

¿Cómo clasificar la calidad de la evidencia y la fuerza de las recomendaciones?

La revista ATENCIÓN PRIMARIA nos ha brindado la oportunidad de publicar la traducción al castellano del artículo sobre la clasificación de la calidad de la evidencia y fuerza de las recomendaciones desarrolladas por el grupo de trabajo GRADE (Grading of Recommendations Assessment, Development and Evaluation Working Group), publicado recientemente en el *British Medical Journal* (BMJ)¹. En este editorial que acompaña a la traducción presentamos los antecedentes históricos sobre la evaluación de la calidad de las evidencias y la asignación de la fuerza de las recomendaciones, situamos el tema en el contexto internacional actual e intentamos aportar y compartir una serie de reflexiones. La posibilidad de difundir estos materiales es importante y, a la vez, oportuna, y creemos que la sensibilidad que los editores han mostrado hacia este tema repercutirá positivamente en el debate acerca de cómo clasificar la calidad de la evidencia y la fuerza de las recomendaciones en nuestro medio.

La separación entre los «niveles de evidencia» para clasificar la validez de los estudios y el «grado de recomendación» que se confiere a una intervención tiene su origen en la Canadian Task Force on Preventive Health Care (CTFPHC), institución creada en el año 1976 a partir de una conferencia de los ministros de sanidad de Canadá². La metodología empleada en su primer informe sobre actividades de prevención en las personas asintomáticas contempló la idea de que la evidencia científica puede presentarse de manera jerárquica y que algunos diseños de estudio están más sujetos a sesgos que otros y, por lo tanto, justifican en menor medida las decisiones clínicas. En esta misma línea, en 1984, el Servicio de Salud Pública del Departamento de Salud de Estados Unidos estableció la United State Preventive Services Task Force (USPSTF) y adaptó la metodología propuesta por la CTFPHC para hacer una revisión extensa de las intervenciones clínicas preventivas³.

Posteriormente, diversas instituciones y sociedades científicas han adaptado y creado nuevos sistemas de clasificación de la calidad de la evidencia, contabilizándose hasta más de cien⁴. De entre estas clasificaciones destaca la de la Agency for Healthcare Research and Quality (AHRQ) (antiguamente denominada US Agency for Health Care Policy and Research y, que actualmente también integra a la USPSTF)⁵. La clasificación de la AHRQ ha sido un referente importante para las instituciones que evalúan tecnologías sanitarias y para contestar preguntas sobre efecti-

vidad de los tratamientos, en cuyo caso el ensayo clínico aleatorizado se acepta como el diseño de estudio más robusto y que presenta un menor riesgo de sesgo en sus resultados.

Otras propuestas más recientes surgen con el objetivo de superar las limitaciones de las clasificaciones orientadas solamente a las intervenciones de tratamiento y prevención y a las que tienen en cuenta únicamente el diseño del estudio y no otros aspectos de la calidad. Así, la clasificación del Centre for Evidence Based Medicine (CEBM) de Oxford se justifica por la necesidad de contemplar los diseños de estudio relacionados con el diagnóstico, el pronóstico, los factores de riesgo y la evaluación económica⁶. Iniciativas como las del Scottish Intercollegiate Guidelines Network (SIGN) contemplan la necesidad de disponer de un sistema de clasificación que facilite la elaboración de las recomendaciones en el contexto de las guías de práctica clínica (GPC) basadas en la evidencia científica⁷. La clasificación de SIGN tiene en cuenta la calidad de los estudios individuales (tablas con el resumen completo de la evidencia) y el juicio ponderado y explícito acerca de la generalización, la aplicabilidad, la consistencia y el impacto clínico de la evidencia disponible, lo que permite crear un nexo claro entre la evidencia disponible y la recomendación. Otros grupos que elaboran GPC, como el New Zealand Guidelines Group y el National Institute for Clinical Excellence (NICE), también han desarrollado sus propios sistemas de clasificación. Muy recientemente, el NICE ha optado por utilizar una adaptación de la clasificación de SIGN para los estudios de intervención y una adaptación de la clasificación de Oxford para los estudios de pruebas diagnósticas⁸.

Diversas revistas médicas del ámbito americano de la atención primaria (*American Family Physician*, *Journal of Family Practice*, *Journal of American Board Family Practice*) han desarrollado una escala de graduación, denominada Strength of Recommendation Taxonomy (SORT)⁹. Los objetivos de esta clasificación son: *a)* que las revistas de atención primaria dispongan (de forma obligatoria u opcional) de una clasificación uniforme; *b)* que los autores puedan evaluar la fuerza de recomendación a partir de un conjunto de evidencias y asignar un nivel de evidencia a los estudios individuales; *c)* que sea entendible y permita evaluar los estudios de cribado, diagnóstico, tratamiento, prevención y pronóstico; *d)* que sea fácil de usar y no consuma demasiado el tiempo de los autores, los revisores y los

redactores que pueden ser expertos en el contenido del tema, pero no en la apreciación de la lectura crítica o en epidemiología clínica, y e) que sea lo suficientemente claro para que los médicos de atención primaria puedan integrar las recomendaciones en la práctica diaria.

Tradicionalmente, en nuestro entorno de atención primaria, las clasificaciones más utilizadas han sido las de CTFPHC y la USPSTF. A modo de ejemplo, el Programa de Actividades Preventivas y de Promoción de la Salud (PAPPS), aunque no utiliza de forma sistemática ninguna clasificación, en su metodología aboga por la clasificación de CTFPHC. Recientemente y por medio de los diferentes grupos que elaboran GPC se han ido introduciendo los nuevos sistemas de clasificación. A modo de ejemplo, en las GPC de Osakidetza Servicio Vasco de Salud realizadas conjuntamente con OSTEBA (Servicio de Evaluación de Tecnologías Sanitarias) se utiliza la clasificación de SIGN. En las GPC del Institut Català de la Salut se utiliza la clasificación de la AHQR. Y en las realizadas por semFYC junto a las diversas sociedades científicas y el Centro Cochrane Iberoamericano se ha utilizado la clasificación del CEBM de Oxford. La revista *Actualización en Medicina de Familia* (AMF) aboga por la clasificación SORT.

La proliferación de múltiples clasificaciones de la evidencia y sus diferentes interpretaciones puede plantear problemas, entre ellos también de comunicación. Asimismo, esta situación entra en contradicción con el propio concepto de medicina basada en la evidencia (MBE), ya que la MBE ha sido creada en parte como instrumento para reducir las inconsistencias y ayudar a estandarizar la práctica clínica¹⁰. Contrariamente, los diferentes requisitos y restricciones de las clasificaciones disponibles, así como las inconsistencias en la nomenclatura, pueden introducir una amplia gama de prácticas basadas en la evidencia científica. En nuestro medio, esta situación, a la vista de las diferentes clasificaciones empleadas, puede crear una confusión todavía mucho mayor.

En el año 2000, con el objetivo de abordar las deficiencias de los actuales sistemas de clasificación, una amplia colaboración de personas pertenecientes a las diferentes instituciones que han desarrollado las actuales clasificaciones y/o elaboran GPC y recomendaciones crearon el grupo de trabajo GRADE. Este grupo ha elaborado una clasificación con un enfoque sistemático y explícito, a partir de la definición de una serie de criterios necesarios para hacer juicios sobre la calidad de la evidencia y la fuerza de las recomendaciones^{1,11}. El grupo también ha revisado los sistemas de clasificación más conocidos¹², ha realizado estudios para su validación¹³ y ha diseñado un *software* (GRADEpro) para facilitar la elaboración de la síntesis de la evidencia. La información sobre GRADE y sus publicaciones está disponible en <http://www.gradeworking-group.org/>. Actualmente, el grupo de trabajo está avanzando en la aplicación de este sistema a los estudios de

diagnóstico y coste-efectividad. Asimismo, está pilotando un estudio para que en las revisiones sistemáticas de la Colaboración Cochrane se elabore una tabla con GRADEpro, con la síntesis de las evidencias y su correspondiente evaluación de la calidad metodológica.

El sistema de clasificación GRADE juzga la calidad de la evidencia para cada variable importante, teniendo en cuenta el diseño del estudio, la calidad, la consistencia y si la evidencia es directa. Para hacer juicios acerca de la fuerza de la recomendación considera el balance entre beneficios y riesgos, la calidad de la evidencia, la aplicabilidad y el riesgo basal de la población de interés. Este sistema de clasificación se puede aplicar a una amplia variedad de intervenciones y contextos, y equilibra la necesidad de sencillez con la necesidad de considerar de forma global y transparente todos los aspectos importantes. El sistema GRADE persigue llevar a cabo juicios más coherentes y que la comunicación de estos juicios pueda respaldar opciones mejor fundamentadas en la atención sanitaria.

Nuestra colaboración con GRADE Working Group empezó de manera informal con la participación en los seminarios de intercambio de experiencias entre los profesionales que elaboran recomendaciones y GPC, impartidos en los diferentes congresos de la Colaboración Cochrane. Desde el año 2002 nos hemos incorporado de forma activa en el grupo y participamos en las diferentes reuniones y actividades que se están desarrollando. Esta participación en GRADE ha sido y es posible gracias al soporte del Centro Cochrane Iberoamericano (CCIB), REDEGUIAS (proyecto incluido en la Red de Investigación Cooperativa sobre MBE financiado por el Instituto de Salud Carlos III) y la semFYC¹⁴. Nuestro compromiso con el grupo de trabajo GRADE y las instituciones que nos dan soporte es facilitar la difusión del sistema GRADE y sus materiales en nuestro entorno.

El artículo que hemos traducido es uno de los materiales más representativos elaborados por GRADE. El proceso de la traducción se ha llevado a cabo en varias etapas. En la fase inicial, y tal como establecen las normas de traducción del BMJ, hicimos una primera traducción ajustándonos al original palabra por palabra. Esta primera versión fue revisada por miembros de REDEGUIAS. A partir de su comentarios, teniendo en cuenta las dificultades para consensuar una serie de términos (*judgement, directness, trade-off, etc.*) y contando, además, con la experiencia de un miembro de la red en el proceso de traducción al castellano del instrumento AGREE, optamos por encargar a terceras personas no conocedoras del tema una traducción inversa de los párrafos con terminología más significativa. En la elaboración de los sucesivos borradores también se han incluido las aportaciones y los comentarios de las personas que han participado en los seminarios GRADE realizados. En la fase final, la traducción ha sido revisada por un grupo de profesionales externos, así como por el grupo editorial de la revista ATENCIÓN PRIMARIA. Fruto

del esfuerzo y del interés de todas las personas que directa o indirectamente han colaborado, disponemos de una traducción de calidad.

Las personas e instituciones que formamos parte de REDEGUIAS somos conscientes de que, antes de proponer un cambio en los sistemas de clasificación que actualmente se están utilizando, es necesario conocer cómo se comporta GRADE en condiciones reales en nuestro medio. Con este objetivo, se ha diseñado un estudio de concordancia aplicando la clasificación de GRADE a 12 tablas de evidencia correspondientes a preguntas de tratamiento, diagnóstico y pronóstico sobre hipertensión arterial, asma y artrosis. El estudio se ha llevado a cabo con 20 profesionales con experiencia en evaluación de evidencia científica y elaboración de GPC con la clasificación de SIGN, los que participaron en el primer seminario de formación en GRADE realizado en abril de 2005 en Vitoria. Asimismo, y aprovechando las experiencias de este grupo de profesionales, en una segunda fase se ha llevado a cabo un estudio cualitativo. Ambos estudios están actualmente en proceso de análisis y esperamos poder difundir sus resultados en un futuro próximo.

Confiamos que toda esta secuencia de realidades aporte elementos para ir tejiendo un núcleo de masa crítica que permita discutir y consensuar cómo clasificar la calidad de la evidencia y la fuerza de las recomendaciones en nuestro medio. La apuesta, como se entrevé a partir de los debates ya iniciados en algunas organizaciones, no es fácil y se complica más cuando abogamos por la participación de atención primaria y de otras instituciones que elaboran GPC y documentos que incluyen recomendaciones.

Con independencia del debate que se establezca entre las diversas instituciones, creemos que en atención primaria es necesario, y en cierta medida urgente, establecer un consenso. Un sistema de clasificación único nos es útil para elaborar e interpretar de manera homogénea los diversos documentos y recursos que se generan por y para los profesionales de atención primaria. El papel de la semFYC en el proceso de consenso podría ser importante, ya que cuenta con profesionales con experiencia en la utilización de los diversos sistemas de clasificación. La creación de un grupo de trabajo podría representar al PAPPS, a los grupos de trabajo de la semFYC que elaboran documentos con recomendaciones y GPC, y a los editores de la Guía Terapéutica y de las diversas revistas de atención primaria. Por otro lado, utilizar un sistema común en el ámbito de atención primaria también facilitaría la comunicación en otros foros, como puede ser la I Conferencia sobre Prevención y Promoción de la Salud en la práctica clínica en España, en la que van a participar profesionales de las sociedades científicas con mayor relevancia en este campo y las administraciones públicas.

En resumen, la clasificación GRADE es un proyecto riguroso y novedoso que tiene como objetivo superar las limitaciones de las anteriores clasificaciones y unificar criterios

para que los distintos protagonistas en la elaboración de documentos con recomendaciones y GPC utilicen lenguajes comunes. El hecho de que presente todavía lagunas, como, por ejemplo, la ausencia de criterios para los aspectos diagnósticos o económicos, impone todavía cierta cautela a la hora de adoptar plenamente esta clasificación. Esperamos que en un futuro próximo el sistema de clasificación GRADE se consolide como sistema de referencia internacional.

M. Marzo-Castillejo^{a-c}, P. Alonso-Coello^{b,d}
y R. Rotaecbe del Campo^{b,c,e}

^aDivisión de Atención Primaria. ICS. ^bGrupo de trabajo GRADE y miembros de la REDEGUIAS. ^cComité científico y grupo MBE de la semFYC. ^dColaborador Centro Cochrane Iberoamericano. ^eCentro de Salud de Alza. Guipúzcoa. España.

Bibliografía

1. 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.
2. Canadian Task Force on Preventive Health Care. History and methods [citado 24 Ago 2005]. Disponible en: www.ctfphc.org
3. US Preventive Services Task Force Ratings: Strength of Recommendations and Quality of Evidence. Guide to Clinical Preventive Services, Third Edition: Periodic Updates, 2000-2003. Agency for Healthcare Research and Quality, Rockville, MD [citado 24 Ago 2005]. Disponible en: <http://www.ahrq.gov/clinic/3rduspstf/ratings.htm>
4. Systems to Rate the Strength of Scientific Evidence. Summary, Evidence Report/Technology Assessment: number 47. AHRQ Publication No. 02-E015, March 2002. Agency for Healthcare Research and Quality, Rockville, MD [citado 24 Ago 2005]. Disponible en: <http://www.ahrq.gov/clinic/epcsums/strength-sum.htm>
5. Agency for Healthcare Research and Quality (AHRQ). EPC Evidence Reports [citado 24 Mar 2005]. Disponible en: <http://www.ahrq.gov/clinic/epcindex.htm#methodology>
6. Centre for Evidence-Based Medicine de Oxford. Levels of Evidence and Grades of Recommendation [citado 24 Mar 2005]. Disponible en: http://www.cebm.net/levels_of_evidence.asp
7. Scottish Intercollegiate Guidelines Network. A guideline developer's handbook. Edinburg: SIGN; 2001; Publication nº 50 [citado 24 Mar 2005]. Disponible en: <http://www.show.scot.nhs.uk/sign/guidelines/fulltext/50/index.html>
8. National Institute for Clinical Excellence (NICE): Guideline Development Methods. Information for National Collaborating Centers and Guideline Developers. London [citado 24 Mar 2005]. Disponible en: <http://www.nice.org.uk/>
9. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman JL, Ewigman B, et al. Simplifying the language of evidence to improve patient care: Strength Of Recommendation Taxonomy (SORT): a patient-centered approach to grading evidence in medical literature. *J Fam Pract*. 2004;53:111-20.
10. Upshur RE. Are all evidence-based practices alike? Problems in the ranking of evidence. *CMAJ*. 2003;169:672-3.
11. Schünemann HJ, Best D, Vist G, Oxman AD, for the GRADE working group. Letters, numbers, symbols, and words: how best to communicate grades of evidence and recommendations? *CMAJ*. 2003;169:677-80.

12. 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.
13. The GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations II: a pilot study of a new system for grading the quality of evidence and the strength of recommendations. *BMC Health Serv Res.* 2005;5:25.
14. Marzo M, Basora J, Rotaecche R, Mascort JJ. La trayectoria científica de semFYC. ¿Hacia dónde queremos avanzar? *Aten Primaria.* 2005;35:447-8.

Nota del Editor

Publicación redundante

Con el fin de introducir elementos de información y reflexión para los autores que publican en la revista ATENCIÓN PRIMARIA, relativos a posibles aspectos conflictivos que se pueden plantear con manuscritos sometidos al proceso editorial, principalmente en la sección de «Originales», el Comité de Redacción ha considerado pertinente difundir los elementos esenciales sobre publicaciones redundantes

Nota del Editor

Las Normas de Vancouver consideran publicación redundante a un artículo que se solapa sustancialmente con otro. El Comité de Ética de las Publicaciones (COPE) la define cuando, sin referencias cruzadas, dos o más artículos comparten las mismas hipótesis, datos, puntos de discusión o conclusiones. Los lectores deben confiar en que los artículos que aparecen en la sección de "Originales" lo son efectivamente y que, cuando no lo sean, exista una comunicación explícita citando el otro trabajo publicado. Por esa razón, los Comités Editoriales no desean recibir manuscritos sobre trabajos que hayan sido publicados en su mayor parte. Las Normas de Vancouver establecen que el autor, al enviar un manuscrito a una revista, debe informar al Consejo de Redacción de todo aquello que pudiera plantear dudas sobre una posible publicación duplicada o redundante del mismo trabajo, incluso si lo que se ha publicado son resultados preliminares del mismo estudio en la misma o en otra revista. Estos artículos deberían citarse en la bibliografía del trabajo que se somete a revisión. Sería recomendable que se adjuntaran copias de dicho material para ayudar al editor a tomar una decisión. En caso de que se detecte una publicación duplicada o redundante, las Normas establecen que el editor debe rechazar el trabajo si todavía se encuentra en proceso de revisión o publicar una nota informativa si el estudio ya ha aparecido en la revista.

Se considera aceptable la publicación secundaria de un trabajo en las siguientes circunstancias:

1. Cuando los autores tengan la aprobación del Consejo de Redacción de ambas revistas.
2. Cuando se respete la prioridad de la publicación primaria.
3. Cuando la publicación secundaria se dirija a un grupo de lectores diferente y por tanto una versión abreviada pueda ser suficiente. Habitualmente, se aplica a artículos que aparecen en diferentes países o idiomas.
4. Cuando refleja fielmente los datos e interpretaciones de la publicación primaria.
5. Cuando se incluya una nota a pie de página del artículo informando que el artículo ha sido publicado total o parcialmente y referenciándolo.
6. El título de la publicación secundaria debería reflejar esta condición.

Noticia sobre publicación redundante

Se han detectado dos publicaciones simultáneas que presentan esencialmente los mismos datos del mismo estudio.

- Pérula-De Torres LA, Fernández García JA, Arias-Vega R, Muriel-Palomino M, Márquez Rebollo E, Ruiz Moral R. Validez del cuestionario AUDIT para la detección de problemas relacionados con el consumo de alcohol en mujeres. *Med Clin (Barc).* 2005;125:727-30.
- Pérula-De Torres LA, Fernández García JA, Arias-Vega R, Muriel-Palomino M, Márquez Rebollo E, Ruiz Moral R. Validación del cuestionario AUDIT para la identificación del consumo de riesgo y de los trastornos del uso de alcohol en mujeres. *Aten Primaria.* 2005;36:499-506.