

SUPPLEMENTARY APPENDIX 1: Methods

2019 American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network Updated Recommendations for the Management of Axial Spondyloarthritis

Methodology Overview

This updated guideline was developed following the American College of Rheumatology (ACR) guideline development process (https://www.rheumatology.org/Portals/0/Files/ACR%20Guideline%20Manual_Appendices_updated%202015.pdf). This process includes the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) methodology (www.gradeworkinggroup.org) [1-3].

Teams Involved

This project was a collaboration between the American College of Rheumatology (ACR), the Spondylitis Association of America (SAA) and the Spondyloarthritis Research and Treatment Network (SPARTAN). The project included a core leadership group, a literature review group, and a voting panel. The Core Leadership Team (3 rheumatologist volunteers members and 3 ACR staff) supervised the project and was responsible for defining the scope, drafting the clinical (Patient/Intervention/Comparator/Outcomes – PICO) questions, coordinating with the Literature Review Team, overseeing the voting process, and drafting the manuscript. The Core Team, together with the Literature Review Team, was comprised of individuals with content and/or methodological expertise; the majority of Core Team and Literature Review Team members had prior experience developing and presenting evidence using GRADE and provided input on the quality assessment of evidence and summary of findings (SoF) tables (provided in Supplementary Appendix 6).

The Literature Review Team (5 members plus a Literature Review Lead was comprised of two rheumatologists, two health services researchers, and two internists with special

expertise in the conduct of literature reviews. The Literature Review Team conducted a systematic search, screened papers for relevance, assessed study quality, extracted data, computed pooled estimates of outcomes, graded the quality of evidence, generated the Summary of Findings (SoF) tables, and compiled an evidence report.

The Voting Panel included 11 rheumatologists, one clinical pharmacist, and two patients. The role of the Voting Panel was to review the evidence report, and participate in the Voting Panel meeting held June 2, 2018, during which they discussed and voted on the recommendations of PICO questions by synthesizing the evidence report based on their expertise and experience as well as patient values and preferences. The Voting Panel used the input from the patient representatives to help guide their votes in balancing tradeoffs between the harms and benefits of the alternative management strategies.

The ACR provided training for everyone involved in the development of this guideline, which included sessions on the ACR guideline process and GRADE methodology. See Supplementary Appendix 2 for team/panel rosters.

Disclosures and Management of Conflicts of Interest

Per ACR policy, those who were intellectually involved in the project (i.e., considered for guideline authorship) disclosed all relationships. Disclosures were compared against a previously drafted list of “affected companies” (i.e., companies or organizations that were considered reasonably likely to be positively or negatively affected by care delivered in accordance with the guideline) to determine which relationships were considered potential conflicts of interest for purposes of this project. Individuals were also asked to explicitly highlight relationships with any companies *not* on the affected companies list that related to the topic of the guideline. Individuals whose primary employment (> 51% of work time/effort) was with a company that manufactured or sold therapeutics or diagnostics were not eligible to

participate. At least 51% of each group (core, literature review, and voting panel) was required to be without relevant conflicts of interest.

The project's principal investigator (PI) and the literature review leader had no relevant conflicts of interest for the full 12 months before this project began and through the course of the project, and the majority of the guideline development team members had no relevant conflicts of interest for the duration of the project. A participant who had any relationship with an affected company was counted as conflicted (i.e., toward the allowed threshold) regardless of the type or subject of the relationship. Intellectual conflicts, such as a prior publication or scientific presentation on axial spondyloarthritis (axSpA) therapy, were recognized as important and were required to be disclosed. Because an intellectual and academic background in axSpA was valued and contributed to informing the decision-making, intellectual conflicts were not counted as conflicted toward the allowed threshold.

Participant disclosures were included in the project plan that was posted online for public comment (see description below). In addition, disclosures of all participants were shared, in writing, with each project participant. At the face-to-face Voting Panel meeting, verbal disclosures were provided before the content discussion began. Updated participant disclosures, as well as ACR committee reviewer disclosures, are included online with this manuscript. Finally, author disclosures are also included in this paper.

Scope and Target Audience

The scope of this project included: the use of nonsteroidal anti-inflammatory drugs, oral small molecules, and biologics (including biosimilars); the role of magnetic resonance imaging and radiography in longitudinal patient management; and the role of a treat-to-target strategy in the care of patients. Since the publication of the 2015 ACR/SAA/SPARTAN AS and nr-axSpA guidelines, additional medications have become available, prompting a need to

reevaluate previous recommendation and incorporate new medications into the recommendations. Consequently, this selective update largely focused on pharmacological treatments, rather than a comprehensive update of all previous recommendations. However, some topics not included in the previous recommendations have been addressed.

AxSpA was defined as a form of chronic inflammatory arthritis characterized by sacroiliitis, extra-articular manifestations, and spinal and peripheral enthesitis; when these progress to sacroiliac joint and spinal fusion the condition is known as ankylosing spondylitis (AS). The target audience for this guideline includes health care providers (principally rheumatologists, primary care clinicians, physiatrists, and physical therapists) and patients with axSpA. The ACR and possibly the SAA and SPARTAN plan to develop derivative products to facilitate implementation of this guideline.

Establishing Key Principles and PICO Development

The GRADE method specifies four elements for each clinical question to be addressed by a recommendation: the Patient (or Population) to whom the recommendation applies; the Intervention; the Comparison (i.e. an alternative intervention or action, which could be no action); and the Outcomes for evaluating the intervention [4]. These PICO elements are used to develop the scope of the literature review and literature search terms, which in turn lead to an estimation of the quantity and quality of evidence supporting each PICO question. The evidence then forms the basis for a recommendation.

The Core Leadership Team developed the initial set of PICO-formatted clinical questions for the guideline update. The core group developed 24 PICO questions for AS in four topic areas: pharmacological therapy, comorbidities, treatment strategies, and imaging. Parallel questions were developed on most topics for nr-axSpA, which expanded the total number of PICO questions to 46. Approximately one-half of the PICO questions were derived from the 2015

ACR/SAA/SPARTAN AS and nr-axSpA guidelines and one-half were developed de novo. The PICO questions were posted for 30 days on the ACR website for public comment and revised accordingly.

Framework and Definitions for the AxSpA Guideline Development

The Core Leadership Team agreed that the scope of the populations to be addressed in the updated guideline would include patients with ankylosing spondylitis (AS; meeting the modified New York criteria [5]) and those with non-radiographic axSpA (meeting the ASAS axSpA criteria [6], but not classified as AS).

After defining the target populations, interventions and comparators were specified for each PICO question (PICO questions appear in Supplementary Appendix 6). The Core Leadership Team agreed that the updated guideline should primarily focus on pharmacologic treatment approaches and elected to include the interventions described above in section “Scope and Target Audience.”

We used the same outcomes framework as in the 2015 ACR/SAA/SPARTAN AS and nr-axSpA guidelines. The framework included five major outcomes: mortality, health status, functional status, serious adverse events, and comorbidities (Supplementary Appendix 3). The critical outcomes included mortality (though this was infrequently reported among the relevant literature) and health status (comprised of symptoms, mental health and quality of life). Functional status was also regarded as an important outcome domain. We used patient-reported outcomes as the primary measures of health status and functional status, because these more directly reflect how the condition impacts the person. When patient-reported outcomes were not available, we used data on other outcomes, such as spinal range of motion. We used this framework for all PICO questions, which implicitly indicates that most

interventions for axSpA have similar goals. For rehabilitation interventions, the outcomes were health status, functional status, and adverse events.

Systematic Synthesis of the Literature

Literature Searches

To identify relevant evidence for the 46 PICO questions, a medical librarian, in collaboration with the Literature Review Team, performed systematic searches of the published English language literature. OVID Medline (since 1946) , PubMed (since its inception in the mid-1960s), and the Cochrane Library, including Cochrane Central Register of Controlled Trials (CENTRAL), were searched from the beginning of each database through September 9, 2017 (Supplementary Appendix 4), and update searches were conducted on February 28, 2018. For PICO questions for which no direct evidence in the axSpA field was found, we informally surveyed indirect evidence, in particular, meta-analyses of randomized trials in non-axSpA populations. For PICO questions for which systematic reviews or Randomized Controlled Trials (RCTs) were not found, observational studies of axSpA populations were sought. Searches for medications were limited to those approved by the U.S. Food and Drug Administration for any indication.

Study Selection

DistillerSR software (<https://www.evidencepartners.com/products/distillersr-systematic-review-software/>) was used to aid screening the literature search results. A team of two independent reviewers performed duplicate screening of each title and abstract, with articles identified as potentially eligible passing to review of full text. We also checked the references of prior systematic reviews, including the Evidence Report from the 2015 ACR/SAA/SPARTAN AS and nr-axSpA guidelines. We excluded articles on children with spondyloarthritis, those in languages other than English, narrative reviews, meeting abstracts,

and case reports. Eligible articles underwent full-text screening by two independent reviewers. The literature review leader had final decision over included studies. Selected manuscripts were then matched to relevant PICO questions based on the population and intervention in question. See Supplementary Appendix 5 for details related to the study selection process.

Data Extraction and Analysis

The GRADE method emphasizes that recommendations be based on the best available evidence that is summarized in estimates of treatment effect whenever possible. These quantitative estimates include pooled odds ratios/relative risks for binary outcomes (analyzed using the Mantel-Haenszel method in a random effects model and reported with 95% confidence intervals) or a mean difference for continuous outcomes (report as pooled mean differences between the intervention and comparator with 95% confidence intervals) [7]. Data from RCTs for each PICO question were extracted into RevMan software (<http://tech.cochrane.org/revman>).

Reviewers also assessed the overall quality of the evidence for each outcome based on the likelihood of bias, degree of imprecision, inconsistency in reported results among studies, indirectness (e.g., the study examined a similar but distinct patient group or intervention), and possible publication bias [8].

In GRADE, randomized controlled trials are assumed to provide higher quality evidence than observational studies [8]. Evidence from trials may be downgraded due to a high risk of bias or other limitations, such as inconsistency or imprecision. Evidence from observational studies may be upgraded when the treatment effect is large, and confounding is judged to be unlikely. In clinical scenarios not addressed by RCT data, data from observational cohort studies was used to estimate relative effects. In situations in which the intervention had not

been tested in axSpA but had been tested in a non-axSpA population, the effect sizes from that study were applied, postulating that the effect was generalizable but downgrading the quality of evidence for indirectness.

Evidence Report Formulation

RevMan files were exported to GRADEpro software to formulate a GRADE Summary of Findings (SoF) table for each PICO question [4]. The quality of evidence for each outcome was evaluated by reviewers using GRADE quality assessment criteria [1]. GRADE specifies four categories in which the quality of evidence may be rated: high, moderate, low, and very low [8]. High quality evidence indicates high confidence in the effect estimate, and new data from future studies are thought unlikely to change the effect. Moderate quality evidence indicates confidence that the true effect is likely to be close to the estimate, but could be substantially different. Low quality evidence implies limited confidence, and the true effect may be substantially different from the estimate. Very low quality evidence implies very little certainty, and the true effect may be quite different from the estimate.

The resulting SoF tables were compiled in an evidence report (Supplementary Appendix 6), which included a comprehensive bibliography of all included studies. The Core Team reviewed the evidence report and addressed possible evidence gaps prior to presentation to the Voting Panel. Evidence from PICOs that were not addressed in this revision were simply carried forward from the Evidence Report used in the 2015 ACR/SAA/SPARTAN AS and nr-axSpA guidelines.

Moving from Evidence to Recommendations

GRADE methodology specifies that panels make recommendations based on a consideration of the balance of benefits and harms of the treatment options under consideration, the quality of the evidence (i.e., confidence in the effect estimates), and patients'

values and preferences. Key to the recommendation is the trade-off between desirable and undesirable outcomes; recommendations require estimating the relative value patients place on the outcomes [1].

A recommendation could be either in favor of or against the proposed intervention and either strong or conditional. According to GRADE, a recommendation is categorized as strong if the panel is very confident that the benefits of an intervention clearly outweigh the harms (or vice versa), that the quality of evidence is high, and that future research will likely not alter the conclusion. Strong recommendations can also be based on less evidence when there is substantial concern for risk of harm. Strong recommendations do not imply large clinical benefits from the intervention, but rather confidence in the evidence base. A conditional recommendation denotes uncertainty regarding the balance of benefits and harms, such as when the evidence quality is low or very low, or when the decision is sensitive to individual patient preferences, or when costs are expected to impact the decision. Thus, conditional recommendations refer to decisions in which incorporation of patient preferences is a particularly essential element of decision making.

Judgments are based on the experience of the clinician panel members in shared decision making with their patients, as well as the experience and perspectives of the two patient panel members. Following ACR policy, the cost of an intervention was not formally considered in developing recommendations.

Consensus Building

The Voting Panel received the evidence report (Supplementary Appendix 6) for review two weeks before it met to discuss and decide on the final recommendations and identify potentially contentious topics. During the face-to-face meeting, held June 2, 2018, the full text of every included study was made available to all members of the Voting Panel. For each PICO

question, the Voting Panel reviewed the evidence, discussed questions, and provided anonymous votes on the direction and strength of the recommendations. The in-person voting process was conducted using Poll Everywhere software (<http://www.polleverywhere.com>). A 70% consensus was used as the threshold for a recommendation; if 70% consensus was not achieved during an initial vote, the panel members held additional discussions before re-voting until at least 70% consensus was achieved.

Consistent with GRADE guidance, in some instances, the Voting Panel chose to provide a strong recommendation despite a low or very low quality rating of evidence [3]. In such cases, a written explanation is provided describing the reasons behind this decision with reference to GRADE guidance on the matter [3].

Subsequent to the face-to-face meeting, two manuscripts were published online assessing the use of ixekizumab in ankylosing spondylitis [9-10]. These manuscripts were provided to the voting panel members on October 30, 2018. The voting panel reviewed the manuscripts and re-voted on the 12 PICOs addressing the use of secukinumab or biologics in general. The primary consideration was whether to substitute “secukinumab or ixekizumab” in places where the PICOs previously referred only to secukinumab.

Final Review and Approval of the Manuscript by the ACR

In addition to journal peer reviews, the manuscript was reviewed by the following committees and subcommittees of the ACR: ACR Guideline Subcommittee; ACR Quality of Care Committee; and ACR Board of Directors. The SAA Medical and Scientific Advisory Board and SPARTAN Board of Directors also reviewed the manuscript. These ACR, SAA, and SPARTAN oversight groups did not mandate that certain recommendations be made within the guideline, but rather, served as peer reviewers.

Moving from Recommendations to Practice

These recommendations are designed to support health care providers who work with patients in selecting therapies. Health care providers and patients must take into consideration not only clinical phenotype and level of disease activity, but also comorbidities, response and tolerance of prior therapies, patient's values and preferences, and patient's functional status and functional goals in choosing the optimal therapy for an individual patient at the given point in treatment.

REFERENCES

1. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-6.
2. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7):719-25.
3. Andrews JC, Schunemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol*. 2013;66(7):726-35.
4. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395-400.
5. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;27:361-8.
6. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, et al. The development of assessment of Spondyloarthritis International Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68:777-83.
7. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, Atkins D, Kunz R, Montori V, Jaeschke R, Rind D, Dahm P, Akl EA, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Schünemann HJ. GRADE guidelines: 11. Making an overall rating of

confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013; 66:151-7.

8. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Jaeschke R, Rind D. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011; 64:383-94.
9. Deodhar A, Poddubnyy D, Pacheco-Tena C, Salvarani C, Lespessailles E, Rahman P, et al. Efficacy and safety of ixekizumab in the treatment of radiographic axial spondyloarthritis: 16 week results of a phase 3 randomized, double-blind, placebo controlled trial in patients with prior inadequate response or intolerance to tumor necrosis factor inhibitors. *Arthritis Rheumatol.* (Published online prior to print). doi:10.1002/art.40753
10. van der Heijde D, Wei JC, Dougados M, Mease P, Deodhar A, Maksymowych WP, et al. Ixekizumab, an interleukin-17A antagonist in the treatment of ankylosing spondylitis or radiographic axial spondyloarthritis in patients previously untreated with biological disease modifying anti-rheumatic drugs (COAST-V): 16 week results of a phase 3 randomised, double-blind, active-controlled and placebo-controlled trial. *Lancet.* (Published online prior to print) [http://dx.doi.org/10.1016/S0140-6736\(18\)31946-9](http://dx.doi.org/10.1016/S0140-6736(18)31946-9)