



# CIRUGÍA ESPAÑOLA

[www.elsevier.es/cirugia](http://www.elsevier.es/cirugia)



## Artículo especial

# Sistema GRADE: clasificación de la calidad de la evidencia y graduación de la fuerza de la recomendación



José Luis Aguayo-Albasini, Benito Flores-Pastor\* y Víctor Soria-Aledo

Servicio de Cirugía General, Hospital General Universitario JM Morales Meseguer, Campus de Excelencia Internacional Mare Nostrum, Universidad de Murcia, Murcia, España

### INFORMACIÓN DEL ARTÍCULO

#### Historia del artículo:

Recibido el 16 de abril de 2013

Aceptado el 19 de agosto de 2013

On-line el 20 de diciembre de 2013

#### Palabras clave:

Calidad de la evidencia

Fuerza de la recomendación

Sistema GRADE

Guías de práctica clínica

### RESUMEN

La adquisición y jerarquización de la evidencia, así como la posterior formulación de recomendaciones, constituyen la base del desarrollo de las guías de práctica clínica. Sistemas de graduación de la calidad de la evidencia y de la fuerza de las recomendaciones han existido muchos y actualmente se va imponiendo el modelo *Grading of Recommendations, Assessment, Development and Evaluation* (GRADE). En el sistema GRADE la calidad de la evidencia se clasifica, inicialmente, en alta o baja, según provenga de estudios experimentales u observacionales; posteriormente, según una serie de consideraciones, la evidencia queda en alta, moderada, baja y muy baja. La fuerza de las recomendaciones se apoya no solo en la calidad de la evidencia, sino en una serie de factores como son el balance entre riesgos y beneficios, los valores y preferencias de pacientes y profesionales, y el consumo de recursos o costes.

© 2013 AEC. Publicado por Elsevier España, S.L. Todos los derechos reservados.

### GRADE system: Classification of quality of evidence and strength of recommendation

### ABSTRACT

The acquisition and classification of scientific evidence, and subsequent formulation of recommendations constitute the basis for the development of clinical practice guidelines. There are several systems for the classification of evidence and strength of recommendations; the most commonly used nowadays is the Grading of Recommendations, Assessment, Development and Evaluation system (GRADE). The GRADE system initially classifies the evidence into high or low, coming from experimental or observational studies; subsequently and following a series of considerations, the evidence is classified into high, moderate, low or very low. The strength of recommendations is based not only on the quality of the evidence, but also on a series of factors such as the risk/benefit balance, values and preferences of the patients and professionals, and the use of resources or costs.

© 2013 AEC. Published by Elsevier España, S.L. All rights reserved.

#### Keywords:

Evidence quality

Strength of recommendation

GRADE system

Clinical Practice guidelines

\* Autor para correspondencia.

Correo electrónico: [florespastorbenito2@gmail.com](mailto:florespastorbenito2@gmail.com), [benitom.flores@carm.es](mailto:benitom.flores@carm.es) (B. Flores-Pastor).

0009-739X/\$ - see front matter © 2013 AEC. Publicado por Elsevier España, S.L. Todos los derechos reservados.

<http://dx.doi.org/10.1016/j.ciresp.2013.08.002>

## Introducción

La medicina basada en la evidencia (MBE) implica que el profesional integre su conocimiento y juicio médicos con la mejor evidencia científica existente. Determinar cuáles son las mejores evidencias exige unas habilidades en la identificación, análisis crítico y jerarquización de las evidencias disponibles en la literatura. Este primer paso es imprescindible, puesto que cualquier recomendación o grado de recomendación que se proponga respecto a una intervención preventiva o terapéutica o respecto a un procedimiento diagnóstico ha de estar directamente relacionada con la calidad (y otros factores) de la evidencia existente.

El mayor interés de la MBE está en su utilización por grupos de expertos que desarrollan guías de práctica clínica (GPC) para el estudio, diagnóstico, tratamiento o prevención de una enfermedad o problema de salud. Se describen hasta 8 fases en el desarrollo de una guía (tabla 1), pero en este artículo únicamente nos ocuparemos de las fases 3 a 6 (formulación de preguntas, adquisición de evidencia, atribución de calidad y formulación de recomendaciones). La obtención de GPC útiles no es tarea sencilla debido a la heterogeneidad de los individuos que constituyen los grupos o comités de expertos que las realizan, a sus diferentes puntos de vista y métodos, y a la misma variabilidad en la información científica disponible sobre un tema concreto<sup>1-3</sup>. Hasta hace unos años estos grupos de expertos seguían una metodología informal para alcanzar los acuerdos, pero recientemente se han mejorado los procedimientos de jerarquización de la evidencia y de establecimiento de las recomendaciones pertinentes. Aquí entra en consideración el sistema *Grading of Recommendations, Assessment, Development and Evaluation* (GRADE).

Existe un amplio número de sistemas refinados de categorización de la evidencia científica, entre los que podemos señalar el modelo inglés *Oxford Centre for Evidence-Based Medicine* (OCEBM), el escocés *Scottish Intercollegiate Guidelines Network* (SIGN) o el estadounidense del *American College of Chest Physicians* (ACCP) que utilizó la misma ACCP en sus guías sobre trombosis venosa hasta su séptima revisión<sup>4-6</sup>. Todos ellos atribuían diferentes niveles de calidad a los estudios sobre un determinado problema, que luego permitían asignar distintos grados de recomendación. Sin embargo, pronto se observaron algunos inconvenientes, como el hecho de que estos sistemas se desarrollaron principalmente por el consenso de opinión de expertos y no fueron validados<sup>7</sup>. Así,

en ocasiones, diferentes sistemas no categorizaban los mismos estudios en los análogos niveles de evidencia. Incluso con un mismo modelo se producían a veces falta de acuerdos. Además, algunos sistemas eran mejores en estimar el nivel de evidencia que en establecer el grado de recomendación, y viceversa. Todo lo anterior hacía que a veces no existiera una buena fiabilidad de las GPC.

En 2004 se comunica la propuesta del grupo de trabajo GRADE, formado por un conjunto internacional y multidisciplinar de metodólogos, expertos en GPC y médicos clínicos, que pretenden obviar los problemas descritos<sup>8,9</sup>. La bondad del sistema estriba en que es un método riguroso y transparente tanto en la clasificación del nivel de evidencia como en la asignación del grado o fortaleza de la recomendación. En este artículo desarrollaremos estos puntos tal como lo hace el sistema GRADE, pero antes comentaremos los pasos que se deben seguir para la elaboración de las preguntas clínicas.

## Formulación de las preguntas clínicas en formato PICO y búsqueda de las respuestas

Una vez establecido el alcance de una GPC es necesario definir una serie de preguntas clínicas que se agruparán en apartados de organización, prevención, diagnóstico, tratamiento, pronóstico, etc. 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 prefiere el método PICO, acrónimo de *Pacientes-Intervención-Comparación-resultados/Outcomes*. Así:

- Paciente: o población, estadios de enfermedad, grupos de edad, comorbilidad, etc.
- Intervención: tratamiento, prueba diagnóstica, agente etiológico, etc.
- Comparación: posible alternativa a la intervención en estudio como tratamiento habitual o placebo, patrón oro de referencia de una prueba diagnóstica, ausencia de agente etiológico, etc.
- Resultados (*outcomes*): variables de resultados relevantes en caso de estudios sobre eficacia, pronóstico o etiología, y estimadores de validez en el caso de pruebas diagnósticas (sensibilidad, especificidad, coeficientes de probabilidad, etc.).

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 literatura. Durante la confección de las preguntas clínicas hay que definir todas las posibles variables de resultado o desenlace, cuestión aún más relevante cuando se utiliza en la elaboración de las recomendaciones el sistema GRADE, donde se califican las variables según su importancia para clínicos y pacientes y se las pondera en una escala que va del 1 al 9. Únicamente las variables con puntuación del 7 al 9 son consideradas clave para afectar una decisión en el sistema GRADE y es a estas variables clave a las que hay que especificar

**Tabla 1 – Elaboración de una guía de práctica clínica: fases**

1. Delimitación del alcance y de los objetivos
2. Creación del grupo elaborador de la GPC
3. Formulación de las preguntas clínicas (PICO)
4. Búsqueda de la evidencia
5. Evaluación y síntesis de la literatura
6. Formulación de recomendaciones
7. Revisión externa
8. Edición

GPC: guía de práctica clínica; PICO.

las preguntas clínicas. Las respuestas a estas preguntas sobre resultados clave serán las que principalmente se utilizarán para graduar las recomendaciones. Las variables con puntuación del 4 al 6 se califican de importantes, pero no clave para la toma de decisiones. Las puntuadas del 1 al 3 se consideran no importantes y no se incluyen en la evaluación, tampoco influirán en las recomendaciones. La selección estricta y acertada de las variables de resultados o desenlaces clave hace que los estudios se seleccionen de igual forma y, así, es posible que los hallazgos que luego se van a considerar para inferir las recomendaciones y, por tanto, la fortaleza de las mismas pueda variar de una GPC a otra<sup>10</sup>.

Encontrar las respuestas que buscamos a las preguntas PICO requiere, en ocasiones, la colaboración de un documentalista. Podemos necesitar consultar otras GPC anteriores, o revisiones sistemáticas actualizadas, o estudios originales. Localizada la evidencia científica, se debe categorizar según su calidad metodológica (validez interna), la importancia de los resultados y su aplicabilidad.

Finalmente, se gradúa la fuerza de las recomendaciones con arreglo a un sistema determinado. Hasta ahora se ha venido utilizando, entre otros, el sistema escocés SIGN para las preguntas relacionadas con tratamiento o pronóstico, y el inglés Oxford CMBE para las preguntas sobre diagnóstico. En la actualidad se va imponiendo el sistema GRADE que será el que comentemos a continuación. El GRADE Working Group propuso un diferente abordaje basado en sistemas previos y que presume de mejor estructura y mayor transparencia e información<sup>8-14</sup>. Tiene la ventajas de 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, unas tablas singulares e imprescindibles que más adelante comentaremos.

## Niveles de evidencia

GRADE define la calidad de la evidencia como el grado de confianza que tenemos en que la estimación de un efecto sea

la adecuada para efectuar una recomendación. Para cada uno de los desenlaces o resultados clave se efectúa una evaluación; así, una misma comparación de una intervención terapéutica o preventiva puede recibir diferentes asignaciones de la calidad de la evidencia. En la clasificación del nivel de calidad de la evidencia, el sistema GRADE establece 4 categorías: alta, moderada, baja y muy baja. Lo que representa o significa cada una de las 4 categorías, en su concepción inicial y en la actual, se expone en la [tabla 2](#).

En una primera etapa el sistema GRADE considera de calidad alta los estudios experimentales (ensayos clínicos aleatorizados) y de calidad baja los estudios observacionales (casos y controles, cohortes). En un segundo paso, para refinar el nivel de calidad, el sistema establece una serie de ítems que hay que considerar y que pueden hacer bajar o subir el escalón o nivel de la calidad inicialmente asignada.

### a. Ítems que bajan la calidad:

- 1) *Limitaciones en el diseño y ejecución del estudio (riesgo de sesgo)*: Así, aleatorización insuficiente o incorrecta, falta de enmascaramiento, pérdidas importantes de seguimiento, análisis sin intención de tratar y ensayos acabados antes de tiempo.
- 2) *Inconsistencia de los resultados (inconsistency)*: cuando los resultados muestran amplia variabilidad o heterogeneidad no explicada. Particularmente si unos estudios muestran beneficios sustanciales y otros ningún efecto o incluso perjuicio.
- 3) *Incertidumbre acerca de que la evidencia sea directa (indirectness)*: Siguiendo el método PICO, puede ocurrir con los pacientes estudiados (diferencias en edad, sexo o estadio clínico); o con la intervención, en caso de que ésta sea parecida pero no idéntica; o con la comparación que se realiza; o con los resultados, si se comparan unos a corto plazo y otros a largo plazo, etc.
- 4) *Imprecisión (imprecision)*: tiene lugar si los intervalos de confianza (IC) son amplios, las muestras pequeñas, o los eventos son pocos.
- 5) *Sesgo de publicación o notificación (publication bias)*: cuando existe alta probabilidad de estudios no reportados, principalmente por ausencia de efectos, o no se han incluido todas las variables relevantes de resultado.

### b. Ítems que suben la calidad:

- 1) *Fuerte asociación*: hallazgos de efectos relativos  $RR > 2$ , o  $< 0,5$  en estudios observacionales sin factores de confusión.

**Tabla 2 – Sistema GRADE: Significado de los 4 niveles de evidencia**

Niveles de calidad	Definición actual	Concepto anterior
Alto	Alta confianza en la coincidencia entre el efecto real y el estimado	La confianza en la estimación del efecto no variará en posteriores estudios
Moderado	Moderada confianza en la estimación del efecto. Hay posibilidad de que el efecto real esté alejado del efecto estimado	Posteriores estudios pueden tener un importante impacto en nuestra confianza en la estimación del efecto
Bajo	Confianza limitada en la estimación del efecto. El efecto real puede estar lejos del estimado	Es muy probable que posteriores estudios cambien nuestra confianza en la estimación del efecto
Muy bajo	Poca confianza en el efecto estimado. El efecto verdadero muy probablemente sea diferente del estimado	Cualquier estimación es muy incierta

**Tabla 3 – Clasificación del nivel de evidencia según el sistema GRADE**

Tipo de estudio	Nivel de calidad a priori	Desciende si	Sube si	Nivel de calidad a posteriori	
Estudios aleatorizados	Alta	Riesgo de sesgo	Efecto	Alta	
		-1 importante	+1 grande		
Estudios observacionales	Baja	-2 muy importante	+2 muy grande	Moderada	
		Inconsistencia	Dosis-respuesta		
		-1 importante	+1 gradiente evidente	Baja	
		-2 muy importante	Todos los factores de confusión:		
		No evidencia directa	+1 reducirían el efecto observado		
		-1 importante	+1 sugerirían un efecto espurio si no hay efecto observado		Muy baja
		-2 muy importante			
		Imprecisión			
Sesgo de publicación					
-1 probable					
-2 muy probable					

- 2) Muy fuerte asociación: hallazgos de efectos relativos  $RR > 5$ ,  $o < 0,2$  basados en estudios sin problemas de sesgo o precisión.
- 3) Existencia de gradiente dosis-respuesta (*dose-response gradient*).
- 4) Evidencia de que todos los posibles factores de confusión o sesgos podrían haber reducido el efecto observado.

Las situaciones que pueden determinar un incremento de la confianza en los resultados de estudios observacionales son infrecuentes. En tales casos se debería considerar este aumento únicamente si no hay limitaciones de diseño o ejecución (que podría disminuir la calidad) y, además, existe un efecto muy importante e inmediato o un cambio radical en el pronóstico tras determinada intervención.

Todos estos ítems determinan, con arreglo a las puntuaciones que se expresan en la [tabla 3](#), un ascenso o un descenso en el nivel de la calidad de la evidencia. Terminado el análisis, los expertos del GRADE resumen toda la evidencia, respecto a las preguntas concretas y las variables de resultado o desenlace previamente escogidas, en una tablas-resumen que denominan perfil de evidencia (*evidence profiles* [GRADE EP]) y sumario de hallazgos (*summary of findings* [GRADE SoF]). Las tablas EP y SoF tienen diferentes propósitos y van dirigidas a colectivos distintos<sup>15,16</sup>.

Las [tablas EP](#) son más amplias, presentan los hallazgos relevantes de cada uno de los resultados o desenlaces clave, ofreciéndolos en líneas donde, en la diferentes columnas, se expresan el número de estudios y el número de pacientes, el diseño (aleatorizado u observacional), las comparaciones efectuadas, los estimadores del efecto observado en términos de efecto relativo RR (con su IC 95%) y efecto absoluto, y, además, incluyen un juicio explícito de los factores que ponderan la calidad de los estudios (limitaciones de diseño, inconsistencia, incertidumbre, sesgo de notificación, etc.); finalmente categorizan la calidad de la evidencia para cada uno de los resultados en forma de cruces positivas (+) de 4 a 1, o sea, calidad alta, moderada, baja o muy baja, respectivamente, con el significado que se expresa en la [tabla 3](#). En los casos de tablas que abordan los niveles de evidencia de cuestiones en relación con pruebas diagnósticas, el formato puede ser

diferente. Los perfiles de evidencia van dirigidos a un colectivo reducido como revisores y creadores de GPC; también a cualquiera que se cuestione o quiera comprobar la bondad de una valoración.

Las [tablas SoF](#) son más escuetas y solo ofrecen los hallazgos relevantes de cada resultado, es decir, como antes señalamos, el número de estudios y el número de pacientes, las comparaciones efectuadas, los estimadores del efecto observado en términos de efecto relativo RR (con su IC 95%) y efecto absoluto; también expresan la calidad asignada. Los sumarios de hallazgos van dirigidos a un colectivo más amplio, principalmente usuarios de GPC y lectores de revisiones sistemáticas. Existe un programa informático o *software* (GRADEpro) que facilita tanto la creación de tablas EP como tablas SoF<sup>14</sup>. Para ampliar conocimientos sobre tablas EP y SoF puede consultarse a Guyatt et al.<sup>16</sup>.

A este respecto, en la guía de la ACCP para *Terapia antitrombótica y prevención de trombosis* (9.ª ed) destacamos una tabla que resume la evidencia para la instauración de profilaxis tromboembólica farmacológica en pacientes quirúrgicos ([tabla 4](#)). Observamos una disminución de la calidad de la evidencia en 2 de los desenlaces de interés (embolismo pulmonar fatal y enfermedad tromboembólica venosa sintomática no fatal). En un caso, por la imprecisión asociada a una posible no obtención de efecto, y, en el otro, por limitaciones en el diseño de uno de los estudios. Si comparamos con la 8.ª edición de dicha guía, observamos que el nivel de evidencia y el grado de recomendación para profilaxis tromboembólica farmacológica en pacientes de riesgo moderado sometidos a cirugía mayor ha disminuido en la nueva guía<sup>17,18</sup>.

## Grado de recomendación

El sistema GRADE establece las recomendaciones basándose en una serie de consideraciones<sup>8,9,11,12</sup>. Así: 1) balance entre riesgos y beneficios: Está basado en el conocimiento de que la mayoría de los clínicos ofrecerán a los pacientes medidas terapéuticas o preventivas siempre que las ventajas de la intervención superen los riesgos y daños colaterales. La

**Tabla 4 – Efecto de profilaxis con heparina de bajo peso molecular respecto a no profilaxis de enfermedad tromboembólica en pacientes quirúrgicos**

Desenlace de interés	N.º de participantes (estudios)	Calidad de la evidencia (GRADE)	Efecto relativo (IC 95%)	Riesgo comparativo (IC 95%)	
				Grupo no profilaxis	Grupo HBPM
EP letal (seguimiento: 7-270 d)	5.142 (5 estudios)	Moderada <sup>a</sup>	RR 0,54 (0,27-1,1)	Población de bajo riesgo	
				3‰	2‰ (1-3)
				Población de riesgo moderado	
Hemorragia letal (seguimiento: 21-270 d)	5.078 (4 estudios)	Moderada		6‰	3‰ (2-7)
				Población de alto riesgo	
				12‰	6‰ (3-13)
ETE no letal sintomática (seguimiento: 21-270 d)	4.890 (3 estudios)	Moderada <sup>b</sup>	RR 0,31 (0,12-0,81)	Población de bajo riesgo	
				1‰	0‰ (0-0)
				Población de alto riesgo	
Hemorragia mayor no letal (seguimiento: 7-270 d)	5.457 (7 estudios)	Alta	RR 2,03 (1,37-3,01)	2‰	0‰ (0-0)
				Población de bajo riesgo	
				15‰	5‰ (2-12)
				Población de riesgo moderado	
				30‰	9‰ (4-24)
				Población de alto riesgo	
				60‰	19‰ (7-49)
				Población de bajo riesgo	
				12‰	24‰ (16-36)
				Población de alto riesgo	
				22‰	45‰ (30-66)

EP: embolismo pulmonar; ETE: enfermedad tromboembólica; HBPM: heparina de bajo peso molecular.  
Fuente: adaptado de Mismetti et al.<sup>17</sup>  
<sup>a</sup> El IC 95% incluye la posibilidad de no efecto (> 1).  
<sup>b</sup> Un estudio presentó limitaciones en su diseño.

certidumbre o incertidumbre con la que el clínico contempla el balance riesgo/beneficio determinará en gran parte la fuerza de la recomendación. 2) Calidad de la evidencia: El segundo factor es la calidad metodológica de los estudios para cada una de las variables de resultados, factor ponderado por los detalles que expusimos y que podían elevar o disminuir el nivel de la evidencia. En general, el grado de la recomendación sigue al nivel de la evidencia, pero no siempre. 3) Han de considerarse también los valores y preferencias de los pacientes. Para ello hay que realizar un juicio de valor y conocer previamente los valores y preferencias de la población de nuestro entorno y de posibles diferencias individuales. 4) Estimación del consumo de recursos o costes.

Todavía faltan estudios apropiados que analicen los valores y preferencias de los pacientes ante situaciones definidas; en cualquier caso, los valores y preferencias fortalecen el grado de la recomendación cuando la concordancia es alta y lo debilitan cuando existe una elevada variabilidad. El análisis de costes suele requerir el concurso de expertos en economía de la salud. En general, se considera que una intervención se puede calificar de muy coste-efectiva si cuesta < una vez la media de la renta *per cápita* de un país o región por año de vida ajustado por calidad (AVAC) ganado. Incluso hasta 3 veces la media de la renta *per cápita* por AVAC

ganado puede ser tolerable. La OMS ha desarrollado tablas de umbrales al respecto<sup>19,20</sup>.

Finalmente, de forma simple, las recomendaciones se gradúan de forma binaria en: fuertes (grado 1) o débiles (grado 2), bien a favor, bien en contra. Una recomendación **fuerte** supone que la inmensa mayoría de los pacientes estarían de acuerdo (o en contra) respecto a la acción recomendada. Los clínicos deberían implementar la acción en la mayor parte de los pacientes y las autoridades sanitarias habrían de adoptar la recomendación como política sanitaria en la mayoría de las situaciones. Una recomendación **débil** supone que la mayoría de los pacientes aceptarían (o rechazarían) la acción recomendada, pero un número importante de ellos no lo estarían. Los clínicos deberían reconocer que hay diferentes opciones apropiadas para distintos pacientes y, en este caso, habría que ayudar a cada paciente a alcanzar la decisión más consistente con sus valores y preferencias; las autoridades sanitarias deberían debatir la puesta en marcha de la recomendación con los grupos de interés.

En el ejemplo que hemos sugerido anteriormente, observamos que la profilaxis farmacológica del tromboembolismo pulmonar reduce significativamente el riesgo de enfermedad tromboembólica venosa (ETE) no fatal, no así el embolismo pulmonar (EP) mortal, con un aumento del riesgo de



hemorragia mayor no letal (balance entre riesgos y beneficios). Por otro lado, respecto a la calidad de la evidencia, vemos como esta disminuye por la imprecisión observada en los resultados de EP letal así como por limitaciones en el diseño de algún estudio. En definitiva, para pacientes con riesgo tromboembólico moderado se establece una recomendación débil a favor del uso de profilaxis farmacológica<sup>18</sup>.

## Limitaciones, futuro, uso en España del sistema GRADE

Se deben conocer algunas limitaciones. Así, en primer lugar, el método se ha desarrollado 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. Segundo, solo considera los pasos 3 a 6 (tabla 1) de elaboración de una GPC. Y tercero, 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.

Para quienes deseen ampliar o profundizar en el método GRADE, como autores de revisiones sistemáticas o de estudios de evaluación de tecnologías sanitarias, panelistas de GPC y metodólogos, existe una amplia y exhaustiva serie de artículos secuenciales al respecto publicada en el *Journal of Clinical Epidemiology* entre 2011 y 2013, todavía no concluida<sup>15,21-32</sup>.

En España, algunas revistas científicas de prestigio y factor impacto notable se han hecho eco del fenómeno GRADE. Así, la *Revista de Atención Primaria*<sup>33</sup>, *Medicina Clínica*<sup>34</sup>, *Archivos de Bronconeumología*<sup>35</sup> y *Revista Española de Cardiología*<sup>13</sup>. Por otra parte, se ha comunicado su uso tanto en la evaluación de tecnologías sanitarias<sup>36</sup> como en el desarrollo de GPC<sup>37-39</sup>.

## Conflicto de intereses

No existe.

## BIBLIOGRAFÍA

- Romero Simó M, Soria Aledo V, Ruiz López P, Rodríguez Cuéllar E, Aguayo Albasini JL. Guidelines and clinical pathways. Is there really a difference? *Cir Esp*. 2010;88:81-4.
- Jaeschke R, Guyatt GH, Dellinger P, Schünemann H, Levy MM, Kunz R, Norris S, Bion J. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ*. 2008;337:a774.
- Tricoci P, Allen JM, Kramer JM, Califf RM, Smith Jr SC. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA*. 2009;301:831-41.
- Phillips B, Ball C, Sackett DL. Levels of evidence and grades of recommendations. Oxford: Centre for Evidence-Based Medicine. 1998.
- Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ*. 2001;323:334-6.
- Guyatt GH, Schünemann HJ, Cook D. Applying the grades of recommendation for antithrombotic and thrombolytic therapy: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;126:1795-875.
- Upshur RE. Are all evidence-based practices alike. *Problems in the ranking of evidence*. *CMAJ*. 2003;169:672-3.
- Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. 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;4:38 [consultado 1 Mar 2013]. Disponible en: <http://www.biomedcentral.com/1472-6963/4/38>
- Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, et al. Systems for grading the quality of evidence and the strength of recommendations II. Pilot study of a new system. *BMC Health Serv Res*. 2005;5:25 [consultado 1 Mar 2013]. Disponible en: <http://www.biomedcentral.com/1472-6963/5/25>
- Arcelus Martínez JI. Las nuevas guías de prevención y terapia antitrombótica del American College of Chest Physicians (ACCP). *Rev Esp Anestesiología Reanim*. 2012. Disponible en: <http://dx.doi.org/10.1016/j.redar.2012.04.024>.
- Grupo de trabajo sobre GPC. Elaboración de guías de práctica clínica en el Sistema Nacional de Salud. Manual metodológico. Guías de Práctica Clínica en el SNS: ICS N° 2006/01. Madrid: Plan Nacional para el SNS del MSC. Instituto Aragonés de Ciencias de la Salud-ICS; 2007.
- Grondin SC, Schieman C. Evidence-based medicine levels of evidence and evaluation systems. En: Ferguson MK, editor. *Difficult decisions in thoracic surgery* London: Springer-Verlag London Limited. 2011.
- Alonso-Coello P, Solà I, Ferreira-González I. La formulación de recomendaciones con GRADE: cuestión de confianza. *Rev Esp Cardiol*. 2013;66:163-7.
- Brozek J, Oxman A, Schünemann HJ. GRADE pro (computer program) version 3.2 for Windows [consultado 22 Mar 2013]. Disponible en: <http://mcmaster.flintbox.com/technology.asp?Page=3993>
- Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: A new series of articles in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol*. 2011;64:380-2.
- Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64:383-94.
- Mismetti P, Laporte S, Darmon JY, Buchmüller A, Decousus H. Meta-analysis of low molecular weight heparin in the prevention of venous thromboembolism in general surgery. *Br J Surg*. 2001;88:913-30.
- Douketis JD, Spyropoulos AC, Spencer FA, Mayr M, Jaffer AK, Eckman MH, et al. Perioperative management of antithrombotic therapy: Antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians. Evidence based clinical practice guidelines. *Chest*. 2012;141(2 suppl):e326S-50S.
- World Health Organization. Threshold values for intervention cost-effectiveness by region. Geneva, Switzerland: World Health Organization. 2010.
- Guyatt HG, Norris SL, Schulman S, Hirsh J, Eckman MH, Akl EA, et al. Methodology for the development of antithrombotic therapy and prevention of thrombosis guidelines. *Chest*. 2012;141:53S-70S.
- 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.

22. Balslem H, Helfand M, Schünemann HJ, Oxman AD, Kunz J, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64:401-6.
23. Guyatt HG, 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.
24. Guyatt HG, 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. Brunotti M, Shelmit I, Pregno S, Vale L, Oxmann AD, Lord J, et al. GRADE guidelines: 10. Rating the quality of evidence for resource use. *J Clin Epidemiol*. 2013;66:140-50.
30. Guyatt GH, 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.
31. 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.
32. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. 13 GRADE guidelines: 13. Preparing summary of findings tables-continuous outcomes. *J Clin Epidemiol*. 2013;66:173-83.
33. Marzo Castillejo M, Montaña Barrientos A. The GRADE system in taking clinical decisions and elaboration of recommendations and clinical practice guidelines. *Aten Primaria*. 2007;39:457-60.
34. Alonso-Coello P, Rigau D, Solà I, Martínez García L. Formulating health care recommendations: The GRADE system. *Med Clin (Barc)*. 2012. doi: 10.1016/j.medcli.2012.10.012.
35. Alonso-Coello P, Rigau D, Sanabria AJ, Plaza V, Miravittles M, Martínez L. Quality and strength. The GRADE system for formulating recommendations in clinical practice guidelines. *Arch Bronconeumol*. 2013. doi:10.1016/arbres.2012.12.001.
36. Ibargoyen-Roteta N, Gutiérrez Ibarluzea I, Rico-Iturrioz R, López-Argumedo M, Reviriego-Rodrigo E, Cabriada Nuño JL, et al. The GRADE approach for assessing new technologies as applied to apheresis devices in ulcerative colitis. *Implement Sci*. 2010. doi:10.1186/1748-59 08-5-48.
37. Marzo Castillejo M, Rotaecche del Campo R, Basora Gallifa J. SemFYC also adopts the GRADE system. *Aten Primaria*. 2010;42:191-3.
38. Gisbert JP, Calvet Calvo X, Ferrándiz Santos J, Mascort Roca JJ, Alonso-Coello P, Marzo Castillejo M. Managing of the patient with dyspepsia. Clinical practice guideline. Update 2012. *Aten Primaria*. 2012;44:728-33.
39. Gomollón F, García-López S, Sicilia B, Gisbert JP, Hinojosa J. Therapeutic guidelines on ulcerative colitis: A GRADE methodology based effort of GETECCU. *Gastroenterol Hepatol*. 2013;36:104-14.