






Using Parallel Streams of Evidence to Inform Guideline Development: The Case of the 2021 American College of Rheumatology Management of Rheumatoid Arthritis Guideline

Sally Yaacoub,¹  Assem M. Khamis,^{1,2}  Mounir Al-Gibbawi,¹ Lara A Kahale,¹ Joan Bathon,³ 
Bryant R England,⁴  Liana Fraenkel,⁵  E. William St. Clair,⁶ Amy S. Turner,⁷  and Elie A. Akl^{1,8,9}

Objective. We aim to describe an evidence synthesis approach using parallel streams of evidence that informed the development of the 2021 American College of Rheumatology (ACR) guideline for the management of rheumatoid arthritis (RA).

Methods. We developed the evidence synthesis approach using parallel streams of evidence in multiple rounds of discussion, piloting, feedback, and revisions. A number of working groups involving ACR staff, content experts, and methodologists coordinated to develop and implement the approach.

Results. We used a major stream of evidence that identified evidence specific to the clinical questions being addressed in the guideline (ie, we were able to match relevant articles to specific questions). We also used additional streams that identified data that applied across multiple questions. We describe in this article the different steps of the major stream, ie, screening and tagging, matching articles to question clusters, matching articles to individual questions, data abstraction and analysis, and Grading of Recommendations Assessment, Development and Evaluation (GRADEing). We then describe how we packaged the parallel streams of evidence into standardized structured tables to facilitate formulating the recommendations. These tables included information for the following factors: desirable effects, undesirable effects, certainty of evidence, valuation of outcomes, cost of interventions, and cost-effectiveness of interventions. The approach allowed us to match eligible articles for 47 of 81 clinical questions. We identified no eligible articles that addressed the remaining 34 questions.

Conclusion. We were successful in using parallel streams of evidence to inform the development of the 2021 ACR guideline for the management of RA.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory polyarticular disease that often leads to joint damage and deformity. Among its many adverse outcomes, it causes fatigue, pain, disability, and reduced quality of life, and if uncontrolled, it may result in a need for joint replacement and other orthopedic surgeries. Many disease-modifying therapies have been approved

for the treatment of RA. These therapies provide rheumatologists with an array of therapeutic options for controlling the disease, limiting the accrual of joint damage, and minimizing the loss of physical function (1). However, appropriate therapy for RA not only has its proven benefits but also its inherent risks, which are balanced in the course of clinical decision-making and often lead to variation in the use of these agents. To support rheumatologists in their decision-making, the American

¹Sally Yaacoub, BSPHarm, MPH, Assem M. Khamis, MD, MPH, Mounir Al-Gibbawi, MD, Lara A. Kahale, RN, PhD, Elie A. Akl, MD, MPH, PhD: Clinical Research Institute, American University of Beirut, Beirut, Lebanon; ²Assem M. Khamis, MD, MPH: Hull York Medical School, University of Hull, Hull, United Kingdom; ³Joan Bathon, MD: American University of Beirut GRADE center, Beirut, Lebanon; ⁴Bryant R England, MD, PhD: Columbia University Irving Medical Center, New York Presbyterian Hospital, New York, New York, United States; ⁵Liana Fraenkel, MD, MPH: University of Nebraska Medical Center and VA Nebraska-Western Iowa Health Care System, Omaha, Nebraska, United States; ⁶E. William St Clair, MD: Berkshire Medical Center, Pittsfield, Massachusetts, and Yale University School of Medicine, New Haven, Connecticut, United States; ⁷Amy S. Turner: Duke University Medical Center, Durham, North Carolina, United States; ⁸Elie A. Akl, MD, MPH, PhD:

American College of Rheumatology, Atlanta, Georgia, United States; ⁹Elie A. Akl, MD, MPH, PhD: Department of Internal Medicine, American University of Beirut Medical Center, Beirut, Lebanon.

Dr. St. Clair has received consulting fees from Bristol-Myers Squibb (less than \$10,000) and receives royalties from UpToDate for his role as section editor. No other disclosures relevant to this article were reported.

Address correspondence to Elie A. Akl, MD, MPH, PhD, American University of Beirut Medical Center, Department of Internal Medicine, PO Box 11-0236 / CRI (E15), Riad-El-Solh Beirut 1107 2020, Lebanon. Email: ea32@aub.edu.lb.

Submitted for publication February 4, 2021; accepted in revised form June 1, 2021.

College of Rheumatology (ACR) publishes updated guidelines that rigorously synthesize the best scientific evidence and the input of clinicians and patients into a series of explicit treatment recommendations.

The ACR last published practice guidelines for the treatment of RA in 2015 (2). Since then, a significant number of relevant studies have been published and new therapies have been approved. Additional clinically relevant questions related to new drugs have also emerged. Consequently, the ACR produced the updated 2021 guideline for the management of RA (3).

The ACR aims to adhere to the Institute of Medicine's standards for developing trustworthy guidelines (4). One of the main standards is clinical practice guideline–systematic review intersection. This requires a continual and close interaction between the systematic review team and the guideline development group regarding to the scope, approach, and output (4).

The development of trustworthy guidelines also requires collecting evidence for different factors that feed into the guideline development process. Those factors include the comparative benefits and harms of the interventions being compared, the cost of those interventions, and how much patients value the outcomes of interest (valuation of outcomes). Although a typical systematic review of randomized controlled trials (RCTs) provides substantial data on benefits and harms (4), it does not provide data on the other factors. In some cases, those systematic reviews are not sufficient for collecting harms data (eg, long-term harms) (5).

Therefore, the objective of this article is to describe an evidence synthesis approach that used parallel streams of evidence to inform the development of the 2021 ACR guideline for the management of RA.

MATERIALS AND METHODS

Groups involved. Five groups were involved in the evidence synthesis for this project: the coordination group, the core team, the methodology team, the literature review team, and the voting panel.

1. The coordination group consisted of two ACR staff (including AST) and two methodologists (SY and EAA). The coordination group was in charge of coordinating the literature review tasks and worked closely with the core team. It held weekly meetings with members of the literature review team to provide guidance on the assigned tasks, address questions, and collect suggestions. Members of the coordination group also met on a weekly basis to discuss progress, time line, and logistics.
2. The core team (led by LF) included a senior ACR staff (AST), four rheumatologists specializing in RA, and two methodologists (SY and EAA). The core team met on a weekly basis to set the overall direction of the guideline project, approve the details of the approach used, and oversee successful and timely project completion.
3. The methodology team consisted of five methodologists (SY, AMK, MA-G, LAK, and EAA). It held daily meetings to develop the details of the approach and discuss challenges.
4. The literature review team consisted of 20 reviewers, nine of whom were junior rheumatologists. They participated in weekly meetings to receive training, review instructions, and provide feedback.
5. The voting panel consisted of 15 members, including two patient representatives. Members of the voting panel reviewed and provided input on the evidence reports generated by the literature review during a 3-week period preceding the voting panel meeting, during which the recommendations were formulated.

Developing the approach. The overall guideline development methodology followed the ACR Policy and Procedure Manual for Clinical Practice Guidelines (6). The coordination group, the core team, and the methodology team collaboratively built the evidence synthesis approach using parallel streams of evidence based on the handbook section “Guideline Development, Phase 2: Development” and developed it to address the specific needs of this project (6).

The methodology team developed the details of the approach through multiple rounds of discussion, piloting, and revisions. Specifically, after implementing a specific step of the review process, we would collect feedback from the literature review team regarding challenges. We would then revise the process with input from the core team and reimplement it with the literature review team. The project was guided by 81 clinically relevant research questions developed by the core team using a consensus process involving a sample of clinicians and patient/consumer representatives, according to the ACR Policy and Procedure Manual for Clinical Practice Guidelines (6). These questions were formulated in the Population, Intervention, Comparator, Outcomes (PICO) format, which represents a standardized way of framing recommendation questions (7). The questions were included in the guideline project protocol and were available on the ACR website for public comment via an online feedback mechanism (8).

The approach aimed to optimize the use of the major stream of evidence (ie, based on the main literature search) as well as the following additional streams of evidence:

- Systematic review of minimal clinically important difference of measurement instruments used in RA: we used this information to guide the judgment of imprecision for continuous variables when rating the certainty of evidence
- A published systematic review on patients' values and preferences for disease-modifying antirheumatic drug (DMARD) treatment in RA (9)
- Retrieval of cost data (average wholesale price) for the US market for all drugs considered in the PICO questions using Lexi-comp data (10)

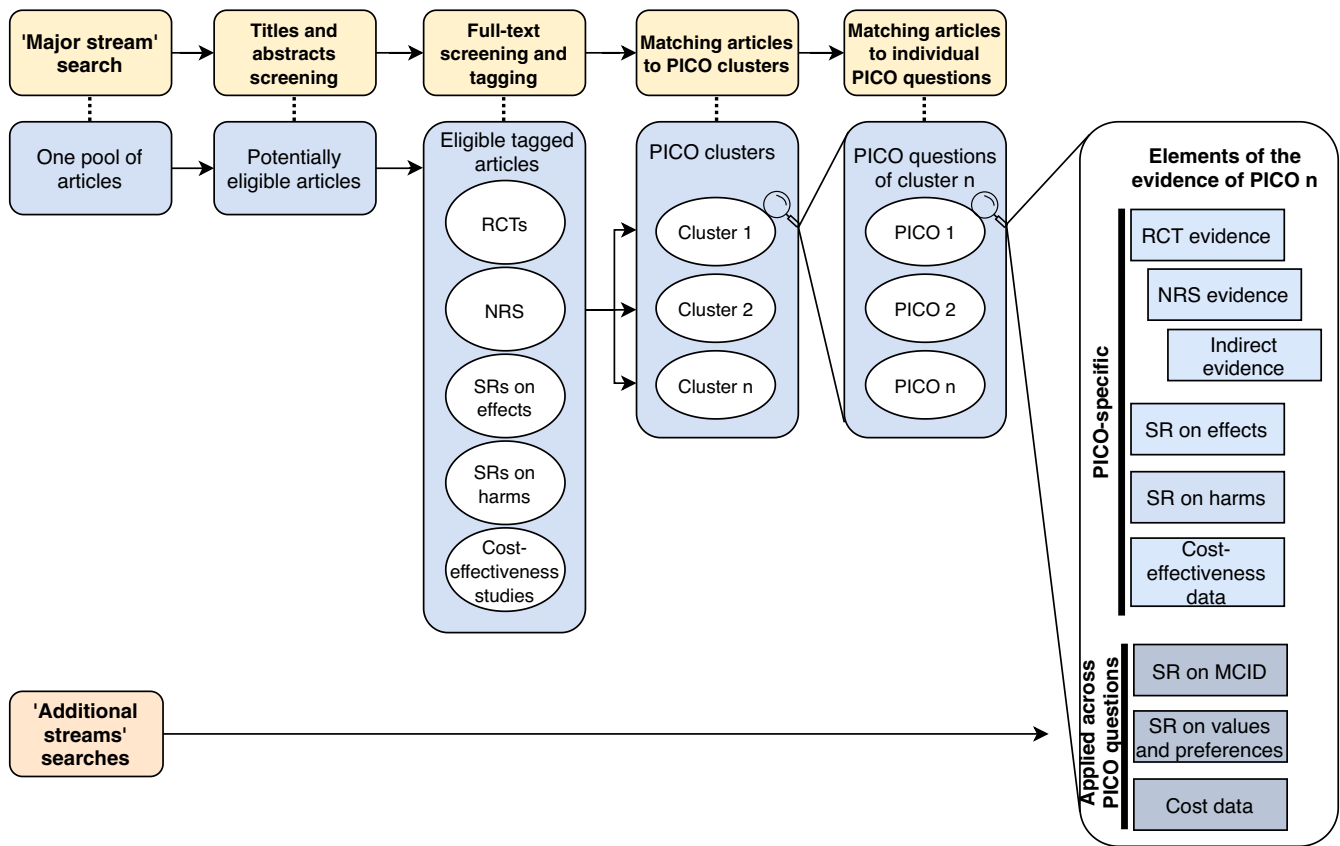


Figure 1. Steps of using parallel streams (major stream and additional streams) of evidence to identify elements of the evidence relevant to the different factors related to the PICO question, including those that are PICO specific and those that are applied across PICO questions. MCID, minimal clinically important difference; NRS, nonrandomized studies; PICO, Population, Intervention, Comparator, Outcomes; RCT, randomized controlled trial; SR, systematic review.

RESULTS

Figure 1 demonstrates how we used the parallel streams to identify evidence relevant to each of the PICO questions. We used the major stream (upper part of Figure 1) to capture the following:

- Published articles of primary studies (RCTs and nonrandomized studies [NRS]) providing both direct and indirect evidence on the comparative benefits and harms of the interventions of interest in the population of interest
- Published systematic reviews providing evidence on the comparative effects (benefits and harms) of the interventions of interest in the population of interest
- Published systematic reviews providing evidence on the comparative harms of the interventions of interest in any population
- Published cost-effectiveness studies on the interventions of interest in the population of interest

Although the major stream identified evidence that is PICO specific (ie, we were able to match the articles to individual PICO questions), the additional streams identified evidence that could be applied across PICO questions.

We describe in the subsequent sections the different steps of the major stream, ie, screening and tagging, matching articles to PICO clusters (eg, all questions on DMARD-naïve population), matching articles to individual PICO questions, data abstraction and analysis, and GRADEing (Grading of Recommendations Assessment, Development and Evaluation). We end with a section describing the packaging of the parallel streams of evidence. Table 1 illustrates the involvement of the different groups in each step of the approach.

Conducting the major stream review involved more steps than a standard review. Whereas in the latter, an article is assessed for eligibility against one PICO question, in the major stream review we had to assess the eligibility of each article against 81 PICO questions. Because a standard approach was not practically feasible, we matched each article in two steps: first to cluster(s) (n = 10) of related PICO questions and then to the individual question(s) within each cluster.

Screening and tagging. After a medical librarian executed the search, we imported the results into DistillerSR (Evidence Partners). Seven pairs of reviewers conducted title and abstract screening followed by full-text screening in a duplicate and

Table 1. Involvement of the different groups in each step of the approach

Step ^a	Coordination group (coordination)	Core team (provision of feedback)	Methodology team	Literature review team	Voting panel (provision of feedback)
Screening and tagging	✓			✓	
Matching articles to PICO clusters	✓			✓	
Matching articles to individual PICO questions	✓	✓		✓	✓
Data abstraction and analysis	✓	✓	✓	✓	
GRADEing	✓	✓	✓		
Packaging the parallel streams of evidence	✓	✓	✓		✓

GRADE, Grading of Recommendations Assessment, Development and Evaluation; PICO, Population, Intervention, Comparator, Outcomes.

^a The literature search was conducted by a medical librarian contracted by the American College of Rheumatology.

independent manner. Any disagreements at the full-text screening step were resolved by discussion or by a third reviewer as needed. The full-text screening form included both eligibility questions and tagging questions (see Appendix 1 in the Supplementary Materials).

At this step, eligibility criteria were broad, ie, whether the study population had RA, whether the intervention was addressed in any of the PICOs, and whether at least one of the outcomes of interest was assessed. Eligible study designs were RCTs, NRS, systematic reviews on effects (ie, comparative benefits and harms), systematic reviews on the comparative harms, and cost-effectiveness studies. To allow the later matching of each eligible article to the relevant cluster(s) of PICO questions, we tagged the articles by answering tagging questions. These included specific questions on the population, the interventions, the comparators, and the study design (RCT, NRS, systematic review, cost-effectiveness study).

Matching articles to PICO clusters. For each included article, we used the answers to the tagging questions to match it to the relevant cluster(s). We defined the 10 clusters according to nonoverlapping categories of population characteristics (eg, methotrexate naïve, methotrexate exposed) and/or intervention characteristics (eg, escalation vs de-escalation). We designed formulas embedded in Microsoft Excel to automate this matching process. An article may have matched to more than one cluster.

Matching articles to individual PICO questions. Within each cluster, seven teams of two reviewers matched each article to one or more (or none of) the individual PICO questions under that cluster. We conducted this process in duplicate and independently. Any disagreements were resolved by discussion or by a third reviewer. Next, content experts (members of the core team) worked in pairs to verify the results of this step in a duplicate and independent manner. We stratified matched articles into those describing primary studies on effects, systematic reviews on effects, systematic reviews on harms, and cost-effectiveness studies.

In addition, the core team asked members of the voting panel to review the accuracy and completeness of the list of included articles for each PICO question when they reviewed the evidence report prior to drafting the guideline recommendations.

Data abstraction and analysis. Once we identified a list of eligible articles to include for each PICO question, we linked articles relating to the same study (typically a trial) to avoid any double counting of data. The study characteristics were abstracted by rheumatologists on the literature team, whereas the risk of bias was assessed by the methodologists. Pairs of reviewers (including one rheumatologist and one methodologist) abstracted statistical data in a duplicate and independent manner and resolved disagreements by discussion.

The methodology team then abstracted additional data from the included studies related to the type of planned modification of treatment and the follow-up time(s) for outcome assessment that would be eligible for analysis. In addition, the methodology team chose the most appropriate outcome measurements (when more than one was reported) from a list of outcomes prioritized earlier by the core team. Also, the methodology team identified study arms eligible for the different comparisons of a PICO question.

We performed a meta-analysis of all outcomes for each comparison under each PICO question. For some of these analyses, we conducted statistical derivations to be able to pool statistical data from different arms.

GRADEing. We rated the certainty of evidence by outcome using the GRADE methodology (11). The GRADEing was conducted by one reviewer and verified by a second reviewer. Any disagreements were resolved by discussion. We used the findings of the systematic review on minimal clinically important difference to inform the judgment of imprecision when grading the evidence.

Packaging the parallel streams of evidence. Figure 2 shows how we packaged, for each PICO question, the parallel streams of evidence into standardized structured tables to facilitate formulating the corresponding recommendation. The structured

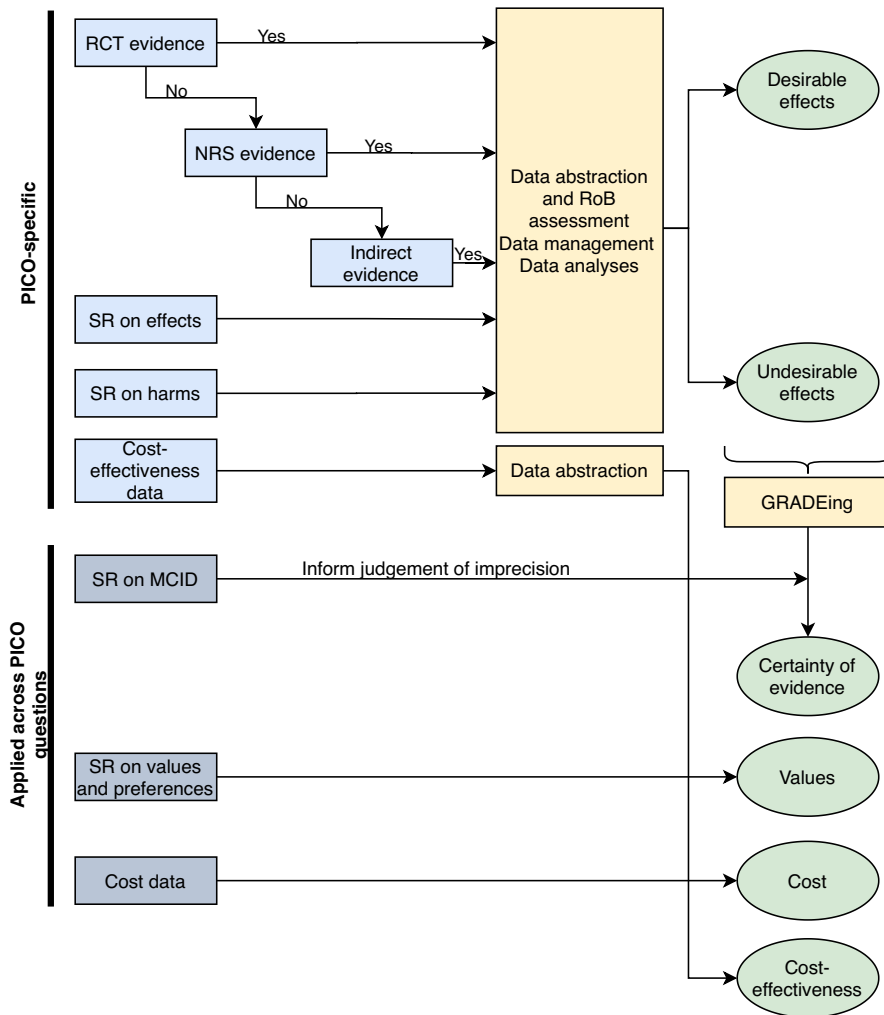


Figure 2. Packaging parallel streams of evidence (represented on the left) into standardized structured tables (represented on the right) to facilitate formulating the recommendations. Undesirable effects, desirable effects, and certainty of evidence are related to the relative health effects of interventions being compared, whereas values, cost, and cost-effectiveness represent contextual factors. GRADE, Grading of Recommendations Assessment, Development and Evaluation; MCID, minimal clinically important difference; NRS, nonrandomized studies; PICO, Population, Intervention, Comparator, Outcomes; RCT, randomized controlled trial; RoB, risk of bias; SR, systematic review.

tables included information for the following factors: desirable effects, undesirable effects, certainty of evidence, valuation of outcomes, cost of interventions, and cost-effectiveness of interventions. Additional data on patients' values and preferences fed into the evidence to consider when formulating recommendations (9). Undesirable effects, desirable effects, and certainty of evidence were related to the relative health effects of interventions being compared, whereas values, cost, and cost-effectiveness represented contextual factors. An example of a standardized structured table is presented in Appendix 2 in the Supplemental Materials.

Output of the approach. The full-text screening identified 815 RCTs, 780 NRS, 245 systematic reviews on effects (ie, comparative benefits and harms), 107 systematic reviews

on the comparative harms, and 168 cost-effectiveness studies for matching consideration. The matching led to the inclusion of a smaller number of articles to summarize the evidence, typically in a standardized structured table: 89 RCTs, 21 NRS, no systematic reviews on effects, 8 systematic reviews on the comparative harms, and 13 studies on cost-effectiveness. These articles were matched to 47 of the 81 PICO questions, leaving 34 PICO questions with no eligible articles, 726 RCTs, 759 NRS, 245 systematic reviews on effects, 99 systematic reviews on the comparative harms, and 155 studies on cost-effectiveness not matching to any PICO question. Although we were not able to directly use any systematic reviews on effects when developing standardized structured tables, those reviews were valuable in verifying that we had captured all eligible studies when

their questions overlapped with one of the guidelines' PICO questions.

DISCUSSION

We developed an evidence synthesis approach in which we used parallel streams of evidence to inform the guideline development. The major stream identified PICO-specific evidence, whereas the additional streams identified evidence that was relevant to several PICO questions. We elaborated on the steps of the major stream. In addition, we elaborated on the packaging of the streams of evidence into standardized structured tables. The approach allowed us to match eligible articles for 47 of the 81 PICO questions. We identified no eligible articles for the remaining 34 PICO questions.

Our approach has several strengths. First, the use of parallel streams of evidence enhanced the trustworthiness of the ensuing recommendations. In addition, we used an extensive search with teams of methodologists and content experts (ie, rheumatologists) to assess eligibility of studies and abstract data. We verified the matching results with senior content experts. In addition, we validated our results by cross-checking the list of studies included in systematic reviews addressing similar questions. Also, and to the best of our knowledge, this is the first article to describe the use of parallel streams of evidence in guideline development. One limitation of our approach is that we did not use full GRADE evidence-to-decision tables (12); however, we did consider evidence on cost, cost-effectiveness, and patients' values and preferences. Also, we could have better engaged stakeholders in the evidence synthesis process (13). Indeed, the project did involve a 10-member patient panel in the process of developing the recommendations but not in the evidence synthesis part.

We encountered several challenges while working on the evidence synthesis approach that uses parallel streams of evidence. First, some studies were described in multiple articles that provided different and sometimes inconsistent results. To avoid double counting and unnecessary duplication of work, we linked the related articles to their corresponding trial(s) after completing the matching to individual PICO questions. This task was challenging because not all articles included the name of the study, some articles included more than one trial, and most trials were published in more than one article. Following the linking, we compared the results reported in the different articles and judged which data were most relevant to include in the analysis.

Another challenge was identifying the type of modification of treatment in each arm during the trial period and judging how it affected the eligibility of the data for the PICO question of interest. One example of modification of treatment included having an escape or rescue treatment for participants who were not responding. Other examples included treatment modification on participant rerandomization, a treat-to-target strategy, or a

cross-over design, ie, switching all participants to the other arm. With input from the core team, we assessed the modification of treatment in each arm of included trials to determine which data were most relevant to include in the analysis.

It was also challenging to meta-analyze data from different studies reporting on the same outcome (eg, disease activity, clinical remission) using different tools or definitions. For example, some studies report disease activity using the Disease Activity Score (DAS) 28, whereas others report it using the DAS-44. Another example is the definition of remission (DAS-28 remission vs ACR–European League Against Rheumatism [EULAR] remission). To address this challenge, the core team decided a priori on a list of preferred measurement instruments for each outcome. Then the methodology team considered which instrument was used by the majority of studies eligible to be included in a meta-analysis.

Although the ACR Policy and Procedure Manual for Clinical Practice Guidelines encourages the development of original systematic reviews, it does allow the use of recently conducted systematic reviews that address a guideline's PICO questions (6). However, despite extensive efforts to identify and match systematic reviews on effects, none of the 245 systematic reviews that we assessed matched in a sufficient way to any of our PICO questions. Consequently, we were not able to use existing systematic reviews to answer our PICO questions or to compare the results of our meta-analyses with those of other systematic reviews. A common reason for their inapplicability was that the population in the PICO question was more narrowly defined than that of any existing systematic review.

This project will help with developing a database of studies relevant to the treatment of RA to make future updates of the guideline a more efficient process and less burdensome. The database would include the different types of studies identified for each PICO question, ie, RCTs, NRS, systematic reviews on effects, systematic reviews on the comparative harms, and cost-effectiveness studies. Our experience shows that pursuing different streams of evidence is feasible and useful and could be replicated in similar guideline development efforts.

There is a need to explore novel methods for making large evidence synthesis projects more efficient, eg, the use of artificial intelligence, particularly when developing guidelines for a large number of questions. Those novel methods would facilitate searching several databases, deduplicating their results, identifying potentially eligible references, and generating a living repository of evidence linked to the different PICO questions. In addition, it would be ideal to compare the results of our approach with those identified using these novel methods.

We were successful in using parallel streams of evidence to inform the development of the 2021 ACR guideline for the management of RA. We were able to produce for the panel standardized structured tables that included information on the different factors for the PICO questions, which facilitated formulating

recommendations. We believe that pursuing this evidence synthesis approach to using parallel streams of evidence is feasible and can be applied in similar guideline development efforts.

ACKNOWLEDGMENTS

The systematic reviews were funded by the ACR.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Yaacoub had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Yaacoub, Akl.

Acquisition of data. Yaacoub, Khamis, Al-Gibbawi, Kahale, Akl.

Analysis and interpretation of data. Yaacoub, Khamis, Al-Gibbawi, Kahale, Bathon, England, Fraenkel, St. Clair, Turner, Akl.

REFERENCES

- England BR, Mikuls TR. Epidemiology of, risk factors for, and possible causes of rheumatoid arthritis. URL: <https://www.uptodate.com/contents/epidemiology-of-risk-factors-for-and-possible-causes-of-rheumatoid-arthritis>. UpToDate. 2020.
- Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol* 2016;68:1–26.
- Fraenkel L, Bathon JM, England BR, St.Clair EW, Arayssi T, Carandang K, et al. 2021 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol* 2021;73:1108–1123. <https://doi.org/10.1002/art.41752>
- Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, editors; Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical practice guidelines we can trust*. Washington (DC): National Academies Press; 2011.
- Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA; on behalf of the Cochrane Non-Randomized Studies of Interventions Methods Group. Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions*. 2nd ed. Chichester (UK): John Wiley & Sons; 2019. p. 595–620.
- American College of Rheumatology. Policy and procedure manual for clinical practice guidelines. 2015 URL: https://www.rheumatology.org/Portals/0/Files/ACR%20Guideline%20Manual_Appendices_updated%202015.pdf.
- Thomas J, Kneale D, McKenzie JE, Brennan SE, Bhaumik S. Determining the scope of the review and the questions it will address. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions*. 2nd ed. Chichester (UK): John Wiley & Sons; 2019. p. 13–31.
- American College of Rheumatology. Updated guideline for the management of rheumatoid arthritis: Project Plan – October 2018. URL: <https://www.rheumatology.org/Portals/0/Files/Rheumatoid-Arthritis-Guideline-Project-Plan.pdf>.
- Durand C, Eldoma M, Marshall DA, Bansback N, Hazlewood GS. Patient preferences for disease-modifying antirheumatic drug treatment in rheumatoid arthritis: a systematic review. *J Rheumatol* 2020;47:176–87.
- Lexicomp Online. Lexi-drugs online. 2019. URL: <https://online.lexi.com>.
- Brožek JL, Akl EA, Compalati E, Kreis J, Terracciano L, Fiocchi A, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 3 of 3. The GRADE approach to developing recommendations. *Allergy* 2011;66:588–95.
- Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1. Introduction. *BMJ* 2016;353:i2016.
- Pollock A, Campbell P, Struthers C, Synnot A, Nunn J, Hill S, et al. Stakeholder involvement in systematic reviews: a scoping review. *Syst Rev* 2018;7:208.