Janus Kinase Inhibitors for Rheumatoid Arthritis: Effectiveness and Value

Research Protocol

July 9, 2019

Institute for Clinical and Economic Review

©Institute for Clinical and Economic Review, 2019

Table of Contents

©Institute for Clinical and Economic Review, 2019

Background, Objectives, and Research Questions

Background

Rheumatoid arthritis (RA) is the most common autoimmune inflammatory arthritis in adults, affecting between 1.3 and 1.8 million Americans.[1,2] RA is more common in women and may occur at any age, with peak incidence occurring at ages 50-60 years.[3] 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.[3] 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.[4] The course of RA may be complicated by cardiac, hematologic, and other extra-articular manifestations.[3] Historically, RA was associated with both progressive disability and a shortened lifespan, although improvements in diagnosis as well as aggressive use of disease-modifying anti-rheumatic drugs (DMARDs) have greatly improved prognosis in the past 20 years.[5]

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).[3] However, only about 50% of patients treated with methotrexate alone will receive sufficient reduction in disease activity or remission of symptoms.[3] Over the past 18 years, the introduction of TIMs has greatly improved prognosis for many RA patients. Agents with indications for RA include inhibitors or antagonists of multiple mediators of the inflammatory cascade, including tumor necrosis factor (TNF), the 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.[6]

Uncertainty remains, however, regarding the relative effectiveness of the different types of TIMs as well as the appropriate sequence of initial and subsequent TIM therapy. In addition, there are long-term safety concerns with chronic use of TIMs in RA that may differ by dose and type of agent.[7] Feedback from patient groups also emphasized the highly individual experience with TIM therapy; some patients see immediate benefit from the first TIM they receive after failure of conventional DMARDs, while others must make multiple attempts before finding an agent that works for them. Therefore, there is a need to seek evidence on patient subgroups, comorbidities, and other factors that can better inform treatment response and selection of appropriate medications.

Objectives

The scope of this project was previously available for public comment and has been revised upon further discussions and input from stakeholders. In accordance with the updated scope, this project will assess both the comparative clinical effectiveness and economic impacts of multiple JAK inhibitors for the treatment of moderately-to-severely active RA, both as monotherapy and in combination with conventional DMARDs. We will also review the clinical and economic evidence for infliximab-dyyb (Inflectra®, Pfizer). The assessment aims to systematically evaluate the existing evidence, taking uncertainty into account. To that aim, the assessment is informed by two research components: a systematic review of the existing evidence and an economic evaluation. This document presents the protocol for the systematic review of existing evidence (i.e., the clinical review).

Research Questions

To inform our review of the clinical evidence, we have developed the following research questions with input from clinical experts, patients, and patient groups:

- In patients with moderately-to-severely active RA with inadequate response to conventional DMARDs, what is the comparative efficacy, safety, and effectiveness of the JAK inhibitors in terms of disease activity, remission, treatment response, quality of life, adverse events, and other key outcomes?
- In patients with moderately-to-severely active RA with inadequate response to conventional DMARDs, what is the comparative efficacy, safety, and effectiveness of the JAK inhibitors versus conventional DMARDs in terms of disease activity, remission, treatment response, quality of life, adverse events, and other key outcomes?
- In patients with moderately-to-severely active RA with inadequate response to conventional DMARDs, what is the comparative efficacy, safety, and effectiveness of infliximab-dyyb versus brand name infliximab in terms of disease activity, remission, treatment response, quality of life, adverse events, and other key outcomes?

PICOTS Criteria

In line with the above research questions, the following specific criteria have been defined utilizing PICOTS (Population, Interventions, Comparisons, Outcomes, Timing, Setting, and Study Design) elements.

Population

The population of focus for the review will be adults ages 18 and older with moderately-to-severely active RA and inadequate response to or intolerance of conventional DMARDs. Level of disease

activity will be defined according to validated and frequently used scales in RA (i.e., Disease Activity Score [DAS28], Clinical Disease Activity Index [CDAI], Simplified Disease Activity Index [SDAI]). Note that this focus will not include children, adolescents, or adults with juvenile forms of RA or other inflammatory arthritis, now collectively known as juvenile idiopathic arthritis (JIA). 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.[8]

We will also seek evidence on key subpopulations and/or data stratifications of interest. Among those suggested by stakeholders during the open input period were a) evaluation of both TIM-naïve patients *and* those with inadequate response to or intolerance of initial TIM therapy; b) use of JAK inhibitors as monotherapy and in combination with conventional DMARDs. Feedback received for the 2017 ICER report on rheumatoid arthritis indicated additional subpopulations or stratifications of interest, including a) presence of comorbidities (e.g., cardiovascular, interstitial lung disease, psychiatric, malignancy); b) both "early" (i.e., within two years of symptom onset) and established RA; c) seropositivity for prognostic markers such as anti-cyclic citrullinated peptide (CCP) antibodies; d) geography, in particular US-based versus non-US settings; and e) study funding (i.e., industry-sponsored vs. other funding sources).

Interventions

The interventions of interest for this review are listed below.

- Tofacitinib (Xeljanz®; Pfizer)
- Baricitinib (Olumiant®; Eli Lilly)
- Upadacitinib ([investigational]; AbbVie)
- Biosimilar exemplar: Inflectra (infliximab-dyyb)

We will seek clinical evidence on all the products listed above. We note, however, that biosimilar data will be presented separately, given differences in study design and intent (i.e., non-inferiority vs. superiority) relative to clinical studies of the originator products. We hope to use these data and the economic analyses to encourage a more general discussion on the role of biosimilars and interchangeability status in RA.

Comparators

We will examine 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, but will also seek head-to-head studies between the JAK inhibitors and TNF inhibitors to evaluate for more contemporary comparisons.

While studies with an active comparator arm are preferred, we will also include placebo-controlled trials as necessary to complete a planned network meta-analysis (NMA) of the effects of treatment on key measures of effectiveness that will combine direct and indirect evidence.

Outcomes

This review will examine key clinical outcomes associated with RA. In conversations held to develop the scoping document for this project, patient organizations advised us that clinical trials are often lacking robust information on patient-reported 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 have adjusted this list considerably 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)
- Other patient-reported outcomes (e.g., patient satisfaction, measures of fatigue, morning joint stiffness)
- Productivity loss and caregiver burden
- Requirements for joint replacement or other surgical intervention
- Utilization of key healthcare resources (e.g., hospitalization, rehabilitation, assisted living)
- Cardiovascular events
- Treatment-related adverse events (e.g., serious infection, malignancy, liver abnormalities)

While we will seek to assess these outcomes quantitatively, some measures may not be widely reported and will necessitate descriptive analysis only. Where possible, we will report 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 will be prioritized for response to therapy, however long-term evidence is preferred for harms.

Setting

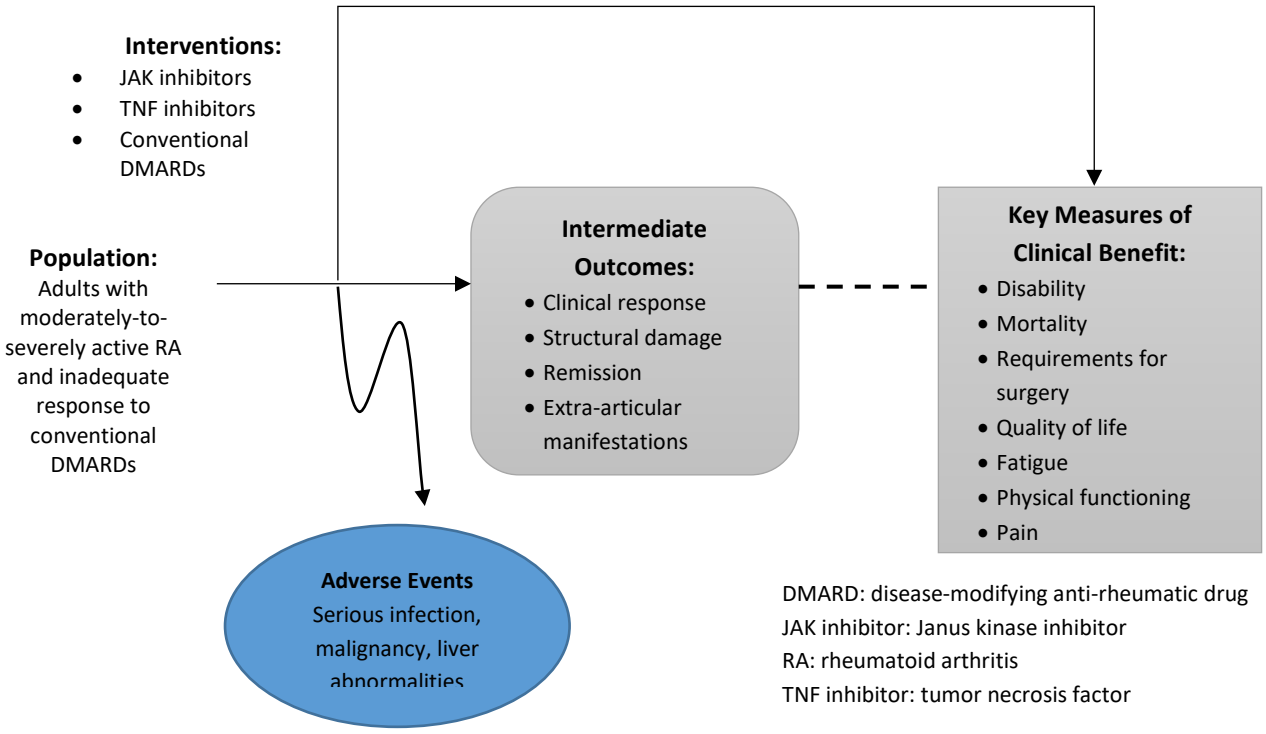
All relevant settings will be considered, including outpatient as well as ambulatory and hospital-based infusion centers. Several stakeholders commented on the importance of geography for our review given differences in treatment guidelines and practice patterns. We will focus attention on studies pertinent to the US setting; however, we recognize that studies conducted outside the US will likely be required for a complete review of the evidence.

Study Design

Randomized controlled trials and non-randomized controlled trials with any sample size will be included. Higher quality comparative observational studies, with sample size greater than 1,000, will also be included.

Analytic Framework

The proposed analytic framework for this project is depicted below:



DMARD: disease-modifying anti-rheumatic drug
JAK inhibitor: Janus kinase inhibitor
RA: rheumatoid arthritis
TNF inhibitor: 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 clinical or health outcomes. Outcomes are listed in the shaded boxes: those within the rounded boxes are intermediate outcomes (e.g., clinical response), and those within the squared-off boxes are key measures of clinical benefit (e.g., disability). The key measures of clinical 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 an action (typically treatment), which are listed within the blue ellipse.

Evidence Review Methods

Search Methods and Data Sources

Procedures for the systematic literature review assessing the evidence on JAK inhibitors for moderately-to-severely active RA will follow established best methods.[9,10] The review will be conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[11] The PRISMA guidelines include a list of 27 checklist items, which are described further in Appendix A.

We will search MEDLINE, EMBASE, Cochrane Database of Systematic Reviews, and Cochrane Central Register of Controlled Trials for relevant studies. Each search will be limited to English language studies of human subjects and will exclude articles indexed as guidelines, letters, editorials, narrative reviews, case reports, or news items. We will include abstracts from conference proceedings identified from the systematic literature search. All search strategies will be generated utilizing the Population, Intervention, Comparator, and Study Design elements described above. The proposed search strategies include a combination of indexing terms (MeSH terms in MEDLINE and EMTREE terms in EMBASE), as well as free-text terms, and are presented in Tables 1-2 on the following page.

To supplement the database searches, we will perform a manual check of the reference lists of included trials and reviews and invite key stakeholders to share references germane to the scope of this project. We will also supplement our review of published studies with data from conference proceedings, regulatory documents, information submitted by manufacturers, and other grey literature when the evidence meets ICER standards (for more information, see http://icer-review.org/methodology/icers-methods/icer-value-assessment-framework/grey-literature-policy/).

Table 1. Search Strategy of Medline 1996 to Present with Daily Update and Cochrane Central Register of Controlled Trials

#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 2. Search Strategy of EMBASE SEARCH

#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

Eligibility Criteria

Selection of Eligible Studies

Subsequent to the literature search and removal of duplicate citations using both online and local software tools, study selection will be accomplished through two levels of screening, at the abstract and full-text level. Two reviewers will independently screen the titles and abstracts of all publications identified using DistillerSR (Evidence Partners, Ottawa, Canada); a third reviewer will work with the initial two reviewers to resolve any issues of disagreement through consensus. No study will be excluded at abstract level screening due to insufficient information. For example, an abstract that does not report an outcome of interest in the abstract would be accepted for further review in full text.

Citations accepted during abstract-level screening will be retrieved in full text for review. Reasons for exclusion will be categorized according to the PICOTS elements during both title/abstract and full-text review.

Data Extraction Strategy

Data will be extracted into evidence tables. The basic design and elements of the extraction forms will follow those used for other ICER reports. Elements include a description of patient populations, sample size, duration of follow-up, funding source, study design features, interventions (agent, dosage, frequency, schedules), concomitant therapy allowed and used (agent, dosage, frequency, schedules), outcome assessments, results, and quality assessment for each study.

The data extraction will be performed in the following steps:

1) One reviewer will extract information from the full articles, and a second reviewer will validate the extracted data.
2) Extracted data will be reviewed for logic, and a random proportion of data will be validated by a third investigator for additional quality assurance.

Quality Assessment Criteria

We will use criteria published by the US Preventive Services Task Force (USPSTF) to assess the quality of clinical trials and cohort studies, using the categories "good," "fair," or "poor."[12]

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 paid to confounders in analysis. In addition, intention to treat analysis is used for RCTs.

Fair: 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: 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 or modified intention to treat (e.g., randomized and received at least one dose of study drug) analysis is lacking.

Publication Bias Assessment

Given the emerging nature of the evidence base for these newer treatments, we will scan the ClinicalTrials.gov site to identify studies completed more than two years ago. Search terms include tofacitinib, baricitinib, upadacitinib, and the biosimilar Inflectra. We will select studies which would have met our inclusion criteria, and for which no findings have been published. We will provide qualitative analysis of the objectives and methods of these studies to ascertain whether there may be a biased representation of study results in the published literature.

Evidence Synthesis

The purpose of the evidence synthesis is to estimate the clinical effectiveness of the interventions being compared. The analysis will be based on the data from all relevant studies identified from the systematic review. This section contains two components: 1) a summary of the evidence base and 2) a synthesis of outcome results.

Summary of Evidence Base

The studies will be summarized in the text and in evidence tables of the Evidence Report. This summary is key to understanding the evidence base pertaining to the topic. Evidence table shells are presented in Appendix B. Relevant data include those listed in the data extraction section. Any key differences between the studies in terms of the study design, patient characteristics, interventions (including dosing and frequency), outcomes (including definitions and methods of assessments), and study quality will be noted in the text of the report.

Synthesis of Results

The results of the studies will be synthesized for each outcome, and described narratively in the report. Analyses to be conducted will reflect the nature and quality of the evidence base (see

below). Key considerations for interpreting the results will be specified and described in the Evidence Report.

All studies deemed sufficiently similar in terms of the key population, intervention, and outcome measures will be included in a quantitative synthesis. For this report, NMAs under a Bayesian framework will be conducted on achievement of low disease activity or remission (as measured by the DAS28, CDAI, and SDAI), as well as ACR 20, 50, and 70 responses. An NMA simultaneously combines both the direct estimates (i.e., estimates obtained from head-to-head comparisons) and indirect estimates (i.e., estimates obtained from common comparator[s]).[13,14] Outcomes will be evaluated at three and six months.

Achievement of low disease activity will be analyzed using a binomial likelihood and logit link. ACR 20, 50, and 70 responses will be analyzed jointly in an ordered multinomial model with probit link.[15] The model assumes the presence of an unobserved underlying continuous variable (e.g., ACR) that has been categorized at different thresholds (e.g., 20%, 50%, 70% response). We will make the following additional assumptions for this model: the thresholds will be fixed across trials, trial-specific treatment effects will be drawn from a common distribution (i.e., random treatment effects model), and the amount of between-study variance (i.e., heterogeneity) will be constant across all treatment comparisons. The base-case model will also include a covariate for placebo response, which will be assumed to be common across all treatments and will provide a control for known and unknown differences between study populations.[16]

For any network where there are "loops" in evidence, we will empirically compare the direct and indirect estimates to assess if the NMA consistency assumption is violated using a node-splitting approach.[15] If there is evidence of inconsistency, the results will be presented for the direct and indirect evidence separately. If there is no evidence of inconsistency, we will present the pooled results.

To explore heterogeneity across studies, we will examine if there are differences in the distribution of key characteristics across studies. For example, it is possible that duration of RA can have an impact on relative effectiveness across treatments. If studies differ with respect to these characteristics, subgroup analyses or meta-regressions may be performed where sufficient data exist. As noted above, the base-case model will include a covariate for placebo response; we will also conduct a separate analysis without the covariate for comparison.

All NMAs will be conducted using JAGS software (Version 4.3.0) via R using the R2jags package.[17] For all analyses we will use noninformative prior distributions for all model parameters. We will initially discard the first 40,000 iterations as "burn-in" and base inferences on an additional 40,000 iterations using three chains. Convergence of chains will be assessed with the Gelman-Rubin statistic and visually using trace plots. If the chains do not converge, an additional 10,000 iterations will be run, sequentially, until convergence. Data included in each analysis along with the

corresponding code will be included in an appendix of the report. Results for all pairwise comparisons will be presented in tabular fashion in terms of a point estimate and 95% credible interval.

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. Crane MM, Juneja M, Allen J, et al. Epidemiology and Treatment of New-Onset and Established Rheumatoid Arthritis in an Insured US Population. *Arthritis Care Res (Hoboken).* 2015;67(12):1646-1655.
3. Huizinga TW, Pincus T. In the clinic. Rheumatoid arthritis. *Annals of internal medicine.* 2010;153(1):ITC1-1-ITC1-15; quiz ITC11-16.
4. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis and rheumatism.* 2010;62(9):2569-2581.
5. 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.
6. Singh JA, Saag KG, Bridges SL, Jr., et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Rheumatol.* 2016;68(1):1-26.
7. Singh JA, Cameron C, Noorbaloochi S, et al. Risk of serious infection in biological treatment of patients with rheumatoid arthritis: a systematic review and meta-analysis. *Lancet (London, England).* 2015;386(9990):258-265.
8. Weiss JE, Ilowite NT. Juvenile idiopathic arthritis. *Rheumatic diseases clinics of North America.* 2007;33(3):441-470, vi.
9. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of internal medicine.* 1997;126(5):376-380.
10. Higgins JP. Cochrane Collaboration Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. 2008.
11. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International journal of surgery (London, England).* 2010;8(5):336-341.
12. Agency for Healthcare Research and Quality. U.S. Preventive Services Task Force Procedure Manual. 2008.
13. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Bmj.* 2005;331(7521):897-900.
14. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in medicine.* 2004;23(20):3105-3124.
15. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in medicine.* 2010;29(7-8):932-944.
16. Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011; last updated April 2012; available from http://www.nicedsu.org.uk. 2011.
17. Yu-sung S, Masanao Y. R2jags: Using R to Run 'JAGS'. R package version 0.5-7. 2015; https://CRAN.R-project.org/package=R2jags.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International journal of surgery (London, England).* 2010;8(5):336-341.

19. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.

Appendix A. PRISMA Checklist

The checklist below is drawn from Moher et al. 2009.[18] Additional explanation of each item can be found in Liberati et al. 2009.[19]

Section/Topic	#	Checklist Item	Reported on Page #
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 and (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., health care 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.	

doi:10.1371/journal.pmed.1000097.t001

Appendix B. Data Extraction Summary Table Shell

Table F1. Study Characteristics

Author & Year of Publication (Trial Name) Quality Rating	Study Sponsor	Study Design and Duration of Follow-Up	Geographic Location of Study	Interventions (n) & Dosing Schedule	Inclusion and Exclusion Criteria	Baseline Patient Characteristics

Table F2. Key Clinical Outcomes

Author & Year of Publication (Trial Name)	Interventions	Treatment Response	Disease Activity	Structural Damage	Function	Laboratory Indices

Table F3. Harms

Author & Year of Publication (Trial Name)	Interventions	Malignancies	Infections	Other Adverse Events	Discontinuation, Serious AE Rate, Deaths

Table F4. Patient-Reported Outcomes

Author & Year of Publication (Trial Name)	Interventions	Health-Related Quality of Life	Pain	Fatigue & Other PROs

Table F5. Non-Health-Care Outcomes

Author & Year of Publication (Trial Name)	Interventions	Requirements for Surgical Intervention	Hospitalization, Rehabilitation, Assisted Living	Productivity Loss	Caregiver Burden	Other Outcomes