

El sistema GRADE para la toma de decisiones clínicas y la elaboración de recomendaciones y guías de práctica clínica

Una intervención clínica puede definirse como cualquier procedimiento que el médico indica o aplica al paciente con el fin de mejorar, mantener o restaurar su salud. Decidir si una determinada intervención clínica resulta adecuada para un paciente concreto equivale a determinar si existe un grado razonable de certeza de que el balance entre los beneficios, por un lado, y los riesgos, los inconvenientes y los costes, por el otro, de dicha intervención es lo suficientemente favorable como para que merezca la pena aplicarla. Dicha decisión es, o al menos debería ser, el resultado final de una serie de juicios secuenciales que, por su complejidad, no resulta práctico que los médicos –y los pacientes– realicen sin ninguna ayuda.

Ésta es la razón fundamental de que en los últimos años las guías de práctica clínica (GPC) hayan cobrado cada vez más importancia como instrumentos de ayuda a la toma de decisiones clínicas: paneles de profesionales con una buena comprensión del problema clínico, acceso a la mejor evidencia disponible y tiempo suficiente para la reflexión, formulan recomendaciones para pacientes tipo –representativos– mediante un procedimiento sistemático. La tarea diaria de los médicos clínicos consiste entonces en decidir sobre la adecuación de tales recomendaciones genéricas para su paciente en concreto.

Para que las GPC sean un instrumento que contribuya realmente a la mejora de la toma de decisiones clínicas, es necesario asegurar que el procedimiento por el que se formulan las recomendaciones sea riguroso. Además, si se quiere que los médicos clínicos –y los pacientes– confíen en ellas, el procedimiento ha de ser también transparente. Hace ya más de 25 años el Canadian Task Force on Preventive Health Care (CTFPHC) intentó satisfacer por primera vez los requerimientos de rigor y transparencia en la elaboración de recomendaciones introduciendo el primer sistema de clasificación de la calidad de la evidencia y la fuerza de las recomendaciones¹. Desde entonces numerosas organizaciones e instituciones han ido desarrollando sus propios sistemas y, actualmente, se contabilizan más de cien². Es indudable que muchos de estos sistemas han contribuido a mejorar el proceso de elaboración de las recomendaciones y a facilitar su interpretación; no obstante, su proliferación también ha generado confusión³.

Baste un ejemplo: hay varias GPC basadas en la evidencia para cáncer de próstata, elaboradas por diversas organizaciones de prestigio. Al compararlas se observan diferencias metodológicas muy importantes, particularmente respecto

a los sistemas utilizados para clasificar la calidad de la evidencia y la fuerza de las recomendaciones⁴. Tales diferencias podrían explicar, al menos en parte, los desacuerdos entre las recomendaciones y también generar desconfianza entre los usuarios de ellas.

La formulación de recomendaciones conlleva la realización de juicios sobre la calidad de la evidencia, acerca del balance entre beneficios y riesgos y, en el caso de que este último sea favorable, sobre si es lo suficientemente favorable como para justificar los inconvenientes y los costes. Con frecuencia tales juicios se hacen de forma implícita² o se mezclan los juicios sobre la calidad de la evidencia con los juicios sobre el balance entre beneficios y riesgos. Ambas circunstancias dificultan o hacen imposible a los usuarios de las GPC el discernimiento de la validez y la aplicabilidad de las recomendaciones. Algunos sistemas de clasificación ni siquiera tienen en cuenta el balance entre beneficios y riesgos y asignan la fuerza de recomendación a partir de la calidad de la evidencia exclusivamente; por ejemplo, es frecuente encontrar una recomendación fuerte sólo por el hecho de estar basada en un ensayo clínico aleatorizado bien diseñado y realizado, sin considerar el balance entre beneficios y riesgos –si el balance entre beneficios y riesgos es incierto, lo adecuado puede ser asignar una recomendación débil, aunque la calidad de la evidencia sea alta–. En definitiva, hacer juicios sobre la calidad de la evidencia y la fuerza de las recomendaciones no es tarea fácil, ya que se han de sopesar numerosos y complejos factores. Desde el año 2000, un grupo internacional, integrado en su mayoría por expertos en metodología y por clínicos, muchos de ellos procedentes de las organizaciones que establecieron las clasificaciones más conocidas o de organizaciones de notable peso tradicional o actual en la formulación de recomendaciones (US Preventive Service Task Force [USPSTF], Scottish Intercollegiate Guidelines Network [SIGN], Oxford Center for Evidence Based Medicine, National Institute for Health and Clinical Excellence [NICE], Organización Mundial de la Salud [OMS]), está trabajando en la iniciativa GRADE (Grading of Recommendations Assessment, Development and Evaluation). Su objetivo es desarrollar un enfoque sistemático y explícito –el sistema GRADE– para realizar los complejos juicios que subyacen al clasificar la calidad de la evidencia y la fuerza de las recomendaciones, un sistema que ayude a prevenir errores y a resolver desacuerdos y que facilite la lectura crítica y la comunicación de la información⁵.

TABLA 1

Calidad de evidencia	Diseño de estudio	Disminuir si*	Aumentar si*
Alta	ECA	Importante (-1) o muy importante (-2) limitación de la calidad del estudio	Asociación fuerte, sin factores de confusión, consistente y directa (+1)**
Moderada		Inconsistencia importante (-1)	Asociación muy fuerte, sin amenazas importantes a la validez (no sesgos) y evidencia directa (+2)
Baja	Estudio observacional	Alguna (-1) o gran (-2)*** incertidumbre acerca de que la evidencia sea directa	
Muy baja	Cualquier otra evidencia	Datos escasos o imprecisos (-1) Alta probabilidad de sesgo de notificación (-1)	Gradiente dosis respuesta (+1) Todos los posibles factores confusores podrían haber reducido el efecto observado (+1)

*1 = subir o bajar un nivel (por ejemplo, de alta a intermedia); 2 = subir o bajar dos niveles (por ejemplo, de alta a baja).

**Un riesgo relativo estadísticamente significativo > 2 (< 0,5), basado en evidencias consistentes en 2 o más estudios observacionales, sin factores de confusión plausibles.

***Un riesgo relativo estadísticamente significativo > 5 (< 0,2), basado en evidencia directa y sin amenazas importantes para la validez.

TABLA 2**Decisiones acerca de la fuerza de las recomendaciones****Factores que pueden debilitar la fuerza de una recomendación**

Calidad de la evidencia baja	Crearé una mayor incertidumbre sobre el tamaño (relativo) de los efectos (beneficios y daños)
Incertidumbre acerca del balance entre beneficios y riesgos e inconvenientes	La incertidumbre sobre el riesgo basal, la prevalencia del problema o el estado de salud, que podrían afectar el tamaño del efecto (absoluto)
Incertidumbre o diferencias en los valores	La incertidumbre sobre la importancia relativa de los beneficios e inconvenientes para los afectados, o diferencias en cuanto a lo importante para las distintas personas, lo que podría afectar al balance entre beneficios y riesgos e inconvenientes
Beneficios e inconvenientes netos marginales	Los beneficios esperados netos o los inconvenientes son pequeños (e inciertos)
Incertidumbre sobre si los beneficios netos valen el coste que suponen	La incertidumbre relacionada con la falta de información sobre el coste o si el gasto de los recursos está justificado para el beneficio esperado

Desde la perspectiva GRADE, la calidad de la evidencia y la fuerza de la recomendación indican, respectivamente, hasta qué punto podemos confiar en que el estimador del efecto de la intervención es correcto y hasta qué punto podemos confiar en que la aplicación de la recomendación conllevará más beneficios que riesgos. En las tablas 1 y 2 se indican los criterios del grupo de trabajo GRADE para evaluar la calidad de la evidencia y la fuerza de las recomendaciones⁶. GRADE clasifica la calidad de la evidencia en alta, moderada, baja o muy baja, teniendo en cuenta el diseño del estudio y su ejecución, si la evidencia es directa o indirecta, la consistencia y la precisión de los resultados. Asimismo, distingue entre recomendaciones fuertes y dé-

biles y hace juicios explícitos sobre los factores que pueden afectar a la fuerza de una recomendación, que van más allá de la calidad de evidencia disponible. El GRADE tiene varias ventajas sobre otros sistemas, incluyendo definiciones explícitas y juicios secuenciales durante el proceso de clasificación; ponderación de la importancia relativa de los resultados de la intervención; criterios detallados para asignar la calidad de la evidencia a cada resultado importante y al conjunto de los resultados que se consideran clave a la hora de formular la recomendación; consideraciones sobre el balance entre los beneficios y los riesgos, los inconvenientes y los costes⁶.

Como puede comprobarse a través del ejemplo que sigue, los juicios secuenciales del sistema GRADE guardan similitud con el proceso de toma de decisiones que el clínico sigue en el día a día de la consulta. Disponemos de evidencia de calidad alta, derivada de ensayos clínicos con asignación aleatoria bien diseñados y ejecutados, que muestran que los anticoagulantes orales administrados durante más de un año reducen el riesgo de recurrencias en pacientes que

han experimentado un episodio de trombosis venosa profunda idiopática⁷. Por otra parte, sabemos que los anticoagulantes orales aumentan el riesgo de hemorragia y presentan inconvenientes tales como tener que tomar la medicación y monitorizar el nivel de anticoagulación, además de los costes asociados sobre todo a los programas de monitorización⁷. Por eso, la recomendación de anticoagular durante más de un año a todos los pacientes es débil, ya que el balance de los beneficios, por un lado, y de los riesgos, inconvenientes y costes, por el otro, es incierto y los pacientes bien informados pueden inclinarse por opciones diferentes: mantener la anticoagulación oral más de un año o no⁶.

En el editorial que acompañaba a la traducción al castellano del artículo sobre la clasificación de la calidad de la evidencia y fuerza de las recomendaciones del GRADE Working Group, señalábamos que en los diferentes proyectos de elaboración de GPC por parte de los grupos de trabajo de la semFYC y del Programa de Actividades Preventivas y de Promoción de la Salud (PAPPS) se utilizan diferentes sistemas de clasificación, y comentábamos la necesidad de consensuar un sistema riguroso y aceptado por todos, a pesar de que la apuesta no era fácil⁸.

A lo largo de estos 2 años el comité científico, integrado por profesionales de perfil clínico y metodológico, referentes en sus respectivas responsabilidades dentro de los proyectos semFYC, y los miembros del Grupo de Medicina Basada en la Evidencia (MBE) hemos ido incorporando la metodología GRADE en algunas de las actividades científicas realizadas, como consensos, recomendaciones, GPC y actividades de formación.

Una de las experiencias ha sido el consenso sobre la administración de suplementos de yodo durante el embarazo. Partíamos de recomendaciones distintas elaboradas por diferentes grupos de la semFYC y de otras organizaciones. Aplicando el sistema GRADE, la recomendación de administrar yodo a todas las mujeres embarazadas es débil, ya que la evidencia para el resultado clave –afectación al desarrollo intelectual y defectos psicomotores– es baja y el balance entre los beneficios y los riesgos, inconvenientes y costes –desconocimiento de todos los posibles riesgos, medicalización de un proceso fisiológico, inconvenientes de tomar la medicación, diferentes niveles basales de yodo en la mujer embarazada– es incierto.

En la última reunión del organismo coordinador del PAPPS se consensuó utilizar la metodología GRADE. El PAPPS publica recomendaciones preventivas basándose en las evidencias científicas, la morbilidad en nuestro país, los recursos disponibles y los resultados de la evaluación de las actividades preventivas de los equipos de atención primaria. Tradicionalmente, el PAPPS ha utilizado el sistema de clasificación de la calidad de la evidencia y la fuerza de las recomendaciones elaborado por el CTFPHC.

Otras experiencias actualmente en curso de aplicación del sistema GRADE son la elaboración de una GPC sobre enfermedad pulmonar obstructiva crónica (EPOC) y otra sobre prevención y tratamiento del tromboembolismo. Ambos son proyectos conjuntos con otras sociedades científicas: el primero con la Sociedad de Neumología y Cirugía Torácica y el segundo junto a otras 22 sociedades científicas. La guía sobre EPOC cuenta exclusivamente con financiación privada y su conclusión se prevé para finales del 2008; la financiación de la de trombosis es mixta: pública (GuíaSalud) y privada (8 entidades), y su conclusión está prevista para finales del 2009.

La competencia en la aplicación del sistema GRADE requiere formación. Con este propósito se han impartido cursos de formación de formadores en las tres últimas edi-

ciones de la Escuela de Verano de la semFYC, dirigidos a los miembros de los grupos de trabajo de la semFYC y del PAPPS. También se ha solicitado una ayuda económica al Ministerio de Sanidad y Consumo para realizar cursos de formación a mayor escala.

Otras iniciativas de aplicación del sistema GRADE están actualmente en curso. En concreto, algunas de las GPC actualmente en fase de desarrollo dentro del Programa de elaboración GPC basadas en la evidencia del Plan de Calidad para el Sistema Nacional de Salud están utilizando el sistema GRADE. Además el Programa se ha planteado generalizar la utilización de GRADE en un futuro. De hecho uno de los proyectos a realizar dentro del objetivo 10.2 del Plan de Calidad –Impulsar la elaboración y el uso de guías de práctica clínica vinculadas a las estrategias de salud, consolidando y extendiendo el Proyecto Guía-Salud y formando a profesionales– consiste en «establecer una metodología válida para el Sistema Nacional de Salud para la elaboración de guías de práctica clínica»¹⁰.

GRADE no es todavía un sistema acabado y consolidado¹⁰. El propio GRADE Working Group no lo considera como tal, y está refinándolo y elaborando un manual de aplicación además de nuevas publicaciones que contribuyan a su difusión –próximamente se publicará una serie de artículos en el *British Medical Journal* sobre el GRADE–. Actualmente está en proceso la mejora del instrumento GRADE profile, una aplicación informática de fácil utilización para la elaboración de los resúmenes –perfiles– de la evidencia.

El GRADE es un proyecto riguroso y novedoso que puede ayudar a superar las limitaciones de los sistemas de clasificación previos. En la actualidad son muchas las organizaciones que se interesan en él: OMS, Colaboración Cochrane, Agency for Healthcare Research and Quality (AHRQ), NICE, UpToDate y diversas sociedades científicas, entre otras. No obstante, algunas organizaciones de prestigio como el USPSTF¹¹, aun habiendo incorporado la filosofía GRADE, mantienen su propio sistema de clasificación. Desde la semFYC, a través de su comité científico y del grupo de MBE, apostamos por avanzar en el uso del sistema GRADE en nuestro contexto y queremos debatir su uso y adaptación con todos los miembros de la semFYC que tengan interés en el tema.

Agradecimientos

Agradecemos las aportaciones y comentarios de Rafa Rotaeché, Pablo Alonso y Josep Basora al documento y a todas las personas que han participado en las diferentes experiencias en la aplicación del sistema GRADE.

Mercé Marzo Castillejo^a
y Antonio Montaña Barrientos^b

^aÀmbit d'Atenció Primària Costa de Ponent. Institut Català de la Salut. Barcelona. España. ^bCentro de Salud Ronda Histórica. Distrito de Atención Primaria Sevilla. Servicio Andaluz de Salud. Sevilla. España.

Bibliografía

1. Anonymous: Canadian Task Force on the Periodic Health Examination. The periodic health examination. *CMAJ*. 1979;121:1193-254.
2. The GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. *BMC Health Serv Res*. 2004;4:38.
3. Guyatt G, Gutterman D, Baumann MH. Grading Strength of Recommendations and Quality of Evidence in Clinical Guidelines. Report from an American College of Chest Physicians Task Force. *Chest*. 2006;129:174-81.
4. Dahm P, Kunz R, Schunemann H. Evidence-based clinical practice guidelines for prostate cancer: the need for a unified approach. *Curr Opin Urol*. 2007;17:200-7.
5. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490.
6. Schunemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. *Health Res Policy Syst*. 2006;4:21.
7. Buller HR, Agnelli G, Hull RD, Hyers TM, Prins MH, Raskob GE. Antithrombotic therapy for venous thromboembolic disease: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;126 Suppl 3:401-8.
8. Marzo-Castillejo M, Alonso-Coello P, Rotaeche del Campo R. ¿Cómo clasificar la calidad de la evidencia y la fuerza de las recomendaciones? *Aten Primaria*. 2006;37:5-8.
9. Plan de Calidad para el Sistema Nacional de Salud. Ministerio de Sanidad y Consumo. Madrid: MSyC; 2006. p. 48.
10. Calderon C, Rotaeche R, Etxebarria A, Marzo M, Rico R, Barandiaran M. Gaining insight into the Clinical Practice Guideline development processes: qualitative study in a workshop to implement the GRADE proposal in Spain. *BMC Health Serv Res*. 2006;6:138.
11. Guirguis-Blake J, Calonge N, Miller T, Siu A, Teutsch S, Whitlock E, U.S. Preventive Services Task Force. Current processes of the U.S. Preventive Services Task Force: refining evidence-based recommendation development. *Ann Intern Med*. 2007;147:117-22.