



Sociedad Española
de Reumatología -
Colegio Mexicano
de Reumatología

Reumatología Clínica

www.reumatologiaclinica.org



Editorial

Sistema GRADE, evaluación sistemática y transparente



GRADE system, systematic and transparent evaluation

Claudia Mendoza Pinto^{a,b} y Mario García Carrasco^{a,b,*}

^a Unidad de Investigación de Enfermedades Autoinmunes Sistémicas, Hospital General Regional N.º 36, IMSS, Puebla, Puebla, México

^b Departamento de Reumatología e Inmunología, Facultad de Medicina, Benemérita Universidad Autónoma de Puebla, Puebla, Puebla, México

En los últimos años el sistema GRADE (por su acrónimo en inglés: *Grade of Recommendation, Assessment, Development, and Evaluation*) se ha utilizado en publicaciones de guías o recomendaciones, principalmente de tratamiento, tanto farmacológicas como no farmacológicas en diferentes áreas de la medicina¹⁻³. Muy recientemente, diversas guías de práctica clínica en el ámbito de la reumatología han utilizado este sistema, como son las guías de polimialgia reumática, artritis reumatoide, fibromialgia, entre otras⁴⁻⁸.

La medicina basada en la evidencia facilita moverse en este terreno incierto y ayuda en la toma de decisiones clínicas. Las guías basadas en evidencia han sufrido una intensa transformación en los últimos años. Existen varios sistemas para graduar la evidencia científica y asignar distintos grados de recomendación, sin embargo, estos sistemas presentan algunos inconvenientes, al no realizar un balance riesgos-beneficios ni tener en cuenta el consumo de recursos o costes, así como el hecho de que se desarrollaron principalmente por el consenso de opinión de expertos y no fueron validados. Miembros del panel de guías de práctica clínica pueden tener fuertes opiniones o sesgos académicos en torno a un área específica en la cual ellos tienen experiencia clínica o pueden tener otros sesgos respecto a interacciones con colegas académicos o compañeros de la industria⁹. En un intento de dirigir esto, GRADE requiere de búsquedas sistemáticas y pragmáticas de la literatura y la producción de sumarios de la evidencia, idealmente basados en efectos de tratamiento agrupados y producidos con miembros del panel sin conflicto de intereses o metodólogos independientes. Esto último ayuda a asegurar una evaluación justa y reproducible de la literatura dirigida a una pregunta clínica específica.

El grupo de trabajo GRADE es una colaboración internacional multidisciplinaria de expertos en desarrollar un sistema común, transparente y razonable para calificar la calidad de la evidencia y la fuerza de las recomendaciones¹⁰. El sistema GRADE ha sido adoptado por más de 80 agencias u organizaciones reconocidas a nivel mundial, como la Organización Mundial de la Salud (OMS), el Grupo de Colaboración Cochrane, el *National Institute for Health and Clinical Excellence* (NICE), entre otros.

En el desarrollo de GRADE, los autores consideraron un rango amplio de preguntas clínicas, incluyendo diagnóstico, tamizaje, prevención y tratamiento. Para pasar de una pregunta clínica genérica a una formulada de forma específica de tal modo que facilite la búsqueda bibliográfica y la elaboración de recomendaciones para cada cuestión, se ha preferido el método PICO, acrónimo de Pacientes-Intervención-Comparación-resultados/*Outcomes*. Al formular las preguntas clínicas en formato PICO queda definido de forma concreta y sin ambigüedad lo que se pretende averiguar y, además, como a cada tipo de pregunta corresponde un tipo de estudio con el diseño adecuado para su respuesta, el formato ayuda a realizar la búsqueda en la literatura¹¹.

Las ventajas del abordaje GRADE son que, a) sopesa la importancia relativa de las variables de resultado o desenlace y elige las que son clave; b) proporciona detalladas descripciones de los criterios de calidad de evidencia respecto a resultados o desenlaces concretos y usa definiciones explícitas y juicios secuenciales durante el proceso de categorización; c) separa la calidad de la evidencia de la fuerza de las recomendaciones; d) además, considera el balance entre beneficios y riesgos, los valores del paciente y el consumo de recursos o costes. También proporciona las tablas de los llamados perfiles de evidencia y sumario de hallazgos. Finalmente, un programa de ordenador, con sus archivos de ayuda asociados, facilita el desarrollo de los cuadros de perfiles de evidencia (EP) y cuadros que resumen los hallazgos (SoF) basados en los EP^{12,13}.

El sistema GRADE se compone de 8 criterios para evaluar la calidad de la evidencia; 5 de los 8 criterios tienen el potencial de disminuir la calidad de la evidencia, incluso de un ensayo clínico controlado (ECC): riesgo de sesgos, inconsistencia de los resultados a través de los estudios, incertidumbre de la evidencia (*indirectness*), imprecisión y sesgos de publicación^{14,15}. Por ejemplo, un ECC realizado con una pobre ocultación de la secuencia y con alta deserción no debería ser considerado equivalente a ECC bien realizado. Además, se han propuesto 3 criterios más con el potencial de incrementar la confianza: una fuerte asociación sin factores de confusión, existencia de gradiente dosis-respuesta basado en estudios sin problemas de sesgo o imprecisiones y la evidencia de que todos los posibles factores de confusión o sesgos podrían haber reducido el efecto observado¹⁶. Por lo tanto, un estudio observacional como es uno de casos y controles con una fuerte asociación,

* Autor para correspondencia.
Correo electrónico: mgc30591@yahoo.com (M. García Carrasco).

que demuestra un gradiente dosis-respuesta, que en otras circunstancias sugeriría un diseño débil, con el abordaje GRADE puede producir evidencia a nivel de un ECC. El método GRADE propone 4 niveles para expresar la calidad de la evidencia: alta, moderada, baja y muy baja¹⁴.

El primer criterio de GRADE, riesgo de sesgos o limitaciones en el diseño, es conceptualmente materia de la validez interna de un estudio científico. El grado del riesgo de sesgo puede estar determinado por una lectura cuidadosa de la sección de métodos de cada estudio original y evaluando qué tan bien se llevaron a cabo la planeación y la ejecución del estudio¹⁷. Existe evidencia en medicina sobre cómo las imperfecciones metodológicas en un ECC pueden influir en el efecto estimado, lo cual usualmente es exagerado¹⁸.

El segundo criterio, inconsistencia de los resultados a través de los estudios incluidos en una revisión sistemática, significa que los resultados se desvían unos con otros, y esto naturalmente conduce a una confianza disminuida de la efectividad de la estimación. Si los estudios originales son clínicamente homogéneos (respondiendo a la misma pregunta de investigación) y son de una alta calidad metodológica, pero hay inconsistencia en los resultados, entonces las pruebas estadísticas probablemente mostrarán que hay heterogeneidad en los resultados¹⁹.

El tercer criterio de GRADE, incertidumbre o ausencia de evidencia directa, se refiere a cualquier desviación en la pregunta de investigación entre los estudios incluidos en la revisión sistemática. En situaciones de ausencia de comparaciones directas entre las intervenciones consideradas, o de diferencias importantes entre los estudios disponibles y la población, las intervenciones o los desenlaces planteados en la pregunta de interés, podemos encontrarnos con que únicamente disponemos de información indirecta. El uso de desenlaces subrogados puede no asociarse con el resultado primario. Esto puede ocasionar problemas con la aplicabilidad²⁰.

El cuarto criterio de GRADE, imprecisión, refleja conceptualmente la variación aleatoria en la estimación del desenlace y es distinta de la validez interna. Si los estudios originales en una revisión sistemática son clínicamente homogéneos y todos tienen un riesgo bajo de sesgos, es apropiado llevar a cabo un metaanálisis y obtener una estimación global. El intervalo de confianza del 95% con frecuencia interpreta que dentro del rango dado se encuentra el valor real de un parámetro con el 95% de certeza. La información sobre la amplitud de este intervalo puede ser usada como base para la inferencia clínica. Por ejemplo, esto permite una conclusión de que los IC han excedido la diferencia mínima clínicamente importante²¹.

El quinto criterio de GRADE que puede disminuir la confianza en los resultados de una revisión sistemática es el sesgo de publicación. El reporte selectivo de desenlaces es materia de la validez interna de un estudio individual y debería incluirse en el criterio para «limitaciones para la calidad del estudio». Cuando los estudios individuales no son publicados pueden emerger sesgos en las revisiones sistemáticas. La existencia de sesgos de publicación es uno de los recursos potenciales de riesgo de sesgos en revisiones sistemáticas. El registro obligatorio de ensayos clínicos ha incrementado las posibilidades de identificar sesgos de publicación²².

Se debe reconocer que el sistema tiene algunas limitaciones; así, en primer lugar, el método se ha desarrollado inicialmente para responder a cuestiones sobre intervenciones alternativas, sobre todo de tratamiento o prevención, no sobre riesgo o pronóstico, y tiene dificultades respecto a pruebas diagnósticas, temas de salud pública o sistemas de salud. Sin embargo, en los últimos años se han realizado adaptaciones de este método para estudios de diagnóstico²³ y de pronóstico²⁴, los cuales ya se han comenzado a utilizar en revisiones sistemáticas. Segundo, aunque el sistema hace juicios de una forma altamente sistemática, transparente y reproducible, no elimina por completo los desacuerdos que pueden existir al valorar una evidencia o al decidir cursos alternativos de acción,

puesto que siempre hay una impregnación subjetiva en todo juicio. Finalmente, cabe señalar que diversos investigadores de revisiones sistemáticas complejas han identificado dificultades en aplicar los criterios de evaluación de la calidad por GRADE en intervenciones complejas²⁵.

Bibliografía

- 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 2 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, Rochweg B, Zhang Y, García 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 nonpharmacologic 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>
- Balshem 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). 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. 2011;64:1283-93.
- 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.