



Sociedad Española  
de Reumatología -  
Colegio Mexicano  
de Reumatología

# Reumatología Clínica

[www.reumatologiaclinica.org](http://www.reumatologiaclinica.org)



## Editorial

### GRADE system, systematic and transparent evaluation<sup>☆</sup>

### Sistema GRADE, evaluación sistemática y transparente

Claudia Mendoza Pinto,<sup>a,b</sup> Mario García Carrasco<sup>a,b,\*</sup>

<sup>a</sup> Unidad de Investigación de Enfermedades Autoinmunes Sistémicas, Hospital General Regional N.º 36, IMSS, Puebla, Puebla, Mexico

<sup>b</sup> Departamento de Reumatología e Inmunología, Facultad de Medicina, Benemérita Universidad Autónoma de Puebla, Puebla, Puebla, Mexico



In recent years, the GRADE system (Grade of Recommendation, Assessment, Development and Evaluation) has been utilized in the publication of guidelines and recommendations, mostly concerning treatment, both with drug and non-drug therapies in different areas of medicine.<sup>1–3</sup> Very recently, a number of clinical practice guidelines in the field of rheumatology have employed this system, including guidelines dealing with polymyalgia rheumatica, rheumatoid arthritis and fibromyalgia, among others.<sup>4–8</sup>

Evidence-based medicine facilitates headway in this uncertain terrain and aid in clinical decision making. Evidence-based guidelines have undergone a profound transformation in recent years. There are a number of systems to evaluate scientific evidence and assign different grades of recommendation. However, these systems have certain drawbacks, as they do not include a balance of risks and benefits, nor do they take into account resource use or costs, and the fact that they were mostly developed by consensus of expert opinion and have not been validated. Clinical practice guideline panel members can have strong opinions or academic biases concerning a specific area in which they have clinical experience or other biases with respect to interactions with academic colleagues or coworkers in the industry.<sup>9</sup> In an attempt to address these problems, GRADE requires systematic and pragmatic searches of the literature and summarizing the evidence, ideally based on grouped treatment effects and produced by panel members with no conflicts of interest or independent methodologists. The latter helps to ensure an impartial and reproducible evaluation of the literature addressing a specific clinical question.

The GRADE working group is a multidisciplinary international collaboration of experts in developing a common, transparent and sensible system for rating quality of evidence and the strength of recommendations.<sup>10</sup> The GRADE system has been adopted by more than 80 agencies and organizations recognized worldwide, such as the World Health Organization (WHO), the Cochrane Collaboration

Group, the United Kingdom National Institute for Health, and the National Institute for Health and Clinical Excellence (NICE), among others.

In the development of GRADE, the authors considered a wide range of clinical questions, including diagnosis, screening, prevention and treatment. To go from a generic clinical question to one specifically formulated in such a way that it facilitates literature searches and the drafting of recommendations for each question, the method referred to as Patients – Intervention – Comparison – Outcome (PICO) was preferred. On drawing up the clinical questions in PICO format, the issue is specifically defined, without ambiguity. Moreover as each type of question corresponds to a type of study in which the design is suitable for the response. The format aids in doing a literature search.<sup>11</sup>

The advantages of the GRADE approach are that it: (a) carefully considers the relative importance of the outcome variables and selects those that are most significant; (b) provides detailed descriptions of the criteria for quality of evidence with respect to specific results or outcomes and uses explicit definitions and sequential judgements during the categorization process; (c) separates the quality of evidence from the strength of the recommendations; and (d) moreover, considers the balance between benefits and risks, patient values and resource use or costs. It also provides tables showing the so-called evidence profiles (EP) and summary of findings (SoF). Finally, software has been developed, with its associated help files, that facilitate the development of EP and SoF tables based on EP.<sup>12,13</sup>

The GRADE system is composed of 8 criteria to evaluate quality of evidence; 5 of the 8 criteria are capable of downgrading quality of evidence, even in a randomized controlled trial (RCT): risk of biases, inconsistency of the results across studies, indirectness, imprecision and publication biases.<sup>14,15</sup> For example, a RCT conducted with inadequate blinding of the sequence and with a high rate of dropouts should not be considered equivalent to a well-performed RCT. Moreover, 3 more criteria were proposed with the potential to increase confidence: a strong association without confounders, the existence of a dose–response gradient based on studies without the problems of biases or imprecisions and the evidence that all the possible confounders or biases could have reduced the observed

<sup>☆</sup> Please cite this article as: Mendoza Pinto C, García Carrasco M. Sistema GRADE, evaluación sistemática y transparente. Reumatol Clin. 2018;14:65–67.

\* Corresponding author.

E-mail address: [mgc30591@yahoo.com](mailto:mgc30591@yahoo.com) (M. García Carrasco).

effect.<sup>16</sup> Thus, an observational report like a case-control study with a strong association, demonstrating a dose-response gradient, which in other circumstances would suggest a weak design, with the GRADE approach would produce evidence at the level of a RCT. The GRADE method proposes 4 levels to express the quality of evidence: high, moderate, low and very low.<sup>14</sup>

The first GRADE criterion, risk of bias or design limitations, is conceptually a matter of the internal validity of a scientific study. The degree of risk of bias can be determined by a careful reading of the methods section of each original study and evaluating how well the authors carried out the planning and performance of the study.<sup>17</sup> There is evidence in medicine that methodological imperfections in a RCT can have an influence on the estimated effect, which is usually exaggerated.<sup>18</sup>

The second criterion, inconsistency of the results across the studies included in a systematic review, signifies that the results deviate from one another, and this naturally leads to decreased confidence in the effectiveness of the estimate. If the original studies are clinically homogeneous (responding to the question posed by the investigation) and the methodological quality is high, but the results are inconsistent, then statistical analyses will probably demonstrate that there is heterogeneity in the results.<sup>19</sup>

The third GRADE criterion, imprecision or absence of direct evidence, refers to any deviation from the research question of the studies included in the systematic review. In cases of the absence of direct comparisons between the interventions being considered, or of substantial differences between the available studies and the population, the interventions or outcomes put forward in the question of interest, we may find that we only have access to indirect information. The use of surrogate outcomes may not be associated with the primary outcome. This may cause problems of applicability.<sup>20</sup>

The fourth GRADE criterion, imprecision, conceptually reflects the random variation in the estimate of the outcome and is different from the internal validity. If the original studies in a systematic review are clinically homogeneous and all of them have a low risk of bias, it is appropriate to conduct a meta-analysis and obtain an overall estimate. The 95% confidence interval (CI) is frequently interpreted to signify that, with a certainty of 95%, the true value of a parameter can be found within the given range. The information on the width of this interval can be used as the basis for clinical inference. For example, this enables the conclusion that the CI have exceeded the clinically important minimal difference.<sup>21</sup>

The fifth GRADE criterion, which can reduce the confidence in the results of a systematic review, is publication bias. The selective report of outcomes is a matter of the internal validity of a given study, and should be included in the criteria for "limitations for study quality". When individual studies are not published, there can be biases in systematic reviews. The existence of publication biases is one of the potential sources of risk of bias in systematic reviews. The obligatory risk of registry of clinical trials has enhanced the possibilities of identifying publication bias.<sup>22</sup>

We must recognize that the system has certain limitations; thus, firstly, the method was initially developed to respond to questions on alternative interventions, especially for treatment or prevention, not for risk or prognosis, and it has problems with respect to diagnostic tests, public health issues and health care systems. However, in recent years, adaptations of this method have been designed for diagnostic<sup>23</sup> and prognostic studies,<sup>24</sup> which are now being used in systematic reviews. Secondly, although the system employs highly systematic, transparent and reproducible judgements, it does not completely eliminate possible disagreements in the evaluation of evidence or in deciding alternative courses, given

that there is always a subjective impregnation in every judgement. Finally, we should point out that a number of researchers analyzing complex systematic reviews have identified difficulties in applying the criteria for the evaluation of quality using GRADE in complex interventions.<sup>25</sup>

## References

- Guyatt GH, Norris SL, Schulman S, Hirsh J, Eckman MH, Akl EA, et al. Methodology for the development of antithrombotic therapy and prevention of thrombosis guidelines: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141 Suppl, 53S–70S.
- Lipsky BA, Berendt AR, Cornia PB, Pile JC, Peters EJG, Armstrong DG, et al. 2012 Infectious Diseases Society of America clinical practice guideline for the diagnosis and treatment of diabetic foot infections. *Clin Infect Dis*. 2012;54:e132–73.
- Raghu G, Rochwerf B, Zhang Y, Garcia CAC, Azuma A, Behr J, et al. An Official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. *Am J Respir Crit Care Med*. 2015;192:e3–19.
- Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of non-pharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res (Hoboken)*. 2012;64:465–74.
- Dejaco C, Singh YP, Perel P, Hutchings A, Camellino D, Mackie S, et al. 2015 recommendations for the management of polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Ann Rheum Dis*. 2015;74:1799–807.
- Ward MM, Deodhar A, Akl EA, Lui A, Ermann J, Gensler LS, et al. American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network 2015. Recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. *Arthritis Rheumatol (Hoboken, NJ)*. 2016;68:282–98.
- Singh JA, Saag KG, Bridges SLJ, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol (Hoboken, NJ)*. 2016;68:1–26.
- Macfarlane GJ, Kronisch C, Dean LE, Atzeni F, Hauser W, Fluss E, et al. EULAR revised recommendations for the management of fibromyalgia. *Ann Rheum Dis*. 2017;76:318–28.
- Guyatt G, Akl EA, Hirsh J, Kearon C, Crowther M, Gutterman D, et al. The vexing problem of guidelines and conflict of interest: a potential solution. *Ann Intern Med*. 2010;152:738–41.
- Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines. 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64:383–94.
- 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.
- Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines. 12. Preparing summary of findings tables—binary outcomes. *J Clin Epidemiol*. 2013;66:158–72.
- Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines. 13. Preparing summary of findings tables and evidence profiles—continuous outcomes. *J Clin Epidemiol*. 2013;66:173–83. <http://dx.doi.org/10.1016/j.jclinepi.2012.08.001>.
- Balslem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines. 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64:401–6.
- Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. 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.
- Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines. 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64:1311–6.
- Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-coello P, et al. GRADE guidelines. 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol*. 2011;64:407–15.
- Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008;336:601–5.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines. 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol*. 2011;64:1294–302.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines. 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol*. 2011;64:1303–10.
- Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-coello P, Rind D, et al. GRADE guidelines. 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol*. 2011;64:1283–93.

22. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines. 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol.* 2011;64:1277–82.
23. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ.* 2008;336:1106–10.
24. Huguet A, Hayden JA, Stinson J, Mcgrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Syst Rev.* 2013;2:71.
25. Movsisyan A, Melendez-Torres CJ, Montgomery P. Users identified challenges in applying GRADE to complex interventions and suggested an extension to GRADE. *J Clin Epidemiol.* 2016;70:191–9.