

For submission to *The Journal of Rheumatology*

GRAPPA Treatment Recommendations: Updates and Methods

Arthur Kavanaugh, Laura C. Coates, Danielle A. van der Windt,

Nadia Corp, Enrique R. Soriano

ABSTRACT: The development and updating of treatment recommendations for optimal treatment approaches for patients with psoriatic arthritis (PsA) has been an important mission of the Group for Research and Assessment of Psoriasis and PsA (GRAPPA) since its inception. Even though the most recent iteration of the GRAPPA PsA recommendations was completed only a few years ago, there have been many significant advances related to therapies and treatment approaches for PsA that have taken place since their publication. Because of these advances, the process to update the recommendations again has begun. The standard approaches to guideline (or treatment recommendation) development have also evolved in recent years. Herein, the basis for the approach that will be taken for the next version of the GRAPPA PsA treatment recommendations is reviewed.

Key Indexing Terms: Psoriasis, Psoriatic Arthritis, GRAPPA, Treatment Recommendations

Author Information: A Kavanaugh, MD, Division of Rheumatology Allergy and Immunology, University of California, San Diego, La Jolla, California, USA, akavanaugh@health.ucsd.edu; LC Coates, MB ChB, PhD, Department of

Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom, laura.coates@ndorms.ox.ac.uk;
DA van der Windt, PhD, Primary Care Centre Versus Arthritis, School of Primary, Community and Social Care, Keele University, Staffordshire, United Kingdom, d.van.der.windt@keele.ac.uk; N Corp, PhD, Primary Care Centre Versus Arthritis, School of Primary, Community and Social Care, Keele University, Staffordshire, United Kingdom, n.corp@keele.ac.uk; ER Soriano, MD, MSc, Rheumatology Unit, Internal Medicine Services, Hospital Italiano de Buenos Aires, and University Institute Hospital Italiano de Buenos Aires, Buenos Aires, Argentina, enrique.soriano@hospitalitaliano.org.ar.

Corresponding Author: Arthur Kavanaugh, MD, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92037-0943, USA, phone 858-246-2383, fax 848246-2401, akavanaugh@health.ucsd.edu.

Running Footline: GRAPPA Treatment Recommendations: Update

Word Count: 2947 words, including text (2442), abstract (126), and references (379, n=13), but excluding title page

The creation of treatment recommendations was considered central to the formation of the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA). GRAPPA's mission statement, as stated in its first guidelines that were published in 2009, is "To develop guidelines, based upon the best scientific evidence, for the optimal treatment of patients with psoriatic arthritis (PsA)".(1) Advances in clinical care drove the need to develop the second set of GRAPPA guidelines, which were published in 2016.(2) Tremendous progress in the therapeutic approach to PsA treatment over the past 5 years again necessitates the revision of these recommendations to best achieve their initial goal.

Key Domains of PsA

While the general approach for creating guidelines has evolved, some core principles remain the same.(3) In a heterogeneous disease such as PsA, the process for creating guidelines begins with delineation of the key domains of disease that must be considered for treatment. In the initial GRAPPA guidelines, the key domains included axial disease, peripheral arthritis, skin and nail psoriasis, enthesitis, and dactylitis.(1) For the second set of guidelines, nail psoriasis and comorbid conditions (because of the growing recognition of the critical role comorbid conditions play in PsA treatment) were each separated into their own domains.(2) For the new set of guidelines, associated conditions (including inflammatory bowel disease (IBD) and uveitis) will be considered as a separate domain due to the clinical relevance of the large amount of evidence related to the treatment of these associated conditions.

Thus, 8 recognised key domains of PsA will be addressed in the new guidelines, each with its own dedicated working/domain group: (1) peripheral arthritis; (2) axial disease; (3) enthesitis; (4) dactylitis; (5) skin disease (psoriasis); (6) nail disease (psoriasis); (7) comorbidities of PsA (including obesity, metabolic syndrome, diabetes mellitus, non-alcoholic fatty liver disease, cardiovascular disease, mood disorders (depression and anxiety), osteoporosis, chronic and serious infections, malignancy, fibromyalgia, and reproductive health); and (8) conditions associated with PsA (including IBD and related autoimmune ophthalmologic disease (e.g., uveitis)).

Key Clinical Questions Informing Optimal Therapy

The next step in revising the recommendations will be to formulate the key clinical questions that inform optimal therapy and that serve as the basis for the systematic searches.(4) Across the different domains, these have included: (1) what is the effect of a given therapy on clinical manifestations (including signs/symptoms, quality of life, functional status, structural integrity, and safety), and (2) what is the magnitude of the effect (effect size, number needed to treat, or number needed to harm)? For comorbidities and associated conditions, this also includes questions about the impact of the condition on response to therapy or therapy-related adverse events. These questions are formulated as Patient/Population-Intervention-Comparison/Comparator-Outcome (PICO) questions that address population, intervention, comparators, and outcomes.(5) Working groups will review the PICOs from the latest GRAPPA guidelines and amend these as they deem necessary. The resulting set of PICOs from each

working group will then form the basis for designing the searches underpinning the evidence synthesis for the updated recommendations.

Guideline Development Process

The development of key clinical questions informing optimal therapy has evolved with the introduction of the GRADE methodology.(6, 7) Over the years, guideline development processes have changed, but the underlying process, where medical literature is reviewed with attention to the quality of the data source, has not changed. Methods for evaluating and synthesizing evidence have evolved. Consensus recommendations are made with various strengths based on the consistency, strength, and quality of evidence.(7) A common approach, and an approach that was used for the first GRAPPA guidelines, grades the data as:

- 1A: evidence from meta-analysis of randomized controlled trials (RCTs);
- 1B: evidence from 1 or more RCTs;
- 2A: evidence from 1 or more controlled trials (without randomization);
- 2B: evidence obtained through other well-designed studies;
- 3: evidence from non-experimental studies (e.g., comparative, correlation, or case-control); or
- 4: expert committee opinions, clinical experience.

For the updated recommendations, given growing evidence for the treatment of PsA, it was decided *a priori* to focus on RCTs. Other study designs will only be considered if there is an absence of evidence for treatment effect or if other designs are relevant in addressing specific PICOs (e.g., cohort designs may be

of relevance when considering the importance of comorbidities or for important topics not yet fully addressed in RCTs such as the use of combination therapies).

While such approaches are still utilized (including the Strength of Recommendation Taxonomy (SORT) process), many recent guideline developers have adopted the GRADE process to assess the quality of the evidence underpinning treatment recommendations, taking into account limitations of studies, inconsistency of results, indirectness of evidence, imprecision, and reporting bias.(6, 8, 9) The GRADE working group began in 2000. GRADE is an informal collaboration of people with an interest in addressing the shortcomings of grading systems in healthcare. The GRADE working group has developed a common sense and transparent approach to grading quality (or certainty) of evidence and strength of recommendations.(6-8)

When using GRADE to grade quality of evidence, the first step is to build PICO questions to guide the literature search. One of the major issues in question development is determining the outcomes that are most important for decision makers. Once the literature search has been performed using PICO questions, evidence is then graded. For simplicity, GRADE classifies the quality of evidence as either high, moderate, low, or very low.(6) This grading of evidence, unlike many other systems, is not only based on a study's design, but also on the study's limitations, inconsistency of results, indirectness of evidence, imprecision, and reporting bias.(6) Thus, although observational studies start with a low-quality rating, in some cases, grading upwards may be warranted and

allowed if, for example, all plausible biases would decrease the magnitude of an apparent treatment effect or the effect is very large.

Evidence is summarized in tables where effect sizes for main outcomes can be calculated. The quality of the evidence is then translated to recommendations with different strengths. GRADE includes 2 strengths of recommendations: “strong” and “weak” that are either against or in favor of the intervention (though guideline panels may prefer terms such as “conditional” or “discretionary” instead of “weak”).(6)

The clear separation between the quality of evidence and the strength of recommendation is a GRADE-system strength. Not all grading systems separate decisions about the quality of evidence from the strength of recommendations, but this creates confusion. High-quality evidence does not necessarily imply strong recommendations, and strong recommendations can arise from low-quality evidence.(6) To move from evidence to recommendations, GRADE considers not only the quality of evidence, but also uncertainty about the balance between desirable and undesirable effects, uncertainty or variability in values and preferences, and uncertainty about whether the intervention represents a wise use of resources.

The GRADE system’s advantages over other systems are that it (1) was developed by a widely representative group of international guideline developers, (2) has a clear separation between quality of evidence and strength of recommendations, (3) includes an explicit evaluation of the importance of outcomes of alternative treatment strategies, (4) has explicit criteria for

downgrading and upgrading quality of evidence ratings, (5) includes an explicit acknowledgment of values and preferences when moving from evidence to recommendations, and (6) includes a clear interpretation of strong versus weak recommendations.

GRADE also has limitations. GRADE's first limitation is its reliability in assessing the quality of evidence.(10) Even experienced evaluators have low interrater reliability when assessing complex bodies of evidence consisting of different study designs(11) (which is the case in many rheumatic diseases or specific outcomes). GRADE's second limitation is its predictive validity of assessments of certainty (in effect estimates), which has been considered a limitation. For example, high-quality evidence has been influenced more by new data than anticipated.(10, 12) GRADE's third limitation, which is particularly relevant for the development of guidelines in PsA, occurs when multiple treatment options are available without data from head-to-head comparisons. In these cases, a hierarchical order of treatments forced by PICO questions and GRADE grading is not always possible and might lead to surprising recommendations. Furthermore, a series of pairwise comparisons and resulting recommendations is challenging for a clinician to employ when deciding between more than two potential therapies, decreasing the relevance to the practicing clinician.

Given the heterogeneity and complexity of PsA, strict adherence to GRADE may not be feasible when summarizing evidence regarding all exigencies. For example, because individual patients can have involvement

across various domains of clinical involvement (e.g., peripheral arthritis and skin involvement), with varying levels of disease activity across domains, existing studies may only provide limited evidence to inform treatment recommendations for such complex presentations. Next, there are multiple treatment options that are now available, many of which work differently across different domains. Furthermore, the many treatments that are available have mostly not been compared across domains in head-to-head studies. For most PICOs, where there is little evidence supporting the superiority of one particular therapy over another, no attempt will be made to recommend specific treatments for the different domains of involvement. Rather, a list of options with their quality of evidence and strength of recommendations will be provided.

As an ongoing quality assurance process throughout the development of the GRAPPA guidelines, and in order to clearly and systematically describe the process of guideline development, we will refer to the AGREE II (Appraisal of Guidelines Research and Evaluation) instrument.⁽¹³⁾ The AGREE II instrument includes six quality-related domains that can be used to assess the process of guideline development and reporting: (1) scope and purpose, (2) stakeholder involvement, (3) rigor of development, (4) clarity of presentation, (5) applicability, and (6) editorial independence.⁽¹³⁾

For the latest version of GRAPPA treatment recommendations, it was considered relevant to update a set of overarching principles that will help guide the approach to therapy (Table 1). These principles were developed with both patient and clinician involvement. They will not only help delineate the process of

treatment recommendation development, but will also serve as reminders in the clinic for clinicians and patients during the course of disease care. These include a combination of evidence-based principles, as far as possible, and experience-based approaches for goals of treatment.

Systematic Searches and Evidence Synthesis

Based on the key questions (PICO), a systematic literature review will be performed, summarizing evidence from the most recent sources and taking into account the quality, consistency, and strength of evidence.

Methodologically, a difference in the newest GRAPPA guidelines' literature search is the engagement and use of methodologists (NC, DvdW) who have expertise in literature searches and evidence synthesis alongside clinical expertise from the GRAPPA working groups, as used in the past). The multidisciplinary core team will work together with GRAPPA members on formulating PICO questions that are central to the GRADE process, searching and selecting studies to inform the guideline, extracting and synthesizing data, and grading the evidence.

For the literature search, results from peer-reviewed RCTs will be considered as the main source of evidence. Other study designs will be considered when a lack of RCT evidence is identified or when other designs are required for specific PICO. In most cases, 3 bibliographic databases will be searched: MEDLINE (Ovid), EMBASE (Ovid), and Cochrane Library. The searches will start from the date the searches underpinning the current recommendations were run (February 19, 2013). In addition, published abstracts

from the European League Against Rheumatism (EULAR), the American College of Rheumatology (ACR), the European Academy of Dermatology and Venereology (EADV), and the American Academy of Dermatology (AAD) will be considered in the evidence synthesis to assure the most up-to-date information is incorporated in the new guideline.

Results from searches will be downloaded into Endnote and deduplicated. Unique references will then be imported into Covidence for screening by pairs of independent reviewers (one methodological expert and one clinical expert). The screening will be performed against agreed selection criteria (e.g., RCT design, appropriate population) with inclusion/exclusion tables designed to aid the process. The core team and the domain groups will work together to extract and summarise data in tables. Using these evidence tables, groups for each domain will work toward providing information to enable the formulation of recommendations and achieving consensus regarding the strength of recommendations for the final guidelines.

Given expected heterogeneity, evidence will be synthesized using a narrative approach. Initially, groups will focus on new evidence that is identified from the updated searches. They will assess whether this new evidence confirms previous treatment recommendations (and, if so, if it alters the strength of the recommendation), or if current recommendations require amendment. For PICOs where new, updated recommendations are needed, new evidence will be integrated with evidence from the previously published evidence tables.

Biosimilars and Tapering or Discontinuing Therapy

Biosimilars and the idea of tapering or discontinuing therapy when patients are doing well clinically are important newer developments and concepts in medicine that are relevant to PsA but that have not yet been well studied in PsA. While relevant globally, these concepts seem to transcend any particular PsA domain. As they are important and relevant to patient care, they are being addressed in these GRAPPA recommendations.

Biosimilars. The introduction of biosimilars has impacted the practice of rheumatology in many jurisdictions worldwide. A subgroup of GRAPPA members that includes patient research partners (PRPs) was convened in 2018 and created a set of ideas that form a position statement on biosimilars that are relevant to PsA. These will be reviewed and refined as necessary with newer data and will be incorporated into the final document.

Tapering or discontinuing therapy. An issue that has emerged and grown in recent years due to greater clinical success (which is often owed to novel therapies and newer treatment paradigms) is the idea of tapering or even discontinuing treatments with patients who are doing well clinically (e.g., in remission). There is a paucity of data on this related to PsA, as well as across domains. It is a topic that has not only scientific and medical implications, but also socioeconomic ramifications, as some payors globally mandate treatment reduction for patients with some rheumatic diseases who have achieved certain goals. As they impact patient care directly, tapering and discontinuing therapy were considered by a subgroup of GRAPPA members and PRPs who suggested

related core principles. These will be reviewed, refined, and added to the final document.

Summary

Because there have been many significant advances related to therapies and treatment approaches for PsA, it is necessary to update the GRAPPA PsA treatment recommendations. This process has begun, with acknowledgement of the evolution in methods to create such documents. Here, we review the basis for the approach that will be taken for the next version of the GRAPPA PsA treatment recommendations.

References

1. Ritchlin CT, Kavanaugh A, Gladman DD, Mease PJ, Helliwell P, Boehncke WH, et al. Treatment recommendations for psoriatic arthritis. *Ann Rheum Dis* 2009;68:1387-94.
2. Coates LC, Kavanaugh A, Mease PJ, Soriano ER, Laura Acosta-Felquer M, Armstrong AW, et al. Group for Research and Assessment of Psoriasis and Psoriatic Arthritis 2015 Treatment Recommendations for Psoriatic Arthritis. *Arthritis Rheumatol* 2016;68:1060-71.
3. Kavanaugh AF, Ritchlin CT, Committee GTG. Systematic review of treatments for psoriatic arthritis: an evidence based approach and basis for treatment guidelines. *J Rheumatol* 2006;33:1417-21.
4. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126:376-80.
5. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007;7:16.
6. 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;336:924-6.
7. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004;4:38.
8. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evid Based Med* 2008;13:162-3.
9. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
10. Norris SL, Bero L. GRADE Methods for Guideline Development: Time to Evolve? *Ann Intern Med* 2016;165:810-1.
11. Berkman ND, Lohr KN, Morgan LC, Richmond E, Kuo TM, Morton S, et al. Reliability Testing of the AHRQ EPC Approach to Grading the Strength of Evidence in Comparative Effectiveness Reviews. Rockville (MD); 2012.
12. Gartlehner G, Dobrescu A, Evans TS, Bann C, Robinson KA, Reston J, et al. The predictive validity of quality of evidence grades for the stability of effect estimates was low: a meta-epidemiological study. *J Clin Epidemiol* 2016;70:52-60.
13. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *J Clin Epidemiol* 2010;63:1308-11.