

El sistema GRADE: un cambio en la forma de evaluar la calidad de la evidencia y la fuerza de recomendaciones

IGNACIO NEUMANN^{1,2,5,6,7}, TOMÁS PANTOJA^{1,3},
BLANCA PEÑALOZA^{1,3}, LORENA CIFUENTES^{1,4}, GABRIEL RADA^{1,2,6,7}

The GRADE system: a change in the way of assessing the quality of evidence and the strength of recommendations

Individual clinicians and organizations making health care decisions should not only consider the magnitude of the benefits and harms of different courses of action (interventions), but also the confidence we can have in those estimates. The Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) approach offers a systematic and transparent way to summarize the evidence, to rate the confidence we can have in the effects of the interventions and to move from evidence to recommendations. The GRADE approach has been adopted by several organizations worldwide, including the World Health Organization and the Cochrane Collaboration. In Chile, this approach has already been used by guidelines produced by the Chilean Ministry of Health. In this paper we describe the core concepts of the GRADE approach to rate the quality of the evidence and to grade the strength of recommendations. As clinicians, being familiar with such concepts may be helpful to make decisions informed by the best available evidence.

(Rev Med Chile 2014; 142: 630-635)

Key words: Evidence-Based Practice; Decision Making; Practice Guidelines as Topic; Review.

¹Programa de Salud Basada en Evidencia.

²Departamento de Medicina Interna.

³Departamento de Medicina Familiar.

⁴Departamento de Pediatría. Facultad de Medicina, Pontificia Universidad Católica de Chile. Santiago, Chile.

⁵Departamento de Epidemiología Clínica y Bioestadística. McMaster University, Hamilton, Canadá.

⁶GRADE Working Group.

⁷Fundación Epistemonikos.

Correspondencia a:
Dr. Ignacio Neumann
ignacio.neumann@gmail.com

Las diferentes personas y grupos que toman decisiones en salud no solo deben considerar la magnitud de los efectos de diferentes cursos de acción (intervenciones) sino también la confianza que es posible tener en dichas estimaciones, ya sea en el contexto de una revisión sistemática o en la elaboración de recomendaciones para una guía de práctica clínica. El concepto “calidad de la evidencia” refleja la confianza que podemos tener en que conocemos los efectos de una intervención. La “fuerza de la recomendación” distingue situaciones en donde la evidencia muestra que una alternativa es claramente superior y,

en consecuencia, como clínicos debiéramos tomar esa alternativa con todos o casi todos nuestros pacientes, de situaciones donde hay incertidumbre respecto de cuál es la mejor alternativa y, por tanto, debiéramos considerar los valores y preferencias de los pacientes y las circunstancias clínicas para tomar una decisión (idealmente utilizando un enfoque de decisiones compartidas).

Hasta hace poco tiempo existían decenas de sistemas para clasificar la calidad de la evidencia y la fuerza de las recomendaciones¹. Muchos de ellos utilizados únicamente por el grupo u organización que los había desarrollado. Afortunadamente,

durante la última década el grupo de trabajo GRADE (por su acrónimo en inglés: ‘*Grades of Recommendation, Assessment, Development, and Evaluation*’) ha elaborado y refinado un método para evaluar la calidad de la evidencia y la fuerza de las recomendaciones que ha sido adoptado por múltiples organizaciones alrededor del mundo, incluyendo la *Organización Mundial de la Salud* y la *Colaboración Cochrane*. Muy probablemente GRADE será el sistema dominante para clasificar la calidad de evidencia y las recomendaciones de guías de práctica clínica en un futuro no muy lejano. En Chile, el sistema GRADE ya se ha comenzado a utilizar^{2,3} y forma parte integral del nuevo manual metodológico para elaboración de guías de práctica clínica que está desarrollando el Ministerio de Salud⁴.

En este artículo se describen los conceptos principales del sistema GRADE para evaluar la calidad de la evidencia y clasificar las recomendaciones de una guía de práctica clínica (Figura 1).

Obteniendo y resumiendo la evidencia

Una revisión sistemática de la evidencia puede ser publicada por sí misma o bien ser parte del proceso de revisión de la evidencia para elaborar una guía de práctica clínica. En ambos casos, la primera etapa del proceso es la formulación de la pregunta

clínica específica⁵. Para ello, se tiene que identificar claramente la población objetivo, la acción clínica que se quiere evaluar (por ej. una intervención terapéutica o un procedimiento diagnóstico), la (s) comparación (es) y todos los desenlaces que sean importantes para los pacientes. Luego, los autores de la revisión o la guía proceden a reunir toda la evidencia disponible, para sintetizarla cuantitativa o cualitativamente⁶ (Figura 1).

Interpretando la evidencia: calidad de la evidencia

Una vez que se ha establecido la pregunta clínica específica y se ha sintetizado la evidencia disponible (idealmente mediante un metanálisis en el caso de la evidencia respecto a los efectos de las intervenciones), los autores de las revisiones sistemáticas y guías de práctica clínica deben analizar apropiadamente los resultados.

Uno de los avances conceptuales del sistema GRADE es que propone una metodología sistemática y transparente para interpretar la evidencia encontrada. Basándose en 3 aspectos, nos permite establecer qué tan confiados podemos estar en que conocemos los efectos de una intervención. Estos aspectos son: diseño del estudio, factores que disminuyen la confianza y factores que aumentan la confianza.

Diseño del estudio

El proceso propuesto por GRADE se inicia por la consideración del tipo de estudio(s) encontrado(s), según sean éstos estudios observacionales o estudios aleatorizados. Uno de los mayores problemas de los estudios observacionales es su inherente sesgo de selección (los participantes que reciben la intervención difieren sistemáticamente de los que no la reciben), por lo que determina una probabilidad distinta de presentar los desenlaces de interés. El proceso de asignar aleatoriamente a los participantes de un estudio a los diferentes grupos, cuando es correctamente realizado, protege de este sesgo. Por esta razón podemos estar más confiados en los resultados de los estudios controlados aleatorizados que en los de estudios observacionales y, consecuentemente, el sistema GRADE asigna una mayor calidad de la evidencia a los estudios aleatorizados que a los estudios observacionales.

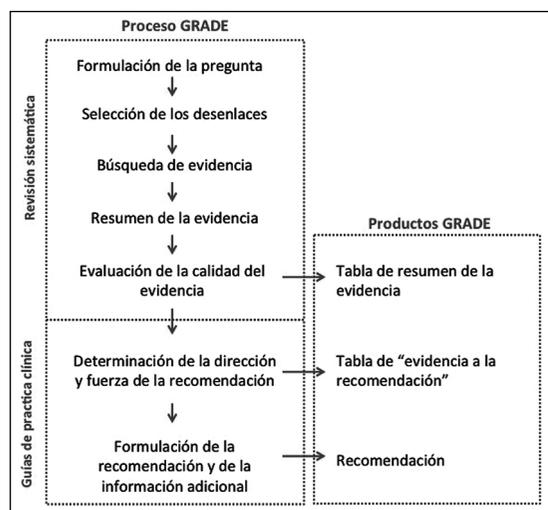


Figura 1. Proceso GRADE.

Factores que disminuyen nuestra confianza

Sin embargo, no solo el diseño del estudio determina la confianza que podemos tener en la evidencia encontrada, dado que diversas limitaciones pueden afectar esta confianza. GRADE reconoce 5 factores que pueden disminuir la calidad de la evidencia: riesgo de sesgo, resultados imprecisos, resultados inconsistentes, problemas de aplicabilidad de la evidencia y la existencia de sesgo de publicación. La presencia de cualquiera de estos factores puede disminuir la calidad de la evidencia.

Factores que aumentan nuestra confianza

Así como existen limitaciones que disminuyen la confianza que podemos tener en la evidencia encontrada, hay ciertas características particulares de la evidencia que, por el contrario, pueden incrementar nuestra confianza. Estas son: un efecto de gran magnitud, existencia de posibles variables “confundentes” (que confunden) que refuercen la conclusión y la existencia de una gradiente dosis-respuesta en el efecto. La presencia de cualquiera de estos factores puede incrementar la calidad de la evidencia.

Integrando los distintos aspectos para establecer la calidad de la evidencia

Integrando sistemáticamente estos 3 aspectos, mediante reglas explícitas, el sistema GRADE establece 4 niveles para calidad de la evidencia: alta, moderada, baja o muy baja. Una calidad de la evidencia alta o moderada refleja que podemos estar relativamente seguros de que conocemos los efectos de la intervención de interés. Una calidad de la evidencia baja o muy baja, por el contrario, refleja que existe incertidumbre respecto de los efectos de la intervención (Tabla 1).

Por ejemplo: una revisión sistemática recientemente publicada en la *Colaboración Cochrane* mostró que la combinación en un mismo comprimido de dosis fijas de aspirina, un antihipertensivo y una estatina, en comparación con los tratamientos por separado, reduce la presión arterial (-7 mm Hg de presión sistólica, y -4 mm Hg de presión diastólica), pero pudiera aumentar el número de eventos cardiovasculares (11 eventos más por cada 1.000 pacientes tratados) y la mortalidad general (2 muertes más por cada 1.000 pacientes tratados)⁷. Para el desenlace “presión arterial”, la calidad de la evidencia fue juzgada como moderada, mientras que para los desenlaces “eventos cardiovasculares” y “mortalidad”, los autores juzgaron la calidad de la evidencia como baja. Esto quiere decir que si aceptamos el juicio de los autores, podemos estar relativamente confiados en que la pastilla de dosis fijas puede reducir la presión arterial en comparación a las drogas por separado. Sin embargo, existe incertidumbre respecto a si la pastilla de dosis fija pudiera aumentar, disminuir o no influir en desenlaces importantes para los pacientes, tales como eventos cardiovasculares y mortalidad.

De la evidencia a las recomendaciones

El producto final de una revisión sistemática adecuadamente elaborada es una estimación del efecto de la intervención en términos absolutos y un juicio respecto a cuán confiados podemos estar en dicha estimación (calidad o certeza de la evidencia). Las guías de práctica clínica usan esta información para elaborar una recomendación para la toma de decisiones en salud.

Otro de los avances conceptuales del sistema GRADE ha sido separar la calidad de la evidencia de la fuerza de las recomendaciones, las cuales

Tabla 1. Niveles de calidad de la evidencia de acuerdo al sistema GRADE

Nivel de evidencia GRADE	Significado
Alto	Estamos relativamente seguros de que el efecto real de la intervención se encuentra cerca de nuestra estimación
Moderado	El efecto real de la intervención probablemente se encuentra cerca de nuestra estimación, pero existe la posibilidad de que sea sustancialmente diferente
Bajo	El efecto real de la intervención puede ser sustancialmente diferente de nuestra estimación
Muy bajo	Es probable que el efecto real de la intervención sea sustancialmente diferente de nuestra estimación

pueden ser clasificadas en fuertes o en débiles (también llamadas condicionales). Las recomendaciones fuertes conllevan el mensaje que la intervención debiera ser ofrecida a todos o casi todos los pacientes si es a favor de la intervención, o que no debiera ser usada en ninguno o casi ninguno de los pacientes si es en contra. Por ejemplo, la guía ARIA (*Allergic Rhinitis and its Impact on Asthma*) emitió una recomendación fuerte a favor del uso de corticoides intranasales en vez de anti-histamínicos para el tratamiento de la rinitis alérgica en adultos⁸. Esto significa que en casi todas las circunstancias resulta apropiado ofrecer corticoides intranasales en vez de antihistamínicos a pacientes con rinitis alérgica. Por el contrario, una recomendación débil conlleva el mensaje que lo que se está proponiendo debe ser considerado a la luz de las circunstancias clínicas y las preferencias de los pacientes (Figura 2 y Tabla 2). Por ejemplo, la guía del *American College of Physicians*

emitió una recomendación débil a favor del uso de memantina en pacientes con demencia⁹. Si bien la memantina puede reducir el deterioro cognitivo de pacientes con demencia, el alto costo del medicamento, los potenciales efectos adversos y la pequeña magnitud del beneficio, condicionan que no sea apropiado utilizarlo en todos los pacientes con demencia. La decisión, por lo tanto, debe ser individualizada para cada paciente.

Existen distintos factores que determinan la fuerza y dirección de las recomendaciones dentro del sistema GRADE (Tabla 2).

La calidad de la evidencia

Cuando la calidad de la evidencia es moderada o alta, podemos esperar recomendaciones fuertes a favor o en contra de la intervención. Sin embargo, esto no siempre es así, ya que la recomendación depende de los otros factores enumerados en esta sección.

Tabla 2. Recomendaciones fuertes y débiles en el sistema GRADE

	Recomendaciones fuertes	Recomendaciones débiles
Significado	La alternativa recomendada puede ser seguida con todos o casi todos los pacientes. Una conversación detallada con el paciente o una revisión cuidadosa de la evidencia que fundamenta la recomendación pudiera no ser necesaria	Si bien la alternativa recomendada es apropiada para la mayoría de los pacientes, la decisión debiera ser individualizada, idealmente mediante un enfoque de decisiones compartidas
Probablemente apropiadas cuando:		
Calidad de la evidencia	Se cumplen TODAS estas condiciones Alta o moderada (o baja o muy baja en circunstancias excepcionales)	Se cumple ALGUNA de estas condiciones Baja o muy baja
Balance de beneficios y riesgos	Una alternativa es claramente superior	El balance de beneficios y riesgos es cercano
Valores y preferencias de los pacientes	Todos o casi todos los pacientes informados toman la misma decisión	Existe variabilidad o incertidumbre respecto de lo que decidirán pacientes informados
Consideraciones de recursos	El costo de la intervención está plenamente justificado	El costo de la intervención pudiera no estar justificado en algunas circunstancias

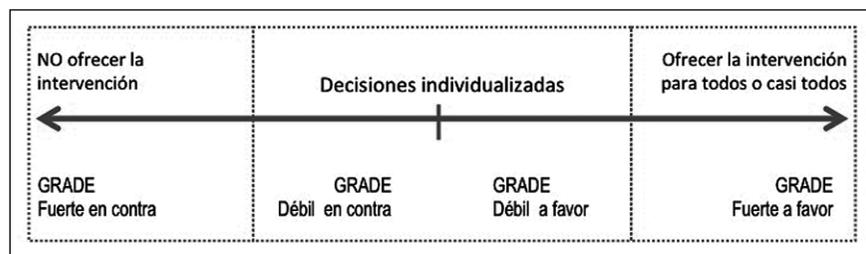


Figura 2. Recomendaciones fuertes y débiles en el sistema GRADE.

Por el contrario, cuando la calidad de la evidencia es baja o muy baja, la mayoría de las veces debiéramos esperar recomendaciones débiles (a favor o en contra de la intervención). Una de las características del sistema GRADE es que permite la elaboración de recomendaciones fuertes en contexto de calidad de la evidencia baja o muy baja, aunque las circunstancias en que esto resulta apropiado son excepcionales¹⁰.

El balance de beneficios y riesgos

El balance de beneficios y riesgos es uno de los principales determinantes de la dirección de la recomendación (a favor o en contra de la intervención). Sin embargo, también es uno de los principales determinantes de la fuerza de la recomendación. Cuando los beneficios superan ampliamente los riesgos (o viceversa), podemos esperar recomendaciones fuertes. Por el contrario, cuando el balance de beneficios y riesgos es cercano podemos esperar recomendaciones débiles.

Los valores y preferencias de los pacientes

Cuando se anticipa que todos o casi todos los pacientes debidamente informados van a tomar la misma opción (por ejemplo, aceptar el uso de aspirina para un infarto agudo al miocardio), podemos esperar recomendaciones fuertes. Por el contrario, cuando existe incertidumbre o variabilidad en lo que los pacientes informados puedan elegir, una recomendación débil resulta más apropiada. Por ejemplo, en el caso del uso de la memantina en pacientes con demencia, la reducción del deterioro cognitivo pudiera no ser deseable para algunos pacientes, especialmente aquellos con demencia avanzada.

Las consideraciones de recursos

Finalmente, cuando los beneficios de una intervención claramente justifican su costo podemos esperar una recomendación fuerte. Cuando los beneficios de una intervención pudieran no justificarlo en todas las circunstancias, una recomendación débil resulta más apropiada.

Integrando los factores

Mediante la integración de los factores descritos, los elaboradores de guías pueden establecer la dirección de la recomendación y su fuerza. Durante los últimos años el *GRADE Working*

Group ha estado trabajando en hacer más explícitos y transparentes los juicios implícitos en la elaboración de las recomendaciones¹¹. Este grupo de trabajo continúa activamente refinando los métodos, simplificando la forma de presentar sus resultados, facilitando su utilización por los clínicos y difundiendo su metodología.

Recomendaciones GRADE

Las recomendaciones GRADE ofrecen una sugerencia o recomendación respecto del curso de acción más deseable, una evaluación explícita de la calidad de la evidencia que sustenta la recomendación y una expresión de la fuerza de la recomendación. Por ejemplo, la siguiente es una recomendación GRADE respecto de la prevención primaria cardiovascular: “Para individuos mayores de 50 años sin un alto riesgo de sangrado, sugerimos el uso de aspirina (75 a 100 mg al día) – Recomendación débil basada en calidad de la evidencia moderada”¹². En este caso, el curso de acción sugerido es el uso de aspirina en dosis bajas, pero además la recomendación nos dice que los elaboradores están relativamente seguros del efecto de la aspirina (calidad de la evidencia moderada), pero nos sugieren individualizar la decisión a las circunstancias clínicas y a los valores y preferencias de nuestro paciente (recomendación débil).

Como profesionales de la salud, debiéramos estar preparados para discutir con nuestros pacientes las recomendaciones débiles. Para ello, las recomendaciones GRADE típicamente ofrecen resúmenes de la evidencia pertinente que pueden ser utilizados en un proceso de decisión compartida con el paciente.

Conclusión

El sistema GRADE ofrece un método sistemático y explícito para clasificar la calidad de la evidencia y la fuerza de las recomendaciones. En los últimos años, un creciente número de revisiones sistemáticas y guías de práctica clínica han comenzado a utilizar GRADE. Familiarizarse con sus conceptos centrales puede ser de ayuda para una toma de decisiones informada por la mejor evidencia disponible.

Referencias

1. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al; GRADE Working Group. 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 (1): 38.
2. Ministerio de Salud de Chile. Guía para el Diagnóstico y Manejo Clínico de casos de Influenza (2013). Disponible en www.minsal.cl (accedido el 15 de mayo del 2014).
3. Neumann I, Cifuentes L. [Toward trustworthy clinical practice guidelines: The Ministry of Health guideline for management of Influenza]. *Rev Med Chile* 2012; 140 (12): 1615-6.
4. Secretaría Técnica AUGE. Desarrollo de guías de práctica clínica: manual metodológico. Ministerio de Salud, Chile; 2013.
5. Claro JC, Lustig N, Soto M, Rada G. [First step: the clinical question]. *Rev Med Chile* 2012; 140 (8): 1067-72.
6. Letelier LM, Manríquez JJ, Rada G. [Systematic reviews and metaanalysis: Are the best evidence?] *Rev Med Chile* 2005; 133 (2): 246-9.
7. de Cates AN, Farr MR, Wright N, Jarvis MC, Rees K, Ebrahim S, Huffman MD. Fixed-dose combination therapy for the prevention of cardiovascular disease. *Cochrane Database Syst Rev* 2014; 4: CD009868.
8. Brozek JL, Bousquet J, Baena-Cagnani CE, Bonini S, Canonica GW, Casale TB, et al; Global Allergy and Asthma European Network; Grading of Recommendations Assessment, Development and Evaluation Working Group. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision. *J Allergy Clin Immunol* 2010; 126 (3): 466-76.
9. Qaseem A, Snow V, Cross JT Jr, Forcica MA, Hopkins R Jr, Shekelle P, et al; American College of Physicians/American Academy of Family Physicians Panel on Dementia. Current pharmacologic treatment of dementia: a clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. *Ann Intern Med* 2008; 148 (5): 370-8.
10. Alexander PE, Bero L, Montori VM, Brito JP, Stoltzfus R, Djulbegovic B, et al. World Health Organization recommendations are often strong based on low confidence in effect estimates. *J Clin Epidemiol* 2014; 67 (6): 629-634.
11. DECIDE Consortium. Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence (DECIDE): protocol and preliminary results. *Implement Sci* 2013; 8: 6.
12. Spencer FA, Guyatt GH. Aspirin in the primary prevention of cardiovascular disease and cancer. In: UpToDate, Basow, DS (Ed), UpToDate, Waltham, MA, 2013.