaike Information Criteria
CDAI	Clinical Disease Activity Index
CI	Confidence interval
CRP	C-reactive protein
CMS	Centers for Medicare & Medicaid Services
DAS28	Disease Activity Score 28
DIC	Deviance information criterion
DMARD	Disease-modifying antirheumatic drug
EQ-5D	EuroQol
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
evLYG	Equal Value of Life Years Gained
FDA	Food and Drug Administration
HAQ	Health Assessment Questionnaire
HAQ-DI	Health Assessment Questionnaire for Rheumatoid Arthritis Disability Index
HCPCS	Healthcare Common Procedure Coding System
IL	Interleukin
IVI	Innovation Value Initiative
JAK	Janus kinase
KM	Kaplan-Meier
LY	Life year
MDHAQ	Multidimensional Health Assessment Questionnaire
MRI	Magnetic resonance imaging
NMA	Network meta-analysis
NNT	Number needed to treat
OR	Odds ratio
PAS	Patient Activity Scale
PICOTS	Population, Intervention, Comparators, Outcomes, Timing, Settings
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROMIS	Patient-Reported Outcomes Measurement Information System
QALY	Quality-adjusted life year
RA	Rheumatoid arthritis
RAPID	Routine Assessment of Patient Index Data
RCT	Randomized controlled trial
RF	Rheumatoid factor
SDAI	Simplified Disease Activity Index
SF-36	Short Form Survey 36
SMD	Standardized mean difference
TIM	Targeted immune modulator
TNF	Tumor necrosis factor
UK	United Kingdom
US	United States
USPSTF	United States Preventive Services Task Force
VAS	Visual analog scale
WAC	Wholesale acquisition cost
WTP	Willingness to pay

Executive Summary

Background

Rheumatoid arthritis (RA) is the most common autoimmune inflammatory arthritis in adults, affecting between 1.3 and 1.8 million Americans.[1] RA is more common in women and may occur at any age, with peak incidence occurring at ages 50-60 years.[2] RA is typically characterized by morning stiffness and symmetrical joint swelling of the feet, hands, and knees, although any joint (and in some cases, internal organs and skin) may be involved.[2]

The chemotherapeutic agent methotrexate is the most widely used conventional disease-modifying antirheumatic drug (DMARD); it is considered an "anchor drug" because of its effectiveness and tolerability as well as its potential to enhance the effectiveness of biologic and non-biologic drugs that are targeted at certain mediators of inflammation in RA, known collectively as targeted immune modulators (TIMs).[2] However, only about 50% of patients treated with methotrexate alone will receive sufficient reduction in disease activity or remission of symptoms.[2]

In our 2017 review (Targeted Immune Modulators for Rheumatoid Arthritis: Effectiveness & Value), the Institute for Clinical and Economic Review (ICER) assessed the relative effectiveness and value of all available TIMs at that time. Since that report two additional Janus kinase (JAK) inhibitors, baricitinib (Olumiant®, Eli Lilly and Company) and upadacitinib (Rinvoq™, AbbVie), received United States Food and Drug Administration (FDA) approval. In this review, ICER is updating the evidence for JAK inhibitors for adults with moderately-to-severely active RA.

In a separate section, we evaluate the evidence supporting infliximab-dyyb (Inflectra®, Pfizer) as a biosimilar for the reference product infliximab (Remicade®, Janssen) for the treatment of RA. Infliximab-dyyb is intended to serve as an exemplar to ground a Policy Roundtable discussion at the public meeting about the role of biosimilars in the United States (US).

Insights Gained from Discussions with Patients and Patient Groups

Patients and advocacy organizations emphasized the long-term nature of the disease and the importance of both the long-term perspective and the variability in disease course and treatment changes, including drug holidays. These groups also highlighted the impact of RA on caregivers and suggested that both caregiver measures and outreach to caregiver groups be part of the review process. Finally, stakeholders highlighted the important progress that has been made through the use of biologics: very few patients progress to disabling joint deformities with current treatments.

Even with the availability of TIMs in addition to conventional DMARDs, patients are not satisfied with their treatment.[3] A recent online survey of 232 patients with RA over the age of 21 who had

been failed by at least one DMARD assessed treatment satisfaction. Almost half of patients (43%) reported daily or almost daily use of prescription pain medication and 44% reported a current flare of their RA. Only 26% reported satisfaction with their treatment, even though 67% were currently on a biologic DMARD.[3] Symptoms that were poorly addressed by current treatments included joint stiffness and sexual dysfunction in 70% to 80% of patients as well as fatigue, sleep disturbance, and joint pain.

The financial burden of RA treatment on patients and their families is also substantial. Patients noted that manufacturers have increased their use of coupons and other copayment assistance programs, but that financial problems remain significant and are not limited to out-of-pocket costs. Issues with coordination of care, navigation of insurance requirements by both patient and provider, lost time at work or school, and other challenges contribute to patient and family impact.

Potential Cost-Saving Measures in RA

Two of the ACR Choosing Wisely® recommendations apply.[4]

Don't perform MRI [magnetic resonance imaging] of the peripheral joints to routinely monitor inflammatory arthritis.

Data evaluating MRI for the diagnosis and prognosis of RA are currently inadequate to justify widespread use of this technology for these purposes in clinical practice.

Don't prescribe biologics for RA before a trial of methotrexate (or other conventional non-biologic DMARDs).

High-quality evidence suggests that methotrexate and other conventional non-biologic DMARDs are effective in many patients with RA. Initial therapy for RA should be a conventional non-biologic DMARDs unless these are contraindicated.

Comparative Clinical Effectiveness

Low disease activity is the most important outcome under the treat-to-target paradigm in which treatment switching is encouraged within three months for patients with ongoing moderate-to-severe disease activity.[5] There are multiple measures of disease activity including Disease Activity Score 28 (DAS28-c-reactive protein or erythrocyte sedimentation rate [CRP or ESR]), Clinical Disease Activity Index (CDAI), and Simple Disease Activity Index (SDAI). In studies that report all four measures, rates of remission and low disease activity with DAS28-ESR are consistently lower than those assessed with DAS-CRP. Disease activity and remission using SDAI or CDAI are usually comparable to each other, but classify fewer patients with low disease activity or remission compared with DAS28 measures. DAS28-CRP may overestimate response to therapy for drugs that lower interleukin (IL)-6 activity (tocilizumab and JAK inhibitors) because IL-6 increases CRP levels.

Because of inconsistent reporting, we were unable to perform a network meta-analysis (NMA) for any of the four measures of disease activity. We were able to perform an NMA for the ACR20, ACR50, and ACR70 (American College of Rheumatology) measures, which reflect the percentage reduction in the number of tender/swollen joints.

Clinical Benefits

The results are organized by indication. First, we consider patients who are predominantly TIM-naïve. Some of these trials included up to 20% of patients who had been failed by a TIM (see Appendix D for details). Since baricitinib is not indicated for this population, no trials of baricitinib were included. Findings from head-to-head studies of JAK inhibitors with adalimumab are also presented for the population of TIM-naïve/mixed population. In the second section, we consider the TIM-experienced population. There were no head-to-head trials for this population. For each JAK inhibitor, we describe results according to their use in combination with conventional DMARDs.

TIM-Naïve/Mixed Populations

Comparisons to Conventional DMARD Therapy

Both upadacitinib and tofacitinib generated superior improvements in disease activity, remission, and ACR response relative to conventional DMARD therapy alone in TIM-naïve/mixed populations at 12 weeks. These results were consistent when reported at 24 and 48 weeks. Radiographic progression was also reduced, but differences in measures used made comparisons across studies difficult. Improvements in function and disability were statistically superior for both upadacitinib and tofacitinib. A greater proportion of patients receiving JAK inhibitors met clinically important thresholds for Health Assessment Questionnaire Disability Index (HAQ-DI) change.

The proportions of patients achieving low disease activity or remission at 12 weeks were substantially greater in JAK inhibitor groups relative to conventional DMARDs alone (Table ES1). Results achieved statistical significance for both upadacitinib and tofacitinib.

Table ES1. Disease Activity Outcomes of JAK Inhibitors and Comparators in TIM-Naïve/Mixed Patients at 12 Weeks/Three Months

Treatment	N	DAS28-ESR or CRP	DAS28 Change from Baseline (Mean)	DAS28 Low Disease Activity (%)	DAS28 Remission (%)	CDAI Low Disease Activity (%)	SDAI Low Disease Activity (%)
SELECT-COMPARE[6]							
Upadacitinib + MTX	651	DAS28-CRP	NR	45*	29*	40*	40*
Adalimumab + MTX	327	DAS28-CRP	NR	29	18	30	30
Placebo + MTX	651	DAS28-CRP	NR	14	6	16	15
SELECT-NEXT[7]							
Upadacitinib + MTX	141	DAS28-CRP	NR	48*	31*	40*	42*
Placebo + MTX	79	DAS28-CRP	NR	17	10	19	19
ORAL Sync[8]							
Tofacitinib + MTX	315	DAS28-ESR	NR	NR	8*	NR	NR
Placebo + MTX	159	DAS28-ESR	NR	NR	0.5	NR	NR
ORAL Standard[9]							
Tofacitinib + MTX	204	DAS28-ESR	NR	NR	NR	NR	NR
Adalimumab + MTX	204	DAS28-ESR	NR	NR	NR	NR	NR
Placebo + MTX	108	DAS28-ESR	NR	NR	NR	NR	NR
ORAL Strategy[10]							
Tofacitinib + MTX	376	DAS28-ESR	NR	NR	NR	NR	NR
Adalimumab + MTX	386	DAS28-ESR	NR	NR	NR	NR	NR
ORAL Scan[11]							
Tofacitinib + MTX	321	DAS28-ESR	NR	NR	NR	NR	NR
Placebo + MTX	160	DAS28-ESR	NR	NR	NR	NR	NR
Pooled Tofacitinib Trials[12]							
Tofacitinib + MTX	1,043	DAS28-ESR	NR	16.6*	7.3*	32.4*	34.6*
Placebo + MTX	638	DAS28-ESR	NR	4.5	2.3	14.3	14.2

CDAI: Clinical Disease Activity Index, CRP: c-reactive protein, DAS28: Disease Activity Score 28, ESR: erythrocyte sedimentation rate, MTX: methotrexate, NR: not reported, SDAI: Simplified Disease Activity Index
*p<0.001.

There were no marked differences in ACR responses between upadacitinib and tofacitinib across the trials, though the changes in HAQ-DI scores were slightly greater for upadacitinib. Again, these comparisons are not head-to-head and are subject to potential selection and measurement bias. The results of our NMA for ACR categories are summarized in Table ES2 on the following page. The results for tofacitinib and adalimumab are very similar and those for upadacitinib are slightly better (more patients in the ACR50 and 70 categories, fewer than ACR20). All three TIMs had markedly better results than continuing conventional DMARDs alone in patients who had been failed by conventional DMARDs.

Table ES2. NMA-Derived Proportions of Patients in Each ACR Response Category for JAK Inhibitors and Comparators in TIM-Naïve/Mixed Patients at 12 Weeks/Three Months

Treatment	ACR<20	ACR20-50	ACR50-70	ACR70-100
Upadacitinib + cDMARD	33.4%	28.7%	12.8%	25.0%
Tofacitinib + cDMARD	40.5%	28.6%	11.5%	19.4%
Adalimumab+ cDMARD	41.1%	28.5%	11.4%	19.0%
Placebo + cDMARD	72.9%	18.2%	4.5%	4.3%

ACR: American College of Rheumatology, cDMARD: conventional disease-modifying antirheumatic drug

TIM-Experienced Population

Studies for all three JAK inhibitors demonstrated statistically and clinically significant improvements in measures of disease activity, ACR response, and HAQ improvement versus conventional DMARDs alone, but there were fewer trials with fewer participants, so the confidence intervals are wider than those for JAK inhibitors in the TIM-naïve population.

The evidence was similar for all three JAK inhibitors: significantly greater proportions of patients randomized to JAK inhibitors plus conventional DMARDs achieved low disease activity and remission at three and six months by multiple measures of disease activity compared with conventional DMARDs alone. They also had greater proportions of patients meeting the ACR20/50/70 response levels and greater improvements in the HAQ-DI compared with conventional DMARDs alone (Table ES3).

Table ES3. Disease Activity Outcomes of JAK Inhibitors and Comparators in TIM-Experienced Patients at 12 Weeks/Three Months

Treatment	N	DAS28-ESR or CRP	DAS28 Change from Baseline (Mean)	DAS28 Low Disease Activity (%)	DAS28 Remission (%)	CDAI Low Disease Activity (%)	SDAI Low Disease Activity (%)
SELECT-BEYOND[13]							
Upadacitinib + MTX	164	DAS28-CRP	NR	43*	NR	32*	34*
Placebo + MTX	169	DAS28-CRP	NR	14	NR	14	14
ORAL Step[14]							
Tofacitinib + MTX	133	DAS28-CRP DAS28-ESR	NR -1.8*	13 14.3†	6 6.7†	NR NR	NR NR
Placebo + MTX	159	DAS28-CRP DAS28-ESR	NR -0.7	4 5	1 1.7	NR NR	NR NR
Pooled Tofacitinib Trials[12]							
Tofacitinib + MTX	258	DAS28-ESR	NR	12.7†	6.6†	29.5*	29.8*
Placebo + MTX	191	DAS28-ESR	NR	5.1	2.3	14.4	13.8
RA-BEACON[15]							
Baricitinib + MTX	174	DAS28-CRP DAS28-ESR	-1.5*	24* 13‡	11† 6‡	24‡	22*
Placebo + MTX	176	DAS28-CRP DAS28-ESR	-0.7	9 4	4 1	11	9

CDAI: Clinical Disease Activity Index, CRP: c-reactive protein, DAS28: Disease Activity Score 28, ESR: erythrocyte sedimentation rate, MTX: methotrexate, NR: not reported, SDAI: Simplified Disease Activity Index

*p <0.001.

†p<0.05 vs. placebo.

‡p<0.01.

Harms

Rates of short-term serious adverse events (within six months) were generally comparable across all treatments, including JAK inhibitors, adalimumab, and conventional DMARDs. Infections (e.g., upper respiratory tract infections, bronchitis, nasopharyngitis) were the most common adverse events during treatment. Based on long-term (one year or more) trial data, upadacitinib, tofacitinib, and baricitinib showed comparable overall safety profiles.

Data on adverse events, discontinuations due to adverse events as well as specific adverse events of interest observed in clinical trials with conventional DMARD controls are presented in Table ES4. The most frequently reported adverse events were mild infections (upper respiratory tract infections, bronchitis, nasopharyngitis). The overall incidence of serious infections, deaths, and all serious adverse events was comparable among treatments, including conventional DMARD therapy.

Table ES4. Harms in TIM-Naïve/Mixed Populations

Intervention	N	Time Point (Weeks)	Any AEs	Serious AEs	AE Leading to D/C	VTE	Malignancy	Death	Serious	Opportunistic	Herpes Zoster Virus
SELECT-COMPARE											
UPA 15 mg + cDMARD	651	26	417 (64.2)	24 (3.7)	23 (3.5)	All: 2 (0.3); PE: 1 (0.2); DVT: 1 (0.2)	0 (0)	0 (0)	12 (1.8)	4 (0.6)	5 (0.8)
ADA 40 mg + cDMARD	327	26	197 (60.2)	14 (4.3)	20 (6.1)	All: 3 (0.9); PE: 3 (0.9); DVT: 0 (0)	1 (0.3)	2 (0.6)	5 (1.5)	1 (0.3)	1 (0.3)
PBO + cDMARD	651	26	347 (53.2)	19 (2.9)	15 (2.3)	All: 1 (0.2); PE: 1 (0.2); DVT: 0 (0)	2 (0.3)	2 (0.3)	5 (0.8)	4 (0.6)	3 (0.5)
SELECT-NEXT											
UPA 15 mg + cDMARD	221	12	125 (56.6)	9 (4.1)	7 (3.2)	0 (0)	0 (0)	0 (0)	1 (<1)	0 (0)	3 (1)
UPA 30 mg + cDMARD	219	12	118 (53.9)	6 (2.7)	13 (5.9)	0 (0)	2 (0.9)	1 (0.5)	3 (1)	3 (1)	6 (3)
cDMARD	221	12	108 (48.9)	5 (2.3)	7 (3.2)	0 (0)	0 (0)	1	1 (<1)	1 (<1)	1 (<1)
ORAL Sync											
TOF 5 mg + cDMARD	388	52	171.9 (152.5-193.8)	6.9 (4.6-10.5)	6.2 (4.0-9.6)	NR	NR	NR	NR	NR	9 (4.2) MTX w/o LEF
PBO (Advanced to TOF 5 mg at 6 Months)	159	52	342.3 (281.1-416.9)	10.9 (4.9-24.2)	5.4 (1.8-16.8)	NR	NR	NR	NR	NR	2 (6.8) MTX alone
ORAL Scan											
TOF 5 mg + cDMARD	321	12	TEAE: 157 (48.9)	12 (3.7)	15 (4.7)	NR	NR	2 (0.6)	2 (0.6)	NR	3 (0.9)

Intervention	N	Time Point (Weeks)	Any AEs	Serious AEs	AE Leading to D/C	VTE	Malignancy	Death	Serious	Opportunistic	Herpes Zoster Virus
PBO (Advanced to TOF 5 mg at 6 Months)	160	12	TEAE: 73 (45.6)	5 (3.1)	5 (3.1)	NR	NR	0 (0)	0 (0)	NR	0 (0)
ORAL Strategy											
TOF 5 mg	384	52	226 (59)	35 (9)	23 (6)	NR	1 (<1)	2 (1)	6 (2)	2 (1)	1/69 (1)
TOF 5 mg + cDMARD	376	52	231 (61)	27 (7)	26 (7)	NR	0 (0)	0 (0)	10 (3)	1 (<1)	2/75 (3)
ADA + cDMARD	386	52	253 (66)	24 (6)	37 (10)	NR	0 (0)	0 (0)	6 (2)	2 (1)	0/27 (0)
ORAL Standard											
TOF 5 mg + cDMARD	204	0-12	135 (66.2)	12 (5.9)	14 (6.9)	NR	NR	1 (<1)	3(1.5)	NR	0 (0)
ADA 40 mg + cDMARD	204	0-12	105 (51.5)	5 (2.5)	10 (4.9)	NR	NR	1 (<1)	0 (0)	NR	0 (0)
PBO Followed by TOF 5 mg + cDMARD	56	0-12	57(52.8)	2 (1.9)	3 (2.8)	NR	NR	NR	1(0.9)	NR	0 (0)
Pooled Tofacitinib											
bDMARD Naïve TOF 5 mg	1071	0-24	NR	131 (12.2)	96 (9.1)	NR	NR	7 (0.6)	32 (3.4)	NR	43 (4.0)
bDMARD Naïve cDMARD	651	0-6	NR	98 (15.0)	66 (10.1)	NR	NR	5 (0.7)	13 (2.0)	NR	13 (2.0)
bDMARD-IR TOF 5 mg	259	0-24	NR	34 (13.0)	39 (14.8)	NR	NR	1(1.2)	6 (2.3)	NR	13 (5.4)
bDMARD-IR cDMARD	193	0-6	NR	37 (19.0)	37 (18.9)	NR	NR	0 (0)	0 (0)	NR	10 (5.4)

ADA: adalimumab, AE: adverse event, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, D/C: discontinuation, LEF: leflunomide, mg: milligram, MTX: methotrexate, N: number, PBO: placebo, PE: pulmonary embolism, TOF: tofacitinib, UPA: upadacitinib, VTE: venous thromboembolism

All three JAK inhibitors carry black-box warnings for serious infections, lymphoma, testing for latent tuberculosis prior to initiating therapy, and monitoring for active tuberculosis. In addition, baricitinib and upadacitinib carry black box warnings regarding thrombosis, including deep venous thrombosis, pulmonary embolism, and arterial thrombosis.

A systematic review and meta-analysis of 21 RCTs of JAK inhibitors including 11,144 patients looked specifically at rates of serious infections and herpes zoster.[16] The review found that the absolute risk of serious infections (hospitalization, death, intravenous [IV] antibiotics) was low and was not significantly higher when compared to the placebo group. Similarly, there was an increase in rates of herpes zoster infections, but the increase was not statistically significant.

Thrombotic events were rare. A systematic review reported that the annual incidence rate for thromboembolic events was 0.6 (0.4-0.9) for tofacitinib and 0.5 (0.3-0.7) for baricitinib.[17] In SELECT-COMPARE, 0.3% of patients in the upadacitinib group had a thrombotic event compared with 0.9% in the adalimumab group.

Controversies and Uncertainties

Across the RCTs identified for this review, only three were based on head-to-head comparisons of the TIMs of interest and none were head-to-head comparisons of JAK inhibitors. The paucity of trial data and the differences in reported measures of disease activity at three months precluded using an NMA to combine direct and indirect evidence of efficacy. Often the studies reported outcome measures at six months even though patients were eligible for rescue therapy and/or treatment-arm crossover 12-24 weeks after randomization. Since guidelines increasingly recommend treatment-switch decisions within three months of initiating therapy, the three-month outcomes would have been more clinically relevant.

Extending trial-based analyses to longer timepoints requires imputation in many instances, which affects the level of confidence in the results no matter how responsibly it is done. In addition, key outcome measures such as disease activity scores, remission criteria, and modified Sharp score have undergone substantial revision and modification over the years, are employed variably in clinical trials, and not measured in others, making cross-trial comparisons problematic.

Finally, although the introduction of TIMs has transformed clinical practice and improved the quality of life and functional capacity of many patients, there are still unanswered questions, including the relationship between levels of disease activity and radiographic evidence of joint damage, whether there are patient or clinical factors that predict response to specific therapies, and the totality of the disease impact on patients, families, and caregivers. As noted in Section 1 of this report, patient groups do not feel that the current tools for patient-reported outcomes sufficiently capture their experience, but to date no new instruments have been accepted into common use in clinical trials.

Summary and Comment

Using the ICER Evidence Rating Matrix, our evidence ratings for selected comparisons of interest are provided in Table ES5 for patients with moderately-to-severely active RA who have had an inadequate response to prior conventional DMARD therapy. As described above, findings of studies using conventional DMARDs as the control indicate clinically and statistically significant improvements in most important disease measures for upadacitinib and tofacitinib combination therapy, so both receive a letter grade of "A" (high certainty of substantial net health benefit) relative to conventional DMARD therapy alone. However, there is a paucity of evidence on baricitinib 2 mg daily in this population and it does not have an FDA indication for this population, so we judge the comparative clinical effectiveness of baricitinib to be insufficient ("I").

Table ES5. Evidence Ratings for Comparative Clinical Effectiveness: Selected Comparisons

Regimen Type/Comparison	Intervention	Comparator	Rating
TIM-Naïve Population			
Compared to cDMARD	Upadacitinib	cDMARDs	A
	Tofacitinib	cDMARDs	A
	Baricitinib	cDMARDs	I
Head-to-Head	Upadacitinib	Adalimumab	B+
	Tofacitinib	Adalimumab	C
	Baricitinib	Adalimumab	I
Biosimilar*	Infliximab-dyyb	Infliximab	C
TIM-Experienced Population			
Combination with cDMARD	Upadacitinib	cDMARDs	B+
	Tofacitinib	cDMARDs	B+
	Baricitinib	cDMARDs	B+

cDMARD: conventional disease-modifying antirheumatic drug, TIM: targeted immune modulator
*Rationale behind the biosimilar evidence rating may be found on Page ES22.

Single RCTs have also evaluated combination therapy regimens of both upadacitinib and tofacitinib plus conventional DMARDs in head-to-head comparison with adalimumab plus conventional DMARDs in the TIM-naïve population. In the SELECT-COMPARE study, upadacitinib plus methotrexate was associated with statistically-significantly but modestly higher rates of disease remission, ACR response, change in pain, and improvement in HAQ-DI. The difference in benefits was smaller than those seen in comparison to conventional DMARDs alone. Rates of serious harm or discontinuation due to adverse events were also similar, so we judge the evidence for combination therapy with upadacitinib versus adalimumab to represent an incremental or better net health benefit ("B+"). There were no significant differences in clinical outcomes between combination regimens using tofacitinib versus adalimumab in two trials, although there was a trend towards more patients randomized to tofacitinib achieving ACR70 and having greater improvements in the HAQ-DI. We therefore assign a net health benefit rating of comparable ("C")

for this comparison. Finally, there is no evidence on baricitinib 2 mg daily versus adalimumab in this population, so we judge the comparative clinical effectiveness of baricitinib to be insufficient ("I").

For each of the JAK inhibitors, there is one randomized trial comparing combination therapy to conventional DMARDs in the TIM-experienced population. All three trials are smaller than those for the TIM-naïve population and the effect sizes are also somewhat smaller with wider confidence intervals. As in the TIM-naïve population, rates of serious harms and discontinuation due to adverse events were low, so we judge the evidence for combination therapy with each of the three JAK inhibitors versus conventional DMARDs alone to represent an incremental or better net health benefit ("B+").

There is much greater uncertainty in assessing the relative comparative clinical effectiveness of JAK inhibitors, which have never been compared head-to-head in a randomized setting. The individual clinical trials had somewhat different patient populations, primary endpoints, timing of assessments, and timing of allowable switching to alternative therapies. As a result, we judge there to be insufficient evidence ("I") to differentiate among JAK inhibitors.

Long-Term Cost Effectiveness

Model Overview

The primary aim of this analysis is to estimate the cost effectiveness of JAK inhibitors for patients with severely active RA using a decision analytic model. The model's objective was to compare each of the three JAK inhibitors, tofacitinib, baricitinib, and upadacitinib, to adalimumab, a TNF inhibitor. While other TIMs indicated for first-line treatment are also appropriate choices for providers and patients, we chose adalimumab as a comparator due to its extensive use in clinical practice for the treatment of RA, and because it was directly compared to JAK inhibitors in clinical trials. Unfortunately, we were unable to directly compare tofacitinib to adalimumab due to inadequate data in the TIM-naïve or TIM-experienced population, and likewise we were unable to compare upadacitinib to adalimumab in the TIM-experienced population. Because the labeled indication of baricitinib is the treatment of patients for whom TNF inhibitors have failed, we attempted to compare it to adalimumab in the TIM-experienced population, but we were unable to do so due to a lack of comparable data. In addition, the goal of evaluating an exemplar biosimilar was to set the stage for a Policy Roundtable discussion, not to perform an exhaustive review of the biosimilars. Thus, we do not evaluate the cost-effectiveness of the biosimilars.

We developed a *de novo* decision analytic model for this evaluation, informed by key clinical trials and prior relevant economic models. While this model was informed by ICER's 2017 review, we made several changes to reflect current clinical practice in RA. The results of this model should not be directly compared to the results obtained in 2017, as the current model includes substantial

changes from the cost-effectiveness model in the prior report. Costs and outcomes in this model were discounted at 3% per year.

The base-case analysis takes a health care sector perspective (i.e., focuses on direct medical care costs only) and uses a one-year time horizon. We model these treatments over a one-year time horizon due to the uncertainty surrounding the number of subsequent lines of TIMs as well as when patients transition to palliative care. The model was developed in hēRo3℠, with some components, such as survival distributions, developed in RStudio (Version 1.1.463).

The primary model focused on a hypothetical cohort of patients with severe RA for whom prior treatment with conventional DMARDs had failed. Upon model entry, the hypothetical patient cohort was initiated on a treatment strategy, with treatment response assessed at three months. In all analyses, a TIM was added to a conventional DMARD, such as methotrexate. Treatment switching was based on disease activity as measured by the DAS28-CRP value, with those in remission and with low disease activity remaining on the same treatment after the first three months, while those with moderate/high disease activity switch to a subsequent market basket average of therapy at the end of the first three-month cycle.

After initiating treatment with a TIM, the model relates the DAS28-based response to the HAQ after three months of therapy. We used a mapping algorithm from EULAR to HAQ, assuming remission as defined by DAS28 as equivalent to "Good" response, low disease activity as equivalent to "Moderate" response, and moderate disease activity and high disease activity as equivalent to "None" on the EULAR scale. The HAQ score was then linked to utility, mortality, hospitalizations, and productivity. Simulated utility scores and mortality were used to calculate the quality-adjusted life years (QALYs) gained, with hospitalization costs and productivity loss costs contributing to the health care sector perspective and societal perspective analyses.

Key Model Characteristics and Assumptions

Our model includes several assumptions, stated below.

Table ES6. Modeling Assumptions

Assumption	Rationale
The TIMs chosen for the second-line market basket were baricitinib (RA-BUILD and RA-BEACON), adalimumab (RA-BEAM), etanercept (APPEAL), tofacitinib (ORAL Step), golimumab (GO-AFTER), and upadacitinib (SELECT-NEXT).[7,14,15,18-21]	The choice of TIMs to include in the market basket average was based on availability of DAS28 data at three months after initiating a TIM. We chose to include data for treatments from line one as well because they could be considered in the second-line market basket of TIMs in the treatment arms where they are not included as line one. However, we included data for these drugs only from trials not informing efficacy in line one.
We assumed that the efficacy of the market basket of TIM treatment was 84% of its calculated average efficacy across the TIMs included.	Prior published data shows a mean 16% reduction in treatment efficacy following failure by current therapy in RA. While this was specific to switching to another TNF inhibitor after failure by a prior one, we apply this reduction to JAK inhibitors following failure.[22,23] We apply this reduced efficacy only once to estimate a readjusted efficacy for the market basket, and do not apply a 16% reduction each time a failure with a line of therapy occurs.
We assumed conversion ratios of 2x and 1.5x for DAS28-ESR to DAS28-CRP to derive the proportions in remission and low disease activity for tofacitinib and its conventional DMARD comparator.	Some trials (SELECT-MONOTHERAPY, RA-BUILD, RA-BEACON, and ORAL Step) simultaneously reported DAS28-ESR and DAS28-CRP outcomes for upadacitinib, baricitinib, and tofacitinib, and their respective comparators. An average of the disease activity proportions using DAS28-CRP and DAS28-ESR data was used to estimate an approximate 2x and 1.5x ratio of DAS28-CRP to DAS28-ESR in the TIM arms, while the conventional DMARD arms showed more variability. In the absence of DAS28-CRP trial data at three months, we applied these ratios to the DAS28-ESR data to derive DAS28-CRP data at three months for tofacitinib and its conventional DMARD comparator. The uncertainty surrounding this was tested in a sensitivity analysis.

Model Inputs

Inputs on the proportion of patients with different levels of disease activity have been derived from individual trials for relevant interventions and comparators. We were unable to make a direct

comparison between tofacitinib and adalimumab in our analyses due to a lack of comparable efficacy data. Our comparison was therefore restricted to only upadacitinib versus adalimumab. However, we comment on the value of tofacitinib and adalimumab based on the cost effectiveness of each of these TIMs when compared to conventional DMARDs in their respective trials.

Table ES7. Treatment Response at Three Months using DAS28

	Proportion of Patients Achieving Different Categories of Disease Activity by DAS28 at Three Months*		
	<2.6 (Remission)	2.6 to ≤3.2 (LDA)	>3.2 (MDA and HDA)
Upadacitinib + cDMARD	29%	16%	55%
Adalimumab + cDMARD	18%	11%	71%
Second-Line Market Basket of TIMs†	22%	14%	64%

cDMARD: conventional disease-modifying antirheumatic drug, DAS28: Disease Activity Score 28, HDA: high disease activity, LDA: low disease activity, MDA: moderate disease activity, TIM: targeted immune modulator

*Mutually exclusive categories.

†Values prior to applying a 0.84 multiplier to reflect lower efficacy after failure by primary treatment.

Table ES8. Treatment Efficacy Estimates for Adalimumab, Tofacitinib, and Their Respective Conventional DMARD Comparators at Three Months

	Proportion of Patients Achieving Different Categories of Disease Activity by DAS28 at Three Months*		
	<2.6 (Remission)	2.6 to ≤3.2 (LDA)	>3.2 (MDA and HDA)
Tofacitinib + cDMARD	15%	14%	71%
Adalimumab + cDMARD	18%	11%	71%
cDMARD†	6%	8%	86%
cDMARD‡	5%	3%	92%
Second-Line Market Basket of TIMs§	22%	14%	64%

cDMARD: conventional disease-modifying antirheumatic drug, DAS28: Disease Activity Score 28, HDA: high disease activity, LDA: low disease activity, MDA: moderate disease activity

*Mutually exclusive categories.

†Versus adalimumab.

‡Versus tofacitinib.

§Values prior to applying a 0.84 multiplier to reflect lower efficacy after failure by primary treatment.

Drug costs in the model are shown in Table ES9. Because upadacitinib was recently approved, we did not find any estimates on its net price, and assumed it would be discounted by 26%, the average discount of the other two JAK inhibitors, to estimate its net price. For the second-line market basket of TIMs, we estimated the cost of TIMs based on a retrospective observational study with prescription market share data that included estimates on eight TIMs.[24] Societal costs in our model are generated from additional costs from unemployment due to RA.

Table ES9. Drug Costs

Drug	WAC per Unit	Discount from WAC	Net Price per Unit	Annual WAC	Annual Net Price
Upadacitinib – 15 mg Tab	$163.89	26%*	$120.56	$59,860	$44,035
Tofacitinib – 5 mg Tab	$74.68	34%	$49.50	$54,552	$36,159
Baricitinib – 2 mg Tab	$71.23	19%	$57.59	$26,017	$21,033
Adalimumab – 40 mg/0.8 ml Sol	$2,587.05	34%	$1,696.21	$67,263	$44,102
Methotrexate Sodium – 2.5 mg Tab	$2.55	--	$2.55	$796	$796

mg: milligram, ml: milliliter, WAC: wholesale acquisition cost

*Discount calculated as the average discount estimated for the other two JAK inhibitors.

Base-Case Results

The base-case results for total cost and outcomes are reported in Table ES10, with incremental cost-effectiveness results in Table ES11. Upadacitinib use resulted in marginally more QALYs gained compared to adalimumab and had higher drug and total costs. The higher total costs are due to patients staying on upadacitinib longer than adalimumab, with the second-line market basket of TIMs being less expensive than upadacitinib or adalimumab.

Table ES10. Results for the Base Case for Upadacitinib versus Adalimumab

Treatment	Drug Cost* (Line One)	Total Cost	LYs	QALYs	Months in Line One Remission
Upadacitinib + cDMARD	$21,400	$48,200	0.985	0.699	2.8
Adalimumab + cDMARD	$15,800	$47,600	0.985	0.693	1.7

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year

*Only costs of TIM; does not include conventional DMARD cost.

The incremental cost-effectiveness ratio for upadacitinib versus adalimumab was estimated to be approximately $92,000 per QALY. The cost per month in remission while on upadacitinib compared to adalimumab was approximately $600.

Table ES11. Incremental Cost-Effectiveness Ratios for Upadacitinib versus Adalimumab

Treatment	Cost per LY Gained*	Cost per QALY Gained	Cost per Month in Line-One Remission
Upadacitinib + cDMARD vs. Adalimumab + cDMARD	--	$92,000	$600

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year

*No difference in LYs was gained between the two TIMs up to the third decimal place, hence this incremental cost-effectiveness ratio was not calculated.

We also compared the outcomes of adalimumab and tofacitinib to their respective conventional DMARD comparators. Results demonstrate that the use of adalimumab or tofacitinib results in similar QALY gains at one year, at a higher cost, compared to conventional DMARDs.

Table ES12. Results for the Base Case for Adalimumab versus Conventional DMARD

Treatment	Drug Cost (Line One)	Total Cost	LYs	QALYs	Months in Line-One Remission
Adalimumab + cDMARD	$15,800*	$47,600	0.985	0.693	1.73
cDMARD	$5,500	$36,500	0.985	0.686	0.58

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year
*Only the costs of TIM; does not include cDMARD cost.

Table ES13. Results for the Base Case for Tofacitinib versus Conventional DMARD

Treatment	Drug Cost (Line One)	Total Cost	LYs	QALYs	Months in Line-One Remission
Tofacitinib + cDMARD	$12,800*	$44,700	0.985	0.692	1.40
cDMARD	$5,500	$38,400	0.985	0.685	0.45

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year
*Only the costs of TIM; does not include cDMARD cost.

Sensitivity and Scenario Analyses

Results from the one-way sensitivity analyses for upadacitinib compared to adalimumab showed that the cost of the second-line TIM market basket was the major driver of results, with probability of achieving remission, the efficacy decrement for line 2+ therapy, and utility estimates having much smaller impacts.

A probabilistic sensitivity analysis was also conducted to assess variation in several parameters with 1,000 Monte Carlo simulations. Approximately 80% of all iterations resulted in cost-utility ratios at or under the $150,000 per QALY threshold. A scatter plot showing the distribution of cost-utility ratios over the 1,000 simulations is presented in Appendix E.

The additional costs of unemployment did not have a substantial impact on the total costs relative to those seen using a health care sector perspective, which is not surprising considering the one-year time horizon.

Summary and Comment

Our base-case findings suggest that upadacitinib provides marginal clinical benefit in comparison to adalimumab, at higher costs. These higher costs are attributed to patients remaining on upadacitinib longer due to better rates of remission and low disease activity relative to those of

adalimumab. Together, these outcomes translate into cost-effectiveness estimates that fall under the upper end of the commonly cited cost-utility threshold of $150,000 per QALY. Results from the modeling comparison of tofacitinib and adalimumab to conventional DMARD suggest that for the similar benefit tofacitinib offers, a price much higher than adalimumab may not be justified.

Our analyses indicate that more comparable data are required on the short- and long-term efficacy of JAK inhibitors. Upadacitinib, for which we have trial-specific data, provided marginal clinical benefit over adalimumab at higher costs, resulting in its incremental cost-utility ratio falling below commonly cited thresholds.

Note that the results presented in this report should not be directly compared to the results of ICER's 2017 report, as the current model includes substantial changes from the cost-effectiveness model in the prior report.

Potential Other Benefits and Contextual Considerations

Our reviews seek to provide information on potential other benefits offered by the interventions to the individual patient, caregivers, the delivery system, other patients, or the public that would not have been considered as part of the evidence on comparative clinical effectiveness. These elements are listed in the tables below.

Potential Other Benefits

Table ES14. Potential Other Benefits

Other Benefits	Description
This intervention offers reduced complexity that will significantly improve patient outcomes.	JAK inhibitors are administered orally, which may be preferable to patients. Other TIMs are administered intravenously or via subcutaneous injection.
This intervention will reduce important health disparities across racial, ethnic, gender, socio-economic, or regional categories.	No.
This intervention will significantly reduce caregiver or broader family burden.	See above; oral administration.
This intervention offers a novel mechanism of action or approach that will allow successful treatment of many patients for whom other available treatments have failed.	No.
This intervention will have a significant impact on improving return to work and/or overall productivity.	The potential impact is likely comparable to other TIMs.
Other important benefits or disadvantages that should have an important role in judgments of the value of this intervention.	No.

Contextual Considerations

Table ES15. Potential Contextual Considerations

Contextual Consideration	Description
This intervention is intended for the care of individuals with a condition of particularly high severity in terms of impact on length of life and/or quality of life.	RA is a condition with a large impact on the length and quality of life, which has been greatly improved with the introduction of biologics and conventional DMARDs.
This intervention is intended for the care of individuals with a condition that represents a particularly high lifetime burden of illness.	See above.
This intervention is the first to offer any improvement for patients with this condition.	No.
Compared to adalimumab, there is significant uncertainty about the long-term risk of serious side effects of this intervention.	No.
Compared to adalimumab, there is significant uncertainty about the magnitude or durability of the long-term benefits of this intervention.	No.
There are additional contextual considerations that should have an important role in judgments of the value of this intervention.	No.

Value-Based Benchmark Prices

Annual value-based price benchmarks (VBPBs) of upadacitinib (vs. adalimumab) are presented in Table ES16. The VBPB for a drug is defined as the price range that would achieve incremental cost-effectiveness ratios between $100,000 and $150,000 per QALY gained. For upadacitinib, price discounts of approximately 25% to 26% from the list price (WAC) would be required to reach the $100,000 to $150,000 per QALY threshold prices, respectively. Note that these discounts are similar to the 26% discount we assumed for upadacitinib in our base case (using the average estimated discount for the other JAK inhibitors).

Table ES16. Value-Based Price Benchmarks for Upadacitinib

	Annual WAC	Annual Price at $100,000 Threshold	Annual Price at $150,000 Threshold	Change from WAC to Reach Threshold Prices
Per QALY Gained	$59,860	$44,144	$44,822	-25% to -26%

QALY: quality-adjusted life year, WAC: wholesale acquisition cost

Potential Budget Impact

Methods

We used the cost-effectiveness model to estimate the total potential budget impact of upadacitinib in adults in the US diagnosed with moderately-to-severely active RA. However, it is important to note that all cost inputs for the budget impact model are generated from our cost-effectiveness model that comprised of a population with severely active RA. Potential budget impact was defined as the total differential cost of using upadacitinib as opposed to adalimumab for the treated population, calculated as differential health care costs (including drug costs) minus any offsets in these costs from averted health care events. All costs were undiscounted and estimated over a five-year time horizon.

We estimated that the eligible population for upadacitinib would consist of adult patients with moderately-to-severely active RA with prior failure by methotrexate (a conventional DMARD). This population aligns with the labeled indication. Although we believe upadacitinib may also be used in a TIM-experienced population, due to a lack of comparable data against other TIMs, we did not model this in our cost-effectiveness analysis, and subsequently did not include this population in our potential budget impact analysis. Additionally, we do not include a prevalent population of TIM users because we believe it is unlikely that these patients would switch to upadacitinib unless they experience inadequate response or intolerance to their current TIM regimen. Thus, our estimates of the eligible population represent only a proportion of the larger eligible population that can be treated with upadacitinib.

The estimated annual incidence of RA in the US is 0.041%.[25] To this incident population, we applied estimates of the proportions with moderately-to-severely active RA (36.1%) and those with RA on biologic therapy (52%).[26] These estimates were applied to the 2019-2023 five-year annual average adult population in the US to arrive at an eligible population size of approximately 20,000 incident patients per year.

Results

For upadacitinib, the annual potential budget impact of treating the entire eligible population across all prices (WAC, assumed net price, and the three cost-effectiveness threshold prices for $50,000, $100,000, and $150,000 per QALY) did not exceed the $819 million threshold. The annual potential budget impacts of treating the entire eligible population using the different price levels are compared to the $819 million annual budget impact threshold in Table ES17. Overall, the greatest potential annual budget impact we estimated was 6% of the $819 million threshold, using its price to reach a threshold of $150,000 per QALY.

Table ES17. Estimated Total Population Annual Potential Budget Impact at Different Prices of Upadacitinib for the Eligible Population of 20,000 per Year

Price	Five-Year Annualized Total Population Budget Impact	Percent of Budget Impact Threshold
$150,000 per QALY Threshold Price	$48.2 million	6%
$100,000 per QALY Threshold Price	$34.3 million	4%
$50,000 per QALY Threshold Price	$20.6 million	3%

QALY: quality-adjusted life year, WAC: wholesale acquisition cost

Biosimilars for RA

In order to set the stage for a Policy Roundtable discussion about the role of biosimilars in the treatment of RA, we provide background about biosimilars and we describe the evidence for one biosimilar, infliximab-dyyb, as an example.

A biosimilar is a biologic drug that is highly similar in structure and function to a licensed reference product. In contrast to a generic (which is a small molecule that can be predictably duplicated), a biosimilar is a larger and more complex molecule that can be sensitive to changes in the manufacturing process. The FDA requires that manufacturers demonstrate that there are no clinically meaningful differences in safety, purity, and potency between the biosimilar and the reference product.[27] The Biologics Price Competition and Innovation Act (BPCI) of 2009 created an abbreviated licensure pathway for products shown to be biosimilar to or interchangeable with an FDA licensed reference product.[28] In Europe, the European Medicines Agency requires that the two products show evidence of similarity in quality, safety, and efficacy.

The FDA approval pathways for biologics and for biosimilars are different. A proposed biosimilar must have the same mechanism of action for the intended condition(s) of use as the reference product. The route of administration, dose form, and strength must also be the same. Of note, once a biosimilar has been approved, it may be prescribed for any indication allowed for the reference product, even if the biosimilar has not been studied in that patient population.[29]

Biosimilar naming follows a standard approach. Each nonproprietary name for a biological product includes the reference product name followed by a hyphen and a four-letter suffix. For example, the biosimilar in this report is infliximab-dyyb. The reference product for infliximab-dyyb is infliximab.[30]

As implied by its name, the BPCI was intended to increase competition in the marketplace and decrease cost, analogous to what occurred when generic drug legislation was passed. Although more than 20 biosimilars have been approved, the uptake of biosimilars has been modest and because of this, significant cost reductions have not been observed in the US.

Example: Infliximab Biosimilar (Inflectra/CT-P13/Infliximab-dyyb)

The PLANETRA (Programme evaluating the autoimmune disease investigational drug CT-P13 in RA patients) trial randomized patients with RA to infliximab-dyyb (CT-P13) or to its reference product infliximab (INX).[31] The investigators randomized 606 patients (83% female) with active RA and an inadequate response to methotrexate.[32] The primary endpoint was ACR criteria for ≥20% clinical improvement response (ACR20) at 30 weeks. Additional endpoints included ACR50, ACR70, DAS28-ESR, DAS28-CRP, SDAI, CDAI, the percentage of patients with response defined according to EULAR criteria, patient-reported outcomes, joint damage progression, safety endpoints and laboratory abnormalities, and immunogenicity endpoints. At week 30, ACR20 responses were 60.9% for CT-P13 and 58.6% for INX (95% confidence interval [CI] -6% to 10%). ACR 50 and ACR 70 responses were also similar at 30-week follow-up (35.1% and 16.6% vs. 34.2% and 15.5% for CT-P13 and INX, respectively). The proportion requiring salvage therapy at 30 weeks was also similar (3.2% for CT-P13 vs. 4.0% for INX). All other outcomes were similar at 30 weeks. The incidence of drug related adverse events was also similar (35.2% for CT-P13 vs. 35.9% for INX) in both groups.

Subsequently, 54-week results were reported. A total of 455 of the 606 patients were treated up to week 54.[33] At 54 weeks, there was no difference between groups who met the primary endpoint (74.7% biosimilar vs. 71.3% reference). The proportion of patients achieving ACR50 and ACR70 at 54 weeks was also comparable between groups.

DAS28-ESR, DAS28-CRP, SDAI, and CDAI scores were similar between groups at 54 weeks follow-up. Mean decreases from baseline of DAS28-ESR, DAS28-CRP, SDAI, and CDAI were 2.4, 2.3, 26.3, and 25.7 compared with 2.4, 2.2, 24.6, and 24.0 for the reference product. The proportion of patients with "Good" and "Moderate" EULAR response was similar between the two groups. With respect to patient-reported outcomes, the Visual Analog Scale (VAS) for patient assessment of pain showed similar reductions from baseline in both groups (30.2 vs. 28.4 at week 54). There were no differences in the immunologic outcomes, including the development of anti-drug antibodies.

Treatment-emergent adverse events at 54 weeks were similarly reported between groups (70.5% vs. 70.3% in the reference product). Twenty-two patients (7.3%) in the infliximab-dyyb group had latent tuberculosis compared with 20 patients (6.7%) in the reference product group. There were no cases of active tuberculosis or lymphoma at 54-week follow-up in either group.

In an extension of the PLANETRA study, 302 of the 455 patients who completed the study enrolled into the extension study.[31] The switch group from the reference product to infliximab-dyyb was compared to the maintenance group continuing infliximab-dyyb after an additional 48 weeks follow-up. Efficacy assessments were made at baseline and at weeks 14, 30, 54, 78, and 102. Efficacy endpoints included the proportion of patients meeting ACR20, ACR50, and ACR70 among others. Response rates at week 102 assessment, for maintenance versus switch groups, were 71.7% versus 71.8% for ACR20, 48.0 % versus 51.4% for ACR50, and 24.3% versus 26.1% for ACR70.

There were no differences in other efficacy endpoints including DAS28 score changes and EULAR response criteria. Similar proportions of patients reported treatment-emergent adverse events in the two groups (53.5% vs. 53.8%). Rates of serious adverse events were similarly reported in the maintenance and the switch groups (7.5% vs. 9.1%). Rates of latent tuberculosis were similar to those seen during the main trial. Active tuberculosis was reported in three patients (1% in the CT-P13 group and no patients in the reference product group).

Based on this evidence, the FDA approved infliximab-dyyb as a biosimilar. PLANETRA was a large study with relatively long follow up. The study found no clinically important or statistically significant differences in benefits or harms between infliximab-dyyb and its reference product, infliximab, at weeks 30, 54, and 102. Observational data also support no important differences between the two therapies. Using the ICER Evidence Matrix, the biosimilar infliximab-dyyb has high certainty of comparable net health benefit ("C") relative to its reference product.

CTAF Votes

The CTAF panel deliberated on key questions raised by ICER's report at a public meeting on December 9, 2019 in Oakland, CA. The results of these votes are presented below, and additional information on the deliberation surrounding the votes can be found in Section 8.

1) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the net health benefit of upadacitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

2) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the net health benefit of upadacitinib plus a conventional DMARD is superior to that provided by adalimumab plus a conventional DMARD?

Yes: 12 votes	No: 2 votes

3) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the net health benefit of tofacitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

4) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the net health benefit of tofacitinib plus a conventional DMARD is superior to that provided by adalimumab plus a conventional DMARD?

Yes: 0 votes	***No: 14 votes***

5) In patients who are naïve to TIMs, is the evidence adequate to distinguish the net health benefit between upadacitinib and tofacitinib?

Yes: 0 votes	***No: 14 votes***

5a) No vote was taken on Question 5a because the panel unanimously voted that the evidence was inadequate to distinguish between upadacitinib and tofacitinib.

6) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the biosimilar infliximab-dyyb produces a net health benefit comparable to that of the originator biologic infliximab?*

Yes: 14 votes	No: 0 votes

7) In patients who are TIM-experienced, is the evidence adequate to demonstrate that the net health benefit of upadacitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

8) In patients who are TIM-experienced, is the evidence adequate to demonstrate that the net health benefit of tofacitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

9) In patients who are TIM-experienced, is the evidence adequate to demonstrate that the net health benefit of baricitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

*The webcast recording of the meeting shows one "no" vote in Question 6. After the conclusion of the meeting, the panelist who voted "no" clarified that they had done so accidentally after misreading the question.

10) Does treating patients with upadacitinib plus a conventional DMARD offer one or more of the following potential "other benefits" in comparison to adalimumab plus a conventional DMARD?

This intervention offers reduced complexity that will significantly improve patient outcomes.	14/14
This intervention will significantly reduce caregiver or broader family burden.	9/14
This intervention offers a novel mechanism of action or approach that will allow successful treatment of many patients for whom other available treatments have failed.	3/14
This intervention will have a significant impact on improving patients' ability to return to work and/or their overall productivity.	6/14
There are other important benefits or disadvantages that should have an important role in judgments of the value of this intervention.	0/14

11) Are any of the following contextual considerations important in assessing the long-term value for money at estimated pricing of upadacitinib?

This intervention is intended for the care of individuals with a condition of particularly high severity in terms of impact on length of life and/or quality of life.	12/14
This intervention is intended for the care of individuals with a condition that represents a particularly high lifetime burden of illness.	12/14
Compared to adalimumab, there is significant uncertainty about the long-term risk of serious side effects of this intervention.	4/14
Compared to adalimumab, there is significant uncertainty about the magnitude or durability of the long-term benefits of this intervention.	5/14
There are additional contextual considerations that should have an important role in judgments of the value of this intervention.	0/14

12) Given the available evidence on comparative effectiveness, incremental cost-effectiveness using the net price estimate of 26% off the WAC, and considering other benefits, disadvantages, and contextual considerations, what is the long-term value for money of treatment with upadacitinib plus a conventional DMARD versus adalimumab plus a conventional DMARD?

Low: 8 votes	Intermediate: 6 votes	High: 0 votes

Key Policy Implications

Following its deliberation on the evidence, the CTAF panel engaged in a moderated discussion with a Policy Roundtable about how best to apply the evidence on TIMs for RA to policy and practice. The Policy Roundtable members included two patient experts, two clinical experts, two payer representatives, and two representatives from manufacturers. The discussion reflected multiple perspectives and opinions, and therefore, none of the statements below should be taken as a consensus view held by all participants. The top-line policy implications are presented below, and additional information can be found in the full report.

Biosimilars

- Clinical societies should educate their clinician members that the evidence behind biosimilars is sound and that whenever they are available at a lower cost than reference products, they should be the preferred option given the benefits of lower costs for patients and the health care system.
- Patient groups should educate their members that biosimilars are as safe and effective as reference products and that starting on a biosimilar, or switching to one, is clinically responsible and may be financially beneficial.
- Manufacturers should not use scare tactics to dissuade responsible switching to biosimilars.
- Health plan sponsors, insurers, PBMs, and provider groups should work together to promote greater use of biosimilars by implementing switching programs that provide broad support to patients while assuring that patients who do not respond well to biosimilars are able to access reference products and/or obtain other targeted treatment options without delay.
- Policymakers should continue work on alternatives to the current rebate system that will allow the market to reward the competitive advantages of lower priced, equally effective biosimilar treatment options. One helpful first step would be to ban the bundling of rebates across multiple indications.

Plan Sponsors, Payers, and PBMs

- Reconsider the need for step therapy and any switching programs if pricing becomes better aligned with clinical value.
- If pricing does not fairly align with patient benefit, plan sponsors may consider the option of benefit designs requiring patients to switch from their current therapy to a lower cost, equally effective option. In all such efforts the clinical effectiveness of the required therapy and the cost overall should be closely monitored.
- Requiring switching for patients who have achieved adequate treatment response on a particular drug raises the risk of irremediable harm should the new treatment not prove equally effective. Therefore, these programs must meet an extremely high bar of evidence and must have ample, efficient means by which clinicians can request exceptions on clinical grounds.

- Increase transparency around the role of discounting and rebate practice in formulary design.
- Design innovative risk-sharing payment agreements, including outcome-based contracts with manufacturers and shared risk agreements with accountable care organizations.

Clinical Societies and Manufacturers

- Establish standardized assessments to allow for rigorous direct and indirect comparisons of evidence across studies and therapeutic alternatives.

Public Policy Decision Makers

- In a dysfunctional market system, in order to protect patients today and improve their future access to innovative therapies, policy makers may need to consider some form of regulatory intervention to ensure that drug prices and price increases do not continue their current upward trajectory, driving prices further from reasonable alignment with the added benefits for patients.
- Public policy decision makers should ban the bundling of rebates across indications.

Research Community

- Researchers and research funding agencies should prioritize studies that identify biomarkers predicting response to currently available therapies or classes of therapies (personalized medicine).
- Researchers should separate outcomes that measure inflammation from those that measure pain.
- Randomized trials and observational studies should include outcomes that are both patient-centered and patient-driven.
- The FDA should require that randomized trials include patients reflecting the average RA patient population.
- The FDA should require that randomized trials of new therapies always include an active comparator.

1. Introduction

1.1 Background

Background

Rheumatoid arthritis (RA) is the most common autoimmune inflammatory arthritis in adults, affecting between 1.3 and 1.8 million Americans.[1] RA is more common in women and may occur at any age, with peak incidence occurring at ages 50-60 years.[2] RA is typically characterized by morning stiffness and symmetrical joint swelling of the feet, hands, and knees, although any joint (and in some cases, internal organs and skin) may be involved.[2] RA is considered a clinical syndrome that encompasses several disease subsets, each of which involves a distinct inflammatory cascade that can lead to joint damage, deformity, and organ dysfunction.[34] The course of RA may be complicated by cardiac, hematologic, and other extra-articular manifestations.[2] Historically, RA was associated with both progressive disability and a shortened lifespan, although improvements in diagnosis as well as aggressive use of disease-modifying antirheumatic drugs (DMARDs) have greatly improved prognosis in the past 20 years.[35]

The chemotherapeutic agent methotrexate is the most widely used conventional DMARD; it is considered an "anchor drug" because of its effectiveness and tolerability as well as its potential to enhance the effectiveness of biologic and non-biologic drugs that are targeted at certain mediators of inflammation in RA, known collectively as targeted immune modulators (TIMs).[2] However, only about 50% of patients treated with methotrexate alone will receive sufficient reduction in disease activity or remission of symptoms.[2] Agents with indications for RA include inhibitors or antagonists of multiple mediators of the inflammatory cascade, including tumor necrosis factor (TNF), B-lymphocyte CD20 antigen, interleukin (IL) 1 and 6, Janus kinase (JAK), and T cells. Guidelines from the American College of Rheumatology (ACR) recommend the use of TIMs in patients with moderate-to-severe disease activity despite the use of conventional DMARDs.[36]

In our 2017 review (Targeted Immune Modulators for Rheumatoid Arthritis: Effectiveness & Value), ICER assessed the relative effectiveness and value of all available TIMs at that time. Since that report two additional JAK inhibitors, baricitinib (Olumiant®, Eli Lilly and Company) and upadacitinib (Rinvoq™, AbbVie), received United States Food and Drug Administration (FDA) approval. Therefore, ICER decided to update the evidence for JAK inhibitors for adults with moderately-to-severely active RA. In addition, to better reflect current clinical practice and guidelines using a treat-to-target approach,[5] we focused on measures of disease activity at three months. Patients not achieving remission or low disease activity after three months of therapy are typically switched to a different TIM.

In a separate section, we evaluate the evidence supporting infliximab-dyyb (Inflectra®, Pfizer) as a biosimilar for the reference product infliximab (Remicade®, Janssen) for the treatment of RA. Infliximab-dyyb is intended to serve as an exemplar to ground a discussion surrounding the relatively low penetrance of biosimilars in the marketplace despite FDA approval of more than 20 biosimilars.

JAK Inhibitors

There are currently three JAK inhibitors approved by the FDA for RA: upadacitinib, tofacitinib (Xeljanz®, Pfizer), and baricitinib. While most TIMs are biologic agents that require subcutaneous injection or intravenous (IV) infusion, JAK inhibitors are small molecules taken orally. They work by inhibiting the JAK enzymes, which mediate intracellular signaling pathways involved in the production of inflammatory cytokines, including IL-2, -4, -6, -7, -9, -15, and -21. There are four JAK subtypes (JAK1, JAK2, JAK3, TYK) that have overlapping functions. The most recently approved JAK inhibitor, upadacitinib, is supposed to be more specific for JAK1, which may influence both its benefits and harms.

Table 1.1. Dosage Forms and Administration Schedules for JAK Inhibitors

JAK Inhibitor	Recommended Dose (mg)	Route of Administration	FDA Approval	Annual WAC
Upadacitinib	15 mg once daily	Oral	08/16/2019	$59,860
Tofacitinib	5 mg twice daily or 11 mg once daily (extended release form)	Oral	11/16/2012	$54,552
Baricitinib	2 mg once daily	Oral	06/01/2018	$26,017

FDA: Food and Drug Administration, JAK: Janus kinase, mg: milligram, WAC: wholesale acquisition cost

1.2 Scope of the Assessment

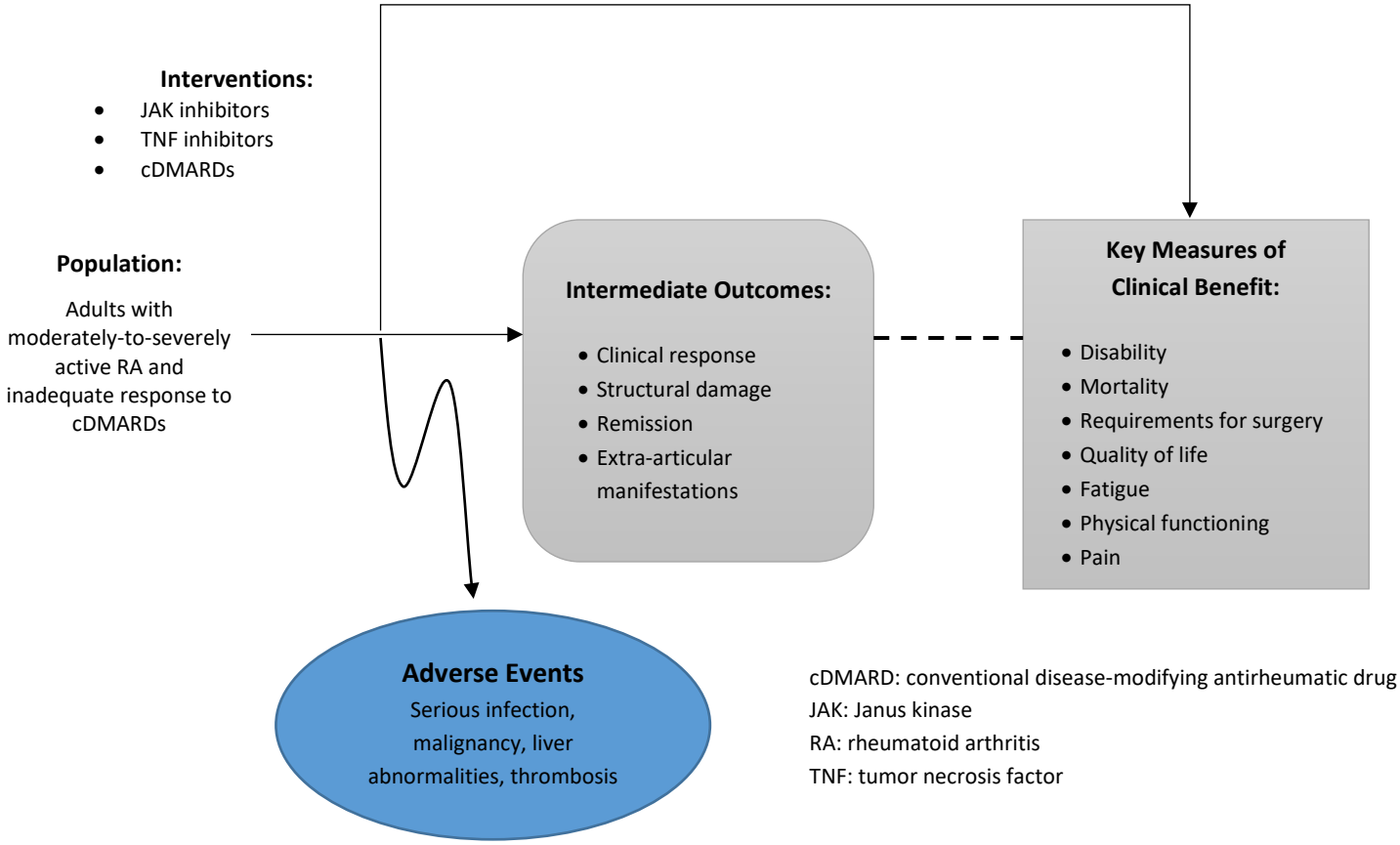
The scope for this assessment is described on the following pages using the PICOTS (Population, Intervention, Comparators, Outcomes, Timing, and Settings) framework. Evidence was collected from randomized controlled trials (RCTs) as well as high-quality systematic reviews; high-quality comparative cohort studies were considered, particularly for long-term outcomes and uncommon adverse events. Our evidence review includes input from patients and patient advocacy organizations, data from regulatory documents, information submitted by manufacturers, and other grey literature when the evidence meets ICER standards (for more information, see https://icer-review.org/methodology/icers-methods/icer-value-assessment-framework/grey-literature-policy/).

Wherever possible, we sought out head-to-head studies of these interventions. We also included studies with an active comparison to conventional DMARDs or TNF inhibitors with or without conventional DMARDs. We used direct and indirect evidence in network meta-analyses (NMA) of selected outcomes. Full details regarding the literature search, screening strategy, data extraction, and evidence synthesis are available in Appendix A.

Analytic Framework

The general analytic framework for the assessment of JAK inhibitors for moderately-to-severely active RA is depicted in Figure 1.1 below.

Figure 1.1. Analytic Framework: JAK Inhibitors for Moderately-to-Severely Active RA



cDMARD: conventional disease-modifying antirheumatic drug
JAK: Janus kinase
RA: rheumatoid arthritis
TNF: tumor necrosis factor

The diagram begins with the population of interest on the left. Actions, such as treatment, are depicted with solid arrows, which link the population to outcomes. For example, a treatment may be associated with specific health outcomes. Outcomes are listed in the shaded grey boxes; those within the rounded boxes are intermediate outcomes (e.g., clinical response), and those within the squared-off boxes are key measures of benefit (e.g., disability). The key measures of benefit are linked to intermediate outcomes via a dashed line, as the relationship between these two types of

outcomes may not always be validated. Curved arrows lead to the adverse events of treatment, which are listed within the blue ellipse.[37]

Populations

The population for the review is adults ages 18 and older with moderately-to-severely active RA and inadequate response or intolerance to conventional DMARDs. Level of disease activity is defined according to validated and frequently used scales in RA (i.e., Disease Activity Score 28 [DAS28], Clinical Disease Activity Index [CDAI], Simplified Disease Activity Index [SDAI]). Note that this review will not include children, adolescents, or adults with juvenile forms of RA or other inflammatory arthritis, now collectively known as juvenile idiopathic arthritis. Feedback from patient groups and clinicians suggested that the clinical presentation and disease trajectory of these patients differs substantially from those with the adult form of RA.[38]

We looked for evidence on key subpopulations and/or data stratifications of interest. Among those suggested by stakeholders during the Open Input Period were 1) evaluation of both TIM-naïve patients *and* those with inadequate response or intolerance to initial TIM therapy and; 2) use of JAK inhibitors as monotherapy and in combination with conventional DMARDs. Feedback received for our prior report indicated additional subpopulations or stratifications of interest, including 1) presence of comorbidities (e.g., cardiovascular, interstitial lung disease, psychiatric, malignancy); 2) both "early" (i.e., within two years of symptom onset) and established RA; 3) seropositivity for prognostic markers such as anti-cyclic citrullinated peptide antibodies; 4) geography, in particular US-based versus non-US settings; and 5) study funding (i.e., industry-sponsored vs. other funding sources).

Interventions

The interventions of interest for this review are listed below.

- Upadacitinib
- Tofacitinib
- Baricitinib
- Biosimilar exemplar: infliximab-dyyb

We sought clinical evidence on all the products listed above. We note, however, that biosimilar data are presented separately, given differences in study design and intent (i.e., non-inferiority vs. superiority) relative to clinical studies of the reference products. We hope these biosimilar data are useful in framing a general discussion about the role of biosimilars and interchangeability status in RA.

Comparators

We examined studies comparing JAK inhibitors to conventional DMARD monotherapy or combination therapy (including triple therapy with the conventional DMARDs methotrexate, sulfasalazine, and hydroxychloroquine) to assess performance versus historical standard treatments as well as head-to-head studies between JAK inhibitors and TNF inhibitors (adalimumab [Humira®, AbbVie] in all cases).

Finally, while studies with an active comparator arm were preferred, we also included placebo-controlled trials as necessary.

Outcomes

This review examines key clinical outcomes associated with RA. Because the recommended treat-to-target paradigm encourages switching therapy within three months for patients who do not achieve remission or low disease activity, we have prioritized measures of disease activity at that timepoint. We also expanded the outcome list based on stakeholder feedback to include additional patient-reported outcomes as well as important clinical and health care utilization measures.

- Mortality
- Treatment response (e.g., ACR20, ACR50, and ACR70, area-under-the-curve analysis)
- Measures of disease activity, remission, and remission loss (e.g., DAS28, CDAI, SDAI)
- Radiographic evidence of structural damage
- Disease-specific and general health-related quality of life (e.g., Health Assessment Questionnaire for Rheumatoid Arthritis Disability Index [HAQ-DI], SF-36 [Short Form Survey])
- Pain (e.g., visual analog scales [VAS])
- Other patient-reported outcomes (e.g., patient satisfaction, measures of fatigue, morning joint stiffness duration and severity)
- Productivity loss and caregiver impact
- Requirements for joint replacement or other surgical intervention
- Utilization of key health care resources (e.g., hospitalization, rehabilitation, assisted living)
- Cardiovascular events
- Treatment-related adverse events (e.g., serious infection, malignancy, liver abnormalities, and both venous and arterial thrombosis).

Whenever possible we reported the absolute risk reduction and number needed to treat in addition to the relative risk reduction for the treatment comparisons.

Timing

Studies of three- and six-months duration were prioritized for response to therapy, but long-term evidence was preferred for harms.

Settings

All relevant settings were considered, including outpatient as well as ambulatory and hospital-based infusion centers. Several stakeholders commented on the importance of geography for our review given differences in treatment guidelines and practice patterns. We focused our attention on studies pertinent to the US setting; however, we recognized that studies conducted outside the US were useful in assessing long-term harms.

1.3 Definitions

ACR Classification Criteria (2010): The scoring algorithm for the determination of definite RA based on 1) number and level of joints involved; 2) diagnostic serology testing; 3) testing for acute-phase reactants; and 4) duration of symptoms.

ACR Response Criteria: Known as ACR20, 50, or 70, representing at least 20%, 50%, or 70% improvement in tender/swollen joint counts as well as at least these levels of improvement in at least three of the following five criteria:

1) Patient global assessment
2) Physician global assessment
3) Pain
4) Disability/function
5) Acute-phase reactant values.

Historically, the ACR20 was the primary endpoint in most clinical trials of RA treatments. With the advent of greater efficacy from treatment with TIMs, and with patient input about clinical significance, the ACR50 and ACR70 are also commonly included as secondary endpoints. With the shift toward treat-to-target approaches, however, measures of disease activity and/or remission are now also commonly used (see below).

Acute-phase reactants: Blood-based biomarkers for systemic inflammation characteristic of RA and other autoimmune diseases, typically c-reactive protein (CRP) and erythrocyte sedimentation rate (ESR).

Anticitrullinated protein antibody (ACPA): A blood test that measures the level of autoantibodies against cyclic citrullinated peptides, which are produced in excess in patients with RA. The test has been used to establish risk for RA as well as to assess disease severity and/or prognosis.

Disease activity measures: Refers to multiple measures of disease activity, which are generally divided into patient-driven, patient/provider composite, and patient/provider/laboratory composite tools. All instruments differentiate low, moderate, and high disease activity.

Patient-driven tools:

- **Patient Activity Scale (PAS):** Scored 0-10 on a continuous scale based on questionnaire items regarding disability (HAQ, see below), pain, and global assessment (VAS). A second version (PAS-II) has been developed using the same format but with a different disability measure (HAQ-II).
- **Routine Assessment of Patient Index Data (RAPID-3):** Scored 0-10 on a continuous scale based on pain and global assessment VAS scales and disability measured via the multidimensional HAQ (MDHAQ).

Patient/provider composite tool:

- **CDAI:** Scored on a 0-76 continuous scale based on tender and/or swollen joint counts (up to 28 each) as well as patient and provider global VAS scores.

Patient/provider/laboratory composite tools:

- **DAS28:** Scored on a 0-9.4 continuous scale based on tender and/or swollen joint counts (up to 28 each), ESR or CRP findings, and patient global VAS score.
- **SDAI:** Scored on a 0-86 continuous scale based on tender and/or swollen joint counts (up to 28 each), CRP findings, provider global and patient global VAS score.

There is no universally accepted measure of disease activity. The ACR recommends using one of six out of the 63 measures of RA disease activity and defines the cut-points to be used to differentiate remission and low, moderate, and high disease activity for seven of the measures described above (PAS, PAS-II, RAPID-3, CDAI, SDAI, DAS28-ESR/CRP).[39] However, disease activity classification is not consistent across all measures. Fleischmann and colleagues reported that DAS28-CRP cut-points yield a higher proportion of patients with remission or low disease activity than DAS28-ESR and that both DAS28 measures underestimate disease activity compared with the SDAI.[40] In the most recent FDA sanctioned trials of therapies for RA, the primary outcomes used DAS28-CRP.

HAQ: A 20-item RA-specific patient questionnaire designed to measure the ability to perform activities of daily living in multiple domains: dressing, standing, eating, walking, hygiene, reach, grip,

other activities, and requirements for assistance from devices or other persons for any of these. Also available in an abbreviated 10-item format (HAQ-II) as well as in the expanded MDHAQ that includes complex activities and psychological status.

Patient-Reported Outcomes Measurement Information System (PROMIS): A relatively new set of person-centered measures that monitors physical, mental, and social health. Early tool development has focused on neurological diseases and sickle cell anemia, and initial validation of general health questionnaires has been conducted in RA samples.[41] Instruments are not yet widely used in clinical trials, however.

Remission: Most commonly defined based on a 0 or minimal score on measures of disease activity (see above), with upper limits ranging from 0.25-1.0 on the 10-point patient-driven scales to 2.6-3.3 on the patient/provider/laboratory composite tools.

Rheumatoid factor (RF): A blood test that measures the presence of an immunoglobulin (most commonly IgM, but can be IgG and/or IgA) that binds to IgG. The test is positive in approximately 80% of patients with RA but is not diagnostic of the disease, as a positive RF can also be seen in other autoimmune and chronic inflammatory diseases as well as in some otherwise healthy older individuals.

Sharp score: The most widely accepted method used to measure radiographic joint damage in RA. Multiple modifications are used, but all focus on both erosion and narrowing of the spaces between joints. The most common modifications include the van der Heijde method, which focuses on 43 areas of the hands and feet (score range: 0-448), and that of Genant, which examines 39 hand/foot areas (score range: 0-290).

1.4 Insights Gained from Discussions with Patients and Patient Organizations

We received valuable input from individual patients and patient organizations throughout the scoping and evidence development process. Patients and patient organizations advised us that health-system challenges with RA are present from the very beginning. Diagnosis is often delayed, due in large part to a shortage of available rheumatologists in many areas of the US. Even after diagnosis, coordination of care across providers and settings is problematic, particularly for patients who self-administer medication and therefore do not receive the opportunity to discuss multiple aspects of their care at an infusion clinic. Perhaps in part because of coordination of care challenges, patients stressed the importance of involving family, informal caregivers, and others as a critical component for successful management of the disease.

Patients and patient organizations emphasized the long-term nature of the disease and the importance of both the long-term perspective and the variability in disease course and treatment

changes, including drug holidays. They highlighted the importance of patient-reported outcomes and offered to help integrate them into the report and model. Multiple stakeholders recommended the inclusion of real-world data for the assessment of evidence on safety, durability of effect, and switching patterns given the widespread availability of such evidence for established therapies. Both clinicians and patient organizations emphasized that their treatment goals are to achieve minimal disease activity; a 50% improvement (e.g., ACR50) is not a successful outcome for a patient with 20 active joints. These groups also highlighted the impact of RA on caregivers and suggested that both caregiver measures and outreach to caregiver groups be part of the review process. Finally, stakeholders highlighted the important progress that has been made through the use of biologics: very few patients progress to disabling joint deformities with current treatments.

Regarding treatment, we were advised that it is not uncommon for patients to cycle through various therapies before finding a treatment option to which they both respond to and tolerate; this mirrored the input received from clinicians. We also received input that "fail-first" or step-therapy insurance policies often require patients to follow a specific sequence of TIM therapies, most commonly requiring a trial of methotrexate followed by multiple attempts with TNF inhibitors. Because of the cyclical nature of the disease and its treatment, patients fear restrictions on access to certain types of drugs as well as more general restrictions (e.g., stopping and re-starting therapy, requirements to repeat step therapy after switching health plans, etc.). We also heard that the high cost of treatments is a large burden for patients who have health plans that do not cover specific medications or have high deductibles.

Even with the availability of TIMs in addition to conventional DMARDs, patients are not satisfied with their treatment.[3] A recent online survey of 232 patients with RA over the age of 21 who had been failed by at least one DMARD assessed treatment satisfaction. Almost half of patients (43%) reported daily or almost daily use of prescription pain medication and 44% reported a current flare of their RA. Only 26% reported satisfaction with their treatment, even though 67% were currently on a biologic DMARD.[3] Symptoms that were poorly addressed by current treatments included joint stiffness and sexual dysfunction in 70% to 80% of patients as well as fatigue, sleep disturbance, and joint pain.

Patients also stressed that RA is a systemic illness that affects organ systems beyond the joints. These include the lungs, the cardiovascular system, and the eyes. There is hope that long-term data will demonstrate improvements in these disease outcomes as well.

The financial burden of RA treatment on patients and their families is also substantial. Patients did mention that manufacturers have increased their recent activity around coupons and other copayment assistance programs, but that the financial problems associated with their care remain significant and are not limited to out-of-pocket costs alone. Issues with coordination of care, navigation of insurance requirements by both patient and provider, lost time at work or school, and other challenges contribute to patient and family impact.

Patient organizations advised us that clinical trials are often lacking robust information on patient-centered outcomes, and suggested a focus on recently developed measures such as those described in the federally-funded PROMIS toolkit (http://www.healthmeasures.net/explore-measurement-systems/promis). We revised our list of possible outcomes considerably based on this feedback. However, patients also felt that much work remains to be done on quantitative, patient-centered measures of treatment success, as many of the recent developments in defining disease remission and treatment response focus primarily on disease activity and not enough on symptom control, activities of daily living, and management of treatment-related side effects. Patients also told us that "point-in-time" measures often fail to capture the lability of RA—the disease burden varies over time, as does the patient's ability to adapt to the realities of the condition.

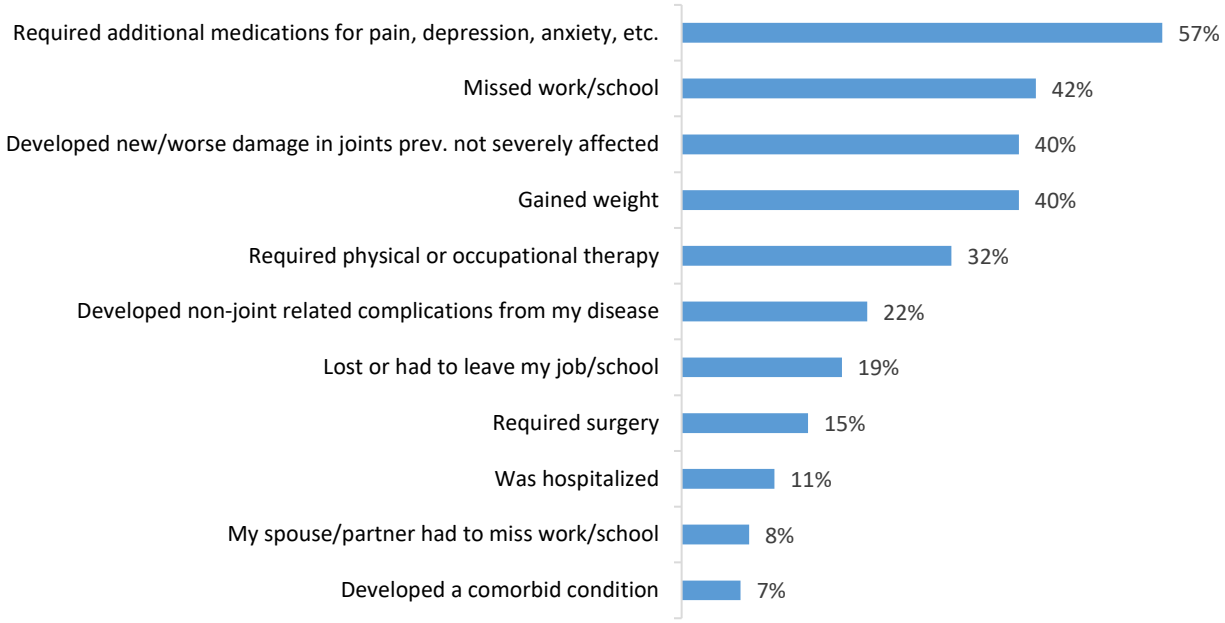
Arthritis Foundation Surveys

Patient Experiences

As part of their engagement with ICER during our initial 2016-2017 review, the Arthritis Foundation, the leading patient advocacy group for patients with RA and other forms of arthritis, deployed an online survey during the first two weeks of November 2016 to gather information about the RA patient experience. Over 3,000 responses were recorded; a total of 1,582 individuals confirmed that they had been diagnosed with RA. The population was comparable to the demographic profile in other US-based RA cohort studies. Eighty-eight percent of RA patients were female, 83% were white (10% were African American or Hispanic), and more than half of the sample was ages 55 or older (mean 59.5). Most respondents reported insurance coverage with a commercial carrier (58%) or Medicare (41%).

Experience with RA was generally longstanding—41% of the sample had been diagnosed 15 or more years ago.[42] The clinical picture for many was complex, with over one-quarter of patients also diagnosed with obesity or depression, and over 10% prevalence of comorbid cancer, heart disease, and diabetes. In addition to clinical complications, RA also has profound lifestyle impacts during periods of greater disease activity. Figure 1.2 on the following page presents impacts experienced during periods when RA was not well-controlled. Nearly 60% of patients required additional medications for pain or mental health concerns, 42% missed some work or school, and nearly one in five had to discontinue work or school because of their condition.

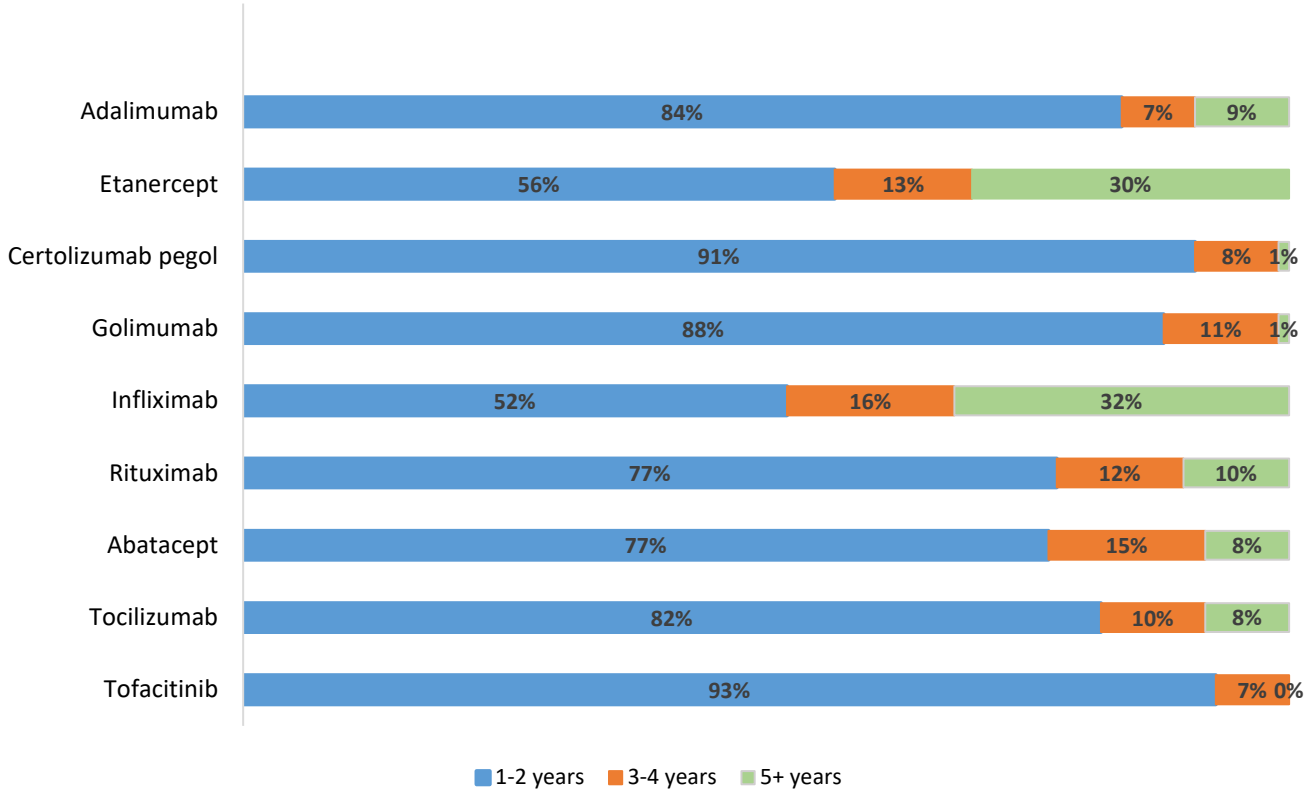
Figure 1.2. Reported Impacts of RA During Periods when Disease was Not Well-Controlled



Source: Arthritis Foundation Survey of Rheumatoid Arthritis Patient Treatment Experiences, November 17, 2016

The survey also indicated that most patients have received multiple TIMs during the course of their disease, without clearly discernible patterns regarding treatment sequence. In addition, changes in medication generally happen relatively early. As shown in Figure 1.3 on the following page, while the proportions vary by TIM, 50-93% of patients are on the same therapy for only one to two years, and relatively small percentages of patients have a course of treatment that is five years or longer. The agents with the greatest proportions of long-duration users were the earliest TIMs approved for RA in the late 1990s (etanercept [Enbrel®, Amgen] and infliximab), which may be at least in part a reflection of their time on market rather than any durability advantage they hold over other TIMs.

Figure 1.3. Duration of Therapy by Type of TIM Therapy



Source: Arthritis Foundation Survey of Rheumatoid Arthritis Patient Treatment Experiences, November 17, 2016

Finally, those surveyed reported few difficulties with accessing treatment facilities or scheduling regular doctor visits, which was surprising given the reported dearth of available rheumatologists. This may be a reflection of the surveyed population (e.g., covered by employer-sponsored health insurance). However, reflecting on our conversations with individual patients and patient organizations, one-third of patients reported problems with access to their medication of choice and restarting a medication they had been using if they stopped for some reason, and over 40% reported problems with care coordination across providers and settings.

Outcomes of Biologic-Naïve versus Biologic-Experienced Patients

The Arthritis Foundation deployed a second survey to assess outcomes of care in RA patients who had been treated with conventional DMARDs only for at least five years (n=222) as well as those who had received at least one TIM during this time period (n=337).[42] While findings are descriptive in nature only (i.e., not adjusted for clinical or demographic differences between groups), they echo those of cross-sectional and other observational studies that have documented the clinical effects of the introduction of TIMs. For example, while substantial proportions of both groups reported that they had experienced some level of joint damage, the proportion was statistically-significantly greater in the TIM-naïve group (90% vs. 65%, p<0.0001). Similarly, the proportion reporting at least

one joint replacement or other major orthopedic surgery (e.g., spinal fusion) was nearly three times greater among TIM-naïve patients (56% vs. 19%, p<0.0001). Finally, while disease impacts were pronounced in both patient subsets, greater percentages of biologic-naïve patients reported hospitalization or emergency room visits due to their condition/symptoms as well as receipt of disability benefits at some point.

1.5 Potential Cost-Saving Measures in RA

ICER now includes in its reports information on wasteful or lower-value services in the same clinical area that could be reduced or eliminated to create headroom in health care budgets for higher-value innovative services (for more information, see https://icer-review.org/final-vaf-2017-2019/). These services are ones that would not be directly affected by therapies for RA (e.g., reduction in disability), as these services would be captured in the economic model. Rather, we sought services used in the current management of RA beyond the potential offsets that would arise from a new intervention. During stakeholder engagement and public comment periods, ICER encouraged all stakeholders to suggest services (including treatments and mechanisms of care) currently used for patients with RA that could be reduced, eliminated, or made more efficient. No suggestions were received, but two of the ACR Choosing Wisely® recommendations apply.[4]

Don't perform MRI [magnetic resonance imaging] of the peripheral joints to routinely monitor inflammatory arthritis.

Data evaluating MRI for the diagnosis and prognosis of RA are currently inadequate to justify widespread use of this technology for these purposes in clinical practice. Although bone edema assessed by MRI on a single occasion may be predictive of progression in certain RA populations, using MRI routinely is not cost effective compared with the current standard of care, which includes clinical disease activity assessments and plain-film radiography.

Don't prescribe biologics for RA before a trial of methotrexate (or other conventional non-biologic DMARDs).

High-quality evidence suggests that methotrexate and other conventional non-biologic DMARDs are effective in many patients with RA. Initial therapy for RA should be a conventional non-biologic DMARDs unless these are contraindicated. If a patient has had an inadequate response to methotrexate with or without other non-biologic DMARDs during an initial three-month trial, then biologic therapy can be considered. Exceptions include patients with high disease activity and poor prognostic features (functional limitations, disease outside the joints, seropositivity, or bony damage), where biologic therapy may be appropriate first-line treatment.

2. Summary of Coverage Policies and Clinical Guidelines

2.1 Coverage Policies

To understand the insurance landscape for treatments for RA, we reviewed publicly available coverage policies from the Centers for Medicare and Medicaid Services (CMS) and Medi-CAL, the five largest national commercial insurers (Aetna, Anthem, Cigna, Humana, and UnitedHealthcare [UHC]), and two California-based insurers (Kaiser Permanente and Health Net). Because upadacitinib was recently approved, we were unable to find comprehensive coverage information for all insurers. We were unable to locate any National Coverage Determinations or Local Coverage Determinations for any of the treatments under review.

Impact of Plan-Level Access Restrictions on Effectiveness of Biologics Among Patients with Rheumatoid or Psoriatic Arthritis[43]

A recent publication by employees of Eli Lilly and Company suggested that some coverage policies may impact the effectiveness of biologics among patients with RA.[43] Remarkably, 66% of patients had no restrictions on any TIM. Using data on 3,993 patients with RA from a claims database, they found non-significant trends towards a reduction in medication adherence (OR 0.90, 0.77-1.05, p=0.182) and treatment effectiveness (OR 0.93, 0.78-1.10, p=0.376) for patients with plans with access restrictions compared to those without access restrictions. In a subgroup analysis, these differences reached statistical significance for patients in plans with step therapy (OR 0.81, 0.67-0.98 for treatment effectiveness; OR 0.81, 0.68-0.96 for medication adherence) versus those without step therapy. However, there was also a trend toward greater adherence and greater effectiveness in patients with plans requiring prior authorization.[43] Given the small effect sizes, the small sample size, the lack of detailed patient-level data, the significant differences among patients in plans without restrictions and those with restrictions, and the lack of a propensity score adjusted analysis, these results should be interpreted with caution.

Upadacitinib

On August 16, 2019, the FDA approved upadacitinib for the treatment of moderate-to-severe RA. We were able to locate some information about upadacitinib from several payers, but the overall coverage landscape is still sparse.

We were unable to find any information related to upadacitinib under Anthem, and there was a limited amount of coverage information under Aetna. As of November 2019, Aetna requires prior authorization to obtain upadacitinib, which is not designated as a "least costly drug."[44]

Cigna, Humana, and UHC have updated policies pertaining to upadacitinib. Effective January 1, 2020, Cigna will designate upadacitinib as a preferred product; patients can access the treatment with documented failure by or contraindication to one conventional DMARD.[45] Humana lists upadacitinib on its specialty tier, which contains high-cost drugs, and as such, access to the treatment requires prior authorization. Patients must demonstrate a documented failure by or contraindication to one conventional DMARD.[46] On November 1, 2019, UHC issued updated converge information that includes prior authorization criteria for upadacitinib. Patients must demonstrate a documented diagnosis of moderately-to-severely active RA and failure by or contraindication to methotrexate.[47]

We were unable to locate comprehensive information about upadacitinib under Kaiser Permanente, an integrated managed-care consortium based in California, but we note that it is listed as a Tier 4 or "specialty-tier drug," which means it has a higher cost share.[48] Health Net, a subsidiary of Centene based in California, issued an update to its clinical policy in November 2019, and included criteria to obtain upadacitinib: patients must be ages 18 years or older and have demonstrated intolerance to or been failed by one conventional DMARD following a three-month trial.[49]

Tofacitinib

To obtain coverage for tofacitinib, most national commercial insurers require prior authorization. To obtain coverage under Aetna and Cigna, patients must be ages 18 years or older, diagnosed with moderately-to-severely active RA, and have documented failure by or inadequate response or contraindication to a conventional DMARD, such as methotrexate, leflunomide, or sulfasalazine. Aetna lists tofacitinib as a one of the "least costly drugs," and neither Aetna nor Cigna require step therapy.[50]

Anthem, Humana, and UHC have more limiting policies that restrict coverage in ways that differ from the labeled indication of tofacitinib. In addition to documented failure by and inadequate response or contraindication to a conventional DMARD, Anthem also requires that patients demonstrate an inadequate response or intolerance to two preferred biologics, which include etanercept, adalimumab, infliximab, and golimumab (Simponi®, Janssen). Humana's policy is comparable: patients seeking tofacitinib must have first tried and been failed by methotrexate and two preferred biologics, which include adalimumab, etanercept, and sarilumab (Kevzara®, Sanofi/Regeneron). UHC lists similar requirements for obtaining tofacitinib. Patients must first try and be failed by two preferred biologics (barring contraindications), which include certolizumab (Cimzia®, UCB), adalimumab, and golimumab. Patients can bypass step therapy requirements with a documented needle phobia.[50]

Kaiser Permanente has a less restrictive policy than most of the surveyed national commercial insurers. To obtain coverage for tofacitinib, patients must be diagnosed with RA and demonstrate

intolerance to or have been failed by a conventional DMARD, such as methotrexate, sulfasalazine, hydroxychloroquine, or leflunomide.[51] Tofacitinib is considered a Tier 4 or "specialty-tier drug" on the formulary, which means it has a higher cost share.[52] Health Net considers tofacitinib medically necessary if the patient is ages 18 years or older, has demonstrated intolerance to or has been failed by methotrexate, and has also been failed by separate three month trials of both adalimumab *and* etanercept.[53]

We were unable to locate a specific utilization management policy for tofacitinib under Medi-Cal, California's Medicaid program, but we note that tofacitinib is listed as a Tier 2 or "non-preferred drug" in its formulary. Tofacitinib requires prior authorization and may be subject to quantity limits.[54]

Baricitinib

The FDA label notes that baricitinib is indicated for the treatment of patients who have demonstrated an inadequate response to one or more TNF inhibitors; to that end, all national commercial insurers require step therapy in order to obtain baricitinib. All insurers first require a trial with methotrexate or a conventional DMARD, but several insurers list different preferred TNF inhibitors and vary in the number that must be tried before obtaining baricitinib.[50]

Anthem has the most restrictive policy of the national commercial insurers surveyed. In the following order, patients are required to 1) try and be failed by one or more TNF inhibitors; 2) try and be failed by one or more non-TNF inhibitors or non-biologics, such as tocilizumab (Actemra®, Genentech), sarilumab, anakinra (Kineret®, Sobi), abatacept (Orencia®, Bristol-Myers Squibb), and tofacitinib, and lastly; 3) try and be failed by two preferred biologic agents, which include etanercept, adalimumab, infliximab, and golimumab.[50]

The policies of the other four national commercial insurers align closer to the labeled indication of baricitinib. Aetna and UHC require that patients try and be failed by at least one TNF inhibitor, while Cigna and Humana require a trial of two preferred products, which include both TNF inhibitors and non-TNF inhibitors.[50]

Kaiser Permanente lists baricitinib as a Tier 4 or "specialty-tier drug" on its formulary, which means it has a higher cost share.[48] To obtain coverage for baricitinib under Health Net, patients must be ages 18 years or older, have demonstrated intolerance to or been failed by methotrexate, and have been failed by separate three month trials of both adalimumab and etanercept, and tofacitinib.[53]

We were unable to locate a specific utilization management policy for baricitinib under Medi-Cal, but we note that baricitinib is listed as a Tier 2 or "non-preferred drug" in its formulary. Baricitinib requires prior authorization and may be subject to quantity limits.[54]

Coverage Comparison of Infliximab and Infliximab-dyyb

To inform a Policy Roundtable discussion about the potential role of the biosimilars in RA treatment, we surveyed policies for infliximab and infliximab-dyyb to see whether any discrepancies in coverage exist between the two products.

All surveyed national commercial insurers, except Aetna and UHC, maintain differing coverage criteria for accessing infliximab versus infliximab-dyyb. As noted above, Aetna treats infliximab and all related products (i.e., infliximab-dyyb) as preferred, least costly brands. The initial criteria for obtaining infliximab and infliximab-dyyb are identical; patients must be ages 18 years or older with a documented diagnosis of active moderate-to-severe RA and must have tried and been failed by methotrexate or a comparable conventional DMARD if methotrexate is contraindicated. Similar to Aetna, UHC designates infliximab and infliximab-dyyb as co-preferred products for all commercial plans.[50]

Anthem, Cigna, and Humana designate infliximab as the preferred product compared to infliximab-dyyb. To access infliximab under the above three insurers, patients must be ages 18 years or older, have a documented diagnosis of moderate-to-severe RA, and must have tried and been failed by a conventional DMARD, such as methotrexate, sulfasalazine, or leflunomide. However, to obtain infliximab-dyyb, the three insurers state more stringent criteria. Patients covered by Anthem must try and be failed by one preferred agent (i.e., infliximab, the reference product). If the patient is currently maintained on infliximab-dyyb, in order to continue on the biosimilar (the non-preferred brand), the patient must have undergone at least one switch between the reference product and a biosimilar. Cigna requires that patients demonstrate intolerance to infliximab and inadequate response or contraindication to one other TNF inhibitor if they are to obtain infliximab-dyyb. Humana's criteria are similar: patients must have been failed by infliximab before accessing infliximab-dyyb.

Kaiser Permanente lists both infliximab and infliximab-dyyb on its Tier 4 or "specialty-drug tier."[52] However, several divisions of Kaiser Permanente, including Kaiser Permanente Colorado and Northwest, have switched patients from infliximab to infliximab-dyyb. Kaiser Permanente Colorado conducted a survey of patients who had switched to the biosimilar between September 1, 2017 and January 31, 2018, and found that 80% of patients were either "satisfied" or "very satisfied" with the switch.[55] As of October 2018, Kaiser Permanente Northwest uses the biosimilar for about 97% of infliximab-related infusions, whereas infliximab biosimilars account for only 2.4% utilization across the US.[56,57]

Health Net offers slightly different criteria for infliximab and infliximab-dyyb. The only restriction for accessing infliximab is a three-month trial and subsequent failure by methotrexate or sulfasalazine, leflunomide, or hydroxychloroquine if methotrexate is contraindicated. If patients are

to obtain infliximab-dyyb, they must first be failed by a three-month trial of either infliximab or golimumab.[53]

Medi-Cal does not differentiate between infliximab and infliximab-dyyb in its coverage policy. To obtain either the biosimilar or the reference product, patients must submit documentation that demonstrates medical necessity and intolerance to or failure by a conventional DMARD.[54]

2.2 Clinical Guidelines

***2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis*[36]**

The ACR guidelines were updated in 2015 and were written prior to the FDA approval of baricitinib and upadacitinib. As an overarching principle, the guidelines note that treatment decisions should be made through a shared decision-making process between the clinician and patient. Any treatment decision should factor in patient preference and comorbidities.

In patients with RA who are naïve to conventional DMARDs, the guidelines recommend the use of conventional DMARDs as monotherapy regardless of disease activity. If disease activity remains moderate or high despite utilization of conventional DMARDs, the use of a TNF inhibitor as monotherapy is recommended over tofacitinib monotherapy; in addition, the use of a TNF inhibitor plus methotrexate is recommended over tofacitinib plus methotrexate.

In patients with established RA who have moderate or high disease activity and are naïve to conventional DMARDs, conventional DMARD monotherapy is recommended over tofacitinib or conventional DMARD combination therapy. If disease activity remains moderate or high despite conventional DMARD monotherapy, the use of combination therapy or the addition of a TNF inhibitor or non-TNF inhibitor, or tofacitinib is recommended in no particular order of preference, rather than continuing conventional DMARD monotherapy. For patients on a TNF inhibitor with persistently moderate or high disease activity, the use of a non-TNF inhibitor is recommended over the use of another TNF inhibitor with or without methotrexate, and tofacitinib with or without methotrexate. For patients on a non-TNF inhibitor with moderate or high disease activity, the use of another non-TNF inhibitor with or without methotrexate is recommended over the use of tofacitinib with or without methotrexate. If patients on multiple sequential TNF inhibitor therapies continue to experience moderate or high disease activity, the use of a non-TNF inhibitor with or without methotrexate is preferred over tofacitinib or another TNF inhibitor with or without methotrexate. If disease activity still remains moderate or high, the use of tofacitinib with or without methotrexate is recommended over another TNF inhibitor with or without methotrexate.

For patients with congestive heart failure, the use of combination DMARDs, non-TNF inhibitors, or tofacitinib is recommended over TNF inhibitors.

EULAR Recommendations for the Management of Rheumatoid Arthritis with Synthetic and Biological Disease-Modifying Antirheumatic Drugs: 2016 Update[58]

The EULAR guidelines were updated in 2016. Similar to the ACR guidelines, the EULAR guidelines state that that treatment should be an individualized and shared decision between clinician and patient. Treatment decisions should be determined by patient-specific factors as well as disease activity.

For patients on conventional DMARDs with poor prognostic factors and who have not met their treatment target, a biologic agent or a JAK inhibitor should be considered as an add-on therapy to the treatment regimen. Biologic agents and JAK inhibitors should be combined with conventional DMARDs for superior clinical effectiveness. As a clarification, however, biologic agents and JAK inhibitors should not be used concomitantly. Patients who are failed by either a biologic agent or a JAK inhibitor should consider switching to different biologic agent or JAK inhibitor.

3. Comparative Clinical Effectiveness

3.1 Overview

To inform our analysis of the comparative clinical effectiveness of JAK inhibitors for patients with moderately-to-severely active RA who experienced an inadequate response to previous methotrexate or other conventional DMARD therapy, we abstracted evidence from available clinical studies, whether in published, unpublished, or abstract form. The drugs and regimens of interest are included in Table 1.1 in Section 1. The evaluation of JAK inhibitors is followed by a description of the evidence for the biosimilar exemplar infliximab-dyyb.

As described in Section 1, we focused on evidence on the comparative clinical effectiveness of JAK inhibitors in the target population (i.e., moderate-to-severe disease with inadequate response or intolerance to conventional DMARDs).

Our review focused on key clinical outcomes common to RA trials as well as patient-reported outcomes, health care system utilization, and work loss where evidence was available.

Clinical Benefits

Trial outcomes preferred at three months:

- Disease activity (DAS28, SDAI, CDAI)
- ACR20/50/70 response
- Function (HAQ-DI)
- Radiographic progression (modified total Sharp score)

Patient-reported outcomes:

- Health-related quality of life (e.g., SF-36)
- Pain
- Fatigue

Non-Clinical Benefits

- Health care system utilization and associated costs
- Productivity

Harms

- Treatment-related adverse events (e.g., deaths, rates of infection, malignancies)
- Treatment tolerability (e.g., discontinuation due to adverse events)

3.2 Methods

Data Sources and Searches

Procedures for the systematic literature review assessing the evidence on JAK inhibitors for RA followed established best research methods.[59,60] We conducted the review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[61] The PRISMA guidelines include a checklist of 27 items, which are described further in Appendix Table A1.

We searched MEDLINE, EMBASE, Cochrane Database of Systematic Reviews, and Cochrane Central Register of Controlled Trials for relevant studies. Each search was limited to English-language studies of human subjects and excluded articles indexed as guidelines, letters, editorials, narrative reviews, case reports, or news items. We included abstracts from conference proceedings identified from the systematic literature search. All search strategies were generated utilizing the Population, Intervention, Comparator, and Study Design elements described above. The proposed search strategies included a combination of indexing terms (MeSH terms in MEDLINE and EMTREE terms in EMBASE) as well as free-text terms.

To supplement the database searches, we performed manual checks of the reference lists of included trials and systematic reviews and invited key stakeholders to share references germane to the scope of this project. We also supplemented our review of published studies with data from conference proceedings, regulatory documents, information submitted by manufacturers, and other grey literature when the evidence met ICER standards (for more information, see http://icer-review.org/methodology/icers-methods/icer-value-assessment-framework/grey-literature-policy/). Where feasible and deemed necessary, we also accepted data submitted by manufacturers "in-confidence," in accordance with ICER's published guidelines on acceptance and use of such data (https://icer-review.org/use-of-in-confidence-data/).

Study Selection

We included evidence from RCTs, comparative observational studies, and high-quality systematic reviews of JAK inhibitors and infliximab-dyyb. We excluded single-arm studies as well as early clinical studies focused on very short-term tolerability; Phase II studies were included if they reported on outcomes of interest and met other specified selection criteria. We required studies to include minimum total sample sizes of 100 and 1,000 for RCTs and observational studies, respectively. Our sample set was further limited to studies with at least three months duration of follow-up for adequate surveillance of outcomes. However, long-term extension studies that evaluated outcomes more than three years after comparator-arm crossover was allowed were excluded, given the challenges with attributing study findings to initial treatment.

Study comparisons must have been between active agents: we excluded trials in which the only comparator was placebo without background methotrexate or another conventional DMARD as well as studies that pooled individual agents into a single treatment arm (e.g., TNF inhibitors). We also excluded studies that only compared combination therapy (JAK inhibitor plus conventional DMARD) to monotherapy with the same TIM, but we did include data on both TIM monotherapy and combination therapy from trials with a third arm that represented conventional DMARD therapy alone. Biosimilar studies of infliximab-dyyb were included if they involved comparisons between the biosimilar and reference product and focused on the outcomes of interest; studies examining only pharmacodynamics or pharmacokinetics were excluded. Finally, we only included data from the FDA-approved dosage(s) for each drug.

In recognition of the evolving evidence base for RA, we supplemented our review of published studies with data from conference proceedings, regulatory documents, information submitted by manufacturers, and other grey literature that met ICER standards for review (for more information, see https://icer-review.org/methodology/icers-methods/icer-value-assessment-framework/grey-literature-policy/). We excluded abstracts that reported data available in peer-reviewed publications as well as abstracts on therapies that have been on the market in the US for at least three years. Further details on the search algorithms, methods for study selection, data extraction, quality assessment, assessment for publication bias, and our approach to meta-analyses are available in Appendix A.

Assessment of Level of Certainty in Evidence

We used the ICER Evidence Rating Matrix to evaluate the level of certainty of a net health benefit in the available evidence among each of the interventions of focus (see Appendix D).[62]

The matrix is meant to be a consistent and transparent system leading to an evidence rating that can guide coverage and formulary placement decisions.

Data Synthesis and Statistical Analyses

Evidence tables were generated based on the data abstracted above and are presented descriptively in the sections that follow (see Appendix D). In addition, because the treatments of interest have not usually been directly compared, we developed quantitative, indirect comparisons among all agents using a Bayesian NMA for ACR response outcomes. There was not sufficient evidence to form a network for measures of disease activity. Consistent with prior published methods,[63] ACR20/50/70 response outcomes from clinical trials were tabulated to create numbers of patients in mutually exclusive categories (i.e., <20, 20-49, 50-69, ≥70); these data were analyzed using a random-effects, multinomial likelihood model to generate proportions of patients in each category. An adjusted model was specified with a covariate for conventional DMARD response

rates as a possible control for between-study heterogeneity and general confounding. The NMA was conducted using JAGS software (Version 4.3.0) via R using the R2jags package.

3.3 Results

Study Selection

Our literature search identified 540 potentially relevant references (see Appendix A, Figure A1), of which 39 met our inclusion criteria. In total, we included 39 reports of 16 RCTs. Primary reasons for study exclusion included the use of regimens or dosing schedules not approved by the FDA, study populations that included patients who were naïve to methotrexate and/or other conventional DMARDs, and smaller sample sizes (<100 for RCTs or <1,000 for observational studies). Additional details of the included references are described in Appendix D.

The 16 RCTs provided data on more than 7,000 patient enrollments. Of these RCTs, six focused on JAK inhibitor combination therapy with methotrexate or other conventional DMARDs in TIM-naïve or predominantly naïve (80% or more) populations and four focused on JAK inhibitor combination therapy with methotrexate or other conventional DMARDs in TIM-experienced patients. The remainder focused on monotherapy.

We identified a total of three RCTs that involved head-to-head comparisons of JAK inhibitors with a TIM, all with adalimumab.

The search identified one randomized trial of infliximab-dyyb for RA and one prospective cohort study.

Quality of Individual Studies

We rated 15 trials to be of good quality (94%) and one of poor (6%) quality using criteria from the US Preventive Services Task Force (USPSTF).[64] Trials of good quality had study arms that were comparable at baseline, employed valid instruments to evaluate outcomes, and did not demonstrate differential attrition. The poor-quality study did not report on randomization methods, allocation concealment, or blinding, and was much smaller than the other studies.

Most of the trials permitted the use of rescue medication as early as three months following randomization, and treatment-arm crossover was often allowed at three months. While these trials had good internal validity during the pre-crossover period, extrapolation to longer-term effects poses challenges. Thus, we have emphasized the three-month outcomes. In addition, because some measures (e.g., radiographic progression) are relatively insensitive to short-term changes, these required imputations due to crossover effects or missing data.

Outcome-Specific Considerations

Our discussion of results is focused on the major clinical and functional outcomes of the available studies, including measures of disease activity and remission, ACR response, radiographic progression, and function or disability. Given the current treat-to-target paradigm, remission or low disease activity at three months is given priority. Specific considerations regarding these measures are described below.

The DAS28-ESR was the most frequent measure of disease activity across all trials, reported in about 80% of the trials that included disease activity measures. Other types of disease activity measures reported less frequently included the DAS28-CRP, CDAI, and SDAI. Most studies used remission rates as one of the study endpoints, defined as DAS28 score ≤2.6, SDAI score ≤3.3, or CDAI score ≤2.8. They also reported low disease activity, defined as DAS28 score ≤3.2, SDAI score ≤11, or CDAI score ≤10. Low disease activity may be the most relevant under the treat-to-target paradigm in which treatment switching is encouraged within three months for patients with ongoing moderate-to-severe disease activity.[5] Given the multiplicity of measures as well as their evolution over time, we opted to describe our findings in descriptive fashion only rather than conduct an NMA. In studies that report all four measures, rates of remission and low disease activity with DAS28-ESR are consistently lower than those assessed with DAS-CRP. Disease activity and remission using SDAI or CDAI are usually comparable to each other, but classify fewer patients with low disease activity or remission compared with the DAS28 measures. The DAS28-CRP may overestimate response to therapy for drugs that lower IL-6 activity (tocilizumab and JAK inhibitors) because IL-6 increases CRP levels.

As noted in Section 1 of this report, the ACR response criteria represent at least 20%, 50%, or 70% improvement in the core measures of RA activity. The primary endpoint in the majority of RCTs included in our analysis set was ACR20. However, ACR20 is generally considered minimal improvement, while ACR50/70 are regarded to be more clinically significant levels of response.[65] We present findings for all levels of response and note where results are similar or inconsistent across these levels.

Structural damage is most commonly assessed using the Sharp score. The Sharp score sums measures of both joint erosion and joint space narrowing across several joints in the hands, wrists, and feet (the Van der Heijde modified Sharp score includes an analysis of several joints in the feet, although other approaches focus solely on the hands). The score has been modified and adapted over time, with iterations from Van der Heijde[66,67] and Genant[68] appearing most commonly in our review.

However, within the studies included in our review, the Genant and Van der Heijde methods were not applied consistently. Maximum possible scores were frequently not specified by trial investigators, and across the studies that did provide detail on the maximum achievable score,

there was considerable variation (e.g., total scores using the Van der Heijde method ranged from 380 to 448).[69,70] Consequently, there is substantial uncertainty in the degree of comparability of results among studies. Furthermore, because radiographic progression occurs gradually over time, this outcome is most frequently reported after at least 12 months of follow-up. Trials that permit early escape and/or crossover must extrapolate how much joint damage would likely occur had the patient continued with the initial treatment. These imputations are often based on a very short duration of observation (e.g., 16 weeks) and may underestimate the true progression that patients would experience had no adjustment to their therapy occurred. Missing or post-rescue therapy data were typically imputed using linear extrapolation of data from baseline and post-baseline radiographic assessment timepoints. Finally, we note that in addition to issues of multiple methods and variants to assess radiographic progression, all such measures rely on clinician interpretation of radiographic data.

The HAQ-DI, a patient completed disability assessment, was the most widely reported measure of function in most studies we identified. HAQ-DI scores range from 0 to 3, with higher scores indicating greater disability. In many published trials, a change of 0.22 in the HAQ-DI score,[71] or a more stringent 0.3,[72] is considered a minimum clinically important difference.

Clinical Benefits

Because our study entry criteria involved patient populations with an inadequate response to conventional DMARD therapy, it is unsurprising that the results of conventional DMARD-controlled studies consistently favored JAK inhibitors for all major outcomes. As noted above, our focus of attention in the report is on the measures of disease activity/remission as well as ACR response, radiographic progression, function/disability, and harms. A summary of other outcomes (e.g., pain, fatigue, quality of life) can be found in Appendix D.

The results are organized by indication. First, we consider patients who are predominantly TIM naïve. Some of these trials included up to 20% of patients who had been failed by a TIM (see Appendix D for details). Since baricitinib is not indicated for this population, no trials of baricitinib were included. Findings from head-to-head studies of JAK inhibitors with adalimumab are also presented for the population of TIM-naïve/mixed population. In the second section, we consider the TIM-experienced population. There were no head-to-head trials for this population. For each JAK inhibitor, we describe results according to their use in combination with conventional DMARDs.

TIM-Naïve/Mixed Populations

Comparisons to Conventional DMARD Therapy

Both upadacitinib and tofacitinib generated superior improvements in disease activity, remission, and ACR response relative to conventional DMARD therapy alone in TIM-naïve/mixed populations at 12 weeks. These results were consistent when reported at 24 and 48 weeks. Radiographic progression was also reduced, but differences in measures used made comparisons across studies difficult. Improvements in function and disability were statistically superior for both upadacitinib and tofacitinib. A greater proportion of patients receiving JAK inhibitors met clinically important thresholds for HAQ-DI change.

A total of six RCTs compared combination therapy with JAK inhibitors plus conventional DMARD therapy with conventional DMARDs alone in TIM-naïve or mixed populations. In addition, a pooled study of randomized trials of tofacitinib reported 12 week outcomes that were not reported in the primary studies, so those results are included here as well.[12] The proportions of patients achieving low disease activity or remission at 12 weeks were substantially greater in JAK inhibitor groups relative to conventional DMARDs alone (Table 3.1). Results achieved statistical significance for both upadacitinib and tofacitinib. It is challenging to compare the results of upadacitinib to tofacitinib as studies of upadacitinib primarily reported the DAS28-CRP, while the studies of tofacitinib used the DAS28-ESR, which generally estimates that a lower proportion of patients achieve remission or low disease activity compared with the DAS28-CRP. In addition, the primary tofacitinib trials rarely reported 12-week outcomes, though they are presented in the pooled study. When measured with the CDAI or SDAI, approximately 40% of patients treated with upadacitinib plus a conventional DMARD achieve at least low disease activity (NNT 2.5) compared with approximately 33% of patients treated with tofacitinib plus a conventional DMARD (NNT 3). Note that these results should not be directly compared as there is likely some degree of selection bias in the patients studied and we were unable to perform an NMA with the available results.

Table 3.1. Disease Activity Outcomes of JAK Inhibitors and Comparators in TIM Naïve/Mixed Patients at 12 Weeks/Three Months

Treatment	N	DAS28-ESR or CRP	DAS28 Change from Baseline (Mean)	DAS28 Low Disease Activity (%)	DAS28 Remission (%)	CDAI Low Disease Activity (%)	SDAI Low Disease Activity (%)
SELECT-COMPARE[6]							
Upadacitinib + MTX	651	DAS28-CRP	NR	45*	29*	40*	40*
Adalimumab + MTX	327	DAS28-CRP	NR	29	18	30	30
Placebo + MTX	651	DAS28-CRP	NR	14	6	16	15
SELECT-NEXT[7]							
Upadacitinib + MTX	141	DAS28-CRP	NR	48*	31*	40*	42*
Placebo + MTX	79	DAS28-CRP	NR	17	10	19	19
ORAL Sync[8]							
Tofacitinib + MTX	315	DAS28-ESR	NR	NR	8*	NR	NR
Placebo + MTX	159	DAS28-ESR	NR	NR	0.5	NR	NR
ORAL Standard[9]							
Tofacitinib + MTX	204	DAS28-ESR	NR	NR	NR	NR	NR
Adalimumab + MTX	204	DAS28-ESR	NR	NR	NR	NR	NR
Placebo + MTX	108	DAS28-ESR	NR	NR	NR	NR	NR
ORAL Strategy[10]							
Tofacitinib + MTX	376	DAS28-ESR	NR	NR	NR	NR	NR
Adalimumab + MTX	386	DAS28-ESR	NR	NR	NR	NR	NR
ORAL Scan[11]							
Tofacitinib + MTX	321	DAS28-ESR	NR	NR	NR	NR	NR
Placebo + MTX	160	DAS28-ESR	NR	NR	NR	NR	NR
Pooled Tofacitinib Trials[12]							
Tofacitinib + MTX	1,043	DAS28-ESR	NR	16.6*	7.3*	32.4*	34.6*
Placebo + MTX	638	DAS28-ESR	NR	4.5	2.3	14.3	14.2

CDAI: Clinical Disease Activity Index, CRP: C-reactive protein, DAS28: Disease Activity Score 28, ESR: erythrocyte sedimentation rate, MTX: methotrexate, NR: not reported, SDAI: Simplified Disease Activity Index
*p<0.001.

The percentages of patients achieving ACR response at 12 weeks were also statistically-significantly greater for JAK inhibitors in combination with conventional DMARDs versus conventional DMARDs alone (Table 3.2 on the following page). This was true not only for ACR20 response (the primary endpoint in most studies), but for ACR50 and 70 as well. There were no marked differences in ACR responses between upadacitinib and tofacitinib across the trials, though the changes in HAQ-DI scores were slightly greater for upadacitinib. Again, these comparisons are not head-to-head and are subject to potential selection and measurement bias. The results of our NMA for ACR categories are summarized in Table 3.3. The results for tofacitinib and adalimumab are very similar and those for upadacitinib are slightly better (more patients in the ACR50 and 70 categories, fewer

than ACR20). All three TIMs showed markedly better results than continuing conventional DMARDs alone in patients who had been failed by conventional DMARDs.

Table 3.2. ACR20/50/70 and HAQ-DI Outcomes of JAK Inhibitors and Comparators in TIM Naïve/Mixed Patients at 12 Weeks/Three Months

Treatment	N	ACR20 (%)	ACR50 (%)	ACR70 (%)	Change in HAQ-DI	HAQ-DI Improved ≥0.22 (%)
SELECT-COMPARE[6]						
Upadacitinib + MTX	651	71*	45*	25*	-0.60*	NR
Adalimumab + MTX	327	63	29	13	-0.49	NR
Placebo + MTX	651	36	15	5	-0.28	NR
SELECT-NEXT[7]						
Upadacitinib + MTX	141	64*	38*	21*	-0.61*	74*
Placebo + MTX	79	36	15	6	-0.26	52
ORAL Sync[8]						
Tofacitinib + MTX	315	56*	27*	9*	-0.44*	NR
Placebo + MTX	159	27	9	2	-0.16	NR
ORAL Standard[9]						
Tofacitinib + MTX	204	61.2†	34.2†	12.1†	-0.55*	NR
Adalimumab + MTX	204	56.5	23.9	8.4	-0.49	NR
Placebo + MTX	108	29	10.8	1.2	-0.24	NR
ORAL Strategy[10]						
Tofacitinib + MTX	376	70.9	40.8	19.3	-0.54	NR
Adalimumab + MTX	386	69.3	37.5	14.3	-0.49	NR
ORAL Scan[11]						
Tofacitinib + MTX	321	NR	NR	NR	-0.40*	NR
Placebo + MTX	160	NR	NR	NR	-0.15	NR
Pooled Tofacitinib Trials[12]						
Tofacitinib + MTX	1,043	60.3*	32.7*	12.9*	NR	52.9*
Placebo + MTX	638	26.5	9.7	2.8	NR	28.7

ACR: American College of Rheumatology, HAQ-DI: Health Assessment Questionnaire Disability Index, MTX: methotrexate, NR: not reported

*p <0.001.

†p<0.05 vs. placebo.

Table 3.3. NMA-Derived Proportions of Patients in Each ACR Response Category for JAK Inhibitors and Comparators in TIM Naïve/Mixed Patients at 12 Weeks/Three Months

Treatment	ACR<20	ACR20-50	ACR50-70	ACR70-100
Upadacitinib + cDMARD	33.4%	28.7%	12.8%	25.0%
Tofacitinib + cDMARD	40.5%	28.6%	11.5%	19.4%
Adalimumab+ cDMARD	41.1%	28.5%	11.4%	19.0%
Placebo + cDMARD	72.9%	18.2%	4.5%	4.3%

ACR: American College of Rheumatology, cDMARD: conventional disease-modifying antirheumatic drug

Head-to-Head Studies of JAK Inhibitors in TIM-Naïve/Mixed Populations

There were no head-to-head studies of JAK inhibitors. However, there were head-to-head studies of JAK inhibitors and adalimumab in the TIM-naïve/mixed population (Tables 3.1 and 3.2 above).

JAK Inhibitors: Upadacitinib versus Adalimumab

In one head-to-head trial, upadacitinib combination therapy was superior to adalimumab combination therapy in rates of disease remission, ACR response, change in pain, and improvement in HAQ-DI after 12 weeks of follow-up. In general, differences observed at 12 weeks were preserved at 24 and 48 weeks of follow-up, although some drug switching occurred between weeks 14 and 24, so randomization was not fully preserved.

We identified one head-to-head study that compared upadacitinib plus methotrexate with adalimumab plus methotrexate conducted in a primarily TIM-naïve population.[6]

Disease Activity and Remission

There were statistically significant differences observed in the proportion of patients achieving DAS28-CRP clinical remission (<2.6) between combination therapy with upadacitinib plus methotrexate versus adalimumab plus methotrexate (29% vs. 18%, p <0.001) as well as DAS28-CRP <3.2 (45% vs. 29%) and CDAI ≤2.6 (13% vs. 8%).[6]

ACR20/50/70

Relative to adalimumab combination therapy, upadacitinib plus methotrexate showed statistical differences at the ACR20 level (71% achieved ACR20 with upadacitinib vs. 63% with adalimumab, p≤0.05), ACR50 (45% vs. 29%, p<0.001), and ACR70 (25% vs. 13%, p<0.001) at 12 weeks of follow-up.[6] These differences were preserved at 26 and 48 weeks of follow-up.

Radiographic Progression

The rate of no radiographic progression was similar for upadacitinib (86%) and adalimumab (88%, p=NS) at 48 weeks.

HAQ-DI

In the trial comparing upadacitinib combination therapy with adalimumab combination therapy, there was a statistically significant difference observed between the mean HAQ-DI change from baseline at 12 weeks in the two groups (-0.60 vs. -0.49, p<0.01).[6]

Other Patient-Reported Outcomes

After 12 weeks of follow-up, patients randomized to upadacitinib experienced greater improvements in quality of life, pain, and fatigue than those randomized to adalimumab therapy.[6]

JAK Inhibitors: Tofacitinib versus Adalimumab

In one head-to-head trial, tofacitinib combination therapy was not statistically different from adalimumab combination therapy in rates of remission achieved, ACR response, and improvement in HAQ-DI after six months of follow-up.

We identified two head-to-head studies that compared tofacitinib plus methotrexate with adalimumab plus methotrexate conducted in a mostly TIM-naïve population.[10,73] The results of the studies are summarized below.

Disease Activity and Remission

There was no statistically significant difference observed in the proportion of patients achieving DAS28-ESR clinical remission or low disease activity between combination therapy with tofacitinib plus methotrexate versus adalimumab plus methotrexate.[10,73]

ACR20/50/70

Relative to adalimumab combination therapy, tofacitinib plus methotrexate showed statistical differences only at the ACR70 level (20% achieved ACR70 with tofacitinib vs. 10% with adalimumab; p≤0.01) at 24 weeks of follow-up in ORAL Standard,[73] but not in ORAL Strategy (25% vs. 21%) at six months.[10]

Radiographic Progression

We did not identify any studies of tofacitinib in comparison to another TIM that reported on radiographic progression.

HAQ-DI

In the trials comparing tofacitinib combination therapy with adalimumab combination therapy, there were no statistically significant differences observed between the mean HAQ-DI change from baseline at 24 weeks or one year between the two groups.[10,73]

Other Patient-Reported Outcomes

After 12 weeks of follow-up, patients experienced comparable improvements in quality of life, pain, and fatigue with combination tofacitinib or adalimumab therapy.[74]

JAK Inhibitors: Baricitinib versus Adalimumab

We identified one head-to-head trial that compared baricitinib plus methotrexate to adalimumab plus methotrexate in mostly TIM-naïve patients.[75] However, this trial used a dose of baricitinib that was not approved by the FDA (4 mg instead of 2 mg), so the trial was excluded.

TIM-Experienced Population

Studies for all three JAK inhibitors demonstrated statistically and clinically significant improvements in measures of disease activity, ACR response, and HAQ improvement versus conventional DMARDs alone, but there were fewer trials with fewer participants, so the confidence intervals are wider than those for JAK inhibitors in the TIM-naïve population.

RCT evidence was more limited in patients with inadequate response to one or more TIMs. A total of three randomized trials were identified plus the pooled study of tofacitinib randomized trials. All of the trials studied combination therapy with conventional DMARDs versus conventional DMARDs alone (see Tables 3.4 and 3.5). There was one RCT for each of the three JAK inhibitors. The evidence was similar for all three: significantly greater proportions of patients randomized to JAK inhibitors plus conventional DMARDs achieved low disease activity and remission at three and six months by multiple measures of disease activity (Appendix D) compared with conventional DMARDs alone. They also had greater proportions of patients meeting the ACR20/50/70 response levels and greater improvements in the HAQ-DI compared with conventional DMARDs alone.

Table 3.4. Disease Activity Outcomes of JAK Inhibitors and Comparators in TIM-Experienced Patients at 12 Weeks/Three Months

Treatment	N	DAS28-ESR or CRP	DAS28 Change from Baseline (Mean)	DAS28 Low Disease Activity (%)	DAS28 Remission (%)	CDAI Low Disease Activity (%)	SDAI Low Disease Activity (%)
SELECT-BEYOND[13]							
Upadacitinib + MTX	164	DAS28-CRP	NR	43*	NR	32*	34*
Placebo + MTX	169	DAS28-CRP	NR	14	NR	14	14
ORAL Step[14]							
Tofacitinib + MTX	133	DAS28-CRP DAS28-ESR	NR -1.8*	13 14.3†	6 6.7†	NR NR	NR NR
Placebo + MTX	159	DAS28-CRP DAS28-ESR	NR -0.7	4 5	1 1.7	NR NR	NR NR
Pooled Tofacitinib Trials[12]							
Tofacitinib + MTX	258	DAS28-ESR	NR	12.7†	6.6†	29.5*	29.8*
Placebo + MTX	191	DAS28-ESR	NR	5.1	2.3	14.4	13.8
RA-BEACON[15]							
Baricitinib + MTX	174	DAS28-CRP DAS28-ESR	-1.5*	24* 13‡	11† 6‡	24‡	22*
Placebo + MTX	176	DAS28-CRP DAS28-ESR	-0.7	9 4	4 1	11	9

CDAI: Clinical Disease Activity Index, CRP: c-reactive protein, DAS28: Disease Activity Score 28, ESR: erythrocyte sedimentation rate, MTX: methotrexate, NR: not reported, SDAI: Simplified Disease Activity Index

*p <0.001.

†p<0.05 vs. placebo.

‡p<0.01.

Table 3.5. ACR20/50/70 and HAQ-DI Outcomes of JAK Inhibitors and Comparators in TIM-Experienced Patients at 12 Weeks/Three Months

Treatment	N	ACR20 (%)	ACR50 (%)	ACR70 (%)	Change in HAQ-DI	HAQ-DI Improved ≥0.22 (%)
SELECT-BEYOND[13]						
Upadacitinib + MTX	164	65*	34*	12†	-0.41*	NR
Placebo + MTX	169	28	12	7	-0.16	NR
ORAL Step[14]						
Tofacitinib + MTX	133	41.7†	26.5*	13.6†	-0.43*	54.2†
Placebo + MTX	159	24.4	8.4	1.5	-0.18	40.5
Pooled Tofacitinib Trials[12]						
Tofacitinib + MTX	258	43.4*	24.4*	9.7†	NR	45.7
Placebo + MTX	191	24.6	10.5	3.1	NR	36.9
RA-BEACON[15]						
Baricitinib + MTX	174	49*	20‡	13*	-0.37*	59‡
Placebo + MTX	176	27	8	2	-0.18	43

ACR: American College of Rheumatology, HAQ-DI: Health Assessment Questionnaire without Disability Index, MTX: methotrexate

*p<0.001.

†p<0.05 vs. placebo.

‡p<0.01.

Harms

Rates of short-term serious adverse events (within six months) were generally comparable across all treatments, including JAK inhibitors, adalimumab, and conventional DMARDs. Infections (e.g., upper respiratory tract infections, bronchitis, nasopharyngitis) were the most common adverse events during treatment. Based on long-term (one year or more) trial data, upadacitinib, tofacitinib, and baricitinib showed comparable overall safety profiles.

Data on adverse events and discontinuations due to adverse events as well as specific adverse events of interest observed in clinical trials are presented in Tables 3.6-3.8. Of note, these represent events as recorded before treatment-arm crossover was permitted. Most adverse events were of mild-to-moderate severity. The most frequently reported adverse events were mild infections (upper respiratory tract infections, bronchitis, nasopharyngitis). The overall incidence of serious infections, deaths, and all serious adverse events was comparable among treatments, including conventional DMARD therapy. As noted below, however, adverse-event rates for tofacitinib were calculated over a 12-week pre-crossover period, versus 24-28 weeks for the other TIMs.

All three JAK inhibitors carry black-box warnings for serious infections, lymphoma, testing for latent tuberculosis prior to initiating therapy, and monitoring for active tuberculosis. In addition, baricitinib and upadacitinib carry black box warnings regarding thrombosis, including deep venous thrombosis, pulmonary embolism, and arterial thrombosis.

The rates of serious infection, serious adverse events, and discontinuation due to adverse events were generally comparable in the head-to-head trials comparing JAK inhibitors with adalimumab (see Appendix D). There was no evidence of material differences in the rates of malignancies or death between treatment groups across trials.

A systematic review and meta-analysis of 21 RCTs of JAK inhibitors including 11,144 patients looked specifically at rates of serious infections and herpes zoster.[16] The review found that the absolute risk of serious infections (hospitalization, death, IV antibiotics) was low and was not significantly higher when compared to the placebo group. Similarly, there was an increase in rates of herpes zoster infections, but the increase was not statistically significant.

Thrombotic events were rare. A systematic review reported that the annual incidence rate for thromboembolic events was 0.6 (0.4-0.9) for tofacitinib and 0.5 (0.3-0.7) for baricitinib.[17] In SELECT-COMPARE, 0.3% of patients in the upadacitinib group had a thrombotic event compared with 0.9% in the adalimumab group.

Table 3.6. Harms in the TIM-Naïve/Mixed Populations

Intervention	N	Time Point (Weeks)	Any AEs	Serious AEs	AE Leading to D/C	VTE	Malignancy	Death	Serious	Opportunistic	Herpes Zoster Virus
SELECT-COMPARE											
UPA 15 mg + cDMARD	651	26	417 (64.2)	24 (3.7)	23 (3.5)	All: 2 (0.3); PE: 1(0.2); DVT: 1 (0.2)	0 (0)	0 (0)	12 (1.8)	4 (0.6)	5 (0.8)
ADA 40 mg + cDMARD	327	26	197 (60.2)	14 (4.3)	20 (6.1)	All: 3 (0.9); PE: 3 (0.9); DVT: 0 (0)	1 (0.3)	2 (0.6)	5 (1.5)	1 (0.3)	1 (0.3)
PBO + cDMARD	651	26	347 (53.2)	19 (2.9)	15 (2.3)	All: 1 (0.2); PE: 1 (0.2); DVT: 0 (0)	2 (0.3)	2 (0.3)	5 (0.8)	4 (0.6)	3 (0.5)
SELECT-NEXT											
UPA 15 mg + cDMARD	221	12	125 (56.6)	9 (4.1)	7 (3.2)	0 (0)	0 (0)	0 (0)	1 (<1)	0 (0)	3 (1)
UPA 30 mg + cDMARD	219	12	118 (53.9)	6 (2.7)	13 (5.9)	0 (0)	2 (0.9)	1 (0.5)	3 (1)	3 (1)	6 (3)
cDMARD	221	12	108 (48.9)	5 (2.3)	7 (3.2)	0 (0)	0 (0)	1	1 (<1)	1 (<1)	1 (<1)
ORAL Sync											
TOF 5 mg + cDMARD	388	52	171.9 (152.5-193.8)	6.9 (4.6-10.5)	6.2 (4.0-9.6)	NR	NR	NR	NR	NR	9 (4.2) MTX w/o LEF
PBO (Advanced to TOF 5 mg at 6 Months)	159	52	342.3 (281.1-416.9)	10.9 (4.9-24.2)	5.4 (1.8-16.8)	NR	NR	NR	NR	NR	2 (6.8) MTX alone
ORAL Scan											
TOF 5 mg + cDMARD	321	12	TEAE: 157 (48.9)	12 (3.7)	15 (4.7)	NR	NR	2 (0.6)	2 (0.6)	NR	3 (0.9)

Intervention	N	Time Point (Weeks)	Any AEs	Serious AEs	AE Leading to D/C	VTE	Malignancy	Death	Serious	Opportunistic	Herpes Zoster Virus
PBO (Advanced to TOF 5 mg at 6 Months)	160	12	TEAE: 73 (45.6)	5 (3.1)	5 (3.1)	NR	NR	0 (0)	0 (0)	NR	0 (0)
ORAL Strategy											
TOF 5 mg	384	52	226 (59)	35 (9)	23 (6)	NR	1 (<1)	2 (1)	6 (2)	2 (1)	1/69 (1)
TOF 5 mg + cDMARD	376	52	231 (61)	27 (7)	26 (7)	NR	0 (0)	0 (0)	10 (3)	1 (<1)	2/75 (3)
ADA + cDMARD	386	52	253 (66)	24 (6)	37 (10)	NR	0 (0)	0 (0)	6 (2)	2 (1)	0/27 (0)
ORAL Standard											
TOF 5 mg + cDMARD	204	0-12	135 (66.2)	12 (5.9)	14 (6.9)	NR	NR	1 (<1)	3(1.5)	NR	0 (0)
ADA 40 mg + cDMARD	204	0-12	105 (51.5)	5 (2.5)	10 (4.9)	NR	NR	1 (<1)	0 (0)	NR	0 (0)
PBO Followed by TOF 5 mg + cDMARD	56	0-12	57(52.8)	2 (1.9)	3 (2.8)	NR	NR	NR	1(0.9)	NR	0 (0)
Pooled Tofacitinib											
bDMARD Naïve TOF 5 mg	1071	0-24	NR	131 (12.2)	96 (9.1)	NR	NR	7 (0.6)	32 (3.4)	NR	43 (4.0)
bDMARD Naïve cDMARD	651	0-6	NR	98 (15.0)	66 (10.1)	NR	NR	5 (0.7)	13 (2.0)	NR	13 (2.0)
bDMARD-IR TOF 5 mg	259	0-24	NR	34 (13.0)	39 (14.8)	NR	NR	1(1.2)	6 (2.3)	NR	13 (5.4)
bDMARD-IR cDMARD	193	0-6	NR	37 (19.0)	37 (18.9)	NR	NR	0 (0)	0 (0)	NR	10 (5.4)

ADA: adalimumab, AE: adverse event, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, D/C: discontinuation, LEF: leflunomide, mg: milligram, MTX: methotrexate, N: number, NR: not reported, PBO: placebo, PE: pulmonary embolism, TOF: tofacitinib, UPA: upadacitinib, VTE: venous thromboembolism

Table 3.7. Harms in Head-to-Head Trials

Inter-vention	N	Time Point (Weeks)	Any AEs	Serious AEs	AE Leading to D/C	VTE	Malignancy	Death	Serious	Oppor-tunistic	Herpes Zoster Virus
SELECT-COMPARE											
UPA 15 mg + cDMARD	651	26	417 (64.2)	24 (3.7)	23 (3.5)	All: 2 (0.3); PE: 1(0.2); DVT: 1 (0.2)	0 (0)	0 (0)	12 (1.8)	4 (0.6)	5 (0.8)
ADA 40 mg + cDMARD	327	26	197 (60.2)	14 (4.3)	20 (6.1)	All: 3 (0.9); PE: 3 (0.9); DVT: 0 (0)	1 (0.3)	2 (0.6)	5 (1.5)	1 (0.3)	1 (0.3)
PBO + cDMARD	651	26	347 (53.2)	19 (2.9)	15 (2.3)	All: 1 (0.2); PE: 1 (0.2); DVT: 0 (0)	2 (0.3)	2 (0.3)	5 (0.8)	4 (0.6)	3 (0.5)
ORAL Strategy											
TOF 5 mg	384	52	226 (59)	35 (9)	23 (6)	NR	1 (<1)	2 (1)	6 (2)	2 (1)	1/69 (1)
TOF 5 mg + cDMARD	376	52	231 (61)	27 (7)	26 (7)	NR	0 (0)	0 (0)	10 (3)	1 (<1)	2/75 (3)
ADA + cDMARD	386	52	253 (66)	24 (6)	37 (10)	NR	0 (0)	0 (0)	6 (2)	2 (1)	0/27 (0)
ORAL Standard											
TOF 5 mg + cDMARD	204	0-12	135 (66.2)	12 (5.9)	14 (6.9)	NR	NR	1 (<1)	3(1.5)	NR	0 (0)
ADA 40 mg + cDMARD	204	0-12	105 (51.5)	5 (2.5)	10 (4.9)	NR	NR	1 (<1)	0 (0)	NR	0 (0)
PBO then TOF 5 mg + cDMARD	56	0-12	57(52.8)	2 (1.9)	3 (2.8)	NR	NR	NR	1(0.9)	NR	0 (0)

ADA: adalimumab, AE: adverse event, cDMARD: conventional disease-modifying antirheumatic drug, D/C: discontinuation, mg: milligram, N: number, NR: not reported, PBO: placebo, PE: pulmonary embolism, TOF: tofacitinib, UPA: upadacitinib, VTE: venous thromboembolism

Table 3.8. Harms in TIM Experienced Populations

Inter-vention	N	Time Point (Weeks)	Any AEs	Serious AEs	AE Leading to D/C	VTE	Malig-nancy	Death	Serious	Oppor-tunistic	Herpes Zoster Virus
SELECT-BEYOND											
UPA 15 mg + cDMARD	164	12	91 (55.4)	8 (4.9)	4 (2.4)	PE: 1 (0.6)	1 (0.6)	0 (0)	1 (0.6)	1 (0.6)	1 (0.6)
UPA 30 mg + cDMARD	165	12	111 (67.3)	12 (7.3)	15 (9.1)	0 (0)	2 (1.2)	1 (0.6)	4 (2.4)	2 (1.2)	4 (2.4)
PBO + cDMARD	169	12	95 (56.2)	0 (0)	9 (5.3)	0 (0)	0 (0)	0 (0)	0	0 (0)	1 (0.6)
ORAL Step											
TOF 5 mg + cDMARD	133	0-12	71 (53.4)	2 (1.5)	8 (6.0)	NR	NR	0 (0)	0 (0)	0 (0)	NR
PBO (Advanced to TOF 5 mg)	132	0-12	75 (56.8)	6 (4.5)	7 (5.3)	NR	NR	0 (0)	0 (0)	0 (0)	NR
Pooled Tofacitinib											
bDMARD Naïve TOF 5 mg	1071	0-24	NR	131 (12.2)	96 (9.1)	NR	NR	7 (0.6)	32 (3.4)	NR	43 (4.0)
bDMARD Naïve cDMARD	651	0-6	NR	98 (15.0)	66 (10.1)	NR	NR	5 (0.7)	13 (2.0)	NR	13 (2.0)
bDMARD-IR TOF 5 mg	259	0-24	NR	34 (13.0)	39 (14.8)	NR	NR	1(1.2)	6 (2.3)	NR	5.4
bDMARD-IR cDMARD	193	0-6	NR	37 (19.0)	37 (18.9)	NR	NR	0 (0)	0 (0)	NR	0 (0)
RA-BEACON											
BAR 2 mg + cDMARD	174	0-12	107 (61.0)	3 (2.0)	7 (4.0)	NR	NR	0 (0)	3 (2)	61 (35.0)	2 (1.0)
cDMARD	176	0-12	96 (55.0)	7 (4.0)	4 (2.0)	NR	NR	0 (0)	3 (2)	35 (20.0)	1 (<1)
BAR 2 mg + cDMARD	174	0-24	123 (71.0)	7 (4.0)	7 (4.0)	NR	NR	0 (0)	4 (2)	76 (44.0)	2 (1.0)
cDMARD	176	0-24	112 (64.0)	13 (7.0)	7 (4.0)	NR	NR	0 (0)	5 (3)	55 (31.0)	2 (1.0)

ADA: adalimumab, AE: adverse event, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, D/C: discontinuation, mg: milligram, N: number, NR: not reported, PBO: placebo, TOF: tofacitinib, UPA: upadacitinib, VTE: venous thromboembolism

Observational Study

In a prospective cohort study analyzing data from the Dutch RA monitoring (DREAM) registry, patients with RA who had prior treatment with at least two conventional DMARDs including methotrexate, starting their first TNF inhibitor (adalimumab, infliximab, or etanercept), were followed for up to five years.[76]

The unadjusted incidence rate of a first serious infection per 100 patient-years was 2.61 (95% CI 2.21 to 3.00) for adalimumab, 3.86 (95% CI 3.33 to 4.40) for infliximab, and 1.66 (95% CI 1.09 to 2.23) for etanercept. Age, year of starting TNF therapy, comorbidities at baseline, and DAS28 score over time were included as confounders. No difference in risk for serious infections was found between adalimumab and infliximab (adjusted HR: 0.90 [95% CI 0.55 to 1.48]), but the risk of serious infections was significantly lower for etanercept than both infliximab (adjusted HR=0.49 [95% CI 0.29 to 0.83]) and adalimumab (adjusted HR=0.55 [95% CI 0.44 to 0.67]).[76] No data were presented for JAK inhibitors.

Controversies and Uncertainties

Across the RCTs identified for this review, only three were based on head-to-head comparisons of the TIMs of interest and none were head-to-head comparisons of JAK inhibitors. The paucity of trial data and the differences in reported measures of disease activity at three months precluded using an NMA to combine direct and indirect evidence of efficacy. Often the studies reported outcome measures at six months even though patients were eligible for rescue therapy and/or treatment-arm crossover 12-24 weeks after randomization. Since guidelines increasingly recommend treatment-switch decisions within three months of initiating therapy, the three-month outcomes would have been both more clinically relevant and methodologically more rigorous.

Extending trial-based analyses to longer timepoints requires imputation in many instances, which affects the level of confidence in the results no matter how responsibly it is done. In addition, key outcome measures such as disease activity scores, remission criteria, and modified Sharp score have undergone substantial revision and modification over the years, are employed variably in clinical trials, and not measured in others, making cross-trial comparisons problematic.

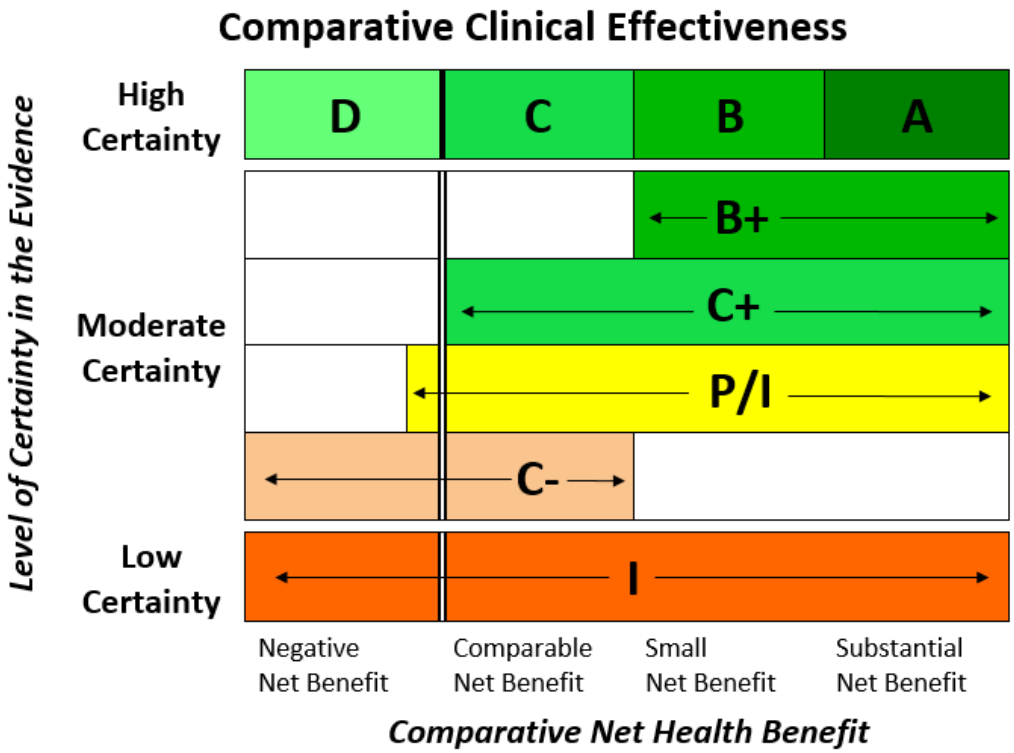
The course of RA may feature multiple periods of remission and flares of symptoms due to the complex and heterogeneous nature of the disease. TIM therapies are chronic, and the long-term effects of prolonged immunomodulation—both clinical benefits and potential harms—are not well-understood for all therapies, particularly for newer classes of TIMs. Evidence is beginning to emerge on the question of whether TIM doses can be modulated or if therapy can be suspended in patients with evidence of durable remission, but early results are limited and mixed. In addition, as noted in Section 1 of this report, some patients may be started on TIM treatment prior to

optimization of conventional DMARD therapy;[77] such challenges are common to other chronic diseases such as diabetes and heart failure as well.

Finally, although the introduction of TIMs has transformed clinical practice and improved the quality of life and functional capacity of many patients, there are still unanswered questions, including the relationship between levels of disease activity and radiographic evidence of joint damage, whether there are patient or clinical factors that predict response to specific therapies, and the totality of the disease impact on patients, families, and caregivers. As noted in Section 1 of this report, patient organizations do not feel that the current tools for patient-reported outcomes sufficiently capture their experience, but to date no new instruments have been accepted into common use in clinical trials.

3.4 Summary and Comment

Figure 3.1. ICER Evidence Rating Matrix



A = "Superior" *- High certainty of a substantial (moderate-large) net health benefit*
B = "Incremental" *- High certainty of a small net health benefit*
C = "Comparable" *- High certainty of a comparable net health benefit*
D = "Negative" *- High certainty of an inferior net health benefit*
B+ = "Incremental or Better" *- Moderate certainty of a small or substantial net health benefit, with high certainty of at least a small net health benefit*
C+ = "Comparable or Better" *- Moderate certainty of a comparable, small, or substantial net health benefit, with high certainty of at least a comparable net health benefit*
P/I = "Promising but Inconclusive" *- Moderate certainty of a comparable, small, or substantial net health benefit, and a small (but nonzero) likelihood of a negative net health benefit*
C- = "Comparable or Inferior" *- Moderate certainty that the point estimate for comparative net health benefit is either comparable or inferior*
I = "Insufficient" *- Any situation in which the level of certainty in the evidence is low*

Using the ICER Evidence Rating Matrix, our evidence ratings for selected comparisons of interest are provided in Table 3.9 for patients with moderately-to-severely active RA who have had an inadequate response to prior conventional DMARD therapy. As described previously, findings of studies using conventional DMARDs as the control indicate clinically and statistically significant improvements in most important disease measures for upadacitinib and tofacitinib combination therapy, so both receive a letter grade of "A" (high certainty of substantial net health benefit) relative to conventional DMARD therapy alone. However, there is a paucity of evidence on baricitinib 2 mg daily in this population and it does not have an FDA indication for this population, so we judge the comparative clinical effectiveness of baricitinib to be insufficient ("I").

Table 3.9. Evidence Ratings for Comparative Clinical Effectiveness: Selected Comparisons

Regimen Type/Comparison	Intervention	Comparator	Rating
TIM-Naïve Population			
Compared to cDMARD	Upadacitinib	cDMARDs	A
	Tofacitinib	cDMARDs	A
	Baricitinib	cDMARDs	I
Head-to-Head	Upadacitinib	Adalimumab	B+
	Tofacitinib	Adalimumab	C
	Baricitinib	Adalimumab	I
Biosimilar*	Infliximab-dyyb	Infliximab	C
TIM-Experienced Population			
Combination with cDMARD	Upadacitinib	cDMARDs	B+
	Tofacitinib	cDMARDs	B+
	Baricitinib	cDMARDs	B+

cDMARD: conventional disease-modifying antirheumatic drug, TIM: targeted immune modulator
*Rationale behind the biosimilar evidence rating may be found in Section 3.5.

Single RCTs have also evaluated combination therapy regimens of both upadacitinib and tofacitinib plus conventional DMARDs in head-to-head comparison with adalimumab plus conventional DMARDs in the TIM-naïve population. In the SELECT-COMPARE study, upadacitinib plus methotrexate was associated with statistically-significantly but modestly higher rates of disease remission, ACR response, change in pain, and improvement in HAQ-DI. The difference in benefits was smaller than those seen in comparison to conventional DMARDs alone. Rates of serious harm or discontinuation due to adverse events were also similar, so we judge the evidence for combination therapy with upadacitinib versus adalimumab to represent an incremental or better net health benefit ("B+"). There were no significant differences in clinical outcomes between combination regimens using tofacitinib versus adalimumab in two trials, although there was a trend towards more patients randomized to tofacitinib achieving ACR70 and having greater improvements in the HAQ-DI. We therefore assign a net health benefit rating of comparable ("C") for this comparison. Finally, there is no evidence on baricitinib 2 mg daily versus adalimumab in this population, so we judge the comparative clinical effectiveness of baricitinib to be insufficient ("I").

For each of the JAK inhibitors, there is one randomized trial comparing combination therapy to conventional DMARDs in the TIM-experienced population. All three trials are smaller than those for the TIM-naïve population and the effect sizes are also somewhat smaller with wider confidence intervals. As in the TIM-naïve population, rates of serious harms and discontinuation due to adverse events were low, so we judge the evidence for combination therapy with each of the three JAK inhibitors versus conventional DMARDs alone to represent an incremental or better net health benefit ("B+").

There is much greater uncertainty in assessing the relative comparative clinical effectiveness of JAK inhibitors, which have never been compared head-to-head in a randomized setting. The individual clinical trials had somewhat different patient populations, primary endpoints, timing of assessments, and timing of allowable switching to alternative therapies. As a result, we judge there to be insufficient evidence ("I") to differentiate among JAK inhibitors.

3.5 Biosimilars for RA

A biosimilar is a biologic drug that is highly similar in structure and function to a licensed reference product. In contrast to a generic (which is a small molecule that can be predictably duplicated), a biosimilar is a larger and more complex molecule that can be sensitive to changes in the manufacturing process. The FDA requires that manufacturers demonstrate that there are no clinically meaningful differences in safety, purity, and potency between the biosimilar and the reference product.[27] The Biologics Price Competition and Innovation Act (BPCI) of 2009 created an abbreviated licensure pathway for products shown to be biosimilar to or interchangeable with an FDA licensed reference product.[28] In Europe, the European Medicines Agency requires that the two products show evidence of similarity in quality, safety, and efficacy.

The FDA approval pathways for biologics and for biosimilars are different. Under the Public Health Service Act, the standard pathway for approval of biologics is described in Section 351 (A). The application requires all information regarding safety and efficacy of a biologic product. It is considered a standalone application and does not depend on any other drug or biologic. Approval via the 351 (K) pathway is for biosimilars. This application is submitted for approval to receive FDA designation as a biosimilar or the more stringent interchangeable designation. The proposed biosimilar must have the same mechanism of action for the intended condition(s) of use. The route of administration, dose form, and strength must be the same as that of the reference product. Applications are also required to provide details about product manufacturing to ensure safety and efficiency of manufacturing plant and process. Of note, once a biosimilar has been approved, it may be prescribed for any indication allowed for the reference product, even if the biosimilar has not been studied in that patient population.[29] Biosimilars may be approved for indications for which the reference product has been approved through the process of extrapolation.[29]

Interchangeability as defined for a biosimilar means that the product may be substituted for the reference product without the intervention of the health care provider who prescribed the reference product.[29] The standards for designation as "interchangeable" include those required to be approved as a biosimilar. In addition, manufacturers must demonstrate that the safety and efficacy of alternating or switching between the interchangeable agent or the reference product is non-inferior to continuing the reference product without switching.

Biosimilar naming follows a standard approach. Each nonproprietary name for a biological product includes the reference product name followed by a hyphen and a four-letter suffix. For example,

the biosimilar in this report is infliximab-dyyb. The reference product for infliximab-dyyb is infliximab.[30]

As implied by its name, the BPCI was intended to increase competition in the marketplace and decrease cost, analogous to what occurred when generic drug legislation was passed. Although more than 20 biosimilars have been approved, the uptake of biosimilars has been modest and because of this, significant cost reductions have not been observed in the US. No biologics have yet been designated as "interchangeable."

Example: Infliximab Biosimilar (Inflectra/CT-P13/Infliximab-dyyb)

The PLANETRA (Programme evaluating the autoimmune disease investigational drug CT-P13 in RA patients) trial randomized patients with RA to infliximab-dyyb (CT-P13) or to its reference product infliximab (INX).[31] The investigators randomized 606 patients (83% female) with active RA and an inadequate response to methotrexate.[32] The primary endpoint was ACR criteria for ≥20% clinical improvement response (ACR20) at 30 weeks. Additional endpoints included ACR50, ACR70, DAS28-ESR, DAS28-CRP, SDAI, CDAI, the percentage of patients with response defined according to EULAR criteria, patient-reported outcomes, joint damage progression, safety endpoints and laboratory abnormalities, and immunogenicity endpoints. At week 30, ACR20 responses were 60.9% for CT-P13 and 58.6% for INX (95% CI -6% to 10%). ACR50 and 70 responses were also similar at 30-week follow up (35.1% and 16.6-34.2% and 15.5% for CT-P13 and INX, respectively). The proportion requiring salvage therapy at 30 weeks was also similar (3.2% for CT-P13 vs. 4.0% for INX). All other outcomes were also similar at 30 weeks. The incidence of drug-related adverse events was also similar (35.2% for CT-P13 vs. 35.9% for INX) in both groups.

Subsequently, 54-week results were reported. A total of 455 of the 606 patients were treated up to week 54.[33] At 54 weeks, there was no difference between groups who met the primary endpoint (74.7% biosimilar vs. 71.3% reference). The proportion of patients achieving ACR50 and ACR70 at 54 weeks was also comparable between groups.

DAS28-ESR, DAS28-CRP, SDAI, and CDAI scores were similar between groups at 54 weeks follow-up. Mean decreases from baseline of DAS28-ESR, DAS28-CRP, SDAI, and CDAI were 2.4, 2.3, 26.3, and 25.7 compared with 2.4, 2.2, 24.6, and 24.0 for the reference product. The proportion of patients with "Good" and "Moderate" EULAR response was similar between the two groups. With respect to patient-reported outcomes, the VAS for patient assessment of pain showed similar reductions from baseline in both groups (30.2 vs. 28.4 at week 54). There were no differences in the immunologic outcomes, including the development of anti-drug antibodies.

Treatment-emergent adverse events at 54 weeks were similarly reported between groups (70.5% vs. 70.3% in the reference product). Twenty-two patients (7.3%) in the infliximab-dyyb group had

latent tuberculosis compared with 20 patients (6.7%) in the reference product group. There were no cases of active tuberculosis or lymphoma at 54-week follow-up in either group.

In an extension of the PLANETRA study, 302 of the 455 patients who completed the study enrolled in the extension study.[31] The switch group from the reference product to infliximab-dyyb was compared to the maintenance group continuing infliximab-dyyb after an additional 48-weeks follow-up. Efficacy assessments were made at baseline and at weeks 14, 30, 54, 78, and 102. Efficacy endpoints included the proportion of patients meeting ACR20, ACR50, and ACR70 among others. Response rates at week 102 assessment for maintenance versus switch groups were 71.7% versus 71.8% for ACR20, 48.0% versus 51.4% for ACR50, and 24.3% versus 26.1% for ACR70. There were no differences in other efficacy endpoints including DAS28 score changes and EULAR response criteria. Similar proportions of patients reported treatment-emergent adverse events in the two groups (53.5% vs. 53.8%). Rates of serious adverse events were similarly reported in the maintenance and the switch groups (7.5% vs. 9.1%). Rates of latent tuberculosis were similar to those seen during the main trial. Active tuberculosis was reported in three patients (1% in the CT-P13 group and no patients in the reference product group).

Finally, in an observational study conducted in Bulgaria, 151 patients with severely active RA (n=81) or ankylosing spondylitis (n=70) were treated with infliximab-dyyb for 24 weeks.[78] The primary outcome for patients with RA was the DAS28-CRP score. Patients with RA had a significant reduction in DAS28-CRP score when compared to baseline. A total of 4.8% of participants reported an adverse event. Two out of seven serious adverse events were considered possibly treatment related. There were no cases of lymphoma or active tuberculosis. CT-P13 was determined to be relatively safe and effective.

Infliximab-dyyb was approved as a biosimilar by the FDA, but it has not been designated as interchangeable; the FDA has not given interchangeability status to any product. Since interchangeability designates whether products can be substituted by a pharmacist, it is uncertain whether products distributed through infusion centers will seek this designation. The FDA draft guidance on interchangeability was published in January 2017 and finalized in May 2019.

PLANETRA was a large study with relatively long follow up. The study found no clinically important or statistically significant differences in benefits or harms between infliximab-dyyb and its reference product infliximab at weeks 30, 54, and 102. Observational data also support no important differences between the two therapies. Using the ICER Evidence Rating Matrix, the biosimilar infliximab-dyyb has high certainty of comparable net health benefit ("C") relative to its reference product.

4. Long-Term Cost Effectiveness

4.1 Overview

The primary aim of this analysis is to estimate the cost effectiveness of JAK inhibitors for patients with severely active RA using a decision analytic model. While the clinical evidence review focuses on patients with moderately-to-severely active disease, mean baseline characteristics from trial data reflected a population with severely active disease. The model's objective was to compare each of the three JAK inhibitors, upadacitinib, tofacitinib, and baricitinib, to adalimumab, a TNF inhibitor. While other TIMs indicated for first-line treatment are also appropriate choices for providers and patients, we chose adalimumab as a comparator due to its extensive use in clinical practice for the treatment of RA, and because it was directly compared to JAK inhibitors in clinical trials. Since the publication of the Modeling Analysis Plan on August 5, 2019, we modified our initial objective to assess the relative value of JAK inhibitors versus adalimumab for line-one treatment after failure by a conventional DMARD. Unfortunately, we were unable to directly compare tofacitinib to adalimumab due to inadequate data in the TIM-naïve or TIM-experienced population, and likewise we were unable to compare upadacitinib to adalimumab in the TIM-experienced population. Because the labeled indication of baricitinib is the treatment of patients for whom TNF inhibitors have failed, we attempted to compare it to adalimumab in the TIM-experienced population, but we were unable to do so due to a lack of comparable data. In addition, the goal of evaluating an exemplar biosimilar was to set the stage for Policy Roundtable discussion, not to perform an exhaustive review of biosimilars. Thus, we do not evaluate the cost effectiveness of the biosimilars.

In all analyses (including the base-case analysis and in scenario analyses), JAK inhibitors and adalimumab are used in combination with conventional DMARD therapy unless otherwise specified. The base-case analysis takes a health care sector perspective (i.e., focuses on direct medical care costs only) and uses a one-year time horizon. We model these treatments over a one-year time horizon due to the uncertainty surrounding the number of subsequent lines of TIMs as well as when patients transition to palliative care. We heard from several stakeholders that patients who were prescribed a TIM typically stay on it for at least one year. We model these treatments using a lifetime time horizon as a scenario analysis, which is described later in this section. We conducted an additional scenario analysis using a modified societal perspective and included productivity impacts. The model was developed in hēRo3℠, with some components, such as survival distributions, developed in RStudio (Version 1.1.463).

hēRo3 is a Web-based, health-economics modeling platform that supports the development of both Markov cohort and partitioned survival models (Policy Analysis Inc., Brookline, MA). Calculations in hēRo3 are performed in the programming language R, using an open-source, health-economics

modeling package, called "heRomod" (https://github.com/PolicyAnalysisInc/heRomod), that runs in a secure, virtual private cloud. heRomod is a modified version of the open-source, health-economics modeling package, HEEMOD (http://cran.r-project.org/package=heemod). An extensive set of unit tests is available to validate calculations of the modeling package. Further details on hēRo3 are available in Appendix E.

4.2 Methods

Model Structure

We developed a *de novo* decision analytic model for this evaluation, informed by key clinical trials and prior relevant economic models. While this model was informed by ICER's 2017 report, we made several changes to reflect current clinical practice in RA. The results of this model should not be directly compared to the results of ICER's 2017 report, as the current model includes substantial changes from the cost-effectiveness model in the prior report. Costs and outcomes in this model were discounted at 3% per year.

The primary model focused on an intention-to-treat and treat-to-target analysis, with a hypothetical cohort of patients with severe RA for whom prior treatment with conventional DMARDs had failed. Upon model entry, the hypothetical patient cohort was initiated on a treatment strategy, with treatment response assessed at three months. In all analyses, a TIM was added to a conventional DMARD, such as methotrexate. Treatment switching was based on disease activity as measured by the DAS28-CRP value (Table 4.1), with those in remission and with low disease activity remaining on the same treatment after the first three months, while those with moderate/high disease activity switch to a subsequent line of therapy at the end of the first three-month cycle. Our model uses DAS28 at three months because of its use in current trials (including those for upadacitinib), the lack of uniformly reported measures across trials, and our understanding from clinicians that this more closely aligns with the clinical strategy used in the recommended treat-to-target approach.[5]

Table 4.1. Disease Activity Based on DAS28 Categories

DAS28	Disease Activity
<2.6	Remission
2.6 to ≤3.2	LDA
>3.2 to ≤5.1	MDA
>5.1	HDA

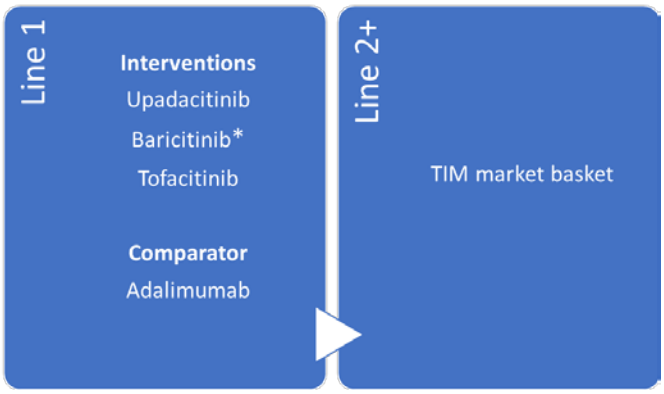
Source: Canhão et al., 2018[79]

DAS28: Disease Activity Score 28, HDA: high disease, LDA: low disease activity, MDA: moderate disease activity, activity

In a real-world clinical setting, patients often cycle through multiple therapies before finding the best treatment option for them. Unfortunately, there is little evidence to guide treatment sequencing for RA patients. Furthermore, the purpose of this analysis was to determine the cost effectiveness of specific treatments and not treatment sequences over time. Thus, treatment switching was assumed to be to a market basket of TIMs with efficacy averaged across the TIMs. The choice of TIMs to include in the market basket average was based on the availability of DAS28 data at three months after initiating a TIM. The TIMs chosen for the second-line market basket were baricitinib (RA-BUILD and RA-BEACON), adalimumab (RA-BEAM), etanercept (APPEAL), tofacitinib (ORAL Step), golimumab (GO-AFTER), and upadacitinib (SELECT-NEXT).[7,14,15,18-21] We chose to include data for treatments from line one because they could be considered in the second-line market basket of TIMs in the treatment arms where they are not included as line one. However, we included data for these drugs only from trials not informing efficacy in line one.

Patients could cycle through multiple lines of the same market basket within the modeled time horizon. After the first three months of using line-one treatment, those with moderate/high disease activity switched to a subsequent line of therapy (market basket TIM), while those in remission/low disease activity stayed on the current treatment. Patients could also switch to a subsequent line of TIM therapy (market basket TIM) due to loss of efficacy, adverse events, patient and clinician preferences, and access restrictions. In order to best compare the relative cost effectiveness of specific products and not the cost effectiveness of different treatment sequences, we standardized treatment sequence beyond line-one TIMs.

Figure 4.1. RA Treatment Sequence (All Treatments Added on to Conventional DMARD)



TIM: targeted immune modulator
*Only in a scenario analysis for a TIM-experienced population.

After initiating treatment with a TIM, the model relates the DAS28-based response to the HAQ after three months of therapy. Other previously published models, including the 2017 ICER report, mapped the ACR response or the EULAR to the HAQ after six months of therapy.[22,80,81] Our model uses a three-month cycle length because we understood from clinicians that this more closely aligns with the clinical strategy used in the recommended treat-to-target approach.[5]

A DAS28-to-HAQ mapping would ideally reflect hospitalization and outcomes such as quality of life and productivity loss, which were modeled as dependent on the HAQ. We found one other published model that related the DAS28 to HAQ score, but at six months of therapy.[82] We found clinical trial data on the proportions of patients within different categories of disease activity based on the DAS28 at three months for all treatments included in line one, but we did not find a robust DAS28-to-HAQ mapping algorithm at three or six months for all treatment strategies included. We hence used a mapping algorithm from EULAR to HAQ (Table 4.2). The EULAR response is divided into three response categories: "Good," "Moderate," and "None," and is based on the baseline DAS28 and the change in DAS28 from baseline at the timepoint measured. Here, we assumed remission as defined by DAS28 as equivalent to "Good" response, low disease activity as equivalent to "Moderate" response, and moderate disease activity and high disease activity as equivalent to "None" on the EULAR scale. While the HAQ-to-EULAR response mapping indicates HAQ change at six months, we assumed this to be the same at three months, which likely overestimates the benefit and biases the results in favor of the TIMs.

Table 4.2. Relationship Between EULAR and HAQ

EULAR Response	Mean HAQ Change	Standard Error
Good	-0.672	0.112
Moderate	-0.317	0.048
None	0	0

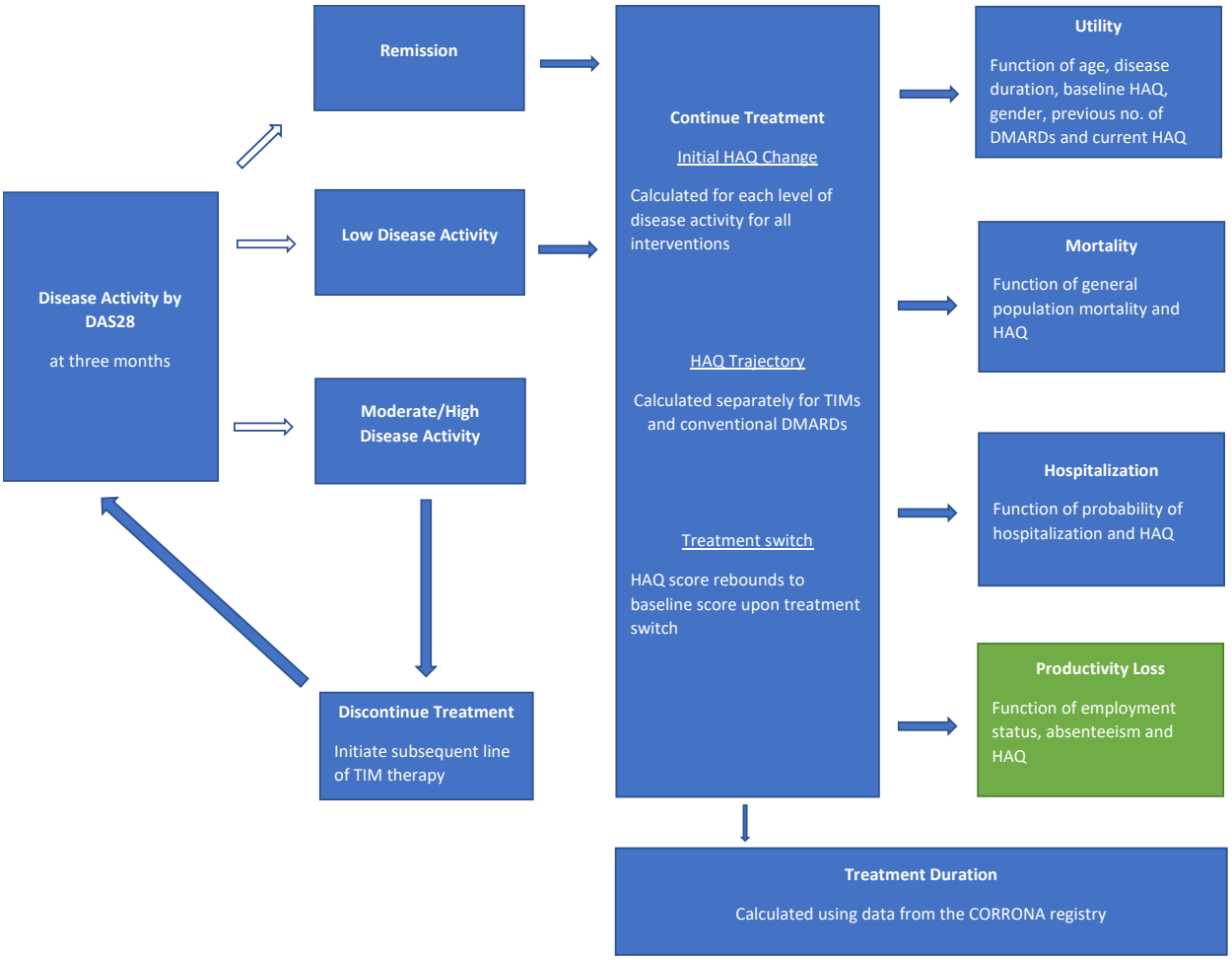
EULAR: European League Against Rheumatism, HAQ: Health Assessment Questionnaire
HAQ change mapping to EULAR response categories was estimated from the British Society for Rheumatology Biologics Register and has been used in the other published economic evaluations.[80]

The HAQ score was then linked to utility, mortality, hospitalizations, and productivity. Simulated utility scores and mortality were used to calculate the quality-adjusted life years (QALYs) gained, with hospitalization costs and productivity loss costs contributing to the health care sector perspective and societal perspective analyses, respectively (Figure 4.2). Long-term HAQ scores were simulated until treatment discontinuation or death, with relevant estimates for long-term HAQ changes applied to those on TIMs and palliative care.

Patients remained in the model until death. All patients were allowed to transition to death from all causes and from RA-related mortality.

Modeled outcomes include costs, life years (LYs), QALYs, and equal value of life years gained (evLYG). An analysis of the incremental cost per evLYG is included in this report to complement the cost per QALY calculations and provide policymakers with a broader view of cost effectiveness. A description of the methodology used to derive the evLYG can be found in Appendix E. Additionally, we measured the duration in remission when on line-one TIM treatment and the incremental cost of remission when on line-one TIM treatment.

Figure 4.2. Model Schematic



DMARD: disease-modifying antirheumatic drug, DAS28: Disease Activity Score 28, HAQ: Health Assessment Questionnaire, TIM: targeted immune modulator
Productivity losses will be measured in the modified societal perspective scenario analysis.

Target Population

The primary population of focus for the economic evaluation included adults in the US with severely active RA with inadequate response to conventional DMARDs and naïve to TIM therapy. The model simulates a hypothetical homogeneous cohort of patients with baseline characteristics consistent with severely active RA similar to those seen in the key trials for all line-one TIM therapies. Because most of these trials had patients with similar baseline characteristics, we used those from the SELECT-COMPARE trial in our modeled population for the base-case analysis, as it included two of the three TIMs assessed in this review.[6] While TIMs are indicated across populations with all levels

of disease activity, we chose to model patients with only severely active RA to match the population from the key clinical trials included in this review.

Table 4.3. Baseline Population Characteristics

	Mean Value	Source
Age	54 years	SELECT-COMPARE[6]
Female (%)	79%	
RA Duration	8 years	
Baseline HAQ	1.6	
Baseline DAS28	5.8	

DAS28: Disease Activity Score 28, HAQ: Health Assessment Questionnaire, RA: rheumatoid arthritis

Treatment Strategies

The list of interventions included in the cost-effectiveness review followed the same PICOTS criteria used for the clinical review and was developed with input from stakeholders. The full list of interventions is as follows:

- Upadacitinib
- Tofacitinib
- Baricitinib

Although there exists clinical trial data on the efficacy of baricitinib in the TIM-naïve population, it is currently approved for use only in the TNF-experienced population. Its cost effectiveness was not analyzed in the TIM-experienced population owing to lack of comparable data to adalimumab.

Comparators

- Adalimumab

In all analyses, the TIMs were considered as add-on therapies to conventional DMARDs versus conventional DMARD therapy alone. Due to a lack of data allowing for a comparison of tofacitinib to adalimumab (head-to-head or through an NMA), we were only able to comment on tofacitinib's value relative to adalimumab via its comparison to conventional DMARDs. This is detailed in Section 4.3. The efficacy of conventional DMARDs in this case was trial and intervention-specific, with estimates for adalimumab derived from the upadacitinib trials (SELECT-COMPARE), and for tofacitinib from the tofacitinib trial.[6,12]

Key Model Characteristics and Assumptions

Our model includes several assumptions, stated below.

Table 4.4. Modeling Assumptions

Assumption	Rationale
A treat-to-target approach was used, with treatment switching dependent on disease activity as measured by the DAS28.	We used a treat-to-target approach to align with real-world clinical practice, using the DAS28 to assess the likelihood of treatment switching.
A three-month cycle length was adopted, rather than the commonly used six-month cycle length seen in several previously published RA economic models,[22,80,82] including the model developed for ICER's 2017 review.	The three-month cycle length more closely aligns with the average length of time clinicians wait before assessing the need for treatment switching using a treat-to-target approach.[5] Additionally, we have clinical trial data on the proportion of patients with different levels of disease activity, as defined by the DAS28, at three months for all treatment strategies included in line one.
We assumed conversion ratios of 2x and 1.5x for DAS28-ESR to DAS28-CRP to derive the proportions in remission and low disease activity for tofacitinib and its conventional DMARD comparator.	Some trials (SELECT-MONOTHERAPY, RA-BUILD, RA-BEACON, and ORAL Step) simultaneously reported DAS28-ESR and DAS28-CRP outcomes for upadacitinib, tofacitinib, and baricitinib and their respective comparators. An average of the disease activity proportions using DAS28-CRP and DAS28-ESR data was used to estimate an approximate 2x and 1.5x ratio of DAS28-CRP to DAS28-ESR in the TIM arms, while the conventional DMARD arms showed more variability. In the absence of DAS28-CRP trial data at three months, we applied these ratios to the DAS28-ESR data to derive DAS28-CRP data at three months for tofacitinib and its conventional DMARD comparator. The uncertainty surrounding this was tested in a sensitivity analysis.
We adopted the EULAR-to-HAQ mapping algorithm for the different DAS28 disease activity categories, assuming remission to reflect "Good" EULAR response, low disease activity to reflect "Moderate" EULAR response, and moderate disease activity and high disease activity to reflect EULAR response of "None."	All trials report DAS28 categories by remission, low disease activity, and a moderate disease activity/high disease activity combination, but we found no robust published evidence mapping DAS28 to HAQ for any included treatment strategies.
No dose increase was assumed for those in the low disease activity category, as measured by the DAS28, at three months after initiation of a new TIM.	Clinical experts indicated that a dose increase for those with low disease activity (but not in remission) is patient-specific and not necessarily uniformly practiced for all drugs.

At three months after initiation of a new therapy (line one) or second-line market basket of TIMs, those with moderate/high disease activity, as measured by the DAS28, were assumed to switch to a market basket of TIM therapy. Similarly, those with low disease activity who discontinue treatment over time switched to this market basket TIM therapy.	Clinical experts reported that they would most likely initiate a new treatment switch if their patients had moderate/high disease activity at the time of assessment. While this switch could be to a different TIM, as stated earlier, in order to isolate the effect of line-one treatments, we assumed the same second-line market basket of TIMs across all included treatment strategies.
The TIMs chosen for the second-line market basket were baricitinib (RA-BUILD and RA-BEACON), adalimumab (RA-BEAM), etanercept (APPEAL), tofacitinib (ORAL Step), golimumab (GO-AFTER), and upadacitinib (SELECT-NEXT).[7,14,15,18-21]	The choice of TIMs to include in the market basket average was based on availability of DAS28 data at three months after initiating a TIM. We chose to include data for treatments from line one as well because they could be considered in the second-line market basket of TIMs in the treatment arms where they are not included as line one. However, we included data for these drugs only from trials not informing efficacy in line one.
We assumed that the efficacy of the market basket of TIM treatment was 84% of its calculated average efficacy across the TIMs included.	Prior published data shows a mean 16% reduction in treatment efficacy following failure by current therapy in RA. While this was specific to switching to another TNF inhibitor after failure by a prior one, we apply this reduction to JAK inhibitors following failure.[22,23] We apply this reduced efficacy only once to estimate a readjusted efficacy for the market basket, and do not apply a 16% reduction each time a failure with a line of therapy occurs.
Upon treatment discontinuation, HAQ rebounds to baseline HAQ.	We are unaware of any robust data on the magnitude of HAQ rebound upon treatment discontinuation. We hence assumed a rebound to baseline HAQ and vary this in the sensitivity analysis.
We assumed the same discontinuation rate among those with remission/low disease activity for all TIMs and conventional DMARDs following the initial three months of therapy.	Prior evaluations attempting class-level economic evaluations have cited errors of confounding in observational studies reporting discontinuation rates. Additionally, because these therapies have been approved over time, there is no consistent comparison of discontinuation rates among treatments.[80,81]
The rate of serious infection is assumed to be the same for all TIMs.	Serious infection measured in the trials does not reflect infection rates beyond the duration of the trials. Real-world estimates of infection rates are difficult to interpret as many patients switch therapies multiple times, which makes a risk attribution to a single therapy difficult. Differences in patient baseline risk for infection further complicate interpretation. In this case, serious infection rates across studies were similar. We therefore chose to use a standardized rate

	of serious infection across therapies, an approach used in other models.[80-82]
For patients in the conventional DMARD arms, we assumed that HAQ degradation over time will be 0.0269 per year for the first 15 years in the model, after which it is assumed to flatten with no additional annual degradation, as long as patients were on the line-one conventional DMARD therapy. This assumption reflects the HAQ trajectory of conventional DMARDs as undertaken in the model for ICER's 2017 review.	Findings from the National Databank on Rheumatic Diseases (NDB) show a degradation of HAQ over time among patients not on TIMs.[83] The HAQ degradation over time reduces in magnitude, but not in a linear manner. We hence assumed a flattening of the HAQ after 15 years in the model for patients remaining on conventional DMARD therapy.
When on TIMs, a long-term HAQ improvement of -0.001 annually was assumed and standardized for all TIMs assessed.	Data from the NDB estimated a long-term improvement in HAQ at -0.001 annually among patients on TIMs.[84]

Model Inputs

Clinical Inputs

Treatment Response

Inputs on the proportion of patients with different levels of disease activity have been derived from individual trials for relevant interventions and comparators (Table 4.5 and Table 4.6). We were unable to make a direct comparison between tofacitinib and adalimumab in our analyses due to a lack of comparable efficacy data. Our comparison was therefore restricted to only upadacitinib versus adalimumab. However, we comment on the value of tofacitinib and adalimumab based on the cost effectiveness of each of these TIMs when compared to conventional DMARDs in their respective trials. The approach and data used for this has been detailed in Appendix E. For the second-line market basket of TIMs, an average of disease activity proportions based on the DAS28 was chosen from trials including baricitinib (RA-BUILD and RA-BEACON), adalimumab (RA-BEAM), etanercept (APPEAL), tofacitinib (ORAL Step), golimumab (GO-AFTER), and upadacitinib (SELECT-NEXT).[15,18,19,7,14,21] We modeled a one-time efficacy decrement of 16% following failure of line-one therapy to reflect reduced efficacy as a result of TIM switching.[22,23]

Table 4.5. Treatment Response at Three Months using DAS28

	Proportion of Patients Achieving Different Categories of Disease Activity by DAS28 at Three Months*		
	<2.6 (Remission)	2.6 to ≤3.2 (LDA)	>3.2 (MDA and HDA)
Upadacitinib + cDMARD	29%	16%	55%
Adalimumab + cDMARD	18%	11%	71%
Second-Line Market Basket of TIMs†	22%	14%	64%

cDMARD: conventional disease-modifying antirheumatic drug, DAS28: Disease Activity Score 28, HDA: high disease activity, LDA: low disease activity, MDA: moderate disease activity, TIM: targeted immune modulator

*Mutually exclusive categories.

†Values prior to applying a 0.84 multiplier to reflect lower efficacy after failure by primary treatment.

Table 4.6. Treatment Efficacy Estimates for Adalimumab, Tofacitinib, and Their Respective Conventional DMARD Comparators at Three Months

	Proportion of Patients Achieving Different Categories of Disease Activity by DAS28 at Three Months*		
	<2.6 (Remission)	2.6 to ≤3.2 (LDA)	>3.2 (MDA and HDA)
Tofacitinib + cDMARD	15%	14%	71%
Adalimumab + cDMARD	18%	11%	71%
cDMARD†	6%	8%	86%
cDMARD‡	5%	3%	92%
Second-Line Market Basket of TIMs§	22%	14%	64%

cDMARD: conventional disease-modifying antirheumatic drug, DAS28: Disease Activity Score 28, HDA: high disease activity, LDA: low disease activity, MDA: moderate disease activity

*Mutually exclusive categories.

†Versus adalimumab.

‡Versus tofacitinib.

§Values prior to applying a 0.84 multiplier to reflect lower efficacy after failure by primary treatment.

Discontinuation

Among those treated with TIMs, a proportion of patients in moderate and high disease activity at three months after initiation of therapy transitioned to a second-line market basket of TIMs. These proportions were estimated from relevant trial data. Patients in remission and with low disease activity at three months after treatment initiation were assumed to continue on initial therapy. We included estimates of later treatment discontinuation due to other reasons such as loss of efficacy, serious adverse events including infections, and physician and patient preferences based on data from an observational study of RA patients in the CORRONA registry.[85] The study included a sample of over 6,000 adult RA patients treated between 2002 and 2011 receiving TIMs, predominantly TNF inhibitors. While these data are not specific to JAK inhibitors, we believe they can be generalized in

the absence of comparable long-term data for these therapies. We digitized the reported Kaplan-Meier (KM) curves and fit relevant parametric distributions to the curves based on Akaike Information Criteria (AIC), and extrapolated the fitted curves over the modeled time horizon. Because the sampled population in the CORRONA registry comprised of patients with moderate disease activity, we adjusted this curve to represent discontinuation among patients with remission and low disease activity using an odds ratio (OR) of 0.52 as reported by Zhang et al.[86] Following Stevenson et al. and the Innovation Value Initiative (IVI) RA modeling group, we assumed the same long-term discontinuation rate for all TIMs due to issues of bias and confounding found in observational studies for specific TIMs.[80-82]

For conventional DMARDs, following the methods adopted by Stevenson et al. and the IVI RA modeling group, we assumed that those who were on conventional DMARD treatment for at least three months had the same treatment duration as those on TIMs.[80-82] This was done only for the comparison of tofacitinib and adalimumab to their respective conventional DMARD comparators when evaluating the relative value of tofacitinib compared to adalimumab.

Mortality

Gender and age-specific mortality were sourced from the Human Mortality Database's US-specific tables.[87] Prior evidence suggests that improved (lower) HAQ scores are associated with lower likelihood of death and that the HAQ was the most significant predictor of mortality in RA patients.[88] The HAQ calculated at the beginning of each cycle for each health state in the model informed the mortality at the end of the cycle. The quantitative relationship between HAQ and mortality was assumed to be the same as that used in ICER's 2017 report and was based on a published US RA cost-effectiveness study.[22] The mortality equation used was:

$$\text{US RA-severity specific mortality rate} = \text{all-cause mortality} * 1.33^{HAQ}$$

Adverse Events

We included adverse events related only to serious infection, aligning with approaches used in prior economic evaluations of RA treatments.[80-82] As stated in Table 4.7, we assumed that the rate of serious infection was uniform across TIMs, as published estimates of serious infection associated with specific TIMs do not represent long-term risk and are likely inaccurately estimated, as mentioned in previously published literature. Estimates on serious infection were sourced from an NMA by Singh et al. (Table 4.7).[89] We assumed the same rate of serious infection with conventional DMARDs as with TIMs. As in ICER's 2017 report and in prior models, we attributed a disutility of 0.156 for a one-month period following a serious infection, along with the relevant costs of treating the infection.

Table 4.7. Adverse Events (Serious Infection)

Parameter	Value (95% CI)*	Source
TIM	0.035 (0.027 – 0.046)	Singh et al., 2011[89]
cDMARD†	0.026 (0.020 – 0.034)‡	

cDMARD: conventional disease-modifying antirheumatic drug, CI: confidence interval, TIM: targeted immune modulator

*Calculated as per person-year.

†The cDMARD value was used only for the comparison of tofacitinib and adalimumab to their respective cDMARD comparators when evaluating the relative value of tofacitinib compared to adalimumab.

‡For the one-way and probabilistic analyses, we assumed upper and lower bounds based on the proportionate variation observed for TIMs.

Health State Utilities

As in ICER's 2017 report, the relationship between HAQ and utility score was based on Wailoo and colleagues' publication.[90] The utility scores from Wailoo and colleagues were based on health state time-tradeoff evaluations made by a US general population sample using the EuroQol (EQ-5D) index, one of the most widely used instruments in health state valuation.[91] We compared the Wailoo et al. utility change from HAQ score 1.0 to 1.5 to the utility change from a more advanced mathematical model.[92] Although the Wailoo et al. relationship produces a higher utility within the HAQ range of 1.0 to 1.5, the change in utility for this HAQ range was approximately 0.1, consistent with the change in the other model.[92] Uncertainty in the Wailoo et al. mapping was evaluated in parameter sensitivity analyses. EQ-5D scores were calculated using this equation:

> EQ-5D score = 1 – 1/(1+ exp(2.0734 + 0.0058*age + 0.0023*disease duration – 0.2004*baseline HAQ – 0.2914*male + 0.0249*# previous DMARDs – 0.8647*current HAQ))

Additionally, a disutility (-0.156) was assigned for one month to individuals who experienced a serious infection.[80]

Drug Utilization

The inputs used to model drug utilization and associated costs are shown in Table 4.8 on the following page.

Table 4.8. Treatment Regimen Recommended Dosage

Treatment	Upadacitinib	Tofacitinib	Baricitinib*	Adalimumab	Methotrexate
Brand Name	Rinvoq	Xeljanz	Olumiant	Humira	Generic
Manufacturer	AbbVie	Pfizer	Eli Lilly	AbbVie	Multiple manufacturers
Route of Administration	Oral	Oral	Oral	Subcutaneous injection	Oral
Dosing	15 mg once daily	5 mg twice daily†	2 mg once daily	40 mg every other week	15 mg weekly‡

mg: milligram

*Included only in a scenario analysis comprising of a TIM-experienced population.

†Extended release version is dosed at 11 mg once daily.

‡Most patients are on an average dose of 15 mg weekly, although the recommended dose is 7.5 mg weekly.

Economic Inputs

Drug costs in the model are current. All other costs were inflated to 2018 values unless otherwise specified.

Drug Acquisition Costs

Drug costs included the cost of acquisition. We obtained net price data from SSR Health[93] that combine information on net US dollar sales with information on unit sales to derive net pricing at the unit level across all payer types. Data on the approved drugs of interest were current through the first quarter of 2019, except for upadacitinib, which was approved on August 16, 2019. We estimated net prices for the TIMs with SSR price data by comparing the four-quarter rolling averages (i.e., second quarter 2018 through first quarter 2019) of both net prices and wholesale acquisition cost (WAC) per unit to arrive at an average discount from WAC for each drug. We then applied this derived discount to the latest WAC[94] of the TIMs of interest. We derived a net price for each TIM using the current WAC and discount from the SSR database.

Because upadacitinib was recently approved, we did not find any estimates on its net price in the SSR dataset. Thus, based on the ICER Reference Case, we assumed its WAC to be discounted by 26%, the average discount of the other two JAK inhibitors as seen in the SSR dataset, to estimate its net price.[93] All annual prices presented in Table 4.9 assume 100% adherence. For the cost of conventional DMARDs, we use the mean WAC of the multiple generic versions of methotrexate, aligning with the ICER Reference Case.

For the second-line market basket of TIMs, we estimated the cost of TIMs based on a retrospective observational study with prescription market share data that included estimates on eight TIMs.[24] This study used administrative claims data from the HealthCore Integrated Research Database with data between July 2009 and January 2013. We weighted the SSR-derived annual net price of these

TIMs by the market share observed in this observational study to arrive at an annual estimated net price of $38,523 for this market basket of TIMs. The eight TIMs included were adalimumab, certolizumab, etanercept, golimumab, infliximab, rituximab, abatacept, and tocilizumab.

Table 4.9. Drug Costs

Drug	WAC per Unit	Discount from WAC	Net Price per Unit	Annual WAC	Annual Net Price
Upadacitinib – 15 mg Tab	$163.89	26%*	$120.56	$59,860	$44,035
Tofacitinib – 5 mg Tab	$74.68	34%	$49.50	$54,552	$36,159
Baricitinib – 2 mg Tab	$71.23	19%	$57.59	$26,017	$21,033
Adalimumab – 40 mg/0.8 ml Sol	$2,587.05	34%	$1,696.21	$67,263	$44,102
Methotrexate Sodium – 2.5 mg Tab	$2.55	--	$2.55	$796	$796

mg: milligram, ml: milliliter, WAC: wholesale acquisition cost

*Discount calculated as the average discount estimated for the other two JAK inhibitors.

Administration and Monitoring Costs

Oral treatments were assumed to have no administration costs. Subcutaneous treatments include the costs for an annual office visit for training on self-administration, which in our analysis is specific only to adalimumab. However, because patients are attributed a physician's office visit every quarter for disease activity assessment, we assumed one of these visits each year to include training for self-administration and we hence do not separately cost out a physician's office visit for training self-administration for subcutaneous injections. For the market basket of TIM treatments, we did not include any TIM-specific administration costs as we felt their impact on the model would be minimal. Administration cost inputs are presented below in Table 4.10. All administration costs represent current 2019 US dollar values.

Table 4.10. Administration Costs

	Cost	Source
Office Visit (HCPCS Code: 99213)	$75.32	CMS[95]

CMS: Centers for Medicare & Medicaid Services, HCPCS: Healthcare Common Procedure Coding System

Drug monitoring costs include costs of quarterly tests including comprehensive metabolic test panel, complete blood cell count, lipid panel, acute hepatitis panel, and an additional annual tuberculosis test. Additionally, we also included a cost attributed to a physician's office visit every quarter. Table 4.11 details monitoring cost inputs.

Table 4.11. Monitoring Costs

	Cost*	Source
Tuberculosis Test (HCPCS Code: 86480)	$84.83	CMS[95]
Comprehensive Metabolic Test Panel (HCPCS Code: 80053)	$11.15	
Complete Blood Cell Count (HCPCS Code: 85025)	$10.65	
Lipid Panel (HCPCS Code: 80061)	$14.61	
Acute Hepatitis Panel (HCPCS Code: 80074)	$65.15	

CMS: Centers for Medicare & Medicaid Services, HCPCS: Healthcare Common Procedure Coding System
*Average Medicare Standardized Payment.

Non-Drug Health Care Utilization Costs

The cost of hospitalization was based on the relationship between HAQ and hospitalization, an approach followed by previously published models.[22,82] As seen in Table 4.12, the number of hospitalization days increases with worsening (increasing) HAQ score. The cost of serious infection was assumed to be the weighted average cost of treating pneumonia and cellulitis, two commonly occurring serious infections in RA patients (Table 4.12). This approach is based on that used in the 2017 ICER RA Review.

Table 4.12. Non-Drug Health Care Utilization Costs

HAQ Range	Hospitalization Days per Year	Cost per Day of Hospitalization	Source
HAQ: 0 to <0.5	0.260	$2,470	Carlson et al., 2015[22], Kaiser Family Foundation, 2018[96], IVI RA model[82]
HAQ: 0.5 to <1	0.130		
HAQ: 1 to <1.5	0.510		
HAQ: 1.5 to <2	0.720		
HAQ: 2 to <2.5	1.860		
HAQ: ≥2.5	4.160		
Cost of Serious Infection*	$9,013		Medicare Provider Utilization and Payment Data, 2016[97]

CI: confidence interval, HAQ: Health Assessment Questionnaire, IVI: Innovation and Value Initiative, RA: rheumatoid arthritis
*Weighted average of costs for pneumonia (2/3) and cellulitis (1/3).

Productivity Costs

Societal costs in our model are generated from additional costs from unemployment due to RA. We apply a HAQ-dependent unemployment rate to the baseline unemployment rate (3.8%) and calculate the costs of unemployment, estimated at $225 per day using an hourly wage of $28.11 for

an eight-hour workday.[98] A 0.25 increase in HAQ is associated with a 30% increased likelihood of unemployed status.[99]

Sensitivity Analyses

We ran one-way sensitivity analyses to identify the key drivers of model outcomes, using available measures of parameter uncertainty (i.e., standard errors) or reasonable ranges for each input described in the model inputs section above. Probabilistic sensitivity analyses were also performed by jointly varying all model parameters over 1,000 simulations, then calculating 95% credible range estimates for each model outcome based on the results. Additionally, we performed a threshold analysis by systematically altering the price of interventions to estimate the maximum prices that would correspond to given willingness-to-pay (WTP) thresholds between $50,000 and $150,000 per QALY.

Scenario Analyses

In addition to the base-case analysis, we conducted the following scenario analyses:

1) Modified societal perspective that included productivity loss costs as a result of unemployment due to RA
2) Using a lifetime time horizon.

Model Validation

We used several approaches to validate the model. First, preliminary methods were presented to manufacturers, and we subsequently shared draft methods and results with clinical expert reviewers and a health economics expert reviewer. Based on feedback from these individuals and groups, we refined data inputs used in the model. Second, we varied model input parameters to evaluate face validity of changes in results. We performed model verification for model calculations using internal reviewers. During the period that the Draft Evidence Report was under review, we shared a working version of the model with several manufacturers who had shown interest in reviewing the model as an additional validation step. Finally, we compared results to other cost-effectiveness models in this therapy area.

4.3 Results

All analyses reported in this section use a one-year time horizon unless otherwise specified.

Base-Case Results

The base-case results for total cost and outcomes are reported in Table 4.13, with incremental cost-effectiveness results in Table 4.14. In the comparison of upadacitinib and adalimumab, we found no difference in the LYs gained up to the fourth decimal place. Upadacitinib use resulted in marginally more QALYs gained compared to adalimumab and had higher drug and total costs. The higher total costs are due to patients staying on upadacitinib longer than adalimumab, with the second-line market basket of TIMs being less expensive than upadacitinib or adalimumab. With no appreciable differences in LYs gained with upadacitinib at one year, we found the evLYG to be the same as the QALYs gained and hence do not report it separately in the tables below. The use of upadacitinib resulted in patients spending approximately one more month on average (over the course of a year) in remission compared to the use of adalimumab.

Table 4.13. Results for the Base Case for Upadacitinib versus Adalimumab

Treatment	Drug Cost* (Line One)	Total Cost	LYs	QALYs	Months in Line-One Remission
Upadacitinib + cDMARD	$21,400	$48,200	0.985	0.699	2.8
Adalimumab + cDMARD	$15,800	$47,600	0.985	0.693	1.7

cDMARD: conventional disease-modifying antirheumatic drug, evLYG: equal value of life years gained, LY: life year, QALY: quality-adjusted life year

*Only costs of TIM; does not include conventional DMARD cost.

The incremental cost-effectiveness ratio for upadacitinib versus adalimumab was estimated to be approximately $92,000 per QALY. The cost per month in remission while on upadacitinib compared to adalimumab was approximately $600.

Table 4.14. Incremental Cost-Effectiveness Ratios for Upadacitinib versus Adalimumab

Treatment	Cost per LY Gained*	Cost per QALY Gained	Cost per Month in Line-One Remission
Upadacitinib + cDMARD vs. Adalimumab + cDMARD	--	$92,000	$600

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year

*No difference in LYs was gained between the two TIMs up to the third decimal place, hence this incremental cost-effectiveness ratio was not calculated.

As stated earlier, we were unable to compare the cost effectiveness of tofacitinib versus adalimumab due to a lack of data. However, we compared the outcomes of the two TIMs relative

to their respective conventional DMARD comparators. The different values noted in the tables for the conventional DMARD comparator arms directly reflect outcomes observed in the adalimumab and tofacitinib clinical trials, respectively.

Table 4.15. Results for the Base Case for Adalimumab versus Conventional DMARD

Treatment	Drug Cost (Line One)	Total Cost	LYs	QALYs	Months in Line-One Remission
Adalimumab + cDMARD	$15,800*	$47,600	0.985	0.693	1.73
cDMARD	$5,500	$36,500	0.985	0.686	0.58

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year
*Only the costs of TIM; does not include cDMARD cost.

Table 4.16. Results for the Base Case for Tofacitinib versus Conventional DMARD

Treatment	Drug Cost (Line One)	Total Cost	LYs	QALYs	Months in Line-One Remission
Tofacitinib + cDMARD	$12,800*	$44,700	0.985	0.692	1.40
cDMARD	$5,500	$38,400	0.985	0.685	0.45

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year
*Only the costs of TIM; does not include cDMARD cost.

Results from Tables 4.15 and 4.16 demonstrate that the use of adalimumab or tofacitinib results in similar QALY gains at one year, at a higher cost, compared to conventional DMARDs.

Sensitivity Analysis Results

Results from the one-way sensitivity analyses are shown in Table 4.17. For upadacitinib compared to adalimumab, cost of the second-line TIM market basket was the major driver of results, with probability of achieving remission, the efficacy decrement for line 2+ therapy, and utility estimates having much smaller impacts.

Table 4.17. One-Way Sensitivity Analyses of Upadacitinib versus Adalimumab for Cost per QALY

Parameter	Low Input Result	Base Case Result	High Input Result
Cost of Second-Line TIMs	$242,000	$92,000	Dominant
Price of Achieving Remission	$97,000	$92,000	$88,000
Second Efficacy Decrease	$87,000	$92,000	$97,000
Utility Values	$88,000	$92,000	$96,000
Cost of Hospitalization	$94,000	$92,000	$90,000

A probabilistic sensitivity analysis was also conducted to assess variation in several parameters with 1,000 Monte Carlo simulations. As shown in Table 4.18, approximately 80% of all iterations

resulted in cost-utility ratios at or under the $150,000 per QALY threshold. A scatter plot showing the distribution of cost-utility ratios over the 1,000 simulations is presented in Appendix E.

Table 4.18. Probabilistic Sensitivity Analysis Results for Upadacitinib versus Adalimumab

	Cost-Effective at $50,000 per QALY	Cost-Effective at $100,000 per QALY	Cost-Effective at $150,000 per QALY
Upadacitinib + cDMARD	26%	54%	80%

cDMARD: conventional disease modifying antirheumatic drug, QALY: quality-adjusted life year

Scenario Analyses Results

Results from the modified societal perspective only impacted total costs in the model. The additional costs of unemployment did not have a substantial impact on the total costs relative to those seen using a health care sector perspective, which is not surprising considering the one-year time horizon of the model and the small proportion of patients expected to experience employment change due to treatment over this time period. Accordingly, these societal costs did not have an impact on the incremental cost-effectiveness ratio (Table 4.19). We also modeled the primary cost-effectiveness analysis using a lifetime time horizon, which showed that upadacitinib's cost-utility ratios exceed a threshold of $150,000 per QALY. Results of this scenario analysis can be found in Appendix E.

Table 4.19. Cost-Effectiveness Results for Upadacitinib versus Adalimumab from a Modified Societal Perspective

Treatment	Total Cost	LYs	QALYs	Cost per QALY Gained	Cost per LY Gained†
Upadacitinib + cDMARD	$48,200	0.985	0.699	$92,000	--
Adalimumab + cDMARD	$47,600	0.985	0.693	--	--

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year
*Only costs of TIM; does not include cDMARD cost.
†No difference in LYs gained between the two TIMs up to the third decimal place, hence this incremental cost-effectiveness ratio was not calculated.

Threshold Analysis Results

Table 4.20 presents results of our threshold analysis for the price of upadacitinib compared to adalimumab. The threshold prices required for upadacitinib are reported as annual prices that would achieve selected cost-effectiveness thresholds.

Table 4.20. Threshold Analysis Results for Upadacitinib versus Adalimumab

	Annual WAC	Annual Net Price*	Annual Price to Achieve Threshold of:		
			$50,000 per QALY	$100,000 per QALY	$150,000 per QALY
Upadacitinib 15 mg Daily	$59,860	$44,035	$43,466	$44,144	$44,822

QALY: quality-adjusted life year, mg: milligram, WAC: wholesale acquisition cost

*The net price was calculated based on the average discount from WAC, with discount based on the average discount seen with other JAK inhibitors.

Model Validation

Model validation followed standard practices in the field. We tested all mathematical functions in the model to ensure they were consistent with the report and supplemental Appendix materials. We also conducted sensitivity analyses with null input values to ensure the model was producing findings consistent with expectations. Further, independent modelers tested the mathematical functions in the model as well as the specific inputs and corresponding outputs.

Additionally, we compared our model results to findings from other published models. We searched the literature to identify models that were similar to ours, with comparable populations, settings, perspective, and treatments. In our literature review, we found several published economic models that evaluated various treatments for RA. Four models were particularly relevant for our review: 1) the model developed for ICER's 2017 report; 2) a patient-level RA simulation model developed by IVI[82]; 3) a United Kingdom (UK)-specific model developed by Stevenson et al. in 2016 (funded by the National Institute of Care and Excellence in Health [NICE][80]); and 4) a later UK-specific model by Stevenson et al. in 2017, which closely mirrored the methods of the 2016 NICE model.[81]

Overall, the most noticeable differences among our current model and the models mentioned above pertain to: 1) treatment sequence: i.e., in order to isolate the true health and economic benefits of the TIMs included, patients in our model transition to a second-line market basket of TIMs that was standardized across all treatment strategies (other models include a treatment sequence comprising of active care with TIMs or conventional DMARDs following failure of line-one TIM treatment); 2) the inclusion of upadacitinib as an intervention of interest in our current model; and 3) time horizon: we use a one-year time horizon while other models were run over a lifetime time horizon. To our knowledge, our model is the first publicly available economic evaluation of upadacitinib, the most recently approved TIM for the treatment of moderately-to-severely active RA. Thus, all comparisons will focus solely on methods and not necessarily on the incremental cost-effectiveness ratios for TIMs as generated by the different models.

While our current model borrows basic structural and parametric model-building approaches from the model developed for ICER's 2017 report, it also differs in the following important ways:

1) Based on feedback from several stakeholders (including clinical experts), our current model measures levels of disease activity using the DAS28-CRP, an absolute measure of disease activity, instead of the ACR, a relative measure of disease activity used in 2017. In 2017 and now, we received mixed feedback on the use of the ACR to measure levels of disease activity. Clinicians noted that the ACR is not as useful in real-world practice for treatment switching decisions. However, most clinical trials employ the ACR as the primary measure of clinical benefit of TIMs. And while the ACR is not the clinically preferred measure of disease activity, we heard from multiple stakeholders that it is used for regulatory approval in several markets. Another notable difference is the categories of disease activity used. In 2017, we used four different levels of disease activity as measured by the ACR, but in this review we used three. In addition, the subsequent HAQ mapping algorithm used with each disease activity measure is different, i.e., for a patient with ACR>70, which is similar to remission using the EULAR response criteria, the respective absolute HAQ changes are different.
2) In the 2017 base-case model, patients in whom line-one TIMs had failed were able to cycle through two additional market baskets of TIMs before transitioning to palliative care. However, in the conventional DMARD comparator arm, patients transitioned directly to palliative care after conventional DMARD failure, which led to a greater magnitude of difference in costs and clinical outcomes between interventions and the comparator. In the current model, patients transition to a market basket of TIMs and never transition to palliative care. We used this approach to better reflect real-world clinical practice and isolate the economic value of JAK inhibitors as line-one therapy.
3) The patient population for the current cost-effectiveness model is limited to patients with severely active RA, which is reflective of the trial populations. In 2017, we classified the population more broadly to include patients with moderately-to-severely active disease.
4) In addition to costs, QALYs, and LYs, outcomes in the current model include duration of remission with line-one treatment to highlight the most preferred clinical benefit expected with the treatments of interest. In 2017, this clinical benefit was assessed more broadly to include responders as defined by the ACR (ACR>20).
5) In our current analysis, we model long-term treatment discontinuation using real-world evidence, though this evidence is generated from data largely representing patients on TNF inhibitors. This is consistent with recent as well as some earlier approaches for modeling treatment discontinuation.[80,82]
6) In our current model, we use adalimumab as the primary comparator to inform more policy-relevant discussions.

Our current model shares some similarities with the recently published RA model by IVI.[82] The IVI model is a patient-level microsimulation model that allows for customization of individual patient characteristics, choice of disease activity measure to assess treatment response, and other customizable inputs. Our current model differs in being cohort-based, but retains key similarities including the use of the same discontinuation rate, rate of adverse events (serious infection), use of a similar mapping algorithm to HAQ change and subsequent health-related quality of life, use of the same long-term HAQ change for TIMs and conventional DMARDs, and use of the same HAQ-dependent hospitalization rate.

A NICE-funded model by Stevenson et al. analyzed the cost effectiveness of several TIMs versus conventional DMARDs in three populations. One population was naïve to conventional DMARD use, while two populations experienced conventional DMARD treatment failure and were eligible for TIM use. In the latter populations, all but one TIM treatment strategy included the use of three lines of TIMs followed by conventional DMARDs, and then palliative care. In one TIM strategy assessing tocilizumab as first-line treatment, only one additional line of treatment with a TIM was assessed followed by conventional DMARDs and then palliative care. This was done to standardize treatment sequencing in the second-line market basket of TIMs. In that model, they assessed disease activity and subsequently treatment switching based on an NMA of EULAR response criteria. Similar to our model, the EULAR response was tied to the initial HAQ change, but at six months. The HAQ then governed utility estimates, mortality, and hospitalization in the UK model. Adhering to NICE guidelines, the UK model used a 3.5% discount rate with population characteristics reflective of a UK population with RA. Treatment costs in the model were specific to the UK, and the long-term HAQ trajectory was modeled using a latent class growth model approach for the conventional DMARD arm. Treatment duration was considered to be the same across TIMs and for conventional DMARDs as long as patients remained on those treatments for at least six months. Discontinuation rates only differed by EULAR response criteria. Serious infections were the only adverse events documented; an approach followed by other models including ours. Mortality was estimated differently as well; in the NICE model, different hazard ratios were used for mortality and were dependent on HAQ ranges. Conversely, our model used a single hazard ratio for mortality, which was dependent on a patient's current HAQ score.

Another model by Stevenson et al.[81] compares seven TIMs to conventional DMARDs in two populations: one with moderate-to-severely active RA and the other with severely active RA. Patients were sampled from the British Society for Rheumatology Biologics Register dataset. The methods adopted in this model closely resemble the model developed by the assessment group for the NICE-funded model. Base-case results in this model also showed results similar to the 2016 NICE model—the cost-utility of each TIM versus conventional DMARD exceeded the commonly cited thresholds of £20,000 to £30,000 per QALY adopted in the UK.

4.4 Summary and Comment

Our aim was to evaluate the cost effectiveness of the three JAK inhibitors versus adalimumab in relevant populations over a one-year time horizon. However, gaps in the literature limited our analysis to only comparing upadacitinib to adalimumab in the TIM-naïve population with prior failure by a conventional DMARD. Our base-case findings suggest that upadacitinib provides marginal clinical benefit in comparison to adalimumab, at higher costs. These higher costs are attributed to patients remaining on upadacitinib longer due to better rates of remission and low disease activity relative to those of adalimumab. Together, these outcomes translate into cost-effectiveness estimates that fall under the upper end of the commonly cited cost-utility threshold of $150,000 per QALY. Results from the modeling comparison of tofacitinib and adalimumab to conventional DMARDs suggest that for the similar benefit tofacitinib offers, a price much higher than adalimumab may not be justified.

The base-case analyses were generally robust to sensitivity analyses. In one-way sensitivity analyses, parameters such as the baseline HAQ, utilities derived from HAQ, probability of remission with upadacitinib, cost of market basket TIM treatment, and hospitalization rate influenced the outcomes the most. In probabilistic analyses, more than three-fourths of all simulations were at or below the $150,000 per QALY threshold.

Results from our scenario analyses evaluating cost outcomes and incremental cost effectiveness from a modified societal perspective were similar to those seen in the base-case health care sector perspective analysis. Analyses using a lifetime time horizon resulted in more unfavorable cost-effectiveness ratios due to smaller differences in outcomes relative to those in the base-case analysis.

Note that the results presented in this report should not be directly compared to the results of ICER's 2017 report, as the current model includes substantial changes from the cost-effectiveness model in the prior report.

Limitations

Our model has several limitations. Treatment sequencing in RA is not standardized and depends on several factors such as individual patient characteristics, treatment effectiveness, patient persistence, physician preference for choice of therapy, and access restrictions. Additionally, there are currently no published standardized guidelines on treatment sequencing. Including different treatment sequences for different initial TIM treatments does not isolate the value of the initial treatments assessed in this analysis. Similarly, assuming a market basket of TIMs as second line of therapy tends to overestimate the efficacy of initial TIM therapies over the lifetime of the model. To counter that, we chose to model the second-line market basket of TIMs over a one-year time

horizon, which we believe to be a time period that patients will remain on TIMs irrespective of multiple switches.

Second, our model is unable to draw a direct comparison of the value between upadacitinib and the older JAK inhibitors, tofacitinib and baricitinib, in the TIM-naïve and TIM-experienced populations due to a lack of published data. This is more a limitation of the data than the model and does not inform what clinicians or policymakers would like to know: the most cost-effective choice among the three JAK inhibitors. Similarly, we could only partially answer a policy-relevant question on the choice between JAK inhibitors or adalimumab following conventional DMARD failure or in TIM-experienced patients with severe RA. This again was due to a lack of comparable data among the included treatments.

A third limitation is the use of the same standardized treatment discontinuation rate for both TIMs and conventional DMARDs. While this may vary in the real world across TIMs and conventional DMARDs, we did not have robust long-term published data for all assessed treatment strategies. Our approach has been used in prior published RA models.[82,100]

Fourth, in the absence of a validated mapping algorithm from DAS28 to HAQ, we assumed that disease activity as measured using the DAS28 could be mapped to EULAR response categories. Furthermore, as indicated by clinical experts and in published literature, the relationship between DAS28 scores and EULAR responses may not be direct in all cases. For instance, patients may still have residual non-inflammatory joint pain despite a "Good" EULAR response.[101] This indirectly affects the utility mapping functions in the model. In addition to disease activity mapping algorithms and patient health-related utility measures, longer-term data are required to provide more accurate mappings of long-term HAQ trajectories and the subsequent utilities that reflect multiple lines of treatment.

Finally, the population of focus in our model is patients with severe RA, with treatment efficacy estimates limited to this group. However, in the real world, a majority of patients have low disease activity as opposed to moderate or high disease activity.[102] The value of the interventions assessed may be different in a trial population with less severe RA.

Conclusions

Our analyses indicate that more comparable data are required on the short and long-term efficacy of JAK inhibitors. Upadacitinib, for which we have trial-specific data, provided marginal clinical benefit over adalimumab at higher costs, resulting in its incremental cost-utility ratio falling below commonly cited thresholds.

5. Potential Other Benefits and Contextual Considerations

Our reviews seek to provide information on potential other benefits offered by the intervention to the individual patient, caregivers, the delivery system, other patients, or the public that would not have been considered as part of the evidence on comparative clinical effectiveness. We also recognize that there may be broader contextual issues related to the severity of the condition, whether other treatments are available, and ethical, legal, or other societal priorities that influence the relative value of illnesses and interventions. These general elements are listed in the table below, and the subsequent text provides detail about the elements that are applicable to the comparison of upadacitinib to adalimumab. We sought input from stakeholders, including individual patients, patient organizations, clinicians, and manufacturers, to inform the contents of this section.

Each ICER review culminates in a public meeting of an independent voting panel of clinicians, patients, and health services researchers. As part of their deliberations, panel members will judge whether a treatment may substantially impact the considerations listed in Table 5.1. The presence of substantial other benefits or contextual considerations may shift a panel member's vote on an intervention's long-term value for money to a different category than would be indicated by the clinical evidence and cost-effectiveness analyses alone. For example, a panel member may initially consider a therapy with an incremental cost-effectiveness ratio of $150,000 per QALY to represent low long-term value for money at current pricing. However, the panel member may vote for a higher value category if they consider the treatment to bring substantial other benefits or contextual considerations. Conversely, disadvantages associated with a treatment may lead a panel member to vote for a lower value category. A panel member may also determine that there are no other benefits or contextual considerations substantial enough to shift their vote. All factors that are considered in the voting process are outlined in ICER's Value Assessment Framework. The content of these deliberations is described in the last chapter of ICER's Final Evidence Report, which is released after the public meeting.

This section as well as the panel's deliberation provides stakeholders with information to inform their decisions on a range of issues, including shared decision-making between patients and clinicians, coverage policy development, and pricing negotiations.

Table 5.1. Potential Other Benefits or Contextual Considerations (Not Specific to Any Disease or Therapy)

Potential Other Benefits
This intervention offers reduced complexity that will significantly improve patient outcomes.
This intervention will reduce important health disparities across racial, ethnic, gender, socio-economic, or regional categories.
This intervention will significantly reduce caregiver or broader family burden.
This intervention offers a novel mechanism of action or approach that will allow successful treatment of many patients for whom other available treatments have failed.
This intervention will have a significant impact on improving return to work and/or overall productivity.
Other important benefits or disadvantages that should have an important role in judgments of the value of this intervention.
Potential Other Contextual Considerations
This intervention is intended for the care of individuals with a condition of particularly high severity in terms of impact on length of life and/or quality of life.
This intervention is intended for the care of individuals with a condition that represents a particularly high lifetime burden of illness.
Compared to adalimumab, there is significant uncertainty about the long-term risk of serious side effects of this intervention.
Compared to adalimumab, there is significant uncertainty about the magnitude or durability of the long-term benefits of this intervention.
There are additional contextual considerations that should have an important role in judgments of the value of this intervention.

5.1 Potential Other Benefits

All three JAK inhibitors are administered orally, which may be preferable to patients. Other TIMs are administered intravenously or via subcutaneous injection.

Biologics and conventional DMARDs in general have improved the natural history of RA, as fewer patients develop disabling joint deformities. It is not clear that JAK inhibitors offer any advantages over the other TIMs.

5.2 Contextual Considerations

RA is a condition with a large impact on the length and quality of life, which has been greatly improved with the introduction of biologics and conventional DMARDs.

6. Value-Based Price Benchmarks

Annual value-based price benchmarks (VBPBs) of upadacitinib (vs. adalimumab) are presented in Table 6.1. The VBPB for a drug is defined as the price range that would achieve incremental cost-effectiveness ratios between $100,000 and $150,000 per QALY gained. For upadacitinib, price discounts of approximately 25% to 26% from the list price (WAC) would be required to reach the $100,000 to $150,000 per QALY threshold prices, respectively (Table 6.1). Note that these discounts are similar to the 26% discount we assumed for upadacitinib in our base case (using the average estimated discount for the other JAK inhibitors).

With no appreciable differences in LYs gained with upadacitinib versus adalimumab at one year, we found the evLYG to be the same as the QALYs gained and hence do not report VBPBs for it in the table below.

Table 6.1. Value-Based Price Benchmarks for Upadacitinib

	Annual WAC	Annual Price at $100,000 Threshold	Annual Price at $150,000 Threshold	Change from WAC to Reach Threshold Prices
Per QALY Gained	$59,860	$44,144	$44,822	-25% to -26%

QALY: quality-adjusted life year, WAC: wholesale acquisition cost

7. Potential Budget Impact

7.1 Overview

We used the cost-effectiveness model to estimate the total potential budget impact of upadacitinib in adults in the US diagnosed with moderately-to-severely active RA. However, it is important to note that all cost inputs for the budget impact model are generated from our cost-effectiveness model that comprised a population with severely active RA.

7.2 Methods

Potential budget impact was defined as the total differential cost of using upadacitinib as opposed to adalimumab for the treated population, calculated as differential health care costs (including drug costs) minus any offsets in these costs from averted health care events. All costs were undiscounted and estimated over a five-year time horizon, although our base-case analysis time horizon was only one year. This was done given the potential for cost offsets to accrue over time and to allow for a more realistic impact on the number of patients treated with the new therapy.

We estimated that the eligible population for upadacitinib would consist of adult patients with moderately-to-severely active RA with prior failure of methotrexate (conventional DMARD). This population aligns with the labeled indication. Although we believe upadacitinib may also be used in a TIM-experienced population, due to a lack of comparable data against other TIMs, we did not model this in our cost-effectiveness analysis, and subsequently did not include this population in our potential budget impact analysis. Additionally, we do not include a prevalent population of TIM users because we believe it is unlikely that these patients would switch to upadacitinib unless they experience inadequate response or intolerance to their current TIM regimen. Thus, our estimates of the eligible population represent only a proportion of the larger eligible population that can be treated with upadacitinib.

The estimated annual incidence of RA in the US is 0.041%, which is based on an observational study that used electronic medical records data to estimate RA incidence.[25] To this incident population, we then applied estimates of the proportions with moderately-to-severely active RA (36.1%) and those with RA on biologic therapy (52%). These estimates were obtained from an observational cohort study that assessed disease severity based on CDAI measures in RA patients on the CORRONA registry.[26] These estimates were applied to the 2019-2023 five-year annual average adult population in the US to arrive at an eligible population size of approximately 20,000 incident patients per year.

ICER's methods for estimating potential budget impact are described in detail elsewhere[103] and have been recently updated. The intent of our revised approach to budget impact is to document

the percentage of patients who could be treated at selected prices without crossing a budget impact threshold that is aligned with overall growth in the US economy. For 2019-2020, the five-year annualized potential budget impact threshold that should trigger policy actions to manage access and affordability is calculated to be approximately $819 million per year for new drugs.

7.3 Results

Table 7.1 illustrates the five-year annualized per-patient potential budget impact of upadacitinib in place of adalimumab in patients in whom conventional DMARDs such as methotrexate have failed, and who are eligible for treatment with a TIM. These results are based on upadacitinib's prices to reach thresholds of $150,000, $100,000, and $50,000 per QALY ($43,466, $44,144, and $44,822 per year, respectively) compared to adalimumab.

Table 7.1. Per-Patient Potential Budget Impact Calculations Over a Five-Year Time Horizon

	Average Annual per Patient Budget Impact		
	$150,000/QALY	$100,000/QALY	$50,000/QALY
Upadacitinib + cDMARD (Annualized Cost)	$47,800	$47,500	$47,300
Adalimumab + cDMARD (Annualized Cost)	$46,900		
Upadacitinib + cDMARD Potential Budget Impact	$900	$600	$300

cDMARD: conventional disease-modifying antirheumatic drug, QALY: quality-adjusted life year

Average potential budget impact at the three cost-effectiveness threshold prices for the drug ranged from approximately $900 per patient using the annual price to achieve $150,000 per QALY to approximately $300 using the annual price to achieve a $50,000 per QALY cost-effectiveness threshold.

For upadacitinib, the annual potential budget impact of treating the entire eligible population across all prices (WAC, assumed net price, and the three cost-effectiveness threshold prices for $50,000, $100,000, and $150,000 per QALY) did not exceed the $819 million threshold. The annual potential budget impacts of treating the entire eligible population using the different price levels are compared to the $819 million annual budget impact threshold in Table 7.2. Overall, the greatest potential annual budget impact we estimated was 6% of the $819 million threshold, using its price to reach a threshold of $150,000 per QALY.

Table 7.2. Estimated Total Population Annual Potential Budget Impact at Different Prices of Upadacitinib for the Eligible Population of 20,000 per Year

Price	Five-Year Annualized Total Population Budget Impact	Percent of Budget Impact Threshold
$150,000 per QALY Threshold Price	$48.2 million	6%
$100,000 per QALY Threshold Price	$34.3 million	4%
$50,000 per QALY Threshold Price	$20.6 million	3%

QALY: quality-adjusted life year, WAC: wholesale acquisition cost

8. Summary of the Votes and Considerations for Policy

8.1 About the CTAF Process

During CTAF public meetings, the panel deliberates and votes on key questions related to the systematic review of the clinical evidence, an economic analysis of the applications of treatments under examination, and the supplementary information presented. Panel members are not pre-selected based on the topic being addressed and are intentionally selected to represent a range of expertise and diverse perspectives.

Acknowledging that any judgment of evidence is strengthened by real-life clinical and patient perspectives, subject matter experts are recruited for each meeting topic and provide input to CTAF panel members before the meeting to help clarify their understanding of the different interventions under analysis in the evidence review. The same clinical experts serve as a resource to the CTAF panel during their deliberation, and help to shape recommendations on ways the evidence can apply to policy and practice.

After the CTAF panel votes, a Policy Roundtable discussion is held with the CTAF panel, clinical experts, patient advocates, payers, and when feasible, manufacturers. The goal of this discussion is to bring stakeholders together to apply the evidence to guide patient education, clinical practice, and coverage and public policies. Participants on Policy Roundtables are selected for their expertise on the specific meeting topic, are different for each meeting, and do not vote on any questions.

At the December 9, 2019 meeting, the CTAF panel discussed issues regarding the application of the available evidence to help patients, clinicians, and payers address important questions related to the use of JAK inhibitors for RA. Following the evidence presentation and public comments (public comments from the meeting can be accessed here, starting at minute 01:11:01), the CTAF panel voted on key questions concerning the comparative clinical effectiveness, comparative value, and potential other benefits and contextual considerations related to JAK inhibitors. These questions are developed by the ICER research team for each assessment to ensure that the questions are framed to address the issues that are most important in applying the evidence to support clinical practice, medical policy decisions, and patient decision-making. The voting results are presented below, along with specific considerations mentioned by CTAF panel members during the voting process.

In its deliberations and votes related to value, the CTAF panel considered the individual patient benefits, and incremental costs to achieve such benefits, from a given intervention over the long term.

There are four elements to consider when deliberating on long-term value for money (see Figure 8.1 below):

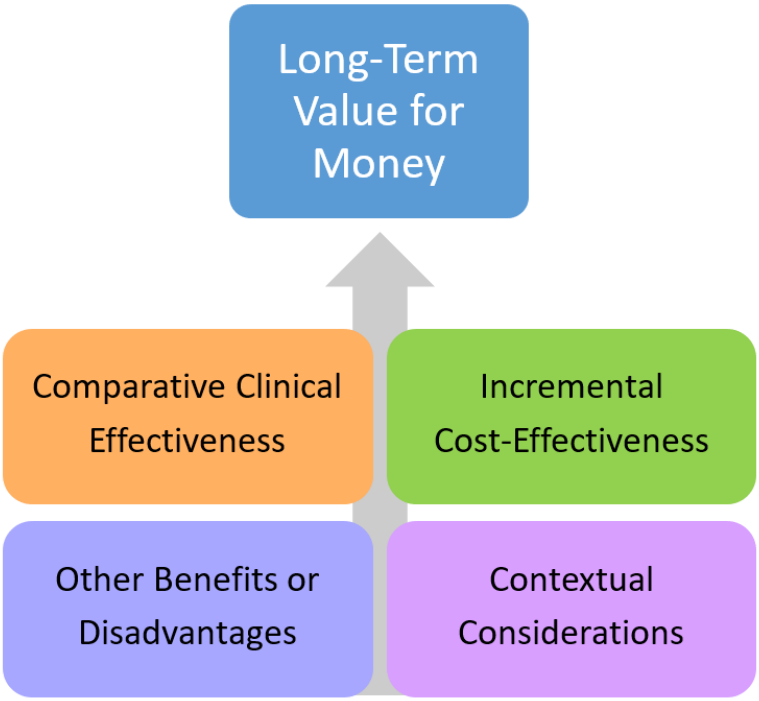
1) Comparative clinical effectiveness is a judgment of the overall difference in clinical outcomes between two interventions (or between an intervention and placebo), tempered by the level of certainty possible given the strengths and weaknesses of the body of evidence. CTAF uses the ICER Evidence Rating Matrix as its conceptual framework for considering comparative clinical effectiveness.

2) Estimated incremental cost-effectiveness is the average incremental cost per patient of one intervention compared to another to achieve a desired "health gain," such as an additional stroke prevented, case of cancer diagnosed, or gain of a year of life. Alternative interventions are compared in terms of cost per unit of effectiveness, and the resulting comparison is presented as a cost-effectiveness ratio. Relative certainty in the cost and outcome estimates continues to be a consideration. As a measure of cost-effectiveness, the CTAF voting panel follows common academic and health technology assessment standards by using cost per quality-adjusted life year (QALY), with formal voting on "long-term value for money" when the base case incremental cost-effectiveness ratio is between $50,000 per QALY and $175,000 per QALY.

3) Potential other benefits refer to any significant benefits or disadvantages offered by the intervention to the individual patient, caregivers, the delivery system, other patients, or the public that would not have been considered as part of the evidence on comparative clinical effectiveness. Examples of potential other benefits include better access to treatment centers, mechanisms of treatment delivery that require fewer visits to the clinician's office, treatments that reduce disparities across various patient groups, and new potential mechanisms of action for treating clinical conditions that have demonstrated low rates of response to currently available therapies. Other disadvantages could include increased burden of treatment on patients or their caregivers. For each intervention evaluated, it will be open to discussion whether potential other benefits or disadvantages such as these are important enough to factor into the overall judgment of long-term value for money. There is no quantitative measure for potential other benefits or disadvantages.

4) Contextual considerations include ethical, legal, or other issues (but not cost) that influence the relative priority of illnesses and interventions. Examples of contextual considerations include whether there are currently any existing treatments for the condition, whether the condition severely affects quality of life or not, and whether there is significant uncertainty about the magnitude of benefit or risk of an intervention over the long term. There is no quantitative measure for contextual considerations.

Figure 8.1. Conceptual Structure of Long-Term Value for Money



8.2 Voting Results

1) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the net health benefit of upadacitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

The panel unanimously determined that the evidence was adequate to demonstrate that upadacitinib plus a conventional DMARD provides a superior net health benefit compared to treatment with conventional DMARDs alone. Panel members cited the favorable results of SELECT-NEXT, in which upadacitinib generated statistically superior improvements in disease activity, remission, and ACR response relative to conventional DMARD therapy at 12, 24, and 48 weeks of follow up.

2) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the net health benefit of upadacitinib plus a conventional DMARD is superior to that provided by adalimumab plus a conventional DMARD?

Yes: 12 votes	No: 2 votes

A majority of the panel judged the evidence adequate to demonstrate that upadacitinib plus a conventional DMARD provides a superior net health benefit in comparison to adalimumab plus a conventional DMARD. Panelists who voted in the affirmative noted the superior results that upadacitinib generated in SELECT-COMPARE, such as statistically significant improvements in rates of disease activity as measured by the DAS28-CRP, ACR response, and HAQ-DI score. In addition, panelists noted that patients randomized to upadacitinib experienced greater improvements in quality of life, pain, and fatigue compared to adalimumab.

The two panelists who voted in the negative expressed concern around the entry criteria of SELECT-COMPARE, namely the requirement that patients display an hsCRP level of 5 mg per liter or greater. These panelists voiced some unease regarding the relationship between CRP elevation and upadacitinib, which operates on the IL-6 pathway. In addition, one panelist who voted "no" pointed to the higher rate of discontinuation in the upadacitinib arm relative to the adalimumab arm in SELECT-COMPARE.

3) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the net health benefit of tofacitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

The panel unanimously voted that the evidence was adequate to demonstrate that tofacitinib plus a conventional DMARD offers a superior net health benefit in comparison to conventional DMARD therapy alone. Panelists cited the positive results of ORAL Sync, ORAL Scan, and several pooled tofacitinib trials, which showed that tofacitinib generated superior improvements in disease activity, remission, and ACR response relative to conventional DMARD therapy alone. Panelists were further convinced because such improvements were sustained at weeks 24 and 48.

4) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the net health benefit of tofacitinib plus a conventional DMARD is superior to that provided by adalimumab plus a conventional DMARD?

Yes: 0 votes	***No: 14 votes***

The panel unanimously concluded that the evidence was inadequate to demonstrate that tofacitinib plus a conventional DMARD offers a superior net health benefit compared to adalimumab plus a conventional DMARD. Panelists underscored the results of ORAL Strategy, in which tofacitinib was not statistically different from adalimumab in rates of remission achieved as measured by the DAS28-ESR, ACR response, and improvement in HAQ-DI. Panelists also cited the lack of a statistically significant difference in HAQ-DI change, and that patients achieved comparable improvements in quality of life, pain, and fatigue with tofacitinib or adalimumab.

5) In patients who are naïve to TIMs, is the evidence adequate to distinguish the net health benefit between upadacitinib and tofacitinib?

Yes: 0 votes	***No: 14 votes***

The panel unanimously determined that the evidence was inadequate to distinguish the net health benefit between upadacitinib and tofacitinib. Panelists cited the lack of head-to-head trials and the insufficiency of data to allow for indirect companions.

5a) No vote was taken on Question 5a because the panel unanimously voted that the evidence was inadequate to distinguish between upadacitinib and tofacitinib.

6) In patients who are naïve to TIMs, is the evidence adequate to demonstrate that the biosimilar infliximab-dyyb produces a net health benefit comparable to that of the reference biologic infliximab?[†]

Yes: 14 votes	No: 0 votes

The panel unanimously determined that the evidence was adequate to demonstrate that infliximab-dyyb produces a net health benefit comparable to that of infliximab. The panel highlighted the preponderance of evidence, including a head-to-head trial, which demonstrated that the biosimilar and the reference product were clinically equivalent and had similar rates of adverse events.

[†]The webcast recording of the meeting shows one "no" vote in Question 6. After the conclusion of the meeting, the panelist who voted "no" clarified that they had done so accidentally after misreading the question.

7) In patients who are TIM-experienced, is the evidence adequate to demonstrate that the net health benefit of upadacitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

The panel unanimously judged that the evidence was adequate to demonstrate that upadacitinib plus a conventional DMARD provides a superior net health benefit compared to conventional DMARD therapy alone. Though trials in this population were smaller than those for TIM-naive patients, the panel expressed certainty, citing the statistically and clinically significant improvements in measures of disease activity, ACR response, and HAQ improvement of upadacitinib plus a conventional DMARD versus conventional DMARD therapy alone.

8) In patients who are TIM-experienced, is the evidence adequate to demonstrate that the net health benefit of tofacitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

The panel unanimously determined that the evidence was adequate to demonstrate that tofacitinib plus a conventional DMARD is superior to a conventional DMARD alone. The panel reiterated the rationale of their unanimous vote for Question 7.

9) In patients who are TIM-experienced, is the evidence adequate to demonstrate that the net health benefit of baricitinib plus a conventional DMARD is superior to that provided by a conventional DMARD alone?

Yes: 14 votes	No: 0 votes

The panel unanimously determined that the evidence was adequate to demonstrate that baricitinib plus a conventional DMARD is superior to a conventional DMARD alone. The panel reiterated the rationale of their unanimous vote for Question 7.

10) Does treating patients with upadacitinib plus a conventional DMARD offer one or more of the following potential "other benefits" in comparison to adalimumab plus a conventional DMARD?

This intervention offers reduced complexity that will significantly improve patient outcomes.	14/14
This intervention will significantly reduce caregiver or broader family burden.	9/14
This intervention offers a novel mechanism of action or approach that will allow successful treatment of many patients for whom other available treatments have failed.	3/14
This intervention will have a significant impact on improving patients' ability to return to work and/or their overall productivity.	6/14
There are other important benefits or disadvantages that should have an important role in judgments of the value of this intervention.	0/14

The panel unanimously acknowledged that upadacitinib offers reduced complexity in comparison to adalimumab. Numerous panelists referred to patient testimony: both patient experts on the Policy Roundtable emphasized the potential benefits offered by oral administration, including improved outcomes for patients with needle phobia, reduction in pain and infection associated with injections, and a possible increase in adherence. Panelists also noted that oral treatments offer consistent delivery, which could potentially mitigate the tapering effects seen with infusions and allow patients to return to work or school. Lastly, several panelists echoed the remarks from patient experts who noted that oral products may have a longer shelf life than injectables, which need to be temperature controlled. Therefore, oral administration may benefit patients who travel or those who were previously unable to do so. A majority of the panel also determined that treatment with upadacitinib could potentially reduce caregiver or broader family burden for many of the same reasons cited above.

11) Are any of the following contextual considerations important in assessing the long-term value for money at estimated pricing of upadacitinib?

This intervention is intended for the care of individuals with a condition of particularly high severity in terms of impact on length of life and/or quality of life.	12/14
This intervention is intended for the care of individuals with a condition that represents a particularly high lifetime burden of illness.	12/14
Compared to adalimumab, there is significant uncertainty about the long-term risk of serious side effects of this intervention.	4/14
Compared to adalimumab, there is significant uncertainty about the magnitude or durability of the long-term benefits of this intervention.	5/14
There are additional contextual considerations that should have an important role in judgments of the value of this intervention.	0/14

A majority of the panel recognized that upadacitinib is intended for the care of individuals with a condition of high severity with a particularly high lifetime burden of illness. The panel again referred to patient testimony, which highlighted the emotional and physical toll

of RA as well as the impact of the disease on quality of life, loved ones and family, and activities of daily living. Several panelists expressed uncertainty about the risk of long-term side effects and durability of long-term benefit, but were largely assuaged by the results of SELECT-COMPARE, in which upadacitinib and adalimumab showed comparable safety profiles.

12) Given the available evidence on comparative effectiveness, incremental cost-effectiveness using the net price estimate of 26% off the WAC, and considering other benefits, disadvantages, and contextual considerations, what is the long-term value for money of treatment with upadacitinib plus a conventional DMARD versus adalimumab plus a conventional DMARD?

Low: 8 votes	Intermediate: 6 votes	High: 0 votes

A slight majority of the panel judged the long-term value for money of treatment with upadacitinib plus a conventional DMARD as "low." Though most panelists who voted this way acknowledged that upadacitinib offers an incremental net health benefit, many voiced concerns not only around uncertainty, but also generalizability. Several panelists noted that a significant portion of patients with RA likely would not have qualified for SELECT-COMPARE based on its entry criteria. Panelists expressed overall uncertainty about whether the results achieved in the trial would extend to patients in the real world. Considering these issues, as well as the small number of QALYs gained and the fact that adalimumab may not be cost effective in and of itself, eight members of the panel were unable to justify an "intermediate" or "high" value vote even though the price of treatment falls under the upper end of commonly cited cost-utility thresholds.

In contrast to the panelists who voted "low," a majority of panelists who voted "intermediate" determined the comparative clinical effectiveness to be superior, as opposed to incremental. Further, panelists who voted "intermediate," highlighted that the price of upadacitinib falls within commonly cited cost-utility thresholds. In addition, many panel members reiterated several contextual considerations and potential other benefits that factored into their vote, such as oral administration, positive impact on caregiver or family burden, and the high lifetime burden and severity of RA.

8.3 Roundtable Discussion and Key Policy Implications

Following its deliberation on the evidence, the CTAF panel engaged in a moderated discussion with a Policy Roundtable about how best to apply the evidence on JAK inhibitors for RA to policy and practice. The Policy Roundtable members included two patients, two clinical experts, two payer representatives, and two manufacturer representatives. The discussion reflected multiple perspectives and opinions, and therefore, none of the statements below should be taken as a consensus view held by all participants. The names of the Policy Roundtable participants are shown below, and COI disclosures for all meeting participants can be found in Appendix G.

Table 8.1. Policy Roundtable Members

Name	Title and Affiliation
Happy Chan, DO, FACR	Medical Director, Medical Care Solutions, Blue Shield of California
Andrew L. Concoff, MD, FACR, CAQSM	Executive Vice President, Chief Value Medical Officer, United Rheumatology
Peggy Ehling	Board Chair, Arthritis Foundation Local Leadership Board Los Angeles
Kirstin Griffing MD, MS, FACR	Senior Medical Advisor, United States Medical Affairs, Rheumatology, Eli Lilly and Company
Marc Jensen, PharmD	Senior Director, Inflammation and Immunology Field Medical, Pfizer
Christopher Phillips, MD	Chair, American College of Rheumatology Insurance Subcommittee; Rheumatologist, Paducah Rheumatology
John S. Yao, MD, MPH, MBA, MPP, CPC, FACP	Regional Vice President and Chief Medical Officer, Anthem Blue Cross

The Policy Roundtable discussion was facilitated by Dr. Steven D. Pearson, MD, MSc, President of ICER. The main themes and recommendations from the discussion are summarized below.

Biosimilars

Biosimilars offer a tremendous unrealized opportunity to bring patients the benefits of the most effective treatments available while controlling costs for patients and the health care system. Looking at the evidence on one representative biosimilar, infliximab-dyyb, CTAF voted unanimously that the evidence was adequate to demonstrate that the biosimilar was clinically equivalent to its reference product, Remicade. Unfortunately, several significant barriers have limited the impact of biosimilars like infliximab-dyyb in the US health care system. At the CTAF meeting, two obstacles were highlighted: 1) a lack of trust among some clinicians and patients in the clinical equivalency of biosimilars, and 2) market barriers caused by current rebate structures that dominate autoimmune formulary placement decisions.

Related to these barriers, the Arthritis Foundation has highlighted three key principles for increasing uptake in biosimilars in a position statement, which garnered the support of many other organizations:

- Patient trust and physician confidence are crucial factors for broader uptake.
- Precise and consistent language is essential to prevent confusion and bias.
- Public policies that prohibit anticompetitive rebate structures, promote affordable out-of-pocket costs for patients, and enhance choice are critical.

Additional information regarding the Arthritis Foundation position on biosimilars may be found here. Position statements from the ACR and United Rheumatology may be found here and here, respectively.

Kaiser is an example of an organization where concerted efforts between the insurer and providers have dramatically increased the use of biosimilars in delivering high quality, guideline-concordant care to patients. Kaiser has focused not only on using lower cost biosimilars for patients beginning autoimmune treatment but has also worked to create a positive environment for switching patients over to biosimilars from reference products. To do this, Kaiser rheumatologists supported an overall process that created a leading role for clinical pharmacists, who were trained in how to present to patients the information about switching to a biosimilar from the reference product. Pharmacists reached out directly to patients, provided information about biosimilars, allayed fears about switching, and answered patients' questions. Kaiser also had their pharmacy outcomes group collect data following switching to biosimilars and demonstrated that patient satisfaction was excellent and that there were no worrisome changes in the DAS28 ESR or CRP levels. This kind of broad approach represents one best practice in efforts to build a system in which clinicians and patients feel enough trust to participate in large-scale switching of patients over to biosimilars.

Unfortunately, it is difficult to identify a "best practice" that can address the negative market incentives for biosimilars created by bundled rebates across indications that can cement preferred formulary status for leading autoimmune drugs like adalimumab (Humira). Indication-specific pricing and formulary development may have some potential to allow biosimilars to compete on a more even playing field, but short-term financial incentives for pharmacy benefit managers (PBMs) and plan sponsors to optimize rebates will continue to bedevil efforts to seek longer-term cost control through biosimilar competition unless major changes are made to the rebate system.

To address the trust and rebate barriers to biosimilar adoption, several policy recommendations are suggested:

1) **Clinical societies should educate their clinician members that the evidence behind biosimilars is sound and that whenever they are available at a lower cost than reference products, they should be the preferred option given the benefits of lower costs for patients and the health care system.**
2) **Patient groups should educate their members that biosimilars are as safe and effective as reference products and that starting on a biosimilar, or switching to one, is clinically responsible and may be financially beneficial.**
3) **Manufacturers should not use scare tactics to dissuade responsible switching to biosimilars.**

4) **Health plan sponsors, insurers, PBMs, and provider groups should work together to promote greater use of biosimilars by implementing switching programs that provide broad support to patients while assuring that patients who do not respond well to biosimilars are able to access reference products and/or obtain other targeted treatment options without delay.**
5) **Policymakers should continue work on alternatives to the current rebate system that will allow the market to reward the competitive advantages of lower priced, equally effective biosimilar treatment options. One helpful first step would be to ban the bundling of rebates across multiple indications.**

Plan Sponsors, Payers, and PBMs

1) **Reconsider the need for step therapy and any switching programs if pricing becomes better aligned with clinical value.**

There should be fair access for fair pricing. Current step therapy is not working to contain prices and is a result of a dysfunctional pricing system and the inability to use other mechanisms to lower costs for equally effective therapies. When developed using evidence-based standards and clinical expert input, step therapy can be a useful tool to steer patients to higher value therapies. The lack of evidence to support treatment sequencing makes evidenced-based step therapy impossible. For example, there is no evidence that the most commonly required first step therapy, an anti-TNF agent, is any more effective or safer than other therapies, such as the JAK inhibitors. At current prices, our analysis indicates that all of the TIMs under review exceed common cost-effectiveness thresholds but have comparable levels of effectiveness. Step therapy is not an unreasonable approach in this case, as the focus can be on the less expensive (or most heavily discounted) TIMs. However, if pricing were to become better aligned with clinical value for these therapies, payers should reconsider whether step therapy remains a necessary option.

Guidelines for the ethical design of step therapy exist in the literature.[104] The key elements include:

- Ensure that short-term cost savings are weighed against long-term outcome and costs.
- Ensure that first step drugs are clinically appropriate.
- There should be high likelihood that the patient will attain treatment goals with step therapy.
- First step therapy should not cause long-term harm.
- Moving from the first step based on clinical groups should be quick and easy.
- Failure of first step therapy should be clearly defined.
- Relevant new evidence should be rapidly reviewed leading to appropriate modifications.
- Rationale and rules should be explicit and transparent.

2) **If pricing does not fairly align with patient benefit, plan sponsors may consider the option of benefit designs requiring patients to switch from their current therapy to a lower cost, equally effective option. In all such efforts, the clinical effectiveness of the required therapy and the cost overall should be closely monitored.**

Traditionally, payers have authorized coverage of continuation of existing therapies when a patient switches to their plan. However, the high cost of TIMs has led some plans to drop this "grandfathering" of prior therapy even when it is not included on their formulary. It is incumbent upon payers to monitor the impact of such policies on patients to ensure that they are not harmed and that the goal of cost containment for the plan is actually met.

3) **Requiring switching for patients who have achieved adequate treatment response on a particular drug raises the risk of irremediable harm should the new treatment not prove equally effective. Therefore, these programs must meet an extremely high bar of evidence and must have ample, efficient means by which clinicians can request exceptions on clinical grounds.**

Patients should never be required to switch back to a prior therapy that they have been failed by. They should stay within the same drug class with the availability of an immediate path back to their initial therapy.

4) **Increase transparency around the role of discounting and rebate practice in formulary design.**

Patients and providers expressed concerns around a lack of transparency regarding the connection between discounting/rebate practice and formulary design, as the terms are typically confidential and current designs do not always lead to cost-effective choices. For example, many of the TIMs in our prior review were more effective than adalimumab in head-to-head studies and less expensive, but are disadvantaged in most formulary designs. Transparency is an important first step in educating the general public on the role of rebates in formulary design as well as how the savings are shared among patients, PBMs, and health plans.

5) **Design innovative risk-sharing payment agreements, including outcome-based contracts with manufacturers and shared risk agreements with accountable care organizations.**

The Policy Roundtable discussion mentioned other efforts that might yield cost savings beyond price reductions and rebate negotiations. Pay-for-performance agreements have increased in popularity; as a relevant example, payers might receive an additional rebate for patients who do not achieve clinical remission during TIM therapy. Payers could also develop contracts with accountable care organizations that share both financial risk and savings for optimized use and sequencing of TIM therapies. Finally, if TIMs with multiple indications beyond RA (e.g., Crohn's disease, psoriasis) show markedly different clinical performance and costs, indication-specific formularies should be developed to recognize these differences.

Clinical Societies and Manufacturers

1) **Establish standardized assessments to allow for rigorous direct and indirect comparisons of evidence across studies and therapeutic alternatives.**

The field of rheumatology should be applauded for its work in developing core sets of clinical and patient-reported outcomes measures with strong internal and external validity in RA clinical studies. However, these measures have proliferated, with variations on the same theme (e.g., multiple disease activity and radiographic progression measures) making cross-study comparisons problematic. Clinical societies and manufacturers should collaborate to develop a standardized *and* limited "must have" list of clinical outcome measures at specific time-points for clinical trials and post-marketing studies.

Public Policy Decision Makers

1) **In a dysfunctional market system, in order to protect patients today and improve their future access to innovative therapies, policy makers may need to consider some form of regulatory intervention to ensure that drug prices and price increases do not continue their current upward trajectory, driving prices further from reasonable alignment with the added benefits for patients.**

The discussion highlighted that the current "marketplace" for RA drugs is not working to align prices with value in a way that would reward new innovative drugs while also reaping the benefits of competition to keep drugs affordable. One result is that patients and clinicians face heavy access restrictions from insurers who lack other effective methods to control unsustainable cost increases. In addition, innovative companies cannot compete on price for RA-specific products in the face of rebate-driven agreements that cover multiple indications beyond RA. If some or all of the measures described above cannot be implemented or are not effective, policymakers should consider some form of regulatory intervention to address rebate structures and price increases in order to increase true competition and better protect patients.

2) **Public policy decision makers should ban the bundling of rebates across indications.**

We understand that rebates are a major impediment to the alignment of overall pricing with outcomes. Bundling and rebates present almost insurmountable obstacles to value-based pricing and represent a wall against innovation for new products trying to enter the market with only a single indication. In addition, bundling leads to step therapy that does not always match the FDA indications for the drugs nor their evidence base or clinical guidelines. Finally, patients' copays are linked to the WAC, so their out-of-pocket costs remain high, even though rebates reduce the cost for payers and PBMs. Eliminating rebates is the first step towards improving value for therapies for RA.

Research Community

1) **Researchers and research funding agencies should prioritize studies that identify biomarkers predicting response to currently available therapies or classes of therapies (personalized medicine).**

The current treatment paradigm is based on trial and error and the majority of patients fail to respond adequately to their first TIM. Clinicians urgently need tools that can help them identify the therapy that will work best for each individual patient that they treat.

2) **Researchers should separate outcomes that measure inflammation from those that measure pain.**

Research should not try to capture all aspects of RA in one measure. There should be several essential outcomes and patient-reported outcomes should complement the more objective outcomes. Therapies that target inflammation, such as TIMs, are best assessed by measures like ESR, CRP, swollen joint, and the prevention of joint destruction. Other therapies, such as those addressing central pain mediators, are best assessed through patient-reported outcomes such as joint pain, joint stiffness, and fatigue.

3) **Randomized trials and observational studies should include outcomes that are both patient-centered and patient-driven.**

Patient-centered outcomes are answered by patients, while patient-driven outcomes are those that are coproduced by patients during the design phase of the studies and reflect values that are meaningful to patients. Both are essential in order to capture the value of treatments to patients.

4) **The FDA should require that randomized trials include patients reflecting the average RA patient population.**

Patients, specialists, and their respective society representatives reported that 80% to 90% of the patients they treat with TIMs would not have been eligible to participate in the pivotal randomized trials, primarily due to lower numbers of inflamed joints and lower levels of inflammatory markers (CRP, ESR). Specialists have concerns about using the estimates of benefits and harms from the clinical trials when counseling their patients about the relative benefits and harms of the therapeutic options and patients do not feel adequately represented in the trials.

5) The FDA should require that randomized trials of new therapies always include an active comparator.

It is unethical to randomize patients to a placebo added to a therapy that they have already been failed by (i.e., placebo + conventional DMARD in a patient population with moderately-to-severely active disease on conventional DMARD therapy) given that multiple TIMs have been demonstrated to be clinically superior to conventional DMARD therapy alone in this population. If the FDA feels that a placebo control is essential to demonstrate clinical efficacy, then the trial should still include a third arm with an active comparator. Head-to-head randomized trial data is essential to support comparative clinical effectiveness analyses, NMAs, and economic analyses with low uncertainty.

***

This is the second ICER review of treatments for RA.

References

1. Helmick CG, Felson DT, Lawrence RC, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis and rheumatism.* 2008;58(1):15-25.
2. Huizinga TW, Pincus T. In the clinic. Rheumatoid arthritis. *Ann Intern Med.* 2010;153(1):ITC1-1-ITC1-15; quiz ITC11-16.
3. Radawski C, Genovese MC, Hauber B, et al. Patient Perceptions of Unmet Medical Need in Rheumatoid Arthritis: A Cross-Sectional Survey in the USA. *Rheumatology and therapy.* 2019;6(3):461-471.
4. Wisely C. American College of Rheumatology: Five Things Physicians and Patients Should Question. 2013; https://www.choosingwisely.org/societies/american-college-of-rheumatology/. Accessed 9/9/19, 2019.
5. Smolen JS, Breedveld FC, Burmester GR, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Annals of the rheumatic diseases.* 2016;75(1):3-15.
6. Fleischmann R, Pangan AL, Song I-H, et al. Upadacitinib versus Placebo or Adalimumab in Patients with Rheumatoid Arthritis and an Inadequate Response to Methotrexate: Results of a Phase 3, Double-Blind, Randomized Controlled Trial. *Arthritis & Rheumatology.* 2019;0(ja).
7. Burmester GR, Kremer JM, Van den Bosch F, et al. Safety and efficacy of upadacitinib in patients with rheumatoid arthritis and inadequate response to conventional synthetic disease-modifying anti-rheumatic drugs (SELECT-NEXT): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet (London, England).* 2018;391(10139):2503-2512.
8. Kremer J, Li Z-G, Hall S, et al. Tofacitinib in combination with nonbiologic disease-modifying antirheumatic drugs in patients with active rheumatoid arthritis: a randomized trial. *Annals of internal medicine.* 2013;159(4):253-261.
9. Vollenhoven R. Tofacitinib or adalimumab versus placebo in rheumatoid arthritis (New England Journal of Medicine (2012). *New England Journal of Medicine.* 2013;369(3):293.
10. Fleischmann R, Mysler E, Hall S, et al. Efficacy and safety of tofacitinib monotherapy, tofacitinib with methotrexate, and adalimumab with methotrexate in patients with rheumatoid arthritis (ORAL Strategy): a phase 3b/4, double-blind, head-to-head, randomised controlled trial. *Lancet (London, England).* 2017;390(10093):457-468.
11. van der Heijde D, Tanaka Y, Fleischmann R, et al. Tofacitinib (CP-690,550) in patients with rheumatoid arthritis receiving methotrexate: twelve-month data from a twenty-four-month phase III randomized radiographic study. *Arthritis and rheumatism.* 2013;65(3):559-570.
12. Charles-Schoeman C, Burmester GNPZCASKKKHTBEFR. Efficacy and safety of tofacitinib following inadequate response to conventional synthetic or biological disease-modifying antirheumatic drugs. *Annals of the rheumatic diseases.* 2016;75(7):1293.
13. Genovese MC, Fleischmann R, Combe B, et al. Safety and efficacy of upadacitinib in patients with active rheumatoid arthritis refractory to biologic disease-modifying anti-rheumatic drugs (SELECT-BEYOND): a double-blind, randomised controlled phase 3 trial. *Lancet (London, England).* 2018;391(10139):2513-2524.
14. Burmester GR, Blanco R, Charles-Schoeman C, et al. Tofacitinib (CP-690,550) in combination with methotrexate in patients with active rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitors: a randomised phase 3 trial. *Lancet (London, England).* 2013;381(9865):451-460.

15. Genovese MC, Kremer J, Zamani O, et al. Baricitinib in Patients with Refractory Rheumatoid Arthritis. *The New England journal of medicine.* 2016;374(13):1243-1252.
16. Bechman K, Subesinghe S, Norton S, et al. A systematic review and meta-analysis of infection risk with small molecule JAK inhibitors in rheumatoid arthritis. *Rheumatology.* 2019.
17. Harigai M. Growing evidence of the safety of JAK inhibitors in patients with rheumatoid arthritis. *Rheumatology (Oxford, England).* 2019;58(Supplement_1):i34-i42.
18. Dougados M, van der Heijde DCYCGMDELJBSWSdlTIGCRTS. Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. *Annals of the rheumatic diseases.* 2017;76(1):88.
19. Taylor PC, Keystone EC, van der Heijde D, et al. Baricitinib versus Placebo or Adalimumab in Rheumatoid Arthritis. *The New England journal of medicine.* 2017;376(7):652-662.
20. Kim SC, Solomon DH, Rogers JR, et al. No difference in cardiovascular risk of tocilizumab versus abatacept for rheumatoid arthritis: A multi-database cohort study. *Seminars in arthritis and rheumatism.* 2018;48(3):399-405.
21. Smolen J, Landewe RB, Mease P, et al. Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the RAPID 2 study. A randomised controlled trial. *Annals of the rheumatic diseases.* 2009;68(6):797-804.
22. Carlson JJ, Ogale S, Dejonckheere F, Sullivan SD. Economic evaluation of tocilizumab monotherapy compared to adalimumab monotherapy in the treatment of severe active rheumatoid arthritis. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2015;18(2):173-179.
23. Karlsson JA, Kristensen LE, Kapetanovic MC, Gulfe A, Saxne T, Geborek P. Treatment response to a second or third TNF-inhibitor in RA: results from the South Swedish Arthritis Treatment Group Register. *Rheumatology (Oxford, England).* 2008;47(4):507-513.
24. Gu T, Shah N, Deshpande G, Tang DH, Eisenberg DF. Comparing Biologic Cost Per Treated Patient Across Indications Among Adult US Managed Care Patients: A Retrospective Cohort Study. *Drugs - real world outcomes.* 2016;3(4):369-381.
25. Myasoedova E, Crowson CS, Kremers HM, Therneau TM, Gabriel SE. Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. *Arthritis and rheumatism.* 2010;62(6):1576-1582.
26. Kavanaugh A, Singh R, Karki C, et al. Disease activity and biologic use in patients with psoriatic arthritis or rheumatoid arthritis. *Clinical rheumatology.* 2018;37(8):2275-2280.
27. United States Public Health Service. CHAPTER 6A-PUBLIC HEALTH SERVICE SUBCHAPTER II-GENERAL POWERS AND DUTIES: Part F-Licensing of Biological Products and Clinical Laboratories: subpart 1-biological products. In: Services DoHaH, ed. uscode.house.gov2019.
28. Biosimilar Implementation Committee. Biologics Price Competition and Innovation Act of 2009. In:2009.
29. Kinzinger A. Public Health Service Act. In: Congress US, ed. legcounsel.house.gov2018.
30. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Nonproprietary Naming of Biological Products: Guidance for Industry. In: Administration FaD, ed. fda.gov: Food and Drug Administration; 2017.
31. Yoo D, Prodanovic NJJMPRELABAWPA-MCOBSSKHLS. Efficacy and safety of CT-P13 (biosimilar infliximab) in patients with rheumatoid arthritis: comparison between switching from reference infliximab to CT-P13 and continuing CT-P13 in the PLANETRA extension study. *Annals of the rheumatic diseases.* 2017;76(2):355.

32. Yoo D, Miranda P, Piotrowski M, et al. A randomized, double-blind, phase 3 study demonstrates clinical equivalence of CT-P13 to infliximab when co-administered with methotrexate in patients with active rheumatoid arthritis. *Annals of the Rheumatic Disease.* 2013;71.
33. Yoo DH, Prodanovic N, Jaworski J, et al. Efficacy and safety of CT-P13 (biosimilar infliximab) in patients with rheumatoid arthritis: Comparison between switching from reference infliximab to CT-P13 and continuing CT-P13 in the PLANETRA extension study. *Annals of the Rheumatic Diseases.* 2016.
34. Smolen JS, Aletaha D, Bijlsma JW, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. *Annals of the rheumatic diseases.* 2010;69(4):631-637.
35. Van Nies JA, de Jong Z, van der Helm-van Mil AH, Knevel R, Le Cessie S, Huizinga TW. Improved treatment strategies reduce the increased mortality risk in early RA patients. *Rheumatology (Oxford, England).* 2010;49(11):2210-2216.
36. 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.
37. McCormick KA, Fleming B. Clinical practice guidelines. The Agency for Health Care Policy and Research fosters the development of evidence-based guidelines. *Health progress (Saint Louis, Mo).* 1992;73(10):30-34.
38. Weiss JE, Ilowite NT. Juvenile idiopathic arthritis. *Rheumatic diseases clinics of North America.* 2007;33(3):441-470, vi.
39. 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 & research.* 2012;64(5):640-647.
40. Fleischmann R, van der Heijde D, Koenig AS, et al. How much does Disease Activity Score in 28 joints ESR and CRP calculations underestimate disease activity compared with the Simplified Disease Activity Index? *Annals of the rheumatic diseases.* 2015;74(6):1132-1137.
41. Bartlett SJ, Orbai AM, Duncan T, et al. Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. *PloS one.* 2015;10(9):e0138543.
42. Arthritis Foundation. Arthritis Foundation Survey of Rheumatoid Arthritis Patient Treatment Experiences. In. Targeted Immune Modulators for Rheumatoid Arthritis: Effectiveness & Value: Institute for Clinical and Economic Review; 2017.
43. Boytsov N, Zhang X, Evans KA, Johnson BH. Impact of Plan-Level Access Restrictions on Effectiveness of Biologics Among Patients with Rheumatoid or Psoriatic Arthritis. *PharmacoEconomics.* 2019.
44. Aetna. Procedures, programs and drugs you must precertify: Participating provider precertification list. 2019; https://member.aetna.com/navigatorSecure/assets/pdfs/Precertification_Authorization.pdf. Accessed 11/15/19, 2019.
45. Cigna. Drug and Biologic Coverage Policy: Immunomodulators - Oral and Subcutaneous (Effective 01/01/20). 2019; https://cignaforhcp.cigna.com/public/content/pdf/coveragePolicies/pharmacy/ph_Future/ph_1903_coveragepositioncriteria_immunomodulators_IFP.pdf. Accessed 11/15/19, 2019.
46. Humana. Prescription Drug Guide Humana Formulary: List of covered drugs 2019; http://apps.humana.com/marketing/documents.asp?file=3450265. Accessed 11/15/19, 2019.
47. UnitedHealthcare. UnitedHealthcare Pharmacy: Clinical Pharmacy Programs. 2019; https://www.uhcprovider.com/content/provider/en/viewer.html?file=%2Fcontent%2Fdam%2Fprovider%2Fdocs%2Fpublic%2Fprior-auth%2Fdrugs-pharmacy%2Fcommercial%2Fr-z%2FCOMM-Rinvoq-Prior-Authorization-Notification.pdf. Accessed 11/15/19, 2019.

48. Kaiser Permanente. 2019 Drug Formulary. 2019; https://wa.kaiserpermanente.org/static/pdf/public/formulary/if-formulary-specialty-2019.pdf. Accessed 09/10/19, 2019.
49. Commercial Rx PA Guidelines. Health Net; 2019. https://www.healthnet.com/portal/common/content/iwc/common/unprotected/pharmacy_info/prior_auth_criteria.action. Accessed 11/15/19.
50. The Speciality Drug Evidence and Coverage (SPEC) Database. Tufts Medical Center; 2018. https://cevr.tuftsmedicalcenter.org/databases/spec-database. Accessed 09/10/19.
51. Kaiser Permanente. Criteria Based Consultation Prescribing Program: CRITERIA FOR DRUG COVERAGE tofacitinib (Xeljanz®, Xeljanz XR®). 2018; https://healthy.kaiserpermanente.org/static/health/pdfs/formulary/nw/Xeljanz.pdf. Accessed 09/10/19, 2019.
52. Kaiser Permanente. 2019 California Marketplace Formulary. 2019; https://healthy.kaiserpermanente.org/static/health/pdfs/formulary/cal/2019_CA_Marketplace_Formulary_Final_EN_REGLEG_NM_06.10.19_NV_ADA.pdf. Accessed 09/10/19, 2019.
53. Health Net. Clinical Policy: Biologic DMARDs. 2019; www.healthnet.com. Accessed 09/10/19, 2019.
54. California Department of Health Care Services. Medi-Cal Medicaid-Approved Drug List. In. https://www.medi-cal.ca.gov/: California Department of Health Care Services; 2019.
55. Chau J, Delate T, Ota T, Bhardwaja B. Patient Perspectives on Switching from Infliximab to Infliximab-dyyb in Patients with Rheumatologic Diseases in the United States. *ACR Open Rheumatology.* 2019;1(1):52-57.
56. Wild D. Bucking Trend, Kaiser Boosts Biosimilar Uptake. *Gastroenterology & Endoscopy News.* April 29, 2019, 2019;In the News.
57. Schumock GT, Stubbings J, Wiest MD, et al. National trends in prescription drug expenditures and projections for 2018. *American Journal of Health-System Pharmacy.* 2018;75(14):1023-1038.
58. Smolen JS, Landewe R, Breedveld FC, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. *Annals of the rheumatic diseases.* 2014;73(3):492-509.
59. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997;126(5):376-380.
60. Higgins JPT GSe. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011].* The Cochrane Collaboration. Available from http://handbook.cochrane.org.; 2011.
61. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg.* 2010;8(5):336-341.
62. Ollendorf DA, Pearson SD. An integrated evidence rating to frame comparative effectiveness assessments for decision makers. *Medical care.* 2010;48(6 Suppl):S145-152.
63. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med.* 2010;29(7-8):932-944.
64. U.S. Preventive Services Task Force. Procedure Manual. Agency for Healthcare Research and Quality. 2008.
65. van Tuyl LH, Vlad SC, Felson DT, Wells G, Boers M. Defining remission in rheumatoid arthritis: results of an initial American College of Rheumatology/European League Against Rheumatism consensus conference. *Arthritis and rheumatism.* 2009;61(5):704-710.
66. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. *The Journal of rheumatology.* 2000;27(1):261-263.

67. van der Heijde D, Simon L, Smolen J, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion. *Arthritis and rheumatism.* 2002;47(2):215-218.
68. Peterfy CG, Wu C, Szechinski J, et al. Comparison of the Genant-modified Sharp and van der Heijde-modified Sharp scoring methods for radiographic assessment in rheumatoid arthritis. *Int J Clin Rheumatol.* 2011;6(1):15-24.
69. O'Dell JR, Mikuls TR, Taylor TH, et al. Therapies for active rheumatoid arthritis after methotrexate failure. *The New England journal of medicine.* 2013;369(4):307-318.
70. van der Heijde D, Tanaka Y, Fleischmann R, et al. Tofacitinib (CP-690,550) in patients with rheumatoid arthritis receiving methotrexate: twelve-month data from a twenty-four-month phase III randomized radiographic study. *Arthritis and rheumatism.* 2013;65(3):559-570.
71. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE, Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis and rheumatism.* 2000;43(7):1478-1487.
72. Ward MM, Guthrie LC, Alba MI. Clinically important changes in individual and composite measures of rheumatoid arthritis activity: thresholds applicable in clinical trials. *Annals of the rheumatic diseases.* 2015;74(9):1691-1696.
73. van Vollenhoven RF, Geborek P, Forslind K, et al. Conventional combination treatment versus biological treatment in methotrexate-refractory early rheumatoid arthritis: 2 year follow-up of the randomised, non-blinded, parallel-group Swefot trial. *The Lancet.* 2012;379(9827):1712-1720.
74. Strand V, van Vollenhoven RF, Lee EB, et al. Tofacitinib or adalimumab versus placebo: patient-reported outcomes from a phase 3 study of active rheumatoid arthritis. *Rheumatology (Oxford, England).* 2016;55(6):1031-1041.
75. Taylor P, Keystone E, Van Der Heijde D, et al. Baricitinib versus placebo or adalimumab in patients with active rheumatoid arthritis (RA) and an inadequate response to background methotrexate therapy: Results of a phase 3 study. *Arthritis and Rheumatology.* 2015;67(no pagination).
76. van Dartel SA, Fransen J, Kievit W, et al. Difference in the risk of serious infections in patients with rheumatoid arthritis treated with adalimumab, infliximab and etanercept: results from the Dutch Rheumatoid Arthritis Monitoring (DREAM) registry. *Annals of the rheumatic diseases.* 2013;72(6):895-900.
77. Rohr MK, Mikuls TR, Cohen SB, Thorne JC, O'Dell JR. Underuse of Methotrexate in the Treatment of Rheumatoid Arthritis: A National Analysis of Prescribing Practices in the US. *Arthritis Care & Research.* 2017;69(6):794-800.
78. Codreanu C, Šírová K, Jarošová K, Batalov A. Assessment of effectiveness and safety of biosimilar infliximab (CT-P13) in a real-life setting for treatment of patients with active rheumatoid arthritis or ankylosing spondylitis. *Current Medical Research and Opinion.* 2018;34(10):1763-1769.
79. Canhão H, Rodrigues AM, Gregório MJ, et al. Common Evaluations of Disease Activity in Rheumatoid Arthritis Reach Discordant Classifications across Different Populations. *Front Med (Lausanne).* 2018;5:40-40.
80. Stevenson M, Archer R, Tosh J, et al. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. *Health technology assessment (Winchester, England).* 2016;20(35):1-610.

81. Stevenson MD, Wailoo AJ, Tosh JC, et al. The Cost-effectiveness of Sequences of Biological Disease-modifying Antirheumatic Drug Treatment in England for Patients with Rheumatoid Arthritis Who Can Tolerate Methotrexate. *The Journal of rheumatology.* 2017;44(7):973-980.
82. Incerti D, Jansen J. *A Description of the IVI-RA Model.* Innovation and Value Initiative;2017.
83. Gibson L, Alava MH, Wailoo A. *Progression of Disease in People with Rheumatoid Arthritis Treated with Non-Biologic Therapies: Report by the Decision Support Unit.* School of Health and Related Research, University of Sheffield;2015.
84. Wolfe F, Michaud K. The loss of health status in rheumatoid arthritis and the effect of biologic therapy: a longitudinal observational study. *Arthritis research & therapy.* 2010;12(2):R35-R35.
85. Strand V, Miller P, Williams SA, Saunders K, Grant S, Kremer J. Discontinuation of Biologic Therapy in Rheumatoid Arthritis: Analysis from the Corrona RA Registry. *Rheumatology and therapy.* 2017;4(2):489-502.
86. Zhang J, Shan Y, Reed G, et al. Thresholds in disease activity for switching biologics in rheumatoid arthritis patients: experience from a large U.S. cohort. *Arthritis care & research.* 2011;63(12):1672-1679.
87. Human Mortality Database. 2016. https://usa.mortality.org/. Accessed 07/15/2018.
88. Wolfe F, Michaud K, Gefeller O, Choi HK. Predicting mortality in patients with rheumatoid arthritis. *Arthritis and rheumatism.* 2003;48(6):1530-1542.
89. Singh JA, Wells GA, Christensen R, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. *The Cochrane database of systematic reviews.* 2011(2):Cd008794.
90. Wailoo AJ, Bansback N, Brennan A, Michaud K, Nixon RM, Wolfe F. Biologic drugs for rheumatoid arthritis in the Medicare program: a cost-effectiveness analysis. *Arthritis and rheumatism.* 2008;58(4):939-946.
91. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical care.* 2005;43(3):203-220.
92. Hernandez Alava M, Wailoo A, Wolfe F, Michaud K. The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. *Rheumatology (Oxford, England).* 2013;52(5):944-950.
93. SSR Health L. Data on File. In:2019.
94. Redbook. 2019. Accessed July 28, 2019.
95. Medicare National HCPCS Aggregate Summary Table CY2016. 2016. https://data.cms.gov/Medicare-Physician-Supplier/Medicare-National-HCPCS-Aggregate-Summary-Table-CY/jtra-d83c/data.
96. Kaiser Family Foundation. 2018 Employer Health Benefits Survey. 2018.
97. Medicare Provider Utilization and Payment Data: Physician and Other Supplier. 2016. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/medicare-provider-charge-data/physician-and-other-supplier.html.
98. National Economic Accounts. 2018. https://www.bea.gov/national/#gdp. Accessed August 10, 2018.
99. Han C, Li N, Peterson S. Minimal Important Difference in HAQ: A Validation from Health Economic Perspectives in Patient with Rheumatoid Arthritis Using Real-World Data. *ACR.* 2015.
100. National Institute for Health and Care Excellence. Adalimumab, etanercept, inflfliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed. 2016.
101. Altawil R, Saevarsdottir S, Wedren S, Alfredsson L, Klareskog L, Lampa J. Remaining Pain in Early Rheumatoid Arthritis Patients Treated With Methotrexate. *Arthritis care & research.* 2016;68(8):1061-1068.

102. Vashisht P, Sayles H, Cannella AC, Mikuls TR, Michaud K. Generalizability of Patients With Rheumatoid Arthritis in Biologic Agent Clinical Trials. *Arthritis care & research.* 2016;68(10):1478-1488.
103. Pearson SD. The ICER Value Framework: Integrating Cost Effectiveness and Affordability in the Assessment of Health Care Value. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2018;21(3):258-265.
104. Nayak RK, Pearson SD. The ethics of 'fail first': guidelines and practical scenarios for step therapy coverage policies. *Health affairs (Project Hope).* 2014;33(10):1779-1785.
105. Excellence NIfHaC. *Tofacitinib for Moderate to Severe Rheumatoid Arthritis.* 2017.
106. Excellence NIfHaC. *Baricitinib for Moderate to Severe Rheumatoid Arthritis.* 2017.
107. CADTH. *Clinical Review Report -Tofacitnib (Xeljanz).* Canadian Agency for Drugs and Technologies in Health (CADTH);2019.
108. Agency for Healthcare Research and Quality. U.S. Preventive Services Task Force Procedure Manual. 2008.
109. Genovese M, Takeuchi TXLIMPCARTEKSJ. Baricitinib dose step-down following disease control in patients with rheumatoid arthritis. *Internal medicine journal.* 2017;47:30.
110. Van Der Heijde D, Schiff M, Tanaka Y, et al. Low rates of radiographic progression of structural joint damage over 2 years of baricitinib treatment in patients with rheumatoid arthritis. *RMD Open.* 2019;5(1).
111. Keystone EC, Taylor PC, Drescher E, et al. Safety and efficacy of baricitinib at 24 weeks in patients with rheumatoid arthritis who have had an inadequate response to methotrexate. *Annals of the rheumatic diseases.* 2015;74(2):333-340.
112. Strand V, Kremer JM, Gruben D, Krishnaswami S, Zwillich SH, Wallenstein GV. Tofacitinib in Combination With Conventional Disease-Modifying Antirheumatic Drugs in Patients With Active Rheumatoid Arthritis: Patient-Reported Outcomes From a Phase III Randomized Controlled Trial. *Arthritis care & research.* 2017;69(4):592-598.
113. Strand V, Burmester GR, Zerbini CA, et al. Tofacitinib with methotrexate in third-line treatment of patients with active rheumatoid arthritis: patient-reported outcomes from a phase III trial. *Arthritis care & research.* 2015;67(4):475-483.
114. Strand V, Mysler EMRJWGDRLZSKINFR. Tofacitinib with and without methotrexate versus adalimumab with methotrexate for the treatment of rheumatoid arthritis: patient-reported outcomes from a phase 3b/4 randomized trial. *Arthritis and rheumatology Conference: american college of rheumatology/association of rheumatology health professionals annual scientific meeting, ACR/ARHP.* 2017;69(Supplement 10) (no pagination).
115. Strand V, Kosinski M, Chen CI, et al. Sarilumab plus methotrexate improves patient-reported outcomes in patients with active rheumatoid arthritis and inadequate responses to methotrexate: results of a phase III trial. *Arthritis research & therapy.* 2016;18:198.
116. Nakamura Y, Suzuki T, Yamazaki H, Kato H. Tofacitinib versus non-tumor necrosis factor biologics for patients with active rheumatoid arthritis. *Archives of Rheumatology.* 2018;33(2):154-159.
117. Kremer JM, Cohen S, Wilkinson BE, et al. A phase IIb dose-ranging study of the oral JAK inhibitor tofacitinib (CP-690,550) versus placebo in combination with background methotrexate in patients with active rheumatoid arthritis and an inadequate response to methotrexate alone. *Arthritis and rheumatism.* 2012;64(4):970-981.
118. Smolen JS, Pangan AL, Emery P, et al. Upadacitinib as monotherapy in patients with active rheumatoid arthritis and inadequate response to methotrexate (SELECT-MONOTHERAPY): a randomised, placebo-controlled, double-blind phase 3 study. *The Lancet.* 2019;393(10188):2303-2311.

119. Strand V, Pope JTNFACHSPAGAFMGDSM. Upadacitinib improves patient-reported outcomes in patients with rheumatoid arthritis and inadequate response to conventional synthetic disease-modifying anti-rheumatic drugs: results from selectnext. *Annals of the rheumatic diseases.* 2018;77:989.
120. van Vollenhoven RF, Emery P, Bingham CO, 3rd, et al. Long-term safety of rituximab in rheumatoid arthritis: 9.5-year follow-up of the global clinical trial programme with a focus on adverse events of interest in RA patients. *Annals of the rheumatic diseases.* 2013;72(9):1496-1502.
121. Emery P, Blanco R, Maldonado Cocco J, et al. Patient-reported outcomes from a phase III study of baricitinib in patients with conventional synthetic DMARD-refractory rheumatoid arthritis. *Rheumatic and Musculoskeletal Diseases.* 2017;3(1):e000410.
122. Smolen JS, Kremer JM, Gaich CL, et al. Patient-reported outcomes from a randomised phase III study of baricitinib in patients with rheumatoid arthritis and an inadequate response to biological agents (RA-BEACON). *Annals of the rheumatic diseases.* 2017;76(4):694-700.
123. Yoo DH, Hrycaj P, Miranda P, et al. A randomised, double-blind, parallel-group study to demonstrate equivalence in efficacy and safety of CT-P13 compared with innovator infliximab when coadministered with methotrexate in patients with active rheumatoid arthritis: the PLANETRA study. *Annals of the rheumatic diseases.* 2013;72(10):1613-1620.
124. Yoo DH, Racewicz A, Brzezicki J, et al. A phase III randomized study to evaluate the efficacy and safety of CT-P13 compared with reference infliximab in patients with active rheumatoid arthritis: 54-week results from the PLANETRA study. *Arthritis research & therapy.* 2016;18:82.
125. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *Jama.* 2016;316(10):1093-1103.
126. Pickard AS, Law EH, Jiang R, et al. United States Valuation of EQ-5D-5L Health States Using an International Protocol. *Value in Health.* 2019;22(8):931-941.

APPENDICES

Appendix A. Search Strategies and Results

Table A1. PRISMA 2009 Checklist

	#	Checklist Item
TITLE		
Title	1	Identify the report as a systematic review, meta-analysis, or both.
ABSTRACT		
Structured Summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
INTRODUCTION		
Rationale	3	Describe the rationale for the review in the context of what is already known.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
METHODS		
Protocol and Registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.
Eligibility Criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.
Information Sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study Selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data Collection Process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data Items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of Bias in Individual Studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary Measures	13	State the principal summary measures (e.g., risk ratio, difference in means).
Synthesis of Results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.

Risk of Bias across Studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional Analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.
RESULTS		
Study Selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study Characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.
Risk of Bias within Studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).
Results of Individual Studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of Results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
Risk of Bias across Studies	22	Present results of any assessment of risk of bias across studies (see Item 15).
Additional Analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).
DISCUSSION		
Summary of Evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
FUNDING		
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

From: Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

Search Strategies for RA

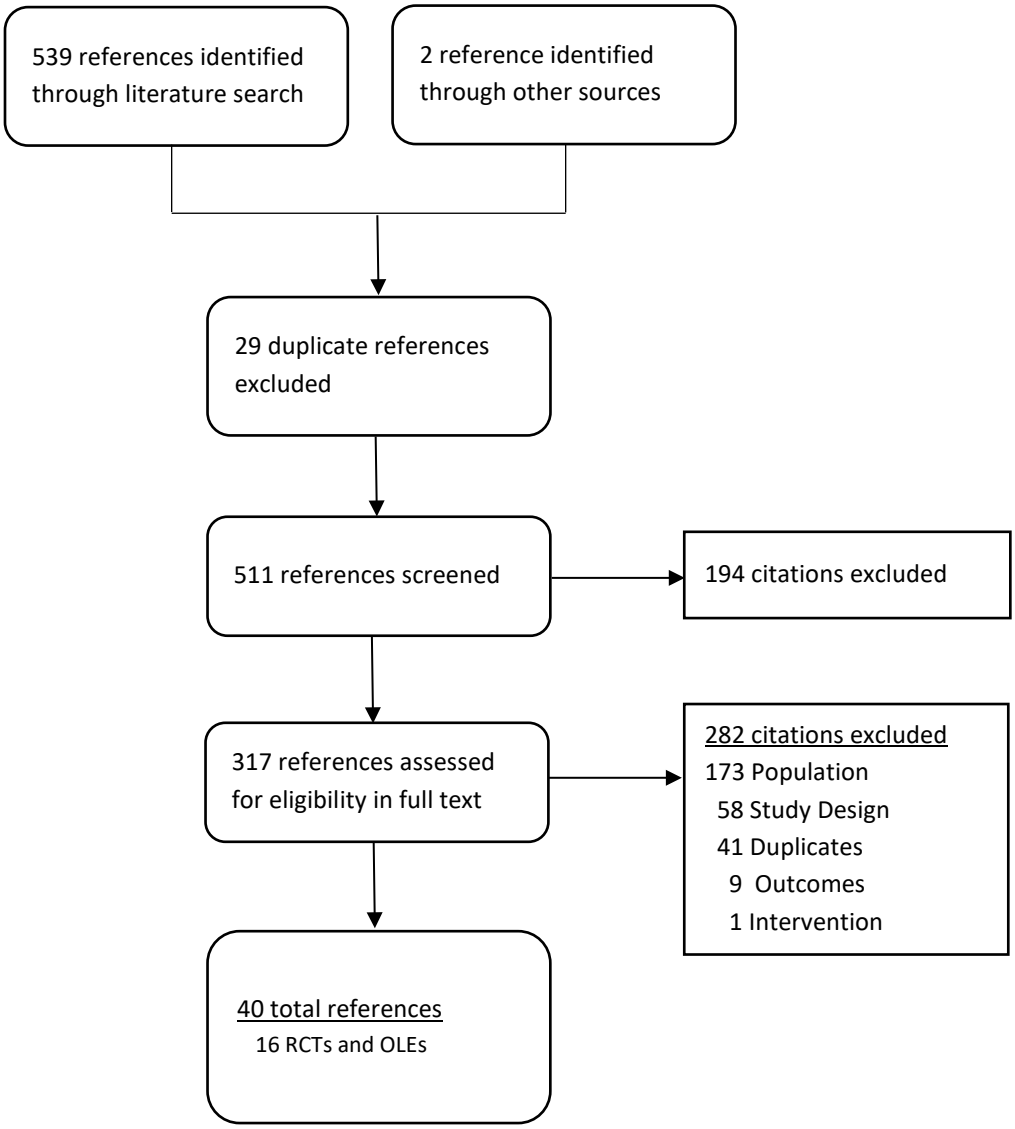
Table A2. Search Strategy of Medline 1996 to Present with Daily Update and Cochrane Central Register of Controlled Trials

No.	Search Terms
#1	((rheumatoid or rheumatic or rheumat$) adj3 (arthrit$ or diseas$ or condition$)).ti,ab.
#2	(tofacitinib or tasocitinib or tofacitinib citrate or Xeljanz).ti,ab
#3	(baricitinib or LY3009104 or INCB028050).ti,ab
#4	(upadacitinib or ABT-494).ti,ab.
#5	(infliximab-dyyb or infliximab dyyb or inflectra or CT-P13 or CT P13).ti,ab
#6	2 or 3 or 4 or 5
#7	1 and 6
#8	(abstract or addresses or autobiography or bibliography or biography or clinical trial, phase I or case report or comment or congresses or consensus development conference or duplicate publication or editorial or guideline or in vitro or interview or lecture or legal cases or legislation or letter or news or newspaper article or patient education handout or periodical index or personal narratives or portraits or practice guideline or review or videoaudio media).pt
#9	7 not 8
#10	(animals not (humans and animals)).sh.
#11	9 not 10
#12	cohort studies/ or longitudinal studies/ or prospective studies/ or retrospective studies/ or comparative study.pt
#13	control Groups/ or (control* adj2 (clinical or group* or trial* or study or studies or design* or arm*)).ti,ab. or ("clinical trial" or "clinical trial, phase ii" or clinical trial, phase iii or clinical trial, phase iv or controlled clinical trial or "multicenter study" or "randomized controlled trial").pt. or (randomi?ed adj6 (study or trial* or (clinical adj2 trial*))).ti,ab. or ((single or doubl*) adj2 blind*).ti,ab.
#14	12 or 13
#15	11 and 14
#16	remove duplicates from 15

Table A3. Search Strategy of EMBASE SEARCH

No.	Search Terms
#1	((rheumatoid OR rheumatic OR rheumat*) NEAR/3 (arthrit* OR diseas* OR condition*)):ab,ti
#2	'tofacitinib'/exp OR tofacitinib:ab,ti OR tasocitinib:ab,ti OR 'tofacitinib citrate':ab,ti OR xeljanz:ab,ti
#3	'baricitinib'/exp OR baricitinib:ab,ti
#4	(‘upadacitinib’ OR ‘ABT-494’):ab,ti
#5	‘inflectra’:ab,ti OR ‘infliximab-dyyb’:ab,ti OR ‘CT-P13’:ab,ti
#6	#2 OR #3 OR #4 OR #5
#7	#1 AND #6
#8	#7 AND ('chapter'/it OR 'conference abstract'/it OR 'conference paper'/it OR 'conference review'/it OR 'editorial'/it OR 'letter'/it OR 'note'/it OR 'review'/it OR 'short survey'/it)
#9	#7 NOT #8
#10	#9 AND [english]/lim
#11	#10 AND [medline]/lim
#12	#10 NOT #11

Figure A1. PRISMA Flow Chart Showing Results of Literature Search for JAKs for RA



Appendix B. Previous Systematic Reviews and Technology Assessments

We identified three completed technology assessments of JAK inhibitors for RA, summarized below: two from the National Institute for Health and Care Excellence (NICE), assessing tofacitinib and baricitinib, and one from the Canadian Agency for Drugs and Technologies in Health (CADTH) accessing tofacitinib. We also identified three ongoing technology assessments: one from NICE for upadacitinib and two from CADTH for baricitinib and upadacitinib.

***NICE: Tofacitinib for Moderate to Severe RA*[105]**

The National Institute for Health and Care Excellence (NICE) recommends tofacitinib, with methotrexate, for treating RA, if the disease is severe (a DAS28 score of more than 5.1) and if the patient has responded inadequately to or cannot tolerate DMARDs, including at least one biological DMARD, or they cannot have rituximab. Further, NICE recommends tofacitinib to be used as monotherapy in adults who cannot take methotrexate. Continued use is only recommended if there is a moderate response using the EULAR criteria, six months after starting therapy. Treatment should be withdrawn if EULAR response is not maintained.

***NICE: Baricitinib for Moderate to Severe RA*[106]**

NICE recommends baricitinib, with methotrexate, as an option for treating RA in lieu of DMARDs, if disease is severe (a DAS28 score of more than 5.1), or the patient cannot tolerate rituximab, only if they have responded inadequately to intensive therapy with a combination of conventional DMARDs. Baricitinib can also be used as monotherapy for patients who cannot take methotrexate.

***CADTH (2014). Tofacitinib (Xeljanz) Clinical Review Report. CADTH Clinical Review Report.*[107]**

CADTH conducted a systematic review of tofacitinib for the treatment of RA. Five RCTs met the inclusion criteria and evaluated the efficacy and safety of 5 mg twice daily or 10 mg twice daily tofacitinib, with one study also evaluating adalimumab versus placebo. Studies evaluated patients who had experienced an inadequate response to TNF inhibitors, non-biologic DMARDs, and/or methotrexate. There were also included studies that evaluated tofacitinib as monotherapy and others, with patients on a background treatment of DMARDs. All five trials assessed the same endpoints including: ACR20, HAQ-DI, and DAS28. CADTH recommends tofacitinib, in combination with methotrexate, be considered as an option for the treatment of RA in patients with moderate to severe disease or as monotherapy in those who are intolerant to methotrexate treatment. Tofacitinib was found to have a safety profile similar to that of biologic DMARDs and the economic analysis concluded that treatment with tofacitinib is more costly than treatment with infliximab, tocilizumab (IV), or tocilizumab (subcutaneous).

NICE: Upadacitinib for Treating Moderate to Severe Rheumatoid Arthritis [ID1400]. Expected publication date 18 March 2020

NICE is currently evaluating the clinical and cost effectiveness of upadacitinib for treating moderate to severe RA. Proposed comparators include combination therapy with conventional DMARDs, conventional DMARDS with dose escalation, and best supportive care for patients who have not responded well to therapy with conventional DMARDs. For patients with severe RA who have not responded well to conventional DMARDS only, comparators include biological DMARDs or JAK inhibitors (baricitinib and tofacitinib). For patients with severe RA who have not responded well to either DMARDs or at least one TNF inhibitor, comparators include rituximab with methotrexate and TNF inhibitors.

Canadian Agency for Drugs and Technologies in Health (CADTH) Common Drug Review: Baricitinib

CADTH recommends reimbursement for baricitinib as monotherapy or in combination with methotrexate (with or without additional conventional DMARDs) for the treatment of adult patients with moderately to severely active RA who have responded inadequately to one or more DMARDs. CADTH requires that patients discontinue therapy if a treatment response (i.e., achievement of ACR20) is not achieved by 12 weeks. Patients should be receiving care from a rheumatologist and their daily dose of baricitinib should be limited to 2 mg. CADTH has placed a pricing condition on baricitinib that stipulates that the drug plan cost of treatment with baricitinib should result in cost-savings compared with the drug plan cost of treatment with the least costly alternative biologic DMARD.

Canadian Agency for Drugs and Technologies in Health (CADTH) Common Drug Review: Upadacitinib [SR0614-000], Expected publication date December 2019

CADTH is currently evaluating the clinical and cost effectiveness of upadacitinib for the treatment of moderate to severe RA.

Appendix C. Ongoing Studies

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Baricitinib (BAR)					
A Study of Baricitinib (LY3009104) in Participants With Rheumatoid Arthritis (RA-BRIDGE) NCT03915964 Eli Lilly and Company	Phase IV, randomized, parallel assignment, open label Enrolled: 2,600	Intervention: • BAR low dose • BAR high dose Comparator: • TNFi	Inclusions: • Documented evidence of VTE prior to study • At least age 60 • Body mass index (BMI) ≥30 • Age 50 to less than 60 years old • BMI 25 to less than 30 Exclusions: • Previous TNFi use • Pregnant or breastfeeding • Multiple VTE • Cancer • Herpes zoster, serious infection, active TB • A live vaccine within 4 weeks of start • Participated in a clinical trial within 4 weeks of start • History of IV drug use, or any other illicit drug use	Time from first dose of study treatment to first event of VTE	February 1, 2026
An Extension Study in Participants with Moderate to Severe Rheumatoid Arthritis (RA-BEYOND) NCT01885078 Eli Lilly and Company	Phase III, multicenter, randomized, parallel assignment, quadruple masking Enrolled: 2,944	Intervention: • BAR 2 mg • BAR 4 mg Comparator: Placebo	Inclusions: • Have completed the final active treatment in study JADV Exclusions: • Have significant uncontrolled cerebro-cardiovascular, respiratory, hepatic, renal, GI, endocrine, hematologic, neuropsychiatric disorders, or abnormal laboratory values • Have a known hypersensitivity to BAR or any component of this investigational product • Had investigational product permanently discontinued at any time during a previous BAR study	Number of participants with one or more drug related adverse events (AEs) or serious adverse events (SAEs)	March 22, 2024

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
			• Had temporary investigational product interruption at the final study visit		
Examination of Efficacy and Safety of Baricitinib in RA Patients NCT03755466 Shinshu University	Phase II, non-randomized, paralleled assignment Enrolled: 90	Intervention: BAR Comparator: • TOF • bDMARDs	Inclusions: • RA patients Exclusions: • RA patients who are allergic to the interventions • Refused study guidelines • Pregnant	1) Assessment of disease activity in RA patients for 1 year treated by BAR, bDMARDs, or TOF 2) The efficacy and AEs of each drug for 1 year in the RA patients	November 20, 2023
Tofacitinib					
Efficacy and Safety of GSK3196165 Versus Placebo and Tofacitinib in Participants With Moderately to Severely Active Rheumatoid Arthritis Who Have an Inadequate Response to Methotrexate NCT03980483 GlaxoSmithKline	52-week, Phase III, randomized, double-blind Enrolled: 1,500	Interventions: • GSK319165 • TOF 5 mg • Comparator: Placebo to TOF • cDMARD	Inclusions: • ≥18 years of age • Has had RA for ≥6 months and was not diagnosed before 16 years of age Has active disease, as defined by having both: • ≥6/68 tender/painful joints, and • ≥6/66 swollen joints • Has at least 1 bone erosion present on hand/wrist or foot radiographs • Has had an inadequate response to cDMARD, despite currently taking cDMARD 15-25 mg/week oral or injected Exclusions: • Has had any active and/or recurrent infections (excluding recurrent fungal infections of the nail bed) or has required management of acute or chronic infections.	Proportion of participants achieving ACR20 at week 12: superiority comparison with placebo	July 3, 2021

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
			• Has received prior treatment with an antagonist of GM-CSF or its receptor or targeted cDMARDs • Has received prior treatment with a bDMARD that was discontinued due to an inadequate response		
Efficacy and Safety of GSK3196165 Versus Placebo and Tofacitinib in Participants With Moderately to Severely Active Rheumatoid Arthritis Who Have an Inadequate Response to Conventional Synthetic (cs)/Biologic (b) Disease Modifying Anti-rheumatic Drugs (DMARDs) NCT03970837 GlaxoSmithKline	52-week, Phase III, randomized, double-blind Enrolled: 1,500	Intervention: • GSK319165 • TOF 5 mg Comparator: • Placebo to GSK31916 • Placebo to TOF • cDMARDs	Inclusions: • ≥18 years of age • Has had RA for ≥6 months and was not diagnosed before 16 years of age • Have active disease Exclusions: • History of other inflammatory rheumatologic or systemic autoimmune disorder, other than Sjogren's syndrome secondary to RA, that may confound the evaluation of the effect of the study intervention • Has had any active and/or recurrent infections or has required management of acute or chronic infections • Has received prior treatment with an antagonist of GM-CSF or its receptor or targeted cDMARDs	Proportion of participants achieving ACR20 at week 12: superiority comparison with placebo	July 2, 2021
Upadacitinib (UPA)					
A Study To Investigate The Safety and Efficacy of ABBV-105 Alone or in Combination With Upadacitinib (ABBV-599 Combination) in Participants With	Phase II, randomized, parallel assignment, quadruple masking Enrolled: 240	Intervention: • UPA • ABBV-105 Comparator: Placebo	Inclusion • Diagnosis of RA for ≥3 months based on 2010 ACR/EULAR classification criteria • ≥ 6 swollen joints and ≥6 tender joints • HsCRP ≥3mg/L • Treated for ≥3 months with ≥1 bDMARD therapy but continue to exhibit active RA	Change from baseline in DAS28 and CRP at 12 weeks	January 15, 2020

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Active Rheumatoid Arthritis NCT03682705 AbbVie			• Receiving cDMARD therapy ≥3 months and on stable dose for ≥4 weeks prior to the first dose of study drug • Participants must have discontinued all bDMARDs prior to first dose of study drug Exclusion: • Participants had prior exposure to JAK inhibitor for greater than two weeks including but not limited to UPA, TOF, BAR, and filgotinib • A washout period of ≥30 days is required for any JAK inhibitor prior to the first dose of study drug		
A Study with Upadacitinib (ABT-494) in Subjects from Chine and Selected Countries with Moderately to Severely Active Rheumatoid Arthritis Who Have Had an Inadequate Response to Conventional Synthetic Disease Modifying Anti-Rheumatic Drugs (cDMARDs) NCT02955212 AbbVie	Phase III, randomized, parallel assignment, quadruple masking Enrolled: 450	Intervention: UPA Comparator: Placebo	Inclusions: • Diagnosis of RA for ≥3 months who also fulfilled the 2010 ACR/EULAR classification criteria for RA • Participants had been receiving cDMARD therapy ≥3 months and on a stable dose for ≥4 weeks prior to first dose of study drug • Participants with prior exposure to at most a bDMARD may be enrolled (up to 20%) • Participants must have discontinued bDMARD therapy prior to the first dose of study drug Participant meet both of the following disease criteria: 1. ≥6 swollen joints and ≥76 tender joints at screening and baseline visits and 2. HsCRP ≥ upper limit of normal at screening visit Exclusions: • Prior exposure to any JAK inhibitor • Participants who are considered inadequate responders to bDMARD therapy • History of arthritis with onset prior to age 17 years or current diagnosis of inflammatory joint disease other than RA	Proportion of participants achieving ACR 20 response	September 10, 2020

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
A Phase 3 Study to Compare Upadacitinib to Abatacept in Subjects with Rheumatoid Arthritis on Stable Dose of Conventional Synthetic Disease-Modifying Antirheumatic Drugs (cDMARDs) Who Have an Inadequate Response of Intolerance to Biologic DMARDs (SELECT-CHOICE) NCT03086343 AbbVie	Phase III, randomized, active-controlled, quadruple masking Enrolled: 614	Intervention: UPA Active Comparator: Abatacept Comparator: Placebo	Inclusions: • Diagnosis of RA • Participants have been treated for ≥3 months with ≥1 bDMARD therapy but continue to exhibit RA or had to discontinue due to intolerability or toxicity • Participants have been receiving cDMARD therapy ≥3 months and on a stable dose for ≥4 weeks prior to first dose of study drug Exclusions: • Prior exposure to JAK inhibitors • Prior exposure to abatacept • History of arthritis with onset prior to age 17 years or current diagnosis of inflammatory joint disease other than RA	Change in DAS28 CRP (non-inferiority)	February 29, 2020
A Phase 2b/3, Randomized, Double-Blind Study Comparing ABT-494 to Placebo in Japanese Subjects With Moderately to Severely Active Rheumatoid Arthritis Who Are on a Stable Dose of Conventional Synthetic Disease-Modifying Anti-	Phase II, randomized, quadruple masking Enrolled: 197	Intervention: ABT-494 (UPA) Comparator: Placebo	Inclusions: • Diagnosis of RA for ≥3 months who also fulfill the 2010 ACR/EULAR classification criteria for RA • Subjects have been receiving cDMARD therapy ≥3 months and on a stable dose for ≥4 weeks prior to the first dose of study drug • Subject has ≥6/66 swollen joints and ≥6/68 tender joints at screening and baseline visits • Subjects with prior exposure to at most bDMARD may be enrolled (up to 20% of total number of subjects) after the required washout period.	Proportion of subjects achieving ACR 20 response at Week 12	July 19, 2020

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Rheumatic Drugs (cDMARDs) and Have an Inadequate Response to cDMARDs (SELECT-SUNRISE) NCT02720523 AbbVie			Exclusions: • Prior exposure to any JAK inhibitor • Subjects who are considered inadequate responders (lack of efficacy) to bDMARD therapy, after minimum 3 months treatment, as determined by the investigator • History of any arthritis with onset prior to age 17 years or current diagnosis of inflammatory joint disease other than RA		
Infliximab-dyyb					
National Observational Study On The Use Of Inflectra An Infliximab Biosimilar In Real Life (ReFLECT) NCT02925338 Pfizer	Observational, cohort, prospective Enrolled: 1,200	QOL questionnaire	Inclusions: • Adult patients treated with Inflectra regardless of treatment phase in Crohn's disease, ulcerative colitis, RA, ankylosing spondylitis, or psoriatic arthritis • Pediatric patients treated with Inflectra regardless of treatment phase time Exclusions: • Patients who refused to access their medical file for collection of their medical data • Patients treated with Inflectra for psoriasis • Patients with past history of hypersensitivity to infliximab • Patients with tuberculosis or any other severe infection such as sepsis, abscess, or opportunistic infection • Patients with moderate to severe heart failure		September 23, 2020

Source: www.ClinicalTrials.gov (NOTE: studies listed on site include both clinical trials and observational studies)

ACR: American College of Rheumatology, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, CRP: C-reactive protein, DAS: Disease Activity Score 28, EULAR: European League Against Rheumatology, IV: intravenous, JAK: Janus kinase, TB: tuberculosis, TNFi: tumor necrosis factor inhibitor, VTE: venous thromboembolism

Appendix D. Comparative Clinical Effectiveness Supplemental Information

We performed screening at both the abstract and full-text level. A single investigator screened all abstracts identified through electronic searches according to the inclusion and exclusion criteria described earlier. We did not exclude any study at abstract-level screening due to insufficient information. For example, an abstract that did not report an outcome of interest would be accepted for further review in full text. We retrieved the citations that were accepted during abstract-level screening for full text appraisal. One investigator reviewed full papers and provided justification for exclusion of each excluded study.

We also included FDA documents related to baricitinib, tofacitinib, upadacitinib, and infliximab-dyyb. These included the prescribing information, manufacturer's submission to the agency, internal FDA review documents, and the transcript of Advisory Committee deliberations and discussions. All literature that did not undergo a formal peer review process is described separately.

We used criteria published by the USPSTF to assess the quality of RCTs and comparative cohort studies, using the categories "good," "fair," or "poor" (see Appendix Table F2)[108] Guidance for quality ratings using these criteria is presented below, as is a description of any modifications we made to these ratings specific to the purposes of this review.

Good: *Meets all criteria: Comparable groups are assembled initially and maintained throughout the study; reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention is paid to confounders in analysis. In addition, intention to treat analysis is used for RCTs.*

Fair: *Studies were graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are addressed. Intention to treat analysis is done for RCTs.*

Poor: *Studies were graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat analysis is lacking.*

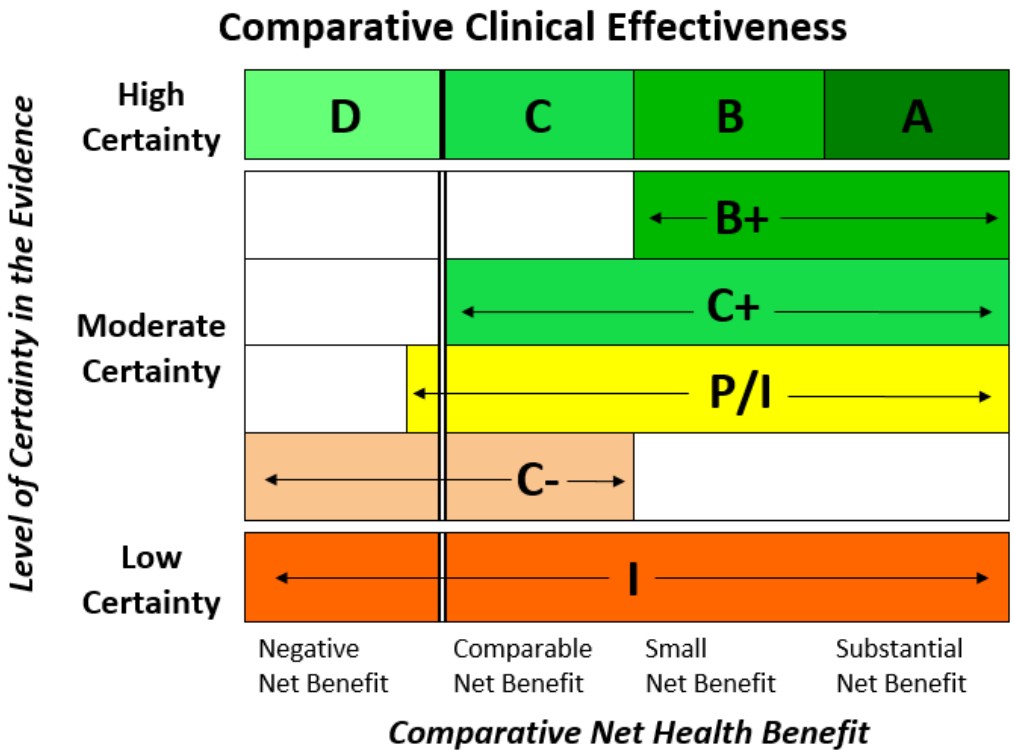
Note that case series are not considered under this rating system – because of the lack of comparator, these are generally considered to be of poor quality.

ICER Evidence Rating

We used the ICER Evidence Rating Matrix (see Figure D1) to evaluate the evidence for a variety of outcomes. The evidence rating reflects a joint judgment of two critical components:

a) The **magnitude** of the difference between a therapeutic agent and its comparator in "net health benefit" – the balance between clinical benefits and risks and/or adverse effects AND
b) The level of **certainty** in the best point estimate of net health benefit.[62]

Figure D1. ICER Evidence Rating Matrix



A = "Superior" *- High certainty of a substantial (moderate-large) net health benefit*
B = "Incremental" *- High certainty of a small net health benefit*
C = "Comparable" *- High certainty of a comparable net health benefit*
D = "Negative" *- High certainty of an inferior net health benefit*
B+ = "Incremental or Better" *- Moderate certainty of a small or substantial net health benefit, with high certainty of at least a small net health benefit*
C+ = "Comparable or Better" *- Moderate certainty of a comparable, small, or substantial net health benefit, with high certainty of at least a comparable net health benefit*
P/I = "Promising but Inconclusive" *- Moderate certainty of a comparable, small, or substantial net health benefit, and a small (but nonzero) likelihood of a negative net health benefit*
C- = "Comparable or Inferior" *- Moderate certainty that the point estimate for comparative net health benefit is either comparable or inferior*
I = "Insufficient" *- Any situation in which the level of certainty in the evidence is low*

Table D1. Included Studies

Trial Name (Author & Year of Publication)	Study Sponsor	Study Design and Duration of Follow-Up	Geographic Location of Study	Interventions (n) & Dosing Schedule	Inclusion and Exclusion Criteria	Primary Outcomes
Baricitinib (BAR)						
RA-BUILD[18] **Dougados 2017**	Eli Lilly & Co.	RCT, Phase III, 12-week follow-up	182 centers in 22 countries	1) BAR 2 mg + cDMARD 2) cDMARD	Concomitant use of cDMARDs, non-steroidal and anti-inflammatory drugs and/or corticosteroids were permitted	ACR20 response at 12 weeks
RA-BEACON[15] **Genovese 2016**	Eli Lilly & Co.	RCT, Phase III, 24-week follow-up	178 centers in 24 countries	1) BAR 2 mg + cDMARD 2) cDMARD	Patients who had received other bDMARDs could participate if bDMARDs were discontinued at least 4 weeks before randomization	ACR20 response at 12 weeks
RA-BEYOND[109,110] **(Abstract) Genovese 2017 Van Der Heijde 2019**	Eli Lilly & Co.	Phase III, multicenter, open-label extension, duration up to 7 years	Multicenter	1) Continued BAR 4 mg 2) Stepped down to BAR 2 mg	Patients were eligible for RA-BEYOND if they had completed one of the originating studies (RA-BUILD, RA-BEGIN, RA-BEAM). Patients were not eligible if they were hypersensitive to BAR or permanently discontinued during the originating study	Disease activity at 12 weeks
Keystone 2015[111]	Eli Lilly & Co.	RCT, Phase IIb, 24-week follow-up	69 centers in 9 countries	1) BAR 2 mg + cDMARD 2) cDMARD	Moderately to severely active RA patients were excluded if they had previously used bDMARDs	ACR20 response at 12 weeks
Tofacitinib (TOF)						
ORAL Sync[8,112] **Kremer 2013 Strand 2017**	Pfizer	RCT, Phase III, double-blind, 1-year trial duration	114 centers in 19 countries	1) TOF 5 mg 2) Placebo (advanced to TOF after 6 months)	Inclusions: active RA diagnosis, an ESR ≥22mm/h or a CRP ≥66.7 nmol/L Exclusions: previous treatment with lymphocyte-depleting therapies within 1 year of randomization or alkylating agents at any time	• ACR20 response at 24 weeks • DAS28-4 (ESR) defined remission ≤2.6 • Change in HAQ-DI score • Safety assessments

Trial Name (Author & Year of Publication)	Study Sponsor	Study Design and Duration of Follow-Up	Geographic Location of Study	Interventions (n) & Dosing Schedule	Inclusion and Exclusion Criteria	Primary Outcomes
ORAL Step[14,113] **Burmester 2013 Strand 2015**	Pfizer	Phase III, double-blind, parallel group study, 24-week follow-up	82 centers in 19 countries	1) TOF 5 mg 2) Placebo (advanced to TOF after 3 months)	Inclusions: ≥18 years with diagnosis of active moderate-to-severe RA with previous inadequate response or intolerance to one or more approved TNFi Exclusions: hemoglobin ≤90.0 g/L	• ACR20 response rate • Mean change from baseline HAQ-DI • Rates of DAS28-4 (ESR) ≤2.6 (all at 12 weeks)
ORAL Scan[11] **Van der Heijde 2013**	Pfizer	RCT, Phase III, double blind, parallel-group, placebo-controlled,12-month interim analysis, 24-month follow up	111 centers in 5 continents	1) TOF 5 mg + cDMARD 2) Placebo (advanced to TOF + cDMARD after 6 months)	Inclusions: ≥18 years with active RA diagnosis Exclusions: hemoglobin <9.0 gm/dl and hematocrit < 30%, white blood cell count <3.0 x 10^9/liter, absolute neutrophil count <1.2 x 10^9/liter, or platelet count <100 x 109/liter	• ACR20 response rate at 24 weeks • Mean change from base in SHS score at 24 weeks • Mean change from baseline in HAQ-DI score at 12 weeks • Rates of remission (DAS28-ESR ≤2.6) at 24 weeks
ORAL Strategy[10,114] **Fleischmann 2017 Strand 2017**	Pfizer	RCT, Phase IIIb/4, head-to-head, non-inferiority, 1-year duration	194 centers in 25 countries	1) TOF 5 mg 2) TOF 5 mg + cDMARD 3) ADA 40 mg + cDMARD	Inclusions: ≥18 years with active RA despite methotrexate therapy Exclusions: patient had previous treatment with TNFi, contraindications for study treatment, history of infections requiring treatment within 2 weeks	ACR50 response rate at 24 weeks
ORAL Standard[9,115] **Van Vollenhoven 2013 Strand 2016**	Pfizer	RCT, Phase III, 12-month follow-up,	multicenter	1) TOF 5 mg + cDMARD 2) ADA 40 mg 3) Placebo followed by TOF 5 mg + cDMARD	Active RA and inadequate response to cDMARD. Excluded if treated with adalimumab previously	• ACR 20 response at 24 weeks • Disease activity at 12 weeks

Trial Name (Author & Year of Publication)	Study Sponsor	Study Design and Duration of Follow-Up	Geographic Location of Study	Interventions (n) & Dosing Schedule	Inclusion and Exclusion Criteria	Primary Outcomes
						• Proportions achieving DAS28-CRP ≤2.6 at 24 weeks
Nakamura 2018[116]	No financial support for research and/or authorship	Prospective, randomized study, 1-year follow-up	Japan	1) TOF 5 mg 2) non-TNF biologics	Moderately to severely active RA and inadequate response to bDMARD or methotrexate	• Disease status at 12 months • Percent change of DAS28-CRP from baseline to 12 months
Charles-Schoeman 2016[12]	Pfizer	Pooled analysis of Phase II and III studies	N/A	<u>bDMARD Naïve</u> 1) TOF 5 mg 2) cDMARD <u>bDMARD-IR</u> 3) TOF 5 mg 4) cDMARD	<u>Phase II studies</u> Patients had to be IR to a bDMARD or cDMARD <u>Phase III studies</u> Patients had an IR to a bDMARD or cDMARD	• ACR20, 50 or 70 response • DAS28 -ESR, CDAI, and SDAI at 3 and 6 months
Kremer 2012[117]	Pfizer	RCT, Phase IIb, double-blind, 24-weeks follow-up	multicenter	1) TOF 5 mg + cDMARD 2) cDMARD	Moderately to severely active RA and receiving methotrexate. Excluded if they had any hematopoietic disorders or comorbidities	• ACR20 response at 12 weeks • Proportion achieving DAS28-CRP <2.6 at 12 weeks
Upadacitinib (UPA)						
SELECT-MONOTHERAPY[118,119] **Smolen 2019** **Strand 2018**	AbbVie	RCT, double-blind, cDMARD arm crossed over to UPA 15mg or 30mg at 14 weeks; patients followed up to 5 years	24 countries	1) UPA 15 mg 2) UPA 30 mg 3) cDMARD	Moderate to severe active RA and inadequate response to methotrexate	• ACR20 response at 14 weeks • Proportion achieving DAS28-CRP ≤3.2 (NRI)

Trial Name (Author & Year of Publication)	Study Sponsor	Study Design and Duration of Follow-Up	Geographic Location of Study	Interventions (n) & Dosing Schedule	Inclusion and Exclusion Criteria	Primary Outcomes
SELECT-COMPARE[6] **Fleischman 2019**	AbbVie	RCT, double-blind, 48 weeks follow-up. At weeks 14, 18, and 22, patients without an improvement of ≥20% in TJC and SJC received rescue therapy, switching from PBO to UPA, UPA to ADA, or ADA to UPA	286 sites in 41 countries	1) UPA 15 mg + cDMARD 2) ADA 40 mg + cDMARD 3) Placebo + cDMARD	Moderate to severe active RA and inadequate response to methotrexate	• ACR20 response at 12 weeks • Proportion achieving DAS28-CRP <2.6 at 12 weeks
SELECT-BEYOND[13,119] **Genovese 2018** **Strand 2018**	AbbVie	RCT, double-blind, 24 weeks randomized treatment, followed by a double-blinded extension of up to 5 years	26 countries	1) UPA 15 mg + cDMARD 2) UPA 30 mg + cDMARD 3) Placebo + cDMARD	Moderate to severe active RA and inadequate response to bDMARDs	• ACR20 response at 12 weeks • Proportion achieving DAS28-CRP ≤3.2 at 12 weeks
SELECT-NEXT[7,119] **Burmester 2018** **Strand 2018**	AbbVie	RCT, Phase III, 12-week study followed by an ongoing double-blind extension of up to 5 years	35 countries	1) UPA 15 mg + cDMARD 2) UPA 30 mg + cDMARD 3) Placebo + cDMARD	Moderate to severely active RA with inadequate response to cDMARDs	• ACR20 response at 12 weeks • Proportion achieving DAS28-CRP ≤3.2 (NRI) at 12 weeks

ACR: American College of Rheumatology, ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, CDAI: clinical disease activity index, cDMARD: conventional disease-modifying antirheumatic drug, CRP: C-reactive protein, DAS: Disease Activity Score 28, ESR: erythrocyte sedimentation rate, EULAR: European League Against Rheumatology, HAQ-DI: Health Assessment Questionnaire without Disability Index, IR: inadequate response, IV: intravenous, JAK: Janus kinase, NRI: non-responder imputation, RA: rheumatoid arthritis, SDAI: simple disease activity index, SHS: Sharp score as modified by van der Heijde, SJC: swollen joint count, TJC: tender joint count, TNFi: tumor necrosis factor inhibitor

Table D2. Quality of Studies

Study	Adequate Randomization	Allocation Concealment	Patient Blinding	Staff Blinding	Outcome Adjudication Blinding	Completeness of Follow Up	Intention to Treat Analysis	Incomplete Data Addressed	Selective Outcome Reporting	Industry Funding	Freedom from Bias	Overall Quality
Baricitinib												
RA-BUILD[18] Dougados 2017	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
RA-Beacon[15] Genovese 2016	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
RA-BEYOND[109,110] Genovese 2017 Van der Heijde 2019	N/A – pooled open label extension study	N/A	No	No	No	No	No	Yes	No	Yes	No	Poor
Keystone 2015[111]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Tofacitinib												
ORAL Sync[8] Kremer 2013 Strand 2017	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
ORAL Step[14] Burmester 2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
ORAL Scan[11] Van der Heijde 2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
ORAL Strategy[10] Fleischmann 2017	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
ORAL Standard[9] Van Vollenhoven 2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Nakamura 2018[116]	Yes	Unclear	Unclear	Unclear	Unclear	No	Yes	No	No	No	Unclear	Poor
Charles-Schoeman 2016[12]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Kremer 2012[117]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good

Study	Adequate Randomization	Allocation Concealment	Patient Blinding	Staff Blinding	Outcome Adjudication Blinding	Completeness of Follow Up	Intention to Treat Analysis	Incomplete Data Addressed	Selective Outcome Reporting	Industry Funding	Freedom from Bias	Overall Quality
Upadacitinib												
SELECT-MONOTHERAPY[118] Smolen 2019	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Good
SELECT-COMPARE[6] Fleischmann 2019	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Good
SELECT-BEYOND[13] Genovese 2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Good
SELECT-NEXT[7] Burmester 2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Good

N/A: not available

Table D3. Baseline Characteristics I

Study	Intervention	N	Female, n (%)	Age, Mean Years (SD)	Disease Duration, Mean Years (SD)	Previous DMARD use, n (%)		
						1	2	≥3
Baricitinib (BAR)								
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	184 (80.0)	52 (12)	8 (8)	104 (45.0)	61 (27.0)	61 (27.0)
RA-BUILD[18] Dougados 2017	cDMARD	228	189 (83.0)	51 (13)	7 (8)	96 (42.0)	81 (36.0)	50 (22.0)
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	137 (79.0)	55 (11)	14 (8)	69 (40.0)‡	55 (32.0)‡	50 (29.0)‡
RA-BEACON[15] Genovese 2016	cDMARD	176	145 (82.0)	56 (11)	14 (10)	81 (46.0)‡	47 (27.0)‡	47 (27.0)‡
RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	N/A					
	Stepped down to BAR 2 mg	146	N/A					
Keystone 2015[111]	BAR 2 mg	52	43 (85.0)	51 (13)	5.5 (4.4)	0 (0)‡	0 (0)‡	0 (0)‡
	cDMARD	98	86 (87.0)	49 (12)	5.4 (4.3)	0 (0)‡	0 (0)‡	0 (0)‡
Tofacitinib (TOF)								
ORAL Sync[8] Kremer 2013	TOF 5 mg	315	264 (83.8)	52.7 (11.7)	8.1 (Range: 0.2-39.9)	NR	NR	NR
ORAL Sync[8] Kremer 2013	Placebo (advanced to TOF 5 mg at 6 months)	79	63 (79.7)	50.8 (11.2)	9.5 (Range: 0.3-39.3)	NR	NR	NR
ORAL Step[14] Burmester 2013	TOF 5 mg	133	113 (85.0)	55.4 (11.5)	13.0 (Range: 1.2-55.0)	84 (63.2)†	37 (27.8)†	11 (8.3)†
ORAL Step[14] Burmester 2013	Placebo	132	106 (80.3)	54.4 (11.3)	11.3 (Range: 0.4-47.0)	86 (65.2)†	37 (28.0)†	9 (6.8)†
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	321	269 (83.8)	53.7 (11.6)	8.9 (Range: 0.3–43.0)	NR	NR	NR
ORAL Scan[11] Van der Heijde 2013	Placebo (advanced to TOF 5 mg at 6 months)	81	65 (80.2)	53.2 (11.5)	8.8 (Range: 0.6–30.8)	NR	NR	NR
ORAL Strategy[10] Fleischmann 2017	TOF 5 mg	384	319 (83.0)	49.7 (12.2)	Median (Range): 6.1 (0.2-41.6)	NR	NR	NR
ORAL Strategy[10] Fleischmann 2017	TOF 5 mg + cDMARD	376	311 (83.0)	50.0 (13.4)	Median (Range): 5.4 (0.0-43.5)	NR	NR	NR
ORAL Strategy[10] Fleischmann 2017	ADA + cDMARD	386	320 (83.0)	50.7 (13.4)	Median (Range): 6.0 (0.3-42.8)	NR	NR	NR

Study	Intervention	N	Female, n (%)	Age, Mean Years (SD)	Disease Duration, Mean Years (SD)	Previous DMARD use, n (%)		
						1	2	≥3
ORAL Standard[9] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	174 (85.3)	53 (11.9)	7.6 (NR)	NR	NR	NR
	ADA 40 mg + cDMARD	204	162 (79.4)	52.5 (11.7)	8.1 (NR)	NR	NR	NR
	Placebo (followed by TOF 5 mg + cDMARD)	56	43 (76.8)	55.5 (13.7)	6.9 (NR)	NR	NR	NR
Nakamura 2018[116]	TOF 5 mg	22	13 (59.1)	67.2 (1.7)	3.4 (0.7)	NR	NR	NR
	Non-TNF biologics	20	17 (85.0)	68.3 (1.2)	3.5 (1.0)	NR	NR	NR
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1071	911 (85.0)	52.6 (11.9)	7.7 (7.4)	0 (0)†	0 (0)†	0 (0)†
	bDMARD naïve: cDMARD	651	539 (82.8)	52.0 (12.5)	8.2 (8.2)	0 (0)†	0 (0)†	0 (0)†
	bDMARD-IR: TOF 5 mg	259	218 (84.2)	54.7 (11.1)	12.1 (9.1)	120 (60.9)†	≥2: 77 (39.1)†	
	bDMARD-IR: cDMARD	193	158 (81.9)	54.0 (11.6)	11.2 (8.6)	97 (62.6)†	≥2: 58 (37.4)†	
Kremer 2012[117]	TOF 5 mg + cDMARD	71	57 (80.3)	52 (12.8)	9.0 (Range: 0.5-46.0)	≥1: 4 (5.6)		
	Placebo + cDMARD	69	56 (81.2)	53 (13.4)	9.2 (Range: 0.5-39.0)	≥1: 2 (2.0)		
Upadacitinib (UPA)								
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	174 (80.2)	54.5 (12.2)	7.5 (8.9)*	NR	NR	NR
	UPA 30 mg	215	170 (79.1)	53.1 (12.7)	6.5 (7.0)*	NR	NR	NR
	MTX	216	179 (82.9)	55.3 (11.1)	5.8 (6.6)*	NR	NR	NR
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	521 (80.0)	54 (12)	8 (8)*	NR	NR	NR
	ADA 40 mg + cDMARD	327	259 (79.2)	54 (12)	8 (8)*	NR	NR	NR
	Placebo + cDMARD	651	512 (78.6)	54 (12)	8 (8)*	NR	NR	NR
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	137 (83.5)	56.3 (11.3)	12.4 (9.4)*	86 (52.4)‡	40 (24.4)‡	38 (23.2)‡
SELECT-BEYOND[13] Genovese 2018	UPA 30 mg + cDMARD	165	138 (84.1)	57.3 (11.6)	12.7 (9.7)*	66 (40.0)‡	51 (30.9)‡	47 (28.5)‡
	Placebo + cDMARD	169	143 (84.6)	57.6 (11.4)	14.5 (9.2)*	83 (49.1)‡	46 (27.2)‡	40 (23.7)‡
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	182 (82.4)	55.3 (11.5)	7.3 (7.9)*	27 (12.2%) previous bDMARD exposure		
SELECT-NEXT[7] Burmester 2018	UPA 30 mg + cDMARD	219	172 (78.5)	55.8 (11.3)	7.3 (7.9)*	28 (12.8%) previous bDMARD exposure		
	Placebo + cDMARD	221	166 (75.1)	56.0 (12.2)	7.2 (7.5)*	29 (13.1%) previous bDMARD exposure		

ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, N: total number, n: number, N/A: not available, NR: not reported, SD: standard deviation

*Reported as time since RA diagnosis.

†Previous TNFi.

‡Previous bDMARD

Table D4. Baseline Characteristics II

Study	Intervention	N	Tender Joint Count-68, Mean (SD)	Swollen Joint Count-66, Mean (SD)	DAS28-ESR, Mean (SD)	DAS28-CRP, Mean (SD)	HAQ-DI, Mean (SD)
Baricitinib (BAR)							
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	24 (14)	14 (9)	6.3 (1.0)	5.6 (1.0)	1.5 (0.62)
RA-BUILD[18] Dougados 2017	cDMARD	228	24 (15)	13 (7)	6.2 (1.0)	5.5. (0.9)	1.5 (0.60)
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	31 (16)	19 (12)	6.7 (1.0)	6.0 (0.9)	1.7 (0.55)
RA-BEACON[15] Genovese 2016	cDMARD	176	28 (16)	17 (11)	6.6 (0.9)	5.9 (0.9)	1.8 (0.57)
RA-BEYOND[109,110] Genovese 2017 (Abstract) RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	N/A				
	Stepped down to BAR 2 mg	146					
Keystone 2015[111]	BAR 2 mg	52	23.0 (12.6)	17.0 (9.3)	6.2 (0.8)	5.4 (0.9)	1.1 (0.7)
	cDMARD	98	22.2 (12.1)	15.8 (8.6)	6.3 (0.8)	5.5 (0.9)	1.2 (0.7)
Tofacitinib (TOF)							
ORAL Sync[8] Kremer 2013	TOF 5 mg	315	25.0 (15.3)	14.5 (10.3)	6.3 (1.0)	NR	1.4 (0.69)
ORAL Sync[8] Kremer 2013	Placebo (advanced to TOF 5 mg at 6 months)	79	27.2 (16.8)	14.6 (9.7)	6.4 (1.0)	NR	1.5 (0.64)
ORAL Step[14] Burmester 2013	TOF 5 mg	133	28.4 (18.3)	16.2 (10.1)	6.5 (1.1)	5.4 (1.0)	1.6 (0.7)
	Placebo	132	28.2 (16.7)	17.2 (10.7)	6.4 (1.1)	5.4 (1.0)	1.6 (0.7)
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	321	24.1 (14.0)	14.1 (8.2)	6.3 (NR)	5.2 (NR)	1.4 (0.7)
ORAL Scan[11] Van der Heijde 2013	Placebo (advanced to TOF 5 mg at 6 months)	81	23.3 (NR)	14.0 (NR)	6.3 (NR)	5.1 (NR)	1.4 (NR)
ORAL Strategy[10] Fleischmann 2017	TOF 5 mg	384	15.4 (6.5)†	11.2 (5.6)†	6.5 (0.9)	5.7 (0.9)	1.6 (0.6)
	TOF 5 mg + cDMARD	376	15.6 (6.5)†	11.8 (5.7)†	6.6 (0.9)	5.8 (0.9)	1.6 (0.6)
	ADA + cDMARD	386	15.2 (6.7)†	11.0 (5.4)†	6.5 (1.0)	5.7 (1.0)	1.6 (0.6)

Study	Intervention	N	Tender Joint Count-68, Mean (SD)	Swollen Joint Count-66, Mean (SD)	DAS28-ESR, Mean (SD)	DAS28-CRP, Mean (SD)	HAQ-DI, Mean (SD)
ORAL Standard[120] Van Vollenhoven 2013 ORAL Standard[120] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	28.5 (NR)	16.7 (NR)	6.6 (NR)	5.4 (NR)	1.5 (NR)
	ADA 40 mg + cDMARD	204	26.7 (NR)	16.4 (NR)	6.4 (NR)	5.3 (NR)	1.5 (NR)
	Placebo followed by TOF 5 mg + cDMARD	56	26.6 (NR)	16.9 (NR)	6.6 (NR)	5.6 (NR)	1.5 (NR)
Nakamura 2018[116]	TOF 5 mg	22	NR	NR	NR	4.2 (0.1)	0.7 (0.2)
	Non-TNF biologics	20	NR	NR	NR	4.2 (0.2)	0.7 (0.2)
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1071	25.1 (14.7)	15.1 (9.2)	6.4 (1.0)	NR	1.4 (0.7)
	bDMARD naïve: cDMARD	651	23.8 (13.9)	15.2 (9.0)	6.3 (1.0)	NR	1.3 (0.7)
	bDMARD-IR: TOF 5 mg	259	28.8 (16.7)	16.1 (9.4)	6.5 (1.0)	NR	1.6 (0.6)
	bDMARD-IR: cDMARD	193	27.8 (16.9)	16.8 (10.6)	6.4 (1.1)	NR	1.6 (0.6)
Kremer 2012[117]	TOF 5 mg + cDMARD	71	21.5 (1.5)	14.1 (0.9)	6.1 (NR)	5.1 (NR)	1.4 (0.1)
	Placebo + cDMARD	69	21.6 (1.6)	15.7 (1.1)	6.1 (NR)	5.3 (NR)	1.2 (0.1)
Upadacitinib (UPA)							
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	24.5 (15.1)	16.4 (10.9)	NR	5.6 (0.9)	1.5 (0.7)
	UPA 30 mg	215	24.8 (15.2)	16.9 (10.2)	NR	5.6 (1.1)	1.5 (0.7)
	cDMARD	216	25.2 (16.0)	16.9 (11.5)	NR	5.6 (1.0)	1.5 (0.7)
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	26 (15.0)	17 (10.0)	6.4 (1.0)	5.8 (1.0)	1.6 (0.6)
	ADA 40 mg + cDMARD	327	26 (15.0)	16 (9.0)	6.5 (1.0)	5.9 (1.0)	1.6 (0.6)
	Placebo + cDMARD	651	26 (14.0)	16 (9.0)	6.5 (1.0)	5.8 (0.9)	1.6 (0.6)
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	27.8 (16.3)	17.0 (10.8)	NR	5.9 (1.0)	1.7 (0.6)
	UPA 30 mg + cDMARD	165	27.3 (15.2)	17.2 (11.4)	NR	5.8 (0.9)	1.6 (0.6)
	Placebo + cDMARD	169	28.5 (15.3)	16.3 (9.6)	NR	5.8 (1.0)	1.6 (0.6)
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	25.2 (13.8)	16.0 (10.0)	NR	5.7 (1.0)	1.5 (0.6)
	UPA 30 mg + cDMARD	219	26.2(14.3)	16.2 (10.6)	NR	5.7 (0.9)	1.5 (0.6)
	Placebo + cDMARD	221	24.7 (15.0)	15.4 (9.2)	NR	5.6 (0.8)	1.4 (0.6)

ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, CRP: C-reactive protein, DAS: Disease Activity Score, ESR: erythrocyte sedimentation rate, GA: global assessment, HAQ-DI: Health Assessment Questionnaire without Disability Index, N: total number, n: number, N/A: not available, NR: not reported, SD: standard deviation, VAS: visual analogue scale

*Combined placebo arms (advanced to TOF 5 mg or 10 mg).

†Out of 28.

Table D5. Baseline Characteristics III

Study	Intervention	N	CDAI, Mean (SD)	SDAI, Mean (SD)	Morning Stiffness Duration, Min (SD)	Morning Stiffness Severity, 0-10 Scale	Concomitant cDMARD Use, n (%)		
							Methotrexate	Hydroxy chloroquine	Sulfa-salazine
Baricitinib (BAR)									
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	37 (13)	38 (13)	NR	NR	111 (48)	NR	NR
	cDMARD	228	36 (12)	37 (12)	NR	NR	109 (48)	NR	NR
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	43 (13)	45 (14)	NR	NR	141 (81.0)	NR	NR
	cDMARD	176	41 (13)	43 (14)	NR	NR	143 (81.0)	NR	NR
RA-BEYOND[109,110] Genovese 2017 (Abstract) RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	N/A						
	Stepped down to BAR 2 mg	146	N/A						
Keystone 2015[111]	BAR 2 mg	52	37.7 (12.2)	38.9 (12.2)	73.1 (42.2)	NR	52 (100)	11 (21.0)	7 (13.0)
	cDMARD	98	37.1 (12.3)	38.6 (12.5)	101.7 (110.7)	NR	98 (100)	16 (16.0)	14 (14)
Tofacitinib (TOF)									
ORAL Sync[8] Kremer 2013	TOF 5 mg	315	NR	NR	NR	NR	250 (79.4)	47 (14.9)	39 (12.4)
	Placebo (advanced to TOF 5 mg at 6 months)	79	NR	NR	NR	NR	61 (77.2)	10 (12.7)	12 (15.2)
ORAL Step[14] Burmester 2013	TOF 5 mg	133	NR	NR	NR	NR	133 (100)	0 (0)	0 (0)
	Placebo	132	NR	NR	NR	NR	132 (100)	0 (0)	0 (0)
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	321	NR	NR	NR	NR	321 (100)	0 (0)	0 (0)

Study	Intervention	N	CDAI, Mean (SD)	SDAI, Mean (SD)	Morning Stiffness Duration, Min (SD)	Morning Stiffness Severity, 0-10 Scale	Concomitant cDMARD Use, n (%)		
							Methotrexate	Hydroxy chloroquine	Sulfa-salazine
	Placebo (advanced to TOF 5mg at 6 months)	81	NR	NR	NR	NR	81 (100)	0 (0)	0 (0)
ORAL Strategy[10] Fleischmann 2017	TOF 5 mg	384	38.6 (12.6)	40.2 (13.0)	NR	NR	0 (0)	0 (0)	0 (0)
	TOF 5 mg + cDMARD	376	39.7 (12.7)	41.6 (13.2)	NR	NR	376 (100)	0 (0)	0 (0)
	ADA + cDMARD	386	38.2 (12.9)	39.8 (13.3)	NR	NR	386 (100)	0 (0)	0 (0)
ORAL Standard[120] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	NR	NR	NR	NR	204 (100)	0 (0)	0 (0)
	ADA 40 mg + cDMARD	204	NR	NR	NR	NR	204 (100)	0 (0)	0 (0)
	Placebo followed by TOF 5 mg + cDMARD	56	NR	NR	NR	NR	56 (100)	0 (0)	0 (0)
Nakamura 2018[116]	TOF 5 mg	22	19.7 (2.2)	NR	NR	NR	19 (86.4)	NR	NR
	Non-TNF biologics	20	18.7 (2.2)	NR	NR	NR	13 (65.0)	NR	NR
Charles-Schoeman 2016[12]	bDMARD Naïve: TOF 5 mg	1071	37.0 (12.3)	38.9 (12.9)	NR	NR	NR	NR	NR
	bDMARD Naïve: cDMARD	651	36.2 (12.8)	38.0 (13.2)	NR	NR	NR	NR	NR
	bDMARD-IR: TOF 5 mg	259	38.6 (12.5)	40.7 (13.4)	NR	NR	NR	NR	NR
	bDMARD-IR: cDMARD	193	38.3 (13.3)	39.9 (13.7)	NR	NR	NR	NR	NR
Kremer 2012[117]	TOF 5 mg + cDMARD	71	NR	NR	NR	NR	71(100)	0 (0)	0 (0)
	cDMARD	69	NR	NR	NR	NR	69 (100)	0 (0)	0 (0)

Study	Intervention	N	CDAI, Mean (SD)	SDAI, Mean (SD)	Morning Stiffness Duration, Min (SD)	Morning Stiffness Severity, 0-10 Scale	Concomitant cDMARD Use, n (%)		
							Methotrexate	Hydroxy chloroquine	Sulfa-salazine
Upadacitinib (UPA)									
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	38.0 (13.1)	39.4 (13.4)	144.2 (215.1)	NR	0 (0)	0 (0)	0 (0)
	UPA 30 mg	215	38.4 (13.8)	40.0 (14.3)	133.9 (152.7)	NR	0 (0)	0 (0)	0 (0)
	cDMARD	216	37.8 (14.4)	39.2 (14.6)	153.0 (221.7)	NR	216 (100)	0 (0)	0 (0)
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	40 (13)	NR	142 (188)	NR	651 (100)	0 (0)	0 (0)
	ADA 40 mg + cDMARD	327	40 (13)	NR	146 (185)	NR	327 (100)	0 (0)	0 (0)
	Placebo + cDMARD	651	40 (13)	NR	142 (170)	NR	651 (100)	0 (0)	0 (0)
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	41.7 (13.3)	43.3 (13.8)	140.4 (189.7)	6.8 (2.1)	137 (83.5)	7 (4.3)	6 (3.7)
	UPA 30 mg + cDMARD	165	40.1 (12.3)	41.7 (12.8)	184.5 (284.9)	6.5 (2.2)	135 (81.8)	14 (8.5)	9 (5.5)
	Placebo + cDMARD	169	41.0 (13.3)	42.6 (13.9)	138.4(178.6)	6.8 (2.3)	139 (82.2)	11 (6.5)	8 (4.7)
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	38.3 (11.9)	39.9 (12.5)	152.4 (241.9)	6.1 (2.4)	169 (76.5)	NR	NR
	UPA 30 mg + cDMARD	219	38.6 (12.7)	40.0 (13.1)	128.6 (156.0)	6.2 (2.2)	175 (79.9)	NR	NR
	Placebo + cDMARD	221	37.8 (11.8)	39.0 (11.9)	138.9 (214.0)	6.1 (2.2)	190 (86.0)	NR	NR

ADA: adalimumab, bDMARD: bDMARD: biologic disease-modifying antirheumatic drug, CDAI: clinical disease activity index, cDMARD: conventional disease-modifying antirheumatic drug, IR: inadequate response, n: number, N: total number, NR: not reported, SDAI: simple disease activity index, TNF: tumor necrosis factor

Table D6. Baseline – Patient Reported Outcomes

Study	Intervention	N	Patients' GA, 0-100 mm VAS (SD)	Physicians' GA, 0-100 mm VAS (SD)	Pain, 0-100 mm VAS (SD)	FACIT-F	SF-36	
							PCS	MCS
Baricitinib (BAR)								
RA-BUILD[18,121] Dougados 2017	BAR 2 mg + cDMARD	229	62 (20)	64 (17)	60 (21)	26.6 (11.5)	32.5 (8.4)	45.0 (11.5)
	cDMARD	228	60 (21)	62 (17)	57 (23)	26.6 (11.5)	32.2 (8.5)	45.7 (11.5)
RA-BEACON[15,122] Genovese 2016	BAR 2 mg + cDMARD	174	67 (19)	67 (17)	62 (22)	22.5 (10.0)	28.7 (8.1)	46.1 (13.1)
	cDMARD	176	66 (19)	67 (19)	65 (19)	22.2 (10.6)	28.2 (7.7)	46.1 (13.7)
RA-BEYOND[109,110] Genovese 2017 Van Der Heijde 2019	Continued BAR 4 mg	147	N/A					
	Stepped down to BAR 2 mg	146						
Keystone 2015[111]	BAR 2 mg	52	NR					
	cDMARD	98						
Tofacitinib (TOF)								
ORAL Sync[8,112] Kremer 2013 Strand 2017	TOF 5 mg	315	59.0 (22.9)	NR	57.1 (23.8)	29.0 (11.1)	32.4 (7.8)	40.9 (12.6)
	Placebo	158	57.9 (23.3)	NR	57.1 (22.8)	29.7 (9.0)	32.7 (7.6)	41.7 (11.6)
ORAL Step[14,113] Burmester 2013 Strand 2015	TOF 5 mg	133	64.7 (23.2)	NR	65.7 (22.8)	27.8 (11.1)	30.7 (9.3)	42.8 (12.7)
	Placebo*	132	61.9 (22.9)	NR	60.7 (23.5)	27.0 (11.5)	30.0 (8.0)	41.3 (13.3)
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	316	58.1 (23.6)	59.4 (15.9)	58.4 (23.1)	NR	NR	NR
	Placebo*	156	54.1 (22.9)	55.9 (17.4)	55.0 (23.9)*	NR	NR	NR
ORAL Strategy[10,114,119] Fleischmann 2017 Strand 2017 Strand 2019	TOF 5 mg	384	60.1 (21.4)	59.7 (17.7)	61.2 (21.7)	27.1 (0.52)	32.4 (0.38)	39.3 (0.59)
	TOF 5 mg + cDMARD	376	61.7 (22.0)	60.7 (18.0)	60.7 (22.4)	26.2 (0.53)	31.7 (0.36)	38.8 (0.57)
	ADA + cDMARD	386	60.2 (23.5)	60.3 (19.6)	60.6 (22.6)	27.1 (0.52)	31.7 (0.38)	39.8 (0.58)
ORAL Standard[115,120] Van Vollenhoven 2013 Strand 2016	TOF 5 mg + cDMARD	198	60.0 (21.4)	NR	59.2 (21.1)	28.2 (10.5)	33.1 (7.7)	39.8 (11.7)
	ADA 40 mg + cDMARD	199	57.1 (22.3)	NR	56.3 (22.0)	27.9 (10.1)	32.7 (6.8)	40.6 (11.7)
	Placebo followed by TOF 5 mg + cDMARD	104	54.3 (21.4)	NR	55.0 (21.4)	30.4 (10.3)	33.1 (6.3)	43.3 (10.7)
Nakamura 2018[116]	TOF 5 mg	22	NR					

	Non-TNF biologics	20	NR					
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1071	59.5 (22.9)	60.3 (16.8)	58.6 (23.4)	NR	NR	NR
	bDMARD naïve: cDMARD	651	56.3 (22.8)	59.1 (17.0)	56.1 (23.2)	NR	NR	NR
	bDMARD-IR: TOF 5 mg	259	64.4 (22.0)	65.1 (17.4)	64.3 (21.9)	NR	NR	NR
	bDMARD-IR: cDMARD	193	61.2 (22.7)	63.4 (16.6)	60.1 (23.5)	NR	NR	NR
Kremer 2012[117]	TOF 5 mg + cDMARD	71	56.5 (2.3)	54.6 (2.8)	54.9 (3.2)	NR	NR	NR
	cDMARD	69	58.3 (1.8)	51.9 (3.2)	51.2 (3.3)	NR	NR	NR
Upadacitinib (UPA)								
SELECT-MONOTHERAPY[118,119] Smolen 2019 Strand 2018	UPA 15 mg	217	62.2 (22.3)	65.7 (18.5)	62.3 (22.5)	NR	33.3 (7.9)	44.1 (11.3)
	UPA 30 mg	215	59.4 (22.8)	62.6 (17.8)	61.9 (22.1)	NR	33.9 (7.8)	44.5 (11.5)
	cDMARD	216	59.6 (21.8)	62.1 (17.5)	62.5 (21.3)	NR	33.3 (7.3)	45.1 (11.0)
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	64 (22.0)	66 (17.0)	66 (21.0)	27 (11)	33 (7)	NR
	ADA 40 mg + cDMARD	327	66 (21.0)	65 (18.0)	66 (21.0)	26 (11)	32 (7)	NR
	Placebo + cDMARD	651	64 (21.0)	66 (18.0)	65 (21.0)	27 (11)	33 (7)	NR
SELECT-BEYOND[13,119] Genovese 2018 Strand 2018	UPA 15 mg + cDMARD	164	67.2 (19.6)	68.7 (16.6)	68.2 (19.8)	NR	30.6 (7.8)	44.0 (11.7)
	UPA 30 mg + cDMARD	165	64.7 (21.1)	66.4 (15.6)	65.3 (20.7)	NR	45.9 (12.3)	45.9 (12.3)
	Placebo + cDMARD	169	66.3 (22.7)	66.9 (16.9)	68.9 (21.0)	NR	31.6 (7.2)	45.9 (12.6)
SELECT-NEXT[7,119] Burmester 2018 Strand 2018	UPA 15 mg + cDMARD	221	63.1 (21.9)	64.3 (16.2)	64.1 (19.5)	28.1 (11.1)	33·4 (7.4)	45.9 (10.9)
	UPA 30 mg + cDMARD	219	62.8 (20.3)	63.0 (18.0)	64.0 (19.8)	27.5 (12.6)	32.6 (7.9)	46.1 (12.0)
	Placebo + cDMARD	221	60.3 (20.5)	64.4 (17.7)	61.5 (20.8)	28.3 (11.5)	33.1 (7.7)	46.5 (11.7)

bDMARD: biologic disease modifying anti-rheumatic drug, cDMARD: conventional disease modifying anti-rheumatic drug, FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue, GA: Global Assessment, IR: inadequate response, MCS: Mental Component Score, N: total number, N/A: not available, NR: not reported, PCS: Physical Component Score, SF: Short Form, TNF: tumor necrosis factor, VAS: visual analogue scale

*Total placebo arm (advanced to either TOF 5 mg or 10 mg at 6 months).

Table D7. Outcomes at Three Months (12-14 Weeks) – ACR and EULAR Response

Study	Interventions	N	ACR20, n (%)	ACR50, n (%)	ACR70, n (%)	EULAR, n (%)
Baricitinib						
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	151 (66.0)†	77 (34.0)†	41 (18.0)†	NR
	cDMARD	228	90 (39.0)	29 (13.0)	7 (3.0)	NR
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	85 (49.0)†	35 (20.0)‡	23 (13.0)‡	NR
	cDMARD	176	48 (27.0)	14 (8.0)	4 (2.0)	NR
RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	NR			
	Stepped down to BAR 2 mg	146				
Keystone 2015[111]	BAR 2 mg + cDMARD	52	28 (54.0)¶¶	9 (17.5) ¶¶	5 (8.5) ¶¶	- Good: 9 (17.0) - Good/moderate: 43 (81.0)†
	cDMARD	98	41 (41.0) ¶¶	11 (11.0) ¶¶	2 (2.5) ¶¶	- Good: 16 (16.0) - Good/moderate: 51 (51.0)
Tofacitinib						
ORAL Sync[8] Kremer 2013	TOF 5 mg + cDMARD	315	177 (56.1)†	86 (27.3)†	27 (8.6)†	NR
	cDMARD	159	43 (27.0)	15 (9.4)	3 (1.9)	NR
ORAL Step[14] Burmester 2013	TOF 5 mg + cDMARD	132	55 (41.7)‡	35 (26.5)*	18 (13.6)*	Remission: 8 (6.1)*
	cDMARD	131	32 (24.4)	11 (8.4)	2 (1.5)	Remission: 0 (0)
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg + cDMARD	321	NR	93 (28.9)†¶¶	38 (11.7)‡¶¶	NR
	Placebo (advanced to TOF 5 mg at 6 months)	81	NR	7 (8.3)¶¶	3 (3.1)¶¶	NR
ORAL Strategy[10] Fleischmann 2017	TOF 5 mg	384	240 (62.5)¶¶	122 (31.7)¶¶	49 (12.7)¶¶	NR
	TOF 5 mg + cDMARD	376	264 (70.2)¶¶	156 (41.4)¶¶	70 (18.5)¶¶	NR
	ADA + cDMARD	386	265 (68.5)¶¶	146 (37.7)¶¶	56 (14.3)¶¶	NR
ORAL Standard[120] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	125 (60.9)¶¶	71 (34.4)¶¶	(11.5)¶¶	NR
	ADA 40 mg + cDMARD	204	115 (56.4)¶¶	49 (23.6)¶¶	(7.8)¶¶	NR
	Placebo followed by TOF 5 mg + cDMARD	56	16 (28.6)¶¶	6 (10.7)¶¶	(1.1)¶¶	NR
Nakamura 2018[116]	TOF group	22	NR	NR	NR	NR
	Non-TNF group	20	NR	NR	NR	NR
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1043	629 (60.3)†	341 (32.7)†	135 (12.9)†	NR

Study	Interventions	N	ACR20, n (%)	ACR50, n (%)	ACR70, n (%)	EULAR, n (%)
	bDMARD naïve: cDMARD	638	169 (26.5)	62 (9.7)	18 (2.8)	NR
	bDMARD-IR: TOF 5 mg	258	112 (43.4)†	63 (24.4)†	25 (9.7)§	NR
	bDMARD-IR: cDMARD	191	47 (24.6)	20 (10.5)	6 (3.1)	NR
Kremer 2012[117]	TOF 5 mg + cDMARD	71	36 (50.7)§	25 (35.2)§	13 (18.0)§	Good: 21 (29.4)§
	Placebo + cDMARD	69	23 (33.3)	12 (17.4)	4 (5.8)	Good: 11 (15.2)
Upadacitinib						
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	147 (67.7)*	91 (41.9)*	49 (22.6)*	NR
	UPA 30 mg	215	153 (71.2)*	112 (52.1)*§	71 (33.0)*	NR
	cDMARD	216	89 (41.2)	33 (15.3)	6 (2.8)	NR
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	462 (71.0)†¶	293 (45.0)†§§	163 (25.0)†§§	NR
	ADA 40 mg + cDMARD‡	327	206 (63.0)	95 (29.1)	43 (13.1)	NR
	Placebo + cDMARD	651	234 (35.9)	98 (15.1)	33 (5.1)	NR
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	106 (64.6)*	56 (34.3)*	19 (11.7)	NR
	UPA 30 mg + cDMARD	165	93 (56.4)*	59 (34.7)*	39 (23.5)*	NR
	cDMARD	169	48 (28.4)	20 (11.8)	11 (6.7)	NR
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	141 (63.8)*	84 (38.0)*	46 (20.8)*	NR
	UPA 30 mg + cDMARD	219	145 (66.2)*	94 (42.9)*	59 (26.9)*	NR
	cDMARD	221	79 (35.7)	33 (14.9)	13 (5.9)	NR

ACR: American College of Rheumatology, ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, EULAR: European League Against Rheumatology, n: number, N: total number, NR: not reported

*p<0.0001.

†p<0.001

‡p<0.01.

§p<0.05 vs. placebo.

§§p<0.001.

¶p<0.05 for upadacitinib vs. adalimumab.

¶¶Data was digitized and should be interpreted with caution.

#p-value not reported for adalimumab vs. placebo.

Table D8. Outcomes at Three Months (12-14 Weeks) – DAS28, SDAI, CDAI

Study	Interventions	N	DAS28-CRP, n (%)		DAS28-ESR, n (%)		CDAI, n (%)		SDAI, n (%)	
			≤3.2	≤2.6	≤3.2	≤2.6	≤10	≤2.8	≤11	≤3.3
Baricitinib										
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	83 (36.0)†	60 (26.0)†	48 (21.0)†	25 (11.0)†	78 (34.0)‡	23 (10.0)†	76 (33.0)‡	21 (9.0)†
	cDMARD	228	39 (17.0)	21 (9.0)	16 (7.0)	5 (2.0)	48 (21.0)	5 (2.0)	46 (20.0)	3 (1.0)
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	42 (24.0)†	20 (11.0)§	23 (13.0)‡	11 (6.0)‡	42 (24.0‡	6 (3.0)	39 (22.0)†	4 (2.0)
	cDMARD	176	16 (9.0)	71 (4.0)	8 (4.0)	2(1.0)	20 (11.0)	4 (2.0)	16 (9.0)	4 (2.0)
RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	Change from baseline: 0.14 (0.70)		Change from baseline: 0.10 (0.73)		136 (92.5)	57 (38.8)	Change from baseline: 0.69 (3.55)	
	Stepped down to BAR 2 mg	146	Change from baseline: 0.36 (0.77)§		Change from baseline: 0.35 (0.89)§		123 (84.2)§	54 (37.0)	Change from baseline: 2.19 (6.33)§	
Keystone 2015[111]	BAR 2 mg	52	12 (23.0)	8 (15.0)§	8 (15.0)	5 (8.0)§	NR	4 (6.0)	NR	4 (6.0)
	cDMARD	98	19 (19.0)	4 (4.0)	10 (10.0)	1 (1.0)	NR	1 (1.0)	NR	1 (1.0)
Tofacitinib										
ORAL Sync[8] Kremer 2013	TOF 5 mg	284	NR	NR	NR	23 (8.4)†	NR	NR	NR	NR
	Placebo	153	NR	NR	NR	1 (0.5)	NR	NR	NR	NR
ORAL Step[14] Burmester 2013	TOF 5 mg	119	NR	NR	17 (14.3)§	8 (6.7)§	NR	NR	NR	8 (6.7)*
	Placebo	120	NR	NR	6 (5.0)	2 (1.7)	NR	NR	NR	0 (0)
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	321	NR	NR	NR	NR	NR	NR	NR	NR
	Placebo advanced to TOF 5 mg (at 6 months)	81	NR	NR	NR	NR	NR	NR	NR	NR
ORAL Strategy[10] Fleischmann 2017	TOF 5 mg	384	NR	NR	NR	NR	NR	NR	NR	NR
	TOF 5 mg + cDMARD	376	NR	NR	NR	NR	NR	NR	NR	NR
	ADA + cDMARD	386	NR	NR	NR	NR	NR	NR	NR	NR

Study	Interventions	N	DAS28-CRP, n (%)		DAS28-ESR, n (%)		CDAI, n (%)		SDAI, n (%)	
			≤3.2	≤2.6	≤3.2	≤2.6	≤10	≤2.8	≤11	≤3.3
ORAL Standard[120] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	NR	NR	NR	NR	NR	NR	NR	NR
	ADA 40 mg + cDMARD	204	NR	NR	NR	NR	NR	NR	NR	NR
	Placebo followed by TOF 5 mg + cDMARD	56	NR	NR	NR	NR	NR	NR	NR	NR
Nakamura 2018[116]	TOF group	22	NR	NR	NR	NR	NR	NR	NR	NR
	non-TNF group	20	NR	NR	NR	NR	NR	NR	NR	NR
Charles-Schoeman 2016[12]	bDMARD Naïve: TOF 5 mg	1043	NR	NR	174 (16.6)†	77 (7.3)†	338 (32.4)†	67 (6.4)†	361 (34.6)†	67 (6.4)†
	bDMARD Naïve: cDMARD	638	NR	NR	29 (4.5)	15 (2.3)	92 (14.3)	5 (0.7)	91 (14.2)	5 (0.7)
	bDMARD-IR: TOF 5 mg	258	NR	NR	33 (12.7)§	17 (6.6)§	77 (29.5)‡	16 (5.9)§	77 (29.8)†	18 (6.8)
	bDMARD-IR: cDMARD	191	NR	NR	10 (5.1)	5 (2.3)	28 (14.4)	3 (1.2)	27 (13.8)	2 (0.6)
Kremer 2012[117]	TOF 5 mg + cDMARD	71	NR	12 (16.4)¶¶	NR	NR	NR	NR	NR	NR
	cDMARD	69	NR	5 (6.1)	NR	NR	NR	NR	NR	NR
Upadacitinib										
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	97 (44.7)*	61 (28.1)*	54 (24.9)†	30 (13.8)†	76 (35.3)*	28 (12.9)*	80 (36.9)	30 (13.8)*
	UPA 30 mg	215	114 (53.0)*	88 (40.9)*	84 (39.1)†	30 (14.0)†	101 (47.0)*	41 (19.1)*	101 (47.0)*	39 (18.1)*
	cDMARD	216	41 (19.0)	17 (7.9)	22 (10.2)	1 (0.5)	54 (25.0)	2 (0.9)	52 (24.1)	2 (0.9)
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	293 (45.0)†§§	189 (29.0)†§§	NR	NR	260 (39.9)†§§	85 (13.1)†§§	260 (39.9)	78 (12.0)
	ADA 40 mg + cDMARD‡	327	95 (29.1)	59 (18.0)	NR	NR	98 (30.0)	26 (8.0)	98 (30.0)	23 (7.0)

Study	Interventions	N	DAS28-CRP, n (%)		DAS28-ESR, n (%)		CDAI, n (%)		SDAI, n (%)	
			≤3.2	≤2.6	≤3.2	≤2.6	≤10	≤2.8	≤11	≤3.3
	Placebo + cDMARD	651	91 (14.0)†	39 (6.0)†	NR	NR	104 (16.0)	20 (3.1)	98 (15.1)	20 (3.1)
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	71 (43.3)*	NR	NR	NR	53 (32.3)†	NR	56 (34.1)*	NR
	UPA 30 mg + cDMARD	165	69 (41.8)*	NR	NR	NR	56 (33.9)*	NR	58 (35.2)*	NR
	cDMARD	169	24 (14.2)	NR	NR	NR	24 (14.2)	NR	24 (14.2)	NR
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	107 (48.4)*	68 (30.8)*	NR	NR	89 (40.3)*	20 (9.0)‡	92 (41.6)*	21 (9.5)‡
	UPA 30 mg + cDMARD	119	105 (47.9)*	62 (28.3)*	NR	NR	92 (42.0)*	26 (11.9)†	99 (45.2)*	27 (12.3)†
	cDMARD	221	38 (17.2)	22 (10.0)	NR	NR	42 (19.0)	7 (3.2)	42 (19.2)	7 (3.2)

ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, CDAI: clinical disease activity index, cDMARD: conventional disease-modifying antirheumatic drug, CRP: C-reactive protein, DAS: Disease Activity Score, ESR: erythrocyte sedimentation rate, SDAI: simple disease activity index, n: number, N: total number, NR: not reported

*p<0.0001.

†p<0.001

‡p<0.01.

§p<0.05 vs. placebo.

§§p<0.001.

¶¶Data was digitized and should be interpreted with caution.

Table D9. Outcomes at Three Months (12-14 Weeks) – HAQ-DI, Sharp Score

Study	Interventions	N	Change in HAQ-DI, Mean (SD)	HAQ-DI Score Improvement, n (%)			SHARP Score (mTSS), n (%)		
				≥0.22	≥0.3	≥0.5	<0	<0.5	<SDC (1.2)
Baricitinib (BAR)									
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	-0.57 (NR)¶¶	158 (69.0)†	138 (60.0)†	NR	NR	NR	NR
	cDMARD	228	-0.36 (NR)¶¶	124 (54.0)	92 (44.0)	NR	NR	NR	NR
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	-0.37 (NR)†¶¶	103 (59.0)‡	84 (48.0)§	NR	NR	NR	NR
	cDMARD	176	-0.18 (NR)¶¶	76 (43.0)	62 (35.0)	NR	NR	NR	NR
RA-BEYOND[109,110] (Abstract) Genovese 2017 Van Der Heijde 2019	Continued BAR 4 mg	147	0.04 (0.26)	NR					
	Stepped down to BAR 2 mg	146	0.06 (0.28)						
Keystone 2015[111]	BAR 2 mg	52	−0.18 (NR)	NR	NR	NR	NR	NR	NR
	cDMARD	98	−0.10 (NR)	NR	NR	NR	NR	NR	NR
Tofacitinib (TOF)									
ORAL Sync[8,112] Kremer 2013 Strand 2017	TOF 5 mg	315	-0.44 (NR)†	NR	NR	NR	NR	NR	NR
	Placebo	159	-0.16 (NR)	NR	NR	NR	NR	NR	NR
ORAL Step[14,113] Burmester 2013 Strand 2015	TOF 5 mg	132	-0.43 (NR)*	71 (54.2)§	NR	47 (35.9)‡	NR	NR	NR
	Placebo	131	-0.18 (NR)	53 (40.5)	NR	27 (20.6)	NR	NR	NR
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	321	-0.4 (NR)	NR	NR	NR	NR	NR	NR
	Placebo advanced to TOF 5 mg (at 6 months)	81	-0.15 (NR)	NR	NR	NR	NR	NR	NR
ORAL Strategy[10,114] Fleischmann 2017 Strand 2017	TOF 5 mg	384	-0.47 (NR)¶¶	NR	NR	NR	NR	NR	NR
	TOF 5 mg + cDMARD	376	-0.54 (NR)¶¶	NR	NR	NR	NR	NR	NR
	ADA + cDMARD	386	-0.49 (NR)¶¶	NR	NR	NR	NR	NR	NR
ORAL Standard[120] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	-0.55 (NR)§¶¶	NR	NR	NR	NR	NR	NR
	ADA 40 mg + cDMARD	204	-0.49 (NR)§	NR	NR	NR	NR	NR	NR

Study	Interventions	N	Change in HAQ-DI, Mean (SD)	HAQ-DI Score Improvement, n (%)			SHARP Score (mTSS), n (%)		
				≥0.22	≥0.3	≥0.5	<0	<0.5	<SDC (1.2)
	Placebo followed by TOF 5 mg + cDMARD	56	-0.24 (NR)	NR	NR	NR	NR	NR	NR
Nakamura 2018[116]	TOF group	22	%-change: -36.2	NR	NR	NR	NR	NR	NR
	Non-TNF group	20	%-change: -28.1	NR	NR	NR	NR	NR	NR
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1043	NR	552 (52.9)†	NR	421 (40.3)†	NR	NR	NR
	bDMARD naïve: cDMARD	638	NR	184 (28.7)	NR	117 (18.2)	NR	NR	NR
	bDMARD-IR: TOF 5 mg	258	NR	118 (45.7)	NR	80 (31.0)§	NR	NR	NR
	bDMARD-IR: cDMARD	191	NR	71 (36.9)	NR	39 (20.1)	NR	NR	NR
Kremer 2012[117]	TOF 5 mg + cDMARD	71	-0.49 (0.1)†	NR	NR	NR	NR	NR	NR
	cDMARD	69	-0.16 (0.1)	NR	NR	NR	NR	NR	NR
Upadacitinib (UPA)									
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	-0.65 (95% CI: -0.73, -0.57)†	140 (65.7)	NR	NR	NR	NR	NR
	UPA 30 mg	215	-0.73 (95% CI: -0.81, -0.64)†	148 (72.5)	NR	NR	NR	NR	NR
	cDMARD	216	-0.32 (95% CI: -0.41, -0.24)	98 (47.8)	NR	NR	NR	NR	NR
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	-0.6 (NR)†‖	NR	NR	NR	NR	NR	NR
	ADA 40 mg + cDMARD‡	327	-0.49 (NR)	NR	NR	NR	NR	NR	NR
	Placebo + cDMARD	651	-0.28 (NR)	NR	NR	NR	NR	NR	NR
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	-0.41 (95% CI -0.50 to -0.33)****	105 (64.1)****	NR	NR	NR	NR	NR
	UPA 30 mg + cDMARD	165	-0.44 (95% CI -0.52 to -0.35)****	92 (55.6)**	NR	NR	NR	NR	NR

Study	Interventions	N	Change in HAQ-DI, Mean (SD)	HAQ-DI Score Improvement, n (%)			SHARP Score (mTSS), n (%)		
				≥0.22	≥0.3	≥0.5	<0	<0.5	<SDC (1.2)
	cDMARD	169	-0.16 (95% CI -0.25 to -0.08)	63 (37.4)	NR	NR	NR	NR	NR
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	-0.61 (NR)	156/212 (73.6)	NR	NR	NR	NR	NR
	UPA 30 mg + cDMARD	219	-0.55 (NR)	148/214 (69.2)	NR	NR	NR	NR	NR
	cDMARD	221	-0.26 (NR)	109/221 (49.3)	NR	NR	NR	NR	NR

ACR: American College of Rheumatology, ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, EULAR: European League Against Rheumatology, N/A: not available, n: number, N: total number, NR: not reported

*p<0.0001.

†p<0.001

‡p<0.01.

§p<0.05 vs. placebo.

§§p<0.001.

¶¶Data was digitized and should be interpreted with caution.

Table D10. Patient Reported Outcomes at Three Months (12-14 Weeks) – Pain, Fatigue, and Quality of Life

Study	Intervention	N	Patients' GA, 0-100 mm VAS (SD)		Physicians' GA, 0-100 mm VAS (SD)		Pain, 0-100 mm VAS (SD)		FACIT-F		SF-36 (PCS / MCS)	
			Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)
Baricitinib (BAR)												
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	-25.3¥¥	NR	-31.7¥¥	NR	-25.5¥¥	NR	NR	NR	NR	NR
	cDMARD	228	-16.8¥¥	NR	-22.0¥¥	NR	-15.6¥¥	NR	NR	NR	NR	NR
RA-BEACON[15,122] Genovese 2016	BAR 2 mg + cDMARD	174	-20.5¥¥	NR	-20.8¥¥	NR	-17.1¥¥	NR	8.3 (NR)*	NR	6.1(NR)‡/ 3.0(NR)	86 (49)‡ / 58(33)
	cDMARD	176	-8.9¥¥	NR	-17.1¥¥	NR	-8.7¥¥	NR	5.2 (NR)	NR	2.7(NR)/ 1.2(NR)	57(32)/ 53(30)
RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	1.2 (14.3)	NR	−0.0 (9.6)	NR	0.7 (14.7)	NR	NR	NR	NR	NR
	Stepped down to BAR 2 mg	146	2.5 (14.5)	NR	2.7 (13.6)††	NR	2.5 (16.4)	NR	NR	NR	NR	NR
Keystone 2015[111]	BAR 2 mg	52	−16.2 (NR)	NR	-25.0 (NR)	NR	−14.2 (NR)	NR	NR	NR	NR	NR
	cDMARD	98	−10.3 (NR)	NR	-19.0 (NR)	NR	−8.8 (NR)	NR	NR	NR	NR	NR
Tofacitinib (TOF)												
ORAL Sync[8,112] Kremer 2013 Strand 2017	TOF 5 mg	315	-24.8 (1.2)‡	216 (68.6)	NR	NR	-24.2 (1.2)‡	214 (67.9)	5.8 (0.5)‡	170 (53.7)	5.9 (0.4)/ 4.4 (0.5)	204 (64.5)/ 184 (58.4)
	Placebo (advanced to TOF 5 mg at 6 mo.)	79	-12.5 (1.7)	39 (49.3)	NR	NR	-11.4 (1.7)	37 (46.0)	2.1 (0.6)	33 (40.8)	2.4 (0.6) /1.6 (0.7)	38 (47.3) /33 (41.8)
ORAL Step[14,113] Burmester 2013 Strand 2015	TOF 5 mg	133	-23.4 (2.4)‡	87 (64.9)†	NR	NR	-27.2 (2.4)‡	93 (69.3)‡	6.3 (1.0)‡	82 (61.5)†	5.65 (0.68)‡/ 3.52 (0.92)§	91 (67.8)/ 72 (54.2)
	Placebo‡‡	132	-9.2 (2.4)	56 (41.9)	NR	NR	-8.3 (2.4)	52 (39.1)	1.1 (1.0)	51 (38.60)	2.0 (0.7) /0.4 (0.9)	65 (49.1)/ 49 (37.1)
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	202	NR									
	Placebo	62										
ORAL-Strategy[10,114] Fleischmann 2017 Strand 2017	TOF 5 mg	384	NR									
	TOF 5 mg + cDMARD	376										
	ADA + cDMARD	386										
ORAL Standard[115,120] Van Vollenhoven 2013 Strand 2016	TOF 5 mg + cDMARD	204	-23.8 (1.7)†	144 (70.3)†	NR	NR	-26.7 (1.6)†	142 (69.2)†	5.9 (0.6)	109 (53.2)§	7.0 (0.5)†/ 3.2 (0.7)	139 (67.9)§/ 102 (50.0)
	ADA 40 mg + cDMARD	204	-21.5 (1.7)†	132 (64.4)†	NR	NR	-22.5 (1.6)†	132 (64.4)†	5.0 (0.6)*	115 (56.0)*	6.3 (0.5)*/ 3.4 (0.7)	136 (66.3)§ /106 (51.9)

Study	Intervention	N	Patients' GA, 0-100 mm VAS (SD)		Physicians' GA, 0-100 mm VAS (SD)		Pain, 0-100 mm VAS (SD)		FACIT-F		SF-36 (PCS / MCS)	
			Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)
	Placebo followed by TOF 5 mg + cDMARD	56	-7.3 (2.3)	20 (35.4)	NR	NR	-9.5 (2.2)	23 (39.6)	1.6 (0.8)	20 (35.4)	3.2 (0.7) /1.8 (0.9)	29 (51.0)/ 24 (42.7)
Nakamura 2018[116]	TOF 5 mg	22	NR									
	Non-TNF biologics	20										
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1071	NR									
	bDMARD naïve: cDMARD	651										
	bDMARD-IR: TOF 5 mg	259										
	bDMARD-IR: cDMARD	193										
Kremer 2012[117]	TOF 5 mg + cDMARD	71	-33.8 (2.3)†	NR	-27.1 (2.7)§	NR	-27.4 (2.8)†	NR	NR	NR	NR	NR
	cDMARD	69	-22.8 (2.3)	NR	-15.1 (2.8)	NR	-13.0 (2.8)	NR	NR	NR	NR	NR
Upadacitinib (UPA)												
SELECT-MONOTHERAPY[118,119] Smolen 2019 Strand 2018	UPA 15 mg	217	-23.4 (95% CI: -27.1, -19.8)§	133 (61.0)	-39.8 (NR)	NR	-26.2 (95% CI: -29.7, -22.6)§	139 (64.0)	NR	NR	8.3 (95% CI: 7.2, 9.4)/ 4.6 (95% CI: 3.3, 5.8)§	141 (65.0) / 109 (50.0)
	UPA 30 mg	215	-29.9 (95% CI: -33.5, -26.3)§	160 (74.0)	-41.9 (NR)	NR	-33.2 (95% CI: -36.7, -29.7)§	162 (75.0)	NR	NR	10.2 (95% CI: 9.1, 11.3)/ 4.7 (95% CI: 3.5, 5.9)§	157 (73.0) / 125 (58.0)
	cDMARD	216	-11.2 (95% CI: -14.9, -7.5)	102 (47.0)	-26.4 (NR)	NR	-13.9 (95% CI: -17.4, -10.3)	98 (45.0)	NR	NR	4.3 (95% CI: 3.2, 5.4)/ 1.9 (95% CI: 0.6, 3.1)	104 (48.0)/ 91 (42.0)
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	-31.0¥†¥¥	NR	-39.8¥†¥¥	NR	-32.1 (NR)†¥	NR	9.0 (NR)†§§	NR	7.9 (NR)†¶¶/ NR	NR
	ADA 40 mg + cDMARD	327	-24.1¥¥	NR	-37.3¥¥	NR	-25.6 (NR)	NR	7.4 (NR)	NR	6.3 (NR) /NR	NR
	Placebo + cDMARD	651	-15.8¥¥	NR	-25.1¥¥	NR	-15.7 (NR)	NR	4.8 (NR)	NR	3.6 (NR) /NR	NR
SELECT-BEYOND[13,119] Genovese 2018 Strand 2018	UPA 15 mg + cDMARD	164	-26.04 (95% CI: -30.16, -21.93)*	NR	-38.9¥¥	NR	-25.91 (95% CI: -30.05, -21.76)*	NR	NR	NR	5.83 (95% CI: 4.60, 7.05)*/ 4.54 (95% CI: 3.22, 5.87)	NR
	UPA 30 mg + cDMARD	165	-29.27 (95% CI: -33.43,	NR	-40.6¥¥	NR	-30.92 (95% CI:	NR	NR	NR	7.02 (95% CI: 5.78, 8.25)*/	NR

Study	Intervention	N	Patients' GA, 0-100 mm VAS (SD)		Physicians' GA, 0-100 mm VAS (SD)		Pain, 0-100 mm VAS (SD)		FACIT-F		SF-36 (PCS / MCS)	
			Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID, n (%)
			-25.12)*				-35.12, -26.72)*				3.37 (95% CI: 2.03, 4.72)	
	Placebo + cDMARD	169	-10.03 (95% CI: -14.22, -5.84)	NR	-25.7¥¥	NR	-10.38 (95% CI: -14.60, -6.16)	NR	NR	NR	2.39 (95% CI: 1.14, 3.64)/ 3.01 (95% CI: 1.65, 4.51)	NR
SELECT-NEXT[7,119] Burmester 2018 Strand 2018	UPA 15 mg + cDMARD	221	-29.7‡ (95% CI: -33.2, -26.2)	157 (71.0)	-38.3 (NR)‡	NR	-29.9§ (95% CI: -33.4, -26.4)	162 (73.0)	7.9§ (95% CI: 6.6, 9.3)	142 (64.0)	7.6§ (95% CI: 6.4, 8.7) /4.7§ (95% CI: 3.4, 6.0)	153 (69.0) / 122 (55.0)
	UPA 30 mg + cDMARD	219	-30.5‡ (95% CI: -34.0, -27.0)	158 (72.0)	-40.2 (NR)‡	NR	-31.7§ (95% CI: -35.2, -28.2)	160 (73.0)	7.7§ (6.4, 9.1)	125 (57.0)	8.0§ (95% CI: 6.8, 9.2)/ 3.7 (95% CI: 2.4, 5.0)	154 (70.0)/ 99 (45.0)
	Placebo + cDMARD	221	-10.4 (95% CI: -13.8, -6.9)	95 (43.0)	-23.2 (NR)	NR	-10.3 (95% CI: -13.7, -6.8)	98 (44.0)	3.0 (95% CI: 1.6, 4.3)	91 (41.0)	3.0 (95% CI: 1.9, 4.2) /2.6 (95% CI: 1.3, 3.9)	106 (48.0)/ 99 (41.0)

bDMARD: biologic disease modifying anti-rheumatic drug, cDMARD: conventional disease modifying anti-rheumatic drug, FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue, GA: Global Assessment, IR: inadequate response, LSM: least square mean, MCID: minimal clinically important difference, MCS: Mental Component Score, n: number, N: total number, N/A: not available, NR: not reported, PCS: Physical Component Score, SE: standard error, SF: Short Form, TNF: tumor necrosis factor, VAS: visual analogue scale

*p<0.01.

†p<0.001.

‡p<0.0001 vs. placebo.

§p<0.5.

§§p<0.5.

¶¶p<0.01.

¥p<0.001.

#p<0.0001 vs. adalimumab.

††p<0.05 vs. baricitinib.

‡‡Total placebo arm (advanced to either TOF 5mg or 10mg at 6 months).

¥¥Data are digitized and should be interpreted with caution.

Table D11. Outcomes at Six Months (24-26 Weeks) – ACR and EULAR

Study	Interventions	N	ACR20, n (%)	ACR50, n (%)	ACR70, n (%)	EULAR, n (%)
Baricitinib (BAR)						
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	140 (61.0)†	95 (41.0)†	58 (25.0)†	NR
	cDMARD	228	96 (42.0)	49 (21.0)	18 (8.0)	NR
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	78 (45.0)†	40 (23.0)§	23 (13.0)†	NR
	cDMARD	176	48 (27.0)	23 (13.0)	6 (3.0)	NR
RA-BEYOND[109,110] (Abstract) Genovese 2017 Van Der Heijde 2019	Continued BAR 4 mg	147	NR			
	Stepped down to BAR 2 mg	146				
Keystone 2015[111]	BAR 2 mg	52	N/A			
	cDMARD	98	N/A			
Tofacitinib (TOF)						
ORAL Sync[8] Kremer 2013	TOF 5 mg	315	164 (52.1)†	105 (33.4)†	41 (12.4)†	NR
	Placebo (advanced to TOF 5 mg at 6 months)	159	49 (30.8)	20 (12.4)†	5 (.03)§	NR
ORAL Step[14] Burmester 2013	TOF 5 mg	133	N/A			
	Placebo (advanced to TOF 5 mg at 3 months)	132				
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	321	81 (51.5)†	103 (32.4)†	45 (14.6)†	NR
	Placebo (advanced to TOF 5 mg at 6 months)	81	21 (25.3)	7 (8.4)	1 (1.3)	NR
ORAL Strategy[10] Fleischmann 2017	TOF 5mg	384	249 (65.0)	147 (38.0)	70 (18.0)	NR
	TOF 5 mg + cDMARD	376	275 (73.0)	173 (46.0)	94 (25.0)	NR
	ADA + cDMARD	386	274 (71.0)	169 (44.0)	80 (21.0)	NR
ORAL Standard[120] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	101/196 (51.5)	77 (37.6)¶¶¥	41 (20.0)¶¶¥	NR
	ADA +cDMARD 40 mg	204	94/199 (47.2)	58 (28.2)¶¶¥	19 (9.0)¶¶¥	
	Placebo followed by TOF 5 mg + cDMARD	56	30/106 (28.3)	8 (13.2)¶¶¥	2 (2.3)¶¶¥	NR
Nakamura 2018[116]	TOF 5 mg	22	NR	NR	NR	NR
	Non-TNF	20	NR	NR	NR	NR
	bDMARD Naïve: TOF 5 mg	1071	557 (51.9)	354 (32.9)	161 (15.0)	NR

Study	Interventions	N	ACR20, n (%)	ACR50, n (%)	ACR70, n (%)	EULAR, n (%)
Charles-Schoeman 2016[12]	bDMARD Naïve: cDMARD	651	NR	NR	NR	NR
	bDMARD-IR: TOF 5 mg	259	117 (45.6)	83 (32.0)	37 (14.8)	NR
	bDMARD-IR: cDMARD	193	NR	NR	NR	NR
Kremer 2012[117]	TOF 5 mg	71	35 (49.0)	24 (33.2)	14 (18.9)§	NR
	Placebo	69	25 (35.3)	16 (22.9)	5 (6.9)§	NR
Upadacitinib (UPA)						
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	N/A			
	UPA 30 mg	215				
	cDMARD	216				
SELECT-COMPARE[6] Fleishmann 2019	UPA 15 mg + cDMARD	651	436 (67.0)†§§	352 (54.1)†§§	NR	NR
	ADA + cDMARD#	327	186 (56.9)	137 (41.9)	NR	NR
	Placebo + cDMARD	651	234 (35.9)	137 (21.0)	NR	NR
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	105 (63.6)¶¶	70 (42.5) ¶¶	37 (22.1) ¶¶	N/A
	UPA 30 mg + cDMARD	165	99 (59.7)¶¶	73 (43.7) ¶¶	40 (24.2) ¶¶	
	Placebo + cDMARD	169	N/A			
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	N/A			
	UPA 30 mg + cDMARD	219				
	cDMARD	221				

ACR: American College of Rheumatology, ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, EULAR: European League Against Rheumatology, N/A: not available, n: number, N: total number, NR: not reported

*p<0.0001.

†p<0.001

‡p<0.01.

§p<0.05 vs. placebo.

§§p<0.001.

¶¶Data was digitized and should be interpreted with caution.

¥With advanced placement therapy.

Table D12. Outcomes at Six Months (24-26 Weeks) – DAS28, SDAI, CDAI

Study	Interventions	N	DAS28-CRP, n (%)		DAS28-ESR, n (%)		CDAI, n (%)		SDAI, n (%)	
			≤3.2	≤2.6	≤3.2	≤2.6	≤10	≤2.8	≤11	≤3.3
Baricitinib (BAR)										
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	106 (46.0)†	71 (31.0)†	67 (29.0)†	33 (14.0)†	104 (45.0)‡	35 (15.0)†	110 (48.0)‡	39 (17.0)†
	cDMARD	228	55 (24.0)	26 (11.0)	23 (10.0)	10 (4.0)	64 (28.0)	10 (4.0)	46 (29.0)	10 (4.0)
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	35 (20.0)§	20 (11.0)	21 (12.0)	9 (5.0)	41 (23.0)	9 (5.0)	39 (22.0)§	9 (5.0)
	cDMARD	176	20 (11.0)	11 (6.0)	13 (7.0)	6 (3.0)	27 (15.0)	6 (3.0)	25 (14.0)	4 (2.0)
RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	NR							
	Stepped down to BAR 2 mg	146								
Keystone 2015[111]	BAR 2 mg	52	N/A							
	cDMARD	98								
Tofacitinib (TOF)										
ORAL Sync[8] Kremer 2013	TOF 5 mg	315	NR	NR	NR	24 (8.5)	NR	NR	NR	NR
	Placebo (advanced to TOF 5 mg at 6 months)	159	NR	NR	NR	4 (2.6)	NR	NR	NR	NR
ORAL Step[14] Burmester 2013	TOF 5mg	122	N/A							
	Placebo (advanced to TOF 5 mg at 3 months)	66								
ORAL Scan[11] Van der Heijde 2013	TOF 5mg	321	NR	NR	45 (14.3)†	225 (7.2)	NR	NR	NR	NR
	Placebo (advanced to TOF 5 mg	81	NR	NR	3 (3.1)	2 (1.6)	NR	NR	NR	NR

Study	Interventions	N	DAS28-CRP, n (%)		DAS28-ESR, n (%)		CDAI, n (%)		SDAI, n (%)	
			≤3.2	≤2.6	≤3.2	≤2.6	≤10	≤2.8	≤11	≤3.3
	at 6 months)									
ORAL Strategy[10] Fleishmann 2017	TOF 5 mg	384	159 (41)	81 (21)	79 (21)	40 (10)	163 (42)	39 (10)	167 (43)	38 (10)
	TOF 5 mg + cDMARD	376	174 (46)	115 (31)	100 (27)	45 (12)	183 (49)	52 (14)	187 (50)	50 (13)
	ADA + cDMARD	386	181 (47)	108 (28)	106 (27)	48 (12)	179 (46)	51 (13)	182 (47)	50 (13)
ORAL Standard[9] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	NR	NR	NR	11/177 (6.2)	NR	NR	NR	NR
	ADA 40 mg + cDMARD	204	NR	NR	NR	12/178 (6.7)	NR	NR	NR	NR
	Placebo followed by TOF 5 mg + cDMARD	56	NR	NR	NR	1/92 (1.1)	NR	NR	NR	NR
Nakamura 2018[116]	TOF 5 mg	22	NR	NR	NR	NR	NR	NR	NR	NR
	Non-TNF	20	NR	NR	NR	NR	NR	NR	NR	NR
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1071	NR	NR	172 (16.3)	750 (7.2)	NR	NR	NR	NR
	bDMARD naïve: cDMARD	651	NR	NR	NR	NR	NR	NR	NR	NR
	bDMARD-IR: TOF 5 mg	259	NR	NR	47 (18.3)	19 (7.1)	NR	NR	NR	NR
	bDMARD-IR: cDMARD	193	NR	NR	NR	NR	NR	NR	NR	NR
Kremer 2012[117]	TOF 5 mg	71	2 (3.00)	21 (28.89)	NR	NR	NR	NR	NR	NR
	cDMARD	69	3 (3.45)	7 (10.31)	NR	NR	NR	NR	NR	NR
Upadacitinib (UPA)										
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	N/A							
	UPA 30 mg	215								
	cDMARD	216								
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	358 (55.0)†§§	267 (41.0)†§§	NR	NR	345 (53.0)†§§	150 (23.0)†§§	NR	NR
	ADA + cDMARD	327	128 (39.1)	88 (26.9)	NR	NR	124 (37.9)	46 (14.1)	NR	NR

Study	Interventions	N	DAS28-CRP, n (%)		DAS28-ESR, n (%)		CDAI, n (%)		SDAI, n (%)	
			≤3.2	≤2.6	≤3.2	≤2.6	≤10	≤2.8	≤11	≤3.3
	Placebo + cDMARD	651	117 (18.0)	59 (9.1)	NR	NR	143 (22.0)	39 (6.0)	NR	NR
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	86 (52.0)	NR		NR	NR	77 (47.0)	NR	81 (49.0)	N/A
	UPA 30 mg + cDMARD	86 (52.0)	NR		NR	NR	73 (44.0)	NR	75 (45.0)	
	Placebo + cDMARD	N/A								
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	N/A							
	UPA 30 mg + cDMARD	219								
	cDMARD	221								

ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, CDAI: clinical disease activity index, cDMARD: conventional disease-modifying antirheumatic drug, CRP: C-reactive protein, DAS: Disease Activity Score, ESR: erythrocyte sedimentation rate, N/A: not available, n: number, N: total number, NR: not reported, SDAI: simple disease activity index

†p<0.001.

‡p<0.01.

§p<0.05 vs. placebo.

§§p<0.001.

Table D13. Outcomes at Six Months (24-26 Weeks) – HAQ-DI, SHARP Score

Study	Interventions	N	Change in HAQ-DI, Mean (SD)	HAQ-DI Improvement, n (%)			Sharp Score, n (%)		
				≥0.22	≥0.3	≥0.5	<0	<0.5	<SDC (1.2)
Baricitinib (BAR)									
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	-0.4 (NR)§§	147 (64.0)‡	133 (58.0)‡	NR	163 (71.6)	187 (81.7	198 (86.5)
	cDMARD	228	-0.6 (NR)§§	96 (42.0)	85 (37.0)	NR	169 (74.2)	177 (77.4)	190 (83.2)
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	-0.2 (NR)§§	87 (50.0)†	72 (41.0)‡	NR	NR	NR	NR
	cDMARD	176	-0.4 (NR)§§	53 (30.0)	43 (24.0)	NR	NR	NR	NR
RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	NR						
	Stepped down to BAR 2 mg	146							
Keystone 2015[111]	BAR 2 mg	N/A							
	cDMARD								
Tofacitinib (TOF)									
ORAL Sync[8,112] Kremer 2013 Strand 2017	TOF 5 mg	315	208	NR	NR	NR	NR	NR	NR
	Placebo (advanced to TOF 5 mg at 6 months)	159	NR	NR	NR	NR	NR	NR	NR
ORAL Step[14,113] Burmester 2013 Strand 2015	TOF 5 mg	N/A							
	Placebo (advanced to TOF 5 mg at 3 months)								
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	321	-0.49 (NR)‡§§	NR	NR	NR	NR	285 (88.9)	NR
	Placebo (advanced to TOF 5 mg at 6 months)	81	-0.16 (NR)	NR	NR	NR	NR	63 (77.7)	NR
ORAL Strategy[10,114] Fleischmann 2017 Strand 2017	TOF 5 mg	384	-0.51 (NR)	NR	NR	NR	NR	NR	NR
	TOF 5 mg + cDMARD	376	-0.58 (NR)	NR	NR	NR	NR	NR	NR
	ADA + cDMARD	386	-0.54 (NR)	NR	NR	NR	NR	NR	NR
ORAL Standard[9] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	-0.63 (NR)	NR	NR	NR	NR	NR	NR
	ADA 40 mg + cDMARD	204	-0.55 (NR)	NR	NR	NR	NR	NR	NR
	Placebo followed by TOF 5 mg + cDMARD	56	-0.47 (NR)	NR	NR	NR	NR	NR	NR
Nakamura 2018[116]	TOF 5 mg	22	NR	NR	NR	NR	NR	NR	NR

Study	Interventions	N	Change in HAQ-DI, Mean (SD)	HAQ-DI Improvement, n (%)			Sharp Score, n (%)		
				≥0.22	≥0.3	≥0.5	<0	<0.5	<SDC (1.2)
	Non-TNF	20	NR	NR	NR	NR	NR	NR	NR
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1071	NR	585 (54.6)	NR	447 (41.7)	NR	NR	NR
	bDMARD naïve: cDMARD	651	NR	NR	NR	NR	NR	NR	NR
	bDMARD-IR: TOF 5 mg	259	NR	123 (47.4)	NR	87 (33.5)	NR	NR	NR
	bDMARD-IR: cDMARD	193	NR	NR	NR	NR	NR	NR	NR
Kremer 2012[117]	TOF 5 mg	71	-0.61 (NR)	NR	NR	NR	NR	NR	NR
	Placebo	69	-0.37 (NR)	NR	NR	NR	NR	NR	NR
Upadacitinib (UPA)									
SELECT-MONOTHERAPY[118] Smolen 2019	UPA 15 mg	217	N/A						
	UPA 30 mg	215							
	cDMARD	216							
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	NR	NR	NR	NR	NR	NR	NR
	ADA + cDMARD‡	327	NR	NR	NR	NR	NR	NR	NR
	Placebo + cDMARD	651	NR	NR	NR	NR	NR	NR	NR
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	-0.44§§		104 (63.2) §§	NR	NR	NR	NR	N/A
	UPA 30 mg + cDMARD	-0.53§§		95 (57.6) §§	NR	NR	NR	NR	
	Placebo + cDMARD	N/A							
SELECT-NEXT[7] Burmester 2018	UPA 15 mg + cDMARD	221	N/A						
	UPA 30 mg + cDMARD	219							
	cDMARD	221							

ADA: adalimumab, bDMARD: biologic disease-modifying antirheumatic drug, cDMARD: conventional disease-modifying antirheumatic drug, CRP: C-reactive protein, DAS: Disease Activity Score, ESR: erythrocyte sedimentation rate, HAQ: Health Assessment Questionnaire, mg: milligram, N/A: not available, n: number, N: total number, NR: not reported, SD: standard deviation, TNF: tumor necrosis factor

*p<0.5.

†p<0.01.

‡p<0.001.

§p<0.0001 vs. placebo.

§§Data are digitized and should be interpreted with caution.

Table D14. Patient Reported Outcomes at Six Months

Study	Intervention	N	Patients' GA, 0-100 mm VAS (SD)		Physicians' GA, 0-100 mm VAS (SD)		Pain, 0-100 mm VAS (SD)		FACIT-F		SF-36 (PCS / MCS)	
			Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)
Baricitinib (BAR)												
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	-27.5***†	NR	-36.4 ***†	NR	-27.4 ***†	NR	NR	NR	NR	NR
	cDMARD	228	-18.8†	NR	-26.3†	NR	-19.7†	NR	NR	NR	NR	NR
RA-BEACON[15,122] Genovese 2016	BAR 2 mg + cDMARD	174	-20.3***†	NR	-28.9 ***†	NR	-18.7 ***†	NR	8.1(NR)*	NR	6.2(NR)‡ /2.8(NR)	63(39)‡/ 46(26)
	cDMARD	176	-8.7†	NR	-19.7†	NR	-8.8†	NR	5.7 (NR)	NR	1.9(NR)/ 1.9(NR)	37(21)/ 39(22)
RA-BEYOND[109,110] Genovese 2017 (Abstract) Van Der Heijde 2019	Continued BAR 4 mg	147	NR									
	Stepped down to BAR 2 mg	146										
Keystone 2015[111]	BAR 2 mg	N/A										
	cDMARD											
Tofacitinib (TOF)												
ORAL Sync[8,112] Kremer 2013 Strand 2017	TOF 5 mg	315	-27.8†	NR	NR	NR	-28.9†	NR	NR	NR	5.7† /4.1†	NR
	Placebo (advanced to TOF 5 mg at 6 months)	79	-20.1†	NR	NR	NR	-19.5†	NR	NR	NR	7.4† /2.3†	NR
ORAL Step[14,113] Burmester 2013 Strand 2015	TOF 5 mg	N/A										
	Placebo¥											
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg	321	-25.8 (1.4)***	NR	-34.4 (1.2)***	NR	-26.4 (1.4)**	NR	NR	NR	NR	NR
	Placebo (advanced to TOF 5 mg at 6 months)	81	-13.6 (2.4)	NR	-23.8 (2.0)	NR	-15.7 (2.4)	NR	NR	NR	NR	NR

Study	Intervention	N	Patients' GA, 0-100 mm VAS (SD)		Physicians' GA, 0-100 mm VAS (SD)		Pain, 0-100 mm VAS (SD)		FACIT-F		SF-36 (PCS / MCS)	
			Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)
ORAL Strategy[10,114,119] Fleischmann 2017 Strand 2017 Strand 2019	TOF 5 mg	384	-35.7 (1.0)	71.7	NR	NR	-26.6 (1.3)	73.1	7.1 (0.5)	62.7	6.7 (0.4) /5.2 (0.5)	71.5/58.3
	TOF 5 mg + cDMARD	376	-38.4 (1.0)¤	76.1	NR	NR	-30.7 (1.3)¤	77.6	7.59 (0.5)¥	66.1**	7.9 (0.4) /5.7 (0.5)¤	74.0/63.6**
	ADA + cDMARD	386	-38.8 (1.0)¤	72.1	NR	NR	-28.1 (1.3)	76.9	6.1 (0.5)	58.5**	7.8 (0.4) /4.4 (0.5)	76.5/53.9**
ORAL Standard[9,115] Van Vollenhoven 2013 Strand 2016	TOF 5 mg + cDMARD	204	-28.6****†	NR	NR	NR	-30.6 ****†	NR	6.8 ****†	NR	8.5****† /5.3†	NR
	ADA 40 mg + cDMARD	204	-25.8****†	NR	NR	NR	-27.1 ****†	NR	6.5***†	NR	7.3***† /3.9†	NR
	Placebo followed by TOF 5 mg + cDMARD	56	-12.5†	NR	NR	NR	-16.0†	NR	2.1†	NR	4.1† /0.8†	NR
Nakamura 2018[116]	TOF 5 mg	22	NR									
	Non-TNF biologics	20										
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1071	NR									
	bDMARD naïve: cDMARD	651										
	bDMARD-IR: TOF 5 mg	259										
	bDMARD-IR: cDMARD	193										
Kremer 2012[117]	TOF 5 mg + cDMARD	71	NR									
	cDMARD	69										
Upadacitinib (UPA)												
SELECT-MONOTHERAPY[118,119] Smolen 2019	UPA 15 mg	217	N/A									
	UPA 30 mg	215										
	cDMARD	216										

Study	Intervention	N	Patients' GA, 0-100 mm VAS (SD)		Physicians' GA, 0-100 mm VAS (SD)		Pain, 0-100 mm VAS (SD)		FACIT-F		SF-36 (PCS / MCS)	
			Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)	Change from Baseline, LSM (SE)	Patients Reporting ≥MCID (%)
Strand 2018												
SELECT-COMPARE[6] Fleischmann 2019	UPA 15 mg + cDMARD	651	-35.7†	NR	-46.1†	NR	-36.5†	NR	9.7 (NR) ***#	NR	9.5 (NR) ***##/ NR	NR
	ADA 40 mg + cDMARD	327	-30.1†	NR	-41.4†	NR	-32.4†	NR	8.2 (NR)	NR	7.8 (NR) /NR	NR
	Placebo + cDMARD	651	-18.0†	NR	-27.6†	NR	-18.9†	NR	5.5 (NR)	NR	4.5 (NR) /NR	NR
SELECT-BEYOND[13,119] Genovese 2018 Strand 2018	UPA 15 mg + cDMARD	164	7.2 (NR)/NR									N/A
	UPA 30 mg + cDMARD	165	8.0 (NR)/NR									
	Placebo + cDMARD	169	NR									
SELECT-NEXT[7,119] Burmester 2018 Strand 2018	UPA 15 mg + cDMARD	221	N/A									
	UPA 30 mg + cDMARD	219										
	Placebo + cDMARD	221										

bDMARD: biologic disease modifying anti-rheumatic drug, cDMARD: conventional disease modifying anti-rheumatic drug, FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue, GA: Global Assessment, IR: inadequate response, LSM: least square mean, MCID: minimal clinically important difference, MCS: Mental Component Score, n: number, N: total number, N/A: not available, NR: not reported, PCS: Physical Component Score, SE: standard error, SF: Short Form, TNF: tumor necrosis factor, VAS: visual analogue scale

*p<0.5, **p<0.01, ***p<0.001, ****p<0.0001 vs. placebo, #p<0.5, ##p<0.01, ###p<0.001, ####p<0.0001 vs. adalimumab, ‡p<0,05 vs. baricitinib, ¥Total placebo arm (advanced to either TOF 5 mg or 10 mg at 6 months), †Data are digitized and should be interpreted with caution, ¤p<0.05 vs. tofacitinib mono therapy.

Table D15. Harms I

Study	Intervention	N	Time Point (Weeks)	Any Adverse Event, n (%)	Serious Adverse Event, n (%)	Adverse Event Leading to Discontinuation, n (%)	Nausea, n (%)	Nasopharyngitis, n (%)	Venous Thromboembolism, n (%)
Baricitinib (BAR)									
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	0-12	122 (53.0)	4 (2.0)	7 (3.0)	NR	NR	NR
	cDMARD	228	0-12	133 (58.0)	8 (4.0)	8 (4.0)	NR	NR	NR
	BAR 2 mg + cDMARD	229	0-24	154 (67.0)	6 (3.0)	10 (4.0)	7 (3.0)	10 (4.0)	NR
	cDMARD	228	0-24	161 (71.0)	11 (5.0)	10 (4.0)	8 (4.0)	18 (8.0)	NR
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	0-12	107 (61.0)	3 (2.0)	7 (4.0)	NR	NR	NR
	cDMARD	176	0-12	96 (55.0)	7 (4.0)	4 (2.0)	NR	NR	NR
	BAR 2 mg + cDMARD	174	0-24	123 (71.0)	7 (4.0)	7 (4.0)	7 (4.0)	12 (7.0)	NR
	cDMARD	176	0-24	112 (64.0)	13 (7.0)	7 (4.0)	5 (3.0)	7 (4.0)	NR
RA-BEYOND[109,110] (Abstract) Genovese 2017 Van Der Heijde 2019	Continued BAR 4 mg	147	12	NR					
	Stepped down to BAR 2 mg	146	12						
Keystone 2015[111]	BAR 2 mg	52	0-12	TEAE: 24 (46.0)	3 (6.0)	5 (5.0)	NR	NR	NR
	cDMARD	98	0-12	TEAE: 45 (46.0)	3 (3.0)	1 (2.0)	NR	NR	NR
	BAR 2 mg	52	0-24	TEAE: 31 (60.0)	3 (6.0)	1 (2.0)	NR	NR	NR
	cDMARD	98	0-24	N/A					
Tofacitinib (TOF)									
ORAL Sync[8] Kremer 2013	TOF 5 mg + cDMARD	388	52	171.9 (152.5-193.8)	6.9 (4.6-10.5)	6.2 (4.0-9.6)	18 (5.5)	23 (7.1)	NR
	Placebo (advanced to TOF 5 mg at 6 months)	159	52	342.3 (281.1-416.9)	10.9 (4.9-24.2)	5.4 (1.8-16.8)	5 (9.0)	12 (21.6)	NR

Study	Intervention	N	Time Point (Weeks)	Any Adverse Event, n (%)	Serious Adverse Event, n (%)	Adverse Event Leading to Discontinuation, n (%)	Nausea, n (%)	Nasopharyngitis, n (%)	Venous Thromboembolism, n (%)
ORAL Step[14] Burmester 2013	TOF 5 mg + cDMARD	133	0-12	71 (53.4)	2 (1.5)	8 (6.0)	4 (3.0)	5 (3.8)	NR
	Placebo (advanced to TOF 5 mg)	132	0-12	75 (56.8)	6 (4.5)	7 (5.3)	9 (6.8)	4 (3.0)	NR
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg + cDMARD	321	12	TEAE: 157 (48.9)	12 (3.7)	15 (4.7)	7 (2.2)	14.4 (4.4)	NR
	Placebo (advanced to TOF 5 mg at 6 months)	160	12	TEAE: 73 (45.6)	5 (3.1)	5 (3.1)	2 (1.3)	1 (0.6)	NR
ORAL Strategy[10] Fleischmann 2017	TOF 5 mg	384	52	226 (59)	35 (9)	23 (6)	11 (3)	22 (6)	NR
	TOF 5 mg + cDMARD	376	52	231 (61)	27 (7)	26 (7)	13 (4)	16 (4)	NR
	ADA + cDMARD	386	52	253 (66)	24 (6)	37 (10)	16 (4)	18 (5)	NR
ORAL Standard[9] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	0-12	135 (66.2)	12 (5.9)	14 (6.9)	3 (1.5)	8 (3.9)	NR
	ADA 40 mg + cDMARD	204	0-12	105 (51.5)	5 (2.5)	10 (4.9)	3 (1.5)	7 (3.4)	NR
	Placebo followed by TOF 5 mg + cDMARD	56	0-12	57(52.8)	2 (1.9)	3 (2.8)	0	0 (0)	NR
Nakamura 2018[116]	TOF 5 mg	22	NR	NR	NR	3(12)	NR	NR	NR
	non-TNF biologics	20	NR	NR	NR	NR	NR	NR	NR
Charles-Schoeman 2016[12]	bDMARD naïve: TOF 5 mg	1071	0-24	NR	131 (12.2)	96 (9.1)	NR	NR	NR
	bDMARD naïve: cDMARD	651	0-6	NR	98 (15.0)	66 (10.1)	NR	NR	NR
	bDMARD-IR: TOF 5 mg	259	0-24	NR	34 (13.0)	39 (14.8)	NR	NR	NR
	bDMARD-IR: cDMARD	193	0-6	NR	37 (19.0)	37 (18.9)	NR	NR	NR
Kremer 2012[117]	TOF 5 mg	71	24	47 (66.2)	4(5.6)	3(4.2)	3(4.2)	5(7.0)	NR
	Placebo	51	24	29 (56.9)	0 (0)	3 (5.9)	1(1.4)	2(2.5)	NR
Upadacitinib (UPA)									
	UPA 15 mg	217	14	103 (47.5)	11 (5.0)	8 (3.7)	NR	NR	1 (<1)*
	UPA 30 mg	215	14	105 (48.8)	6 (2.8)	6 (2.8)	NR	NR	0 (0)

Study	Intervention	N	Time Point (Weeks)	Any Adverse Event, n (%)	Serious Adverse Event, n (%)	Adverse Event Leading to Discontinuation, n (%)	Nausea, n (%)	Nasopharyngitis, n (%)	Venous Thromboembolism, n (%)
SELECT-MONOTHERAPY[118] Smolen, 2019	cDMARD	216	14	102 (47.2)	6 (2.8)	6 (2.8)	NR	NR	0 (0)
SELECT-COMPARE[6] Fleischmann, 2018	UPA 15 mg + cDMARD	651	26	417 (64.2)	24 (3.7)	23 (3.5)	NR	NR	All*: 2 (0.3); PE: 1(0.2); DVT: 1 (0.2)
	ADA 40 mg + cDMARD	327	26	197 (60.2)	14 (4.3)	20 (6.1)	NR	NR	All*: 3 (0.9); PE: 3 (0.9); DVT: 0 (0)
	PBO + cDMARD	651	26	347 (53.2)	19 (2.9)	15 (2.3)	NR	NR	All*: 1 (0.2); PE: 1 (0.2); DVT: 0 (0)
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	12	91 (55.4)	8 (4.9)	4 (2.4)	NR	NR	PE: 1 (0.6)
	UPA 30 mg + cDMARD	165	12	111 (67.3)	12 (7.3)	15 (9.1)	NR	NR	0 (0)
	Placebo + cDMARD	169	12	95 (56.2)	0 (0)	9 (5.3)	NR	NR	0 (0)
SELECT-NEXT[7] Burmester, 2018	UPA 15 mg + cDMARD	221	12	125 (56.6)	9 (4.1)	7 (3.2)	16 (7.2)	12 (5.4)	0 (0)
	UPA 30 mg + cDMARD	219	12	118 (53.9)	6 (2.7)	13 (5.9)	3 (1.4)	13 (5.9)	0 (0)
	cDMARD	221	12	108 (48.9)	5 (2.3)	7 (3.2)	7 (3.2)	9 (4.1)	0 (0)

N/A: not available, n: number, N: total number, NR: not reported, TEAE: treatment emergent adverse event

*(Adjudicated) pulmonary embolism deemed unrelated to study drug.

Table D16. Harms II

Study	Intervention	N	Time Point (Weeks)	Upper Resp. Infection, n (%)	Malignancy, n (%)	Death, n (%)	Headache, n (%)	Infections, n (%)			
								Serious	Opportunistic	Herpes Zoster Virus	TB
Baricitinib (BAR)											
RA-BUILD[18] Dougados 2017	BAR 2 mg + cDMARD	229	0-12	NR	0 (0)	0 (0)	NR	1 (<1)	45 (20.0)	3 (1.0)	0 (0)
	cDMARD	228	0-12	NR	0 (0)	2 (0.9)	NR	3 (1.0)	53 (23.0)	0 (0)	0 (0)
	BAR 2 mg + cDMARD	229	0-24	14 (6.0)	0 (0)	0 (0)	15 (7.0)	2 (<1)	70 (31.0)	4 (2.0)	0 (0)
	cDMARD	228	0-24	18 (8.0)	0 (0)	2 (0.9)	8 (4.0)	4 (2.0)	79 (35.0)	0 (0)	0 (0)
RA-BEACON[15] Genovese 2016	BAR 2 mg + cDMARD	174	0-12	NR	NR	0 (0)	NR	3 (2)	61 (35.0)	2 (1.0)	NR
	cDMARD	176	0-12	NR	NR	0 (0)	NR	3 (2)	35 (20.0)	1 (<1)	NR
	BAR 2 mg + cDMARD	174	0-24	16 (9.0)	NR	0 (0)	17 (10.0)	4 (2)	76 (44.0)	2 (1.0)	NR
	cDMARD	176	0-24	8 (5.0)	NR	0 (0)	11 (6.0)	5 (3)	55 (31.0)	2 (1.0)	NR
RA-BEYOND[109,110] (Abstract) Genovese 2017 Van Der Heijde 2019	Continued BAR 4 mg	147	0-12	NR							
	Stepped down to BAR 2 mg	146	0-12								
Keystone 2015[111]	BAR 2 mg	52	0-12	NR	NR	NR	NR	NR	NR	NR	NR
	cDMARD	98	0-12	NR	NR	NR	NR	NR	NR	NR	NR
	BAR 2 mg	52	0-24	NR	NR	NR	NR	NR	NR	NR	NR
	cDMARD	98	N/A								
Tofacitinib (TOF)											
ORAL Sync[8] Kremer 2013	TOF 5 mg + cDMARD	315	52	40 (12.3)	NR	NR	16 (4.9)	NR	NR	9 (4.2) MTX w/o leflunomide	NR
	Placebo (advanced to TOF 5 mg at 6 months)	79	52	7 (12.6)	NR	NR	8 (14.4)	NR	NR	2 (6.8) MTX alone	NR
ORAL Step[14] Burmester 2013	TOF 5 mg + cDMARD	133	0-12	5 (3.8)	NR	0 (0)	3 (2.3)	0 (0)	0 (0)	NR	NR
	Placebo	132	0-12	4 (3.0)	NR	0 (0)	1 (0.8)	0 (0)	0 (0)	NR	NR

Study	Intervention	N	Time Point (Weeks)	Upper Resp. Infection, n (%)	Malignancy, n (%)	Death, n (%)	Headache, n (%)	Infections, n (%)			
								Serious	Opportunistic	Herpes Zoster Virus	TB
	(advanced to TOF 5 mg)										
ORAL Scan[11] Van der Heijde 2013	TOF 5 mg + cDMARD	321	0-12	9 (2.8)	NR	2 (0.6)	18 (5.6)	2 (0.6)	NR	3 (0.9)	NR
	Placebo (advanced to TOF 5 mg at 6 months)	160	0-12	5 (3.1)	NR	0 (0)	3 (1.9)	0 (0)	NR	0 (0)	NR
ORAL Strategy[10] Fleischmann 2017	TOF 5 mg	384	52	25 (7)	1 (<1)	2 (1)	NR	6 (2)	2 (1)	1/69 (1)	0 (0)
	TOF 5 mg + cDMARD	376	52	37 (10)	0 (0)	0 (0)	NR	10 (3)	1 (<1)	2/75 (3)	2 (1)
	ADA + cDMARD	386	52	29 (8)	0 (0)	0 (0)	NR	6 (2)	2 (1)	0/27 (0)	0 (0)
ORAL Standard[9] Van Vollenhoven 2013	TOF 5 mg + cDMARD	204	12	9 (4.4)	NR	1 (<1)	8 (3.9)	3(1.5)	NR	0 (0)	NR
	ADA 40 mg + cDMARD	204	12	7 (3.4)	NR	1 (<1)	5 (2.5)	0 (0)	NR	0 (0)	NR
	Placebo followed by TOF 5 mg + cDMARD	56	12	1 (0.9)	NR	NR	2 (1.9)	1(0.9)	NR	0 (0)	NR
Nakamura 2018[116]	TOF 5 mg	22	NR	NR	NR	NR	NR	NR	NR	NR	NR
	Non-TNF biologics	20	NR	NR	NR	NR	NR	NR	NR	NR	NR
Charles-Schoeman 2016[12]	bDMARD Naïve: TOF 5 mg	1071	0-24	NR	NR	7 (0.6)	NR	32 (3.4)	NR	43 (4.0)	0 (0)
	bDMARD Naïve: cDMARD	651	0-6	NR	NR	5 (0.7)	NR	13 (2.0)	NR	13 (2.0)	0 (0)
	bDMARD-IR: TOF 5 mg	259	0-24	NR	NR	1(1.2)	NR	6 (2.3)	NR	5.4	0 (0)
	bDMARD-IR: cDMARD	193	0-6	NR	NR	0 (0)	NR	0 (0)	NR	0 (0)	0 (0)
Kremer 2012[117]	TOF 5 mg	71	0-24	5(7.0)	NR	0 (0)	2(2.8)	1(1.4)	16 (22.5)*	NR	NR
	Placebo	69	0-12	2(2.9)	NR	0 (0)	1(1.4)	0 (0)	3 (5.9)*	NR	NR
Upadacitinib (UPA)											
	UPA 15 mg	217	14	NR	2 (0.9)	1 (0.5)	NR	1 (0.5)	0 (0)	3 (1.4)	0 (0)

Study	Intervention	N	Time Point (Weeks)	Upper Resp. Infection, n (%)	Malignancy, n (%)	Death, n (%)	Headache, n (%)	Infections, n (%)			
								Serious	Opportunistic	Herpes Zoster Virus	TB
SELECT-MONTHERAPY[118] Smolen, 2019	UPA 30 mg	215	14	NR	0 (0)	0 (0)	NR	0	3 (1.4)	6 (2.8)	0 (0)
	cDMARD	216	14	NR	1 (0.5)	0 (0)	NR	1 (0.5)	1 (0.5)	1 (0.5)	0 (0)
SELECT-COMPARE[6] Fleischmann, 2018	UPA 15 mg + cDMARD	651	26	NR	0 (0)	0 (0)	NR	12 (1.8)	4 (0.6)	5 (0.8)	NR
	ADA 40 mg + cDMARD	327	26	NR	1 (0.3)	2 (0.6)	NR	5 (1.5)	1 (0.3)	1 (0.3)	NR
	PBO + cDMARD	651	26	NR	2 (0.3)	2 (0.3)	NR	5 (0.8)	4 (0.6)	3 (0.5)	NR
SELECT-NEXT[7] Burmester, 2018	UPA 15 mg + cDMARD	221	12	12 (5)	0 (0)	0 (0)	9 (4)	1 (<1)	0 (0)	3 (1)	0 (0)
	UPA 30 mg + cDMARD	219	12	12 (6)	2 (<1)	1 (0.5)	7 (3)	3 (1)	3 (1)	6 (3)	0 (0)
	cDMARD	221	12	9 (4)	0 (0)	1	12 (5)	1 (<1)	1 (<1)	1 (<1)	0 (0)
SELECT-BEYOND[13] Genovese 2018	UPA 15 mg + cDMARD	164	12	NR	1 (0.6)	0 (0)	NR	1 (0.6)	1 (0.6)	1 (0.6)	NR
	UPA 30 mg + cDMARD	165	12	NR	2 (1.2)	1 (0.6)	NR	4 (2.4)	2 (1.2)	4 (2.4)	NR
	Placebo + cDMARD	169	12	NR	0 (0)	0 (0)	NR	(0)	0 (0)	1 (0.6)	NR
SELECT-NEXT[7] Burmester, 2018	UPA 15 mg + cDMARD	221	12	12 (5.4)	0 (0)	0 (0)	9 (4.1)	1 (0.5)	0 (0)	1 (0.5)	0 (0)
	UPA 30 mg + cDMARD	219	12	12 (5.5)	2 (0.9)	0 (0)	7 (3.2)	3 (1.4)	3 (1.4)	2 (0.9)	0 (0)
	cDMARD	221	12	9 (4.1)	0 (0)	0 (0)	12 (5.4)	1 (0.5)	1 (0.5)	1 (0.5)	0 (0)

N/A: not available, n: number, N: total number, NR: not reported, Resp.: respiratory, TB: tuberculosis *Infections and infestations.

Table D17. Included Studies

Trial Name (Author & Year of Publication)	Study Sponsor	Study Design and Duration of Follow-Up	Geographic Location of Study	Interventions (n) & Dosing Schedule	Inclusion and Exclusion Criteria	Primary Outcomes
PLANETRA 2013[123]	CELLTRION Inc, Incheon, Republic of Korea	RCT, multicenter, double-blind, Phase III Follow-up: 30 weeks	100 centers across 19 countries in Europe, Asia, Latin America, and Middle East	1) CT-P13+MTX (n=302) 2) INX+MTX (n=304)	Active RA according to 1987 ACR criteria for ≥1 year prior to screening; ≥6 swollen and ≥6 tender joints; at least two of the following: morning stiffness lasting ≥45 min; serum CRP concentration >2.0 mg/dl and ESR >28 mm/h despite MTX therapy for ≥3 months (stable dose of 12.5–25 mg/week for ≥4 weeks prior to screening).	Demonstrate equivalent effect of CT-P13 to INX at week 30 as determined by ACR20 response
PLANETRA Extension Study 2017[31]	CELLTRION Inc, Incheon, Republic of Korea	Open-label single-arm extension study following a 52-week RCT Follow-up: 1 year	69 centers in 16 countries in Europe, Asia, Latin America, and the Middle East	1) IFX-bio-maintenance group (n=158) 2) IFX-bio-switch group (n=144)	18-75 years old with active RA for ≥1 year; inadequate response to ≥3 months use of MTX and received stable dose of MTX for ≥4 weeks before study.	Proportion of patients meeting ACR20, ACR50 and ACR70 criteria
Codreanu 2018[78]	Egis Pharmaceuticals PLC	Observational, multicenter Follow-up: 24 weeks	Selected countries in eastern and central Europe	CT-P13	Adults aged 18 and older with RA diagnosed by ACR or EULAR classification criteria	DAS28-CRP at weeks 12 and 24

RCT: randomized-controlled trial, INX: infliximab, ACR: American College of Rheumatology, EULAR: European League Against Rheumatism, RA: rheumatoid arthritis, MTX: methotrexate

Table D18. Study Quality of Inflectra Studies[31,78,123,124]

Study	Adequate Randomization	Allocation Concealment	Patient Blinding	Staff Blinding	Outcome Adjudication Blinding	Completeness of Follow Up	Intention to Treat Analysis	Incomplete Data Addressed	Selective Outcome Reporting	Industry Funding	Freedom from Bias	Overall Quality
Biosimilar												
PLANETRA 2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Good
PLANETRA Extension Study 2017	No	No	No	No	No	No	NA	Yes	No	Yes	No	Poor
Codreanu 2018	No	No	No	No	No	Yes	NA	No	No	No	No	Poor

Table D19. Baseline Characteristics[31,78,123,124]

Study	PLANETRA 2013		PLANETRA OLE 2016		Codreanu 2018
Intervention	**CT-P13**	**INX**	**Maintenance**	**Switch**	**RA**
N	302	304	158	144	81
Age, Median (range)	50 (18-75)	50 (21-74)	50 (18-73)	49 (23-74)	55 (NR)
Female, no(%)	245 (81.1)	256 (84.2)	125 (79.1)	122 (84.7)	55.8 ± 11.4*
Weight (kg)	69 (36.5-134)	68 (36-136)	71.0 (43.-134.)	68.3 (44.3-125)	NR
Duration of Prior MTX therapy (weeks)	97.7 (141.2)	89.4 (96.5)	NR	NR	NR
DAS28-CRP (SD)	5.9 (0.8)	5.8 (0.9)	5.8 (3.4-8.1)	5.8 (2.9-7.9)	5.8 (1.0)
HAQ	1.6 (0.6)	1.6 (0.6)	NR	NR	NR
Patient's Assessment of Pain	65.9 (17.4)	65.5 (17.2)	NR	NR	NR
Patient's Global Assessment of Disease Activity	65.7 (17.2)	65.4 (17)	NR	NR	NR
Physician Global Assessment of Disease Activity	64.7 (14.3	65.0 (13.5)	NR	NR	NR

INX: infliximab, kg: kilogram, MTX: methotrexate, NR: not reported, OLE: open-label extension, SD: standard deviation

*Mean (SD).

Table D20. Outcomes[31,78,123,124]

Study	PLANETRA 2013				PLANETRA OLE 2016			
Intervention	CT-P13	INX	CT-P13	INX	Maintenance (ITT)	Switch (ITT)	Maintenance (ITT)	Switch (ITT)
N	302	304	302	304	158	144	158	144
Timepoint	30 weeks		14 weeks		102 Weeks		54 weeks	
ACR20 (%)	60.9 (ITT) 73.4 (PP)	58.6 (ITT) 69.7 (PP)	ITT NR 72.6 (PP)	ITT NR 65.3 (PP)	71.7	71.8	77.0	77.5
ACR50 (%)	42.3 (PP)	40.6 (PP)	39.5 (PP)	33.9 (PP)	48.0	51.4	46.1	50.0
ACR70 (%)	20.2 (PP)	17.9 (PP)	16.5 (PP)	13.5 (PP)	24.3	26.1	22.4	23.9
CDAI, % Remission	NR	NR	NR	NR	11.8	16.9	10.5	13.4
DAS28-ESR, % Remission	36	27	24	18	13.8	12.7	14.5	12.0
DAS28-CRP, % Remission	61	56	46	44	27.0	31.7	28.3	29.6
HAQ, Mean Change from Baseline	-0.6 (0.6)	-0.5 (0.5)	-0.6 (0.6)	-0.5 (0.5)	-0.64	-0.63	-0.63	-0.58
Patient Assessment of Pain, Mean Change from Baseline	-29.5 (25.5)	-27.8 (24.9)	-29.5 (23.2)	-27.2 (23.2)	-32.4¥	-35.0¥	-31.3¥	-32.4¥
Patient Global Assessment of Disease Activity, Mean Change from Baseline	-28.1 (25.9)	-27.0 (25.6)	-29.5 (22.1)	-25.5 (24.4)	-31.8¥	-34.0¥	-31.4¥	-30.4¥
Physician Global Assessment of Disease Activity, Mean Change from Baseline	-35.6 (20.6)	-35.3 (21.2)	-35.4 (19.3)	-33.7 (19.5)	-38.6¥	-39.6¥	-39.5¥	-39.4¥

ACR: American College of Rheumatology, CDAI: Clinical Disease Activity Index, DAS: Disease Activity Score, HAQ: Health Assessment Questionnaire, INX: infliximab, ITT: intent-to-treat, PP: per protocol

¥Change from 52-week assessment.

Table D21. Harms[31,78,123,124]

Study	PLANETRA		PLANETRA OLE		Codreanu 2018
Intervention	CT-P13	INX	Maintenance	Switch	Entire Study Population
N	302	304	158	144	151
SAEs	30 (10)	21 (7)	12 (7.5)	13 (9.1)	7 (NR)
Deaths	0 (0)	0 (0)	NR	NR	NR
Overall TEAEs	181 (60.1)	183 (60.8)	85 (53.5)	77 (53.8)	NR
TEAEs Related to Treatment	106 (35.2)	108 (35.9)	35 (22)	27 (18.9)	5 (0.7)
TEAEs Leading to Discontinuation	NR	NR	16 (10.1)	8 (5.6)	7 (NR)
Latent Tuberculosis	13 (4.3)	14 (4.7)	9 (5.7)	4 (2.8)	NR
Upper Respiratory Tract Infection	4 (1.3)	4 (1.3)	6 (3.8)	3 (2.1)	1 (NR)
Urinary Tract Infection	4 (1.3)	4 (1.3)	2 (1.3)	2 (1.4)	NR
Nasopharyngitis	6 (2.0)	4 (1.3)	NR	NR	NR
Herpes Zoster	1 (0.3)	3 (1.0)	NR	NR	NR
Infusion-Related Reaction	20 (6.6)	25 (8.3)	11 (6.9)	4 (2.8)	NR
Headache	4 (1.3)	6 (2.0)	NR	NR	NR
Flare in RA Activity	7 (2.3)	4 (1.3)	NR	NR	NR

INX: infliximab, kg: kilogram, MTX: methotrexate, NR: not reported, OLE: open-label extension, RA: rheumatoid arthritis, SD: standard deviation

*Mean(SD).

Appendix E. Comparative Value Supplemental Information

Table E1. Impact Inventory

Sector	Type of Impact (Add Additional Domains, as Relevant)	Included in This Analysis from… Perspective?		Notes on Sources (if Quantified), Likely Magnitude & Impact (if Not)
		Health Care Sector	Societal	
Formal Health Care Sector				
Health Outcomes	Longevity effects	X	X	
	Health-related quality of life effects	X	X	
	Adverse events	X	X	
Medical Costs	Paid by third-party payers	X	X	
	Paid by patients out-of-pocket	☐	☐	
	Future related medical costs	☐	☐	
	Future unrelated medical costs	☐	☐	
Informal Health Care Sector				
Health-Related Costs	Patient time costs	NA	☐	
	Unpaid caregiver-time costs	NA	☐	
	Transportation costs	NA	☐	
Non-Health Care Sectors				
Productivity	Labor market earnings lost	NA	X	
	Cost of unpaid lost productivity due to illness	NA	X	
	Cost of uncompensated household production	NA	☐	
Consumption	Future consumption unrelated to health	NA	☐	
Social Services	Cost of social services as part of intervention	NA	☐	
Legal/Criminal Justice	Number of crimes related to intervention	NA	☐	
	Cost of crimes related to intervention	NA	☐	
Education	Impact of intervention on educational achievement of population	NA	☐	
Housing	Cost of home improvements, remediation	NA	☐	
Environment	Production of toxic waste pollution by intervention	NA	☐	
Other	Other impacts (if relevant)	NA	☐	

Adapted from Sanders et al. 2016[125]

NA: not applicable

hēRo3

hēRo3 compiles information and data that users enter into a browser describing the structure and estimated parameters of a model, sends it to the cloud platform where necessary calculations are performed in heRomod, and then parses information received from the modeling package to various output displays, including Markov traces, bar charts, area charts, tornado diagrams, waterfall charts, efficiency frontiers, and hexbin and contour plots, as well as tabular displays. hēRo3 effectively allows users to build and run models in the programming language R, even if they have had limited or no experience programming in R. hēRo3 also generates an Excel workbook with every model that provides a detailed listing of all input variables, intermediate calculations, and final output on a cycle-by-cycle basis to facilitate model checking and auditing.

Description of the evLYG Calculations

The cost per evLYG considers any extension of life at the same "weight" no matter what treatment is being evaluated. Below are the stepwise calculations used to derive the evLYG.

1) First, we attribute a utility of 0.851, the age- and gender-adjusted utility of the general population in the US that is considered healthy.[126]
2) For each cycle (Cycle I) in the model where using the intervention results in additional years of life gained, we multiply this general population utility with the additional life years gained (ΔLYG).
3) We sum the product of the LYs and average utility (cumulative LYs/cumulative QALYs) for Cycle I in the comparator arm with the value derived in Step 2 to derive the equal value of life years (evLY) for that cycle.
4) If no life years were gained using the intervention versus the comparator, we use the conventional utility estimate for that Cycle I.
5) The total evLY is then calculated as the cumulative sum of QALYs gained using the above calculations for each arm.
6) We use the same calculations in the comparator arm to derive its evLY.
7) Finally, the evLYG is the incremental difference in evLY between the intervention and the comparator arms.

DAS28-ESR to DAS28-CRP Conversion for Tofacitinib and Conventional DMARD

We did not find any published data on DAS28-CRP outcomes at three months for tofacitinib and its conventional DMARD comparator. Some trials (SELECT-MONOTHERAPY, RA-BUILD, RA-BEACON, and ORAL Step) simultaneously reported DAS28-ESR and DAS28-CRP outcomes for upadacitinib, baricitinib, and tofacitinib, and their respective comparators. A simple average of the disease activity proportions using DAS28-CRP and DAS28-ESR data resulted in an approximate 2x and 1.5x

ratio of CRP to ESR in the TIM arms, while the conventional DMARD arms showcased more variability. In the absence of DAS28-CRP trial data at three months, we applied these ratios to the DAS28-ESR data to derive DAS28-CRP data to tofacitinib and its conventional DMARD comparator (Table E2). We multiplied by 2x for remission and 1.5x for LDA, with the MDA/HDA proportion derived as the remainder when remission and LDA proportions were summed.

Table E2. DAS28-ESR to DAS28-CRP Conversion for Tofacitinib and Conventional DMARD at Three Months

	Proportion of Patients Achieving Different Categories of Disease Activity by DAS28 at Three Months*					
	Tofacitinib			cDMARD		
	<2.6 (Remission)	2.6 to ≤3.2 (LDA)	>3.2 (MDA and HDA)	<2.6 (Remission)	2.6 to ≤3.2 (LDA)	>3.2 (MDA and HDA)
DAS28-ESR	7%	9%	83%	2%	2%	96%
Adjusted DAS28-CRP	15%	14%	71%	5%	3%	92%

cDMARD: conventional disease-modifying antirheumatic drug, CRP: c-reactive protein, DAS28: Disease Activity Score 28, ESR: erythrocyte sedimentation rate, HDA: high disease activity, LDA: low disease activity, MDA: moderate disease activity

*Mutually exclusive categories.

Undiscounted Outcomes

Table E3. Key Undiscounted Health and Economic Outcomes for Upadacitinib versus Adalimumab

Treatment	Drug Cost* (Line One)	Total Cost	LYs	QALYs
Upadacitinib + cDMARD	$21,600	$48,700	0.996	0.71
Adalimumab + cDMARD	$15,900	$48,100	0.996	0.70

cDMARD: conventional disease-modifying antirheumatic drug, evLYG: equal value of life years gained, LY: life year, QALY: quality-adjusted life year

*Only the costs of TIM; does not include cDMARD cost.

Table E4. Key Undiscounted Incremental Cost-Effectiveness Ratios for Upadacitinib versus Adalimumab

Treatment	Cost per LY Gained*	Cost per QALY Gained
Upadacitinib vs. Adalimumab	--	$92,000

LY: life year, QALY: quality-adjusted life year

*No difference up to three decimal places in LYs between the two TIMs led to implausible incremental cost-effectiveness ratios.

Table E5. Key Undiscounted Health and Economic Outcomes for Adalimumab versus Conventional DMARD

Treatment	Drug Cost (Line One)	Total Cost	LYs	QALYs
Adalimumab + cDMARD	$15,900*	$48,100	0.996	0.70
cDMARD	$5,600	$37,000	0.996	0.69

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year
*The cost of TIM alone.

Table E6. Key Undiscounted Health and Economic Outcomes for Tofacitinib versus Conventional DMARD

Treatment	Drug Cost (Line One)	Total Cost	LYs	QALYs
Tofacitinib + cDMARD	$12,900*	$45,200	0.996	0.70
cDMARD	$5,600	$38,800	0.996	0.69

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year
*The cost of TIM alone.

One-Way Sensitivity Analyses

Table E7. One-Way Sensitivity Analyses of Upadacitinib versus Adalimumab for QALY Outcomes

Parameters	Low Input	High Input	Low Value	High Value
Baseline HAQ	1.76	1.44	0.664	0.732
Utility (Mapped from HAQ)	Multiple	Multiple	0.683	0.715
Probability of Remission with Upadacitinib at Three Months	26.5%	31.5%	0.698	0.701
Subsequent TIM Efficacy Decrement	0.75	0.92	0.698	0.701
Rate of Serious Infection with TIMs	0.027	0.046	0.699	0.700
Disutility of Serious Infection	0.14	0.17	0.699	0.700

HAQ: Health Assessment Questionnaire, TIM: targeted immune modulator

Table E8. One-Way Sensitivity Analyses of Upadacitinib versus Adalimumab for Total Cost Outcomes

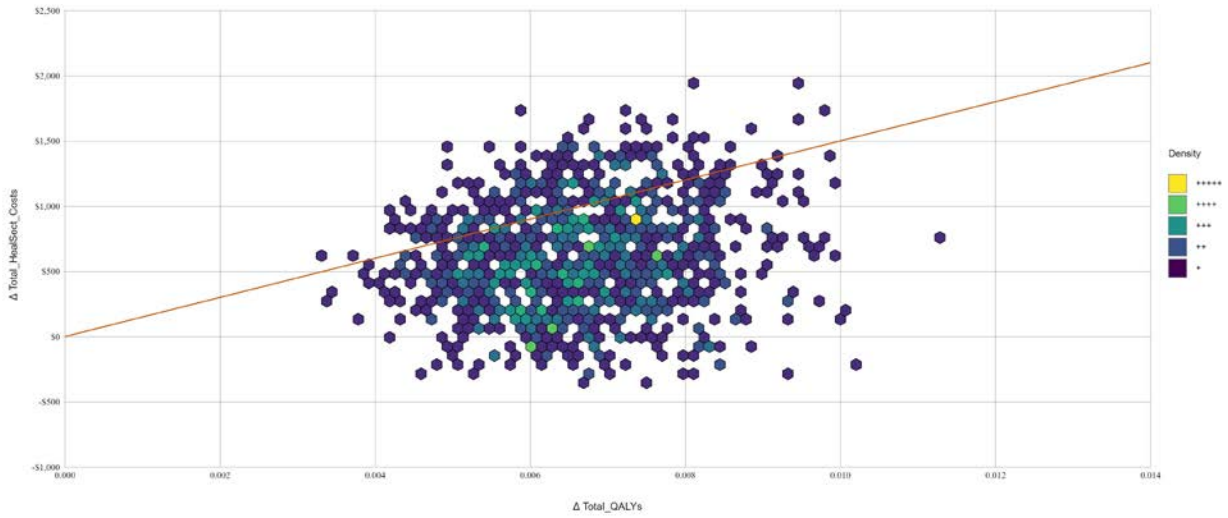
Parameters	Low Input	High Input	Low Value	High Value
Cost of Market Basket TIM Treatment	--	--	$44,300	$52,000
Hospitalization Rate (HAQ Range Dependent)	Multiple	Multiple	$47,000	$51,300
Baseline HAQ	1.76%	1.44%	$48,400	$47,900
Cost of Hospitalization	$1,976	$2,964	$47,900	$48,400
Probability of Remission with Upadacitinib at Three Months	26.50%	31.50%	$48,100	$48,300
Cost of Physician's Office Visit	$60.26	$90.38	$48,100	$48,200
Cost of Hepatitis Panel	$52.12	$78.18	$48,100	$48,200
Cost of TB Screening	$67.86	$101.80	$48,100	$48,200
Subsequent TIM Efficacy Decrement	0.75	0.92	$48,200	$48,100
Cost of Lipid Panel	$11.69	$17.53	$48,100	$48,200
Cost of Metabolic Panel	$8.92	$13.38	$48,100	$48,200
Cost of CBC	$8.52	$12.78	$48,100	$48,200

CBC: complete blood count, HAQ: Health Assessment Questionnaire, TB: tuberculosis, TIM: targeted immune modulator

Probabilistic Analyses

Scatter Plot

Figure E1. Probabilistic Analyses: Upadacitinib versus Adalimumab: Incremental Cost-Utility Ratio HexBin



Cost-Effectiveness Plane

Upadacitinib versus adalimumab over 1,000 probabilistic simulation runs:

- More costly, more effective: 92.9%
- More costly, less effective: 0%
- Less costly, less effective: 0%
- Less costly, more effective: 7.1%

Scenario Analyses

Table E9. Cost-Effectiveness Results for Upadacitinib versus Adalimumab from a Modified Societal Perspective

Treatment	Drug Cost* (Line One)	Total Cost	LYs	QALYs	Cost per QALY Gained	Cost per LY Gained†
Upadacitinib + cDMARD	$72,100	$124,000	0.985	0.699	$92,000	--
Adalimumab + cDMARD	$48,500	$97,900	0.985	0.693	--	--

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year
*Only the costs of TIM; does not include cDMARD cost.
†No difference up to three decimal places in LYs between the two TIMs led to implausible incremental cost-effectiveness ratios.

Lifetime Time Horizon

Table E10. Cost-Effectiveness Results Using a Lifetime Time Horizon

Treatment	Total Cost	LYs	QALYs	Incremental Cost per LY Gained*	Incremental Cost per QALY Gained
Upadacitinib + cDMARD	$808,000	17.44	13.23	$1.1 million	$240,000
Adalimumab + cDMARD	$805,000	17.43	13.22		

cDMARD: conventional disease-modifying antirheumatic drug, LY: life year, QALY: quality-adjusted life year

Appendix F. Public Comments

This section includes summaries of the public comments prepared for the CTAF Public Meeting on Monday, December 9 in Oakland, CA. These summaries were prepared by those who delivered the public comments at the meeting and are presented in order of delivery. One speaker did not submit a summary of their public comment. A video recording of all comments can be found here, beginning at minute 01:11:01

Marc Jensen, PharmD, Pfizer
Senior Director, Field Medical, Inflammation & Immunology

In this public comment, we would like to emphasize that the medical value of Xeljanz has been demonstrated by numerous Phase III and Phase IV clinical trials, as well as real-world evidence collected during the seven years that Xeljanz has been prescribed to more than 280,000 patients around the world.

In the public responses to ICER, Pfizer and other external stakeholders have commented on the methodological flaws in the cost-effectiveness model used and the inconsistencies in interpretations drawn by ICER from the model results. The current cost-effectiveness model required ICER to make substantial assumptions not supported by the literature resulting in an inability to compare tofacitinib to adalimumab. ICER's conclusion that there are not sufficient data to compare tofacitinib to adalimumab is not correct as demonstrated by peer reviewed published cost-effectiveness models comparing tofacitinib to adalimumab. Moreover, there is a wealth of published data on tofacitinib, including several studies with direct comparisons to adalimumab, just not for the particular timepoint ICER selected to assess efficacy.

Beyond these issues with the cost-effectiveness model, what we would like to emphasize is the wealth of data supporting the medical value of tofacitinib, which was approved in the US in 2012 as the first oral JAK inhibitor to treat RA. ICER did highlight this in the report. Tofacitinib has been studied in over 50 clinical trials totaling 36,000 patient years to date. Within the report, tofacitinib demonstrated and incremental net benefit compared to conventional DMARDs and the comparable net health benefit versus adalimumab. Both these ratings had a with a high degree of certainty according to the report.

While there are no head-to-head data comparing tofacitinib and upadacitinib, AbbVie recently published a matching adjusted indirect comparison. Once those results are adjusted for multiplicity, there are no significant differences between ACR20, ACR50, and ACR70 at three months between upadacitinib and tofacitinib. There are also no significant differences between products in SDAI clinical remission, CDAI clinical remission, or DAS28-CRP clinical remission at six months.

Furthermore, tofacitinib has also been widely studied in the real-world setting. These data were not covered fully in the ICER report. In partnership with the CORRONA registry, several real-world studies have looked at the safety and effectiveness of tofacitinib in patients with RA. These results demonstrate that the efficacy and safety seen in the clinical trial program are consistent in a real-world setting. Given that tofacitinib has been available longest of the JAK inhibitors and used by over 280,000 patients worldwide, it is the only JAK inhibitor that can demonstrate robust real-world effectiveness and true impact on patients outside of the tightly controlled clinical trial setting. This is very important if we would like to take a patient-centric perspective.

Many patients have benefited from treatment with tofacitinib in both the clinical trial and real-world setting with a clear demonstration of its medical value. Given the limitations in the ICER cost-effectiveness analysis, no decisive conclusions can be drawn. Tofacitinib is an important treatment option for appropriate patients suffering from this devastating disease.

Appendix G. Conflict of Interest Disclosures

Tables G1 through G3 contain conflict of interest (COI) disclosures for all participants at the December 9, 2019 public meeting of CTAF.

Table G1. ICER Staff and Consultants and COI Disclosures

Name	Organization	Disclosures
Pamela Bradt, MD, MPH	Institute for Clinical and Economic Review	*
Rick Chapman, PhD, MS	Institute for Clinical and Economic Review	*
Laura Cianciolo	Institute for Clinical and Economic Review	*
Monica Frederick	Institute for Clinical and Economic Review	*
Serina Herron-Smith	Institute for Clinical and Economic Review	*
Steven D. Pearson, MD, MSc	Institute for Clinical and Economic Review	*
Jeffrey A. Tice, MD	University of California, San Francisco	*
Tia Tilson	Institute for Clinical and Economic Review	*
Judith Walsh, MD, MPH	University of California, San Francisco	*

*No relevant conflicts of interest to disclose, defined as more than $10,000 in health care company stock or more than $5,000 in honoraria or consultancies during the previous year from health care manufacturers or insurers.

Table G2. CTAF Panel Member Participants and COI Disclosures

Name	Organization	Disclosures
Ralph G. Brindis, MD, MPH, MACC, FSCAI, FAHA	University of California, San Francisco	*
Felicia Cohn, PhD	Kaiser Permanente	*
Robert Collyar	Independent Patient Advocate	*
Rena K. Fox, MD	University of California, San Francisco	*
Kimberly Gregory, MD, MPH	Cedars-Sinai Medical Center	*
Paul Heidenreich, MD, MS	Stanford University	*
Jeffrey Hoch, PhD	University of California, Davis	*
Neal Kohatsu, MD, MPH, FACPM	University of California, Davis	*
Annette Langer-Gould, MD, PhD	Kaiser Permanente	*
Sei Lee, MD	University of California, San Francisco	*
Elizabeth J. Murphy, MD, DPhil	University of California, San Francisco	*
Rita F. Redberg, MD, MSc, FACC	University of California, San Francisco	*
Robert Rentschler, MD	Beaver Medical Group	*
Richard Seiden, JD	Foley & Lardner LLP	*†
Alexander Smith, MD, MPH	University of California, San Francisco	*

*No relevant conflicts of interest to disclose, defined as more than $10,000 in healthcare company stock or more than $5,000 in honoraria or consultancies during the previous year from health care manufacturers or insurers.

†Richard Seiden did not participate in voting due to his prior treatment with one of the agents under review.

Table G3. Policy Roundtable Participants and COI Disclosures

Name	Organization	Disclosures
Happy Chan, DO, FACR	Blue Shield of California	Employee of Blue Shield of California.
Andrew L. Concoff, MD, FACR, CAQSM	United Rheumatology	Served as a consultant and/or speaker to Flexion Pharmaceuticals, Exagen, Inc., and UCB.
Peggy Ehling	Arthritis Foundation Local Leadership Board Los Angeles	No conflicts of interest to disclose.
Julie Eller	Arthritis Foundation	The Arthritis Foundation receives more than 25% of its funding from health care companies.
Kirstin Griffing MD, MS, FACR	Eli Lilly and Company	Employee of Eli Lilly and Company.
Marc Jensen, PharmD	Pfizer	Employee of Pfizer.
Christopher Phillips, MD	American College of Rheumatology; Paducah Rheumatology	Served as a PI on several upadacitinib trials conducted through an independent multi-specialty clinical research center.
John S. Yao, MD, MPH, MBA, MPP, CPC, FACP	Anthem Blue Cross	Employee of Anthem Blue Cross.