

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas

www.evidenciasenpediatria.es

Fundamentos de Medicina Basada en la Evidencia

Valoración de la calidad de la evidencia y fuerza de las recomendaciones (I). El sistema GRADE

Martín Muñoz P¹, González de Dios J²

¹UGC La Plata. Distrito A.P. Sevilla (España).

²Departamento de Pediatría. Hospital General Universitario de Alicante (España).

Correspondencia: Pedro Martín Muñoz, pedromartinm@telefonica.net

Palabras clave en inglés: evidence-based medicine; evidence-based practice; practice guidelines as topic.

Palabras clave en español: medicina basada en pruebas; práctica clínica basada en la evidencia; guías de práctica clínica como asunto.

Fecha de recepción: 21 de agosto de 2010 • Fecha de aceptación: 24 de agosto de 2010

Fecha de publicación en Internet: 1 de septiembre de 2010

Evid Pediatr. 2010;6:63.

CÓMO CITAR ESTE ARTÍCULO

Martín Muñoz P, González de Dios J. Valoración de la calidad de la evidencia y fuerza de las recomendaciones (I). El sistema GRADE.

Evid Pediatr. 2010;6:63.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín por medio del ETOC en <http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2010;3;63>

©2005-10 • ISSN: 1885-7388

Valoración de la calidad de la evidencia y fuerza de las recomendaciones (I). El sistema GRADE

Martín Muñoz P¹, González de Dios J²

¹UGC La Plata. Distrito A.P. Sevilla (España).

²Departamento de Pediatría. Hospital General Universitario de Alicante (España).

Correspondencia: Pedro Martín Muñoz, pedromartinm@telefonica.net

DE LA EVIDENCIA A LA RECOMENDACIÓN: UNA TAREA PENDIENTE

Una aspiración irrenunciable de la medicina actual, reclamada por todos sus protagonistas (enfermos, profesionales y administraciones), es que los actos médicos se sustenten en conocimientos científicos obtenidos de procesos de investigación clínica rigurosa. Decidir si una intervención clínica resulta adecuada para un paciente determinado equivale a establecer 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 resulta lo suficientemente favorable como para que merezca la pena aplicarla. Los conceptos de **calidad (nivel) de la evidencia y fuerza (grado) de las recomendaciones** constituyen un pilar fundamental de la práctica basada en la evidencia, en su intento por estandarizar y proporcionar a los clínicos reglas para analizar la literatura científica, determinar su validez y considerar su utilidad en la asistencia sanitaria.

Cada vez toma más cuerpo el tomar decisiones médicas que estén fundamentadas en el mejor nivel de evidencia (indica hasta qué punto nuestra confianza en la estimación de un efecto es adecuada para apoyar una recomendación) y la mayor fuerza de recomendación (indica hasta qué punto podemos confiar si poner en práctica la recomendación conllevará más beneficios que riesgos).

La **calidad (nivel) de evidencia** se ha relacionado, generalmente, con el diseño del estudio (estudios descriptivos o analíticos, observacionales o experimentales) y la calidad de los mismos. La meta de la investigación es la agudeza en la medición, lo que implica precisión (limitar el error aleatorio) y validez (limitar el error sistemático). En este sentido, por las características propias de cada diseño, el "nivel" de evidencia será mayor en los estudios analíticos que en los descriptivos, y superior en los estudios experimentales (ejemplo, ensayo clínico) que en los observacionales (ejemplo, estudios de cohortes y estudios de casos y controles). Sin embargo, no toda pregunta clínica se puede abordar con el mismo diseño científico: el ensayo clínico es el patrón oro para intervenciones terapéuticas, pero no será el diseño apropiado para preguntas sobre diagnóstico o pronóstico.

Se establecen unos criterios de calidad propios para cada tipo de diseño. Así, podemos considerar cinco criterios de calidad

en el ensayo clínico (definición clara de la población de estudio, intervención y resultado de interés; correcta aleatorización; adecuado enmascaramiento; seguimiento completo –menos del 20% de pérdidas–; análisis correcto –análisis por intención de tratar y control de covariables no equilibradas con la aleatorización–), que serán diferentes a los criterios de calidad barajados en el caso de estudios de valoración de pruebas diagnósticas (comparación con un patrón de referencia válido; muestra representativa; descripción completa de los métodos de realización de la prueba diagnóstica; control de sesgos –comparación ciega e independiente–; control de sesgos de incorporación, verificación diagnóstica y revisión; análisis correcto –datos que permitan calcular indicadores de validez–) o de cohortes (cohortes representativas de la población con y sin exposición, libres del efecto o enfermedad de interés; medición independiente, ciega y válida de exposición y efecto; seguimiento suficiente –superior al 80%–, completo y no diferencial; control de la relación temporal de los acontecimientos –exposición/efecto– y de la relación entre nivel de exposición y grado de efecto –dosis/respuesta–; análisis correcto –control de factores de confusión y modificadores de efecto–), por ejemplo.

La fuerza (grado) de las recomendaciones indica hasta qué punto podemos confiar en que poner en práctica la recomendación conllevará más beneficio que riesgo. En la elaboración de las recomendaciones se debe tener en cuenta, en primer lugar, el nivel de evidencia, pero también otras consideraciones: balance entre beneficios y riesgos, consistencia de los estudios, aplicabilidad práctica en mi paciente o población (incluyendo el riesgo basal en mi población), valores y preferencias de la población diana a la cual va dirigida, costes, etc. Establecer una recomendación, a favor o en contra de una intervención, no significa que todos los pacientes deban ser tratados de la misma manera, pues en la toma de decisión la evidencia procedente de la investigación es sólo uno de los cuatro círculos en una toma de decisiones basada en pruebas (figura 1).

Ambos conceptos, aunque relacionados y complementarios, se ocupan de aspectos distintos. Aunque la fuerza de una recomendación se apoya, decisivamente, en la calidad de la evidencia que la sustenta, ello puede no resultar suficiente de ser por ejemplo muy pequeña la magnitud del efecto sobre las variables primarias, tener poca precisión la estimación realizada o ser irrelevante desde el punto de vista clínico el resultado medido (diferencia entre significación estadística e importancia clínica).

FIGURA 1. Modelo actualizado en la toma de decisiones basada en pruebas



Por último, el elemento clave para decidir el grado de recomendación se obtiene al considerar el binomio beneficio/perjuicio neto para la salud, consecuencia del análisis de varios factores (magnitud del efecto y daño, disponibilidad social y coste).

El primer intento serio de introducir rigor y transparencia en la jerarquización de la evidencia fue realizado hace ya más de 30 años por la Canadian Task Force on Preventive Health Care (CTFPHC)¹, adaptado posteriormente por la United State Preventive Services Task Force (USPSTF)². Desde entonces numerosas organizaciones e instituciones, entre las que destacan el Centre for Evidence-Based Medicine (CEBM) de Oxford³, el Scottish Intercollegiate Guidelines Network (SIGN)⁴, el National Institute for Health and Clinical Excellence (NICE)⁵ o la U.S. Agency for Health Research and Quality (AHRQ)⁶, han

ido desarrollando sus propios sistemas jerárquicos y, actualmente, se contabilizan más de cien herramientas, 19 sistemas para evaluar la calidad y 7 para graduar las recomendaciones⁷. En síntesis, las escalas pueden utilizar letras (ej. A, B, C, etc.), números (ej. I, II, III, etc.) o una combinación de ambos (ej. Ia, Ib, IIa, etc.). Sin embargo, la situación a la que se ha llegado dista de ser satisfactoria^{8,9}. La comparación entre las distintas propuestas existentes (tabla 1) pone de manifiesto diferencias sustanciales en los criterios de gradación, con una baja sensibilidad y reproducibilidad de los mismos, múltiples posibilidades para evaluar y estructurar la evidencia y diferentes interpretaciones de los grados de recomendación. Además, la proliferación de escalas genera confusión y dudas en los usuarios, constatándose la inexistencia, hasta ese momento, de un modelo adecuado que pudiera ser universalmente aceptado¹⁰⁻¹².

TABLA 1. Nivel de evidencia y fuerza de la recomendación

| Organización | Nivel de evidencia más alto para artículos de tratamiento | Condiciones para considerar una recomendación como Grado A |
|---|--|--|
| Canadian Task Force on Preventive Health Care | I = Al menos un ECA | Un estudio (incluidos RS o MA) metodológicamente bien diseñado |
| Centre for Evidence-Based Medicine (Oxford) | Ia = RS con homogeneidad de ECA Ib = ECA con IC estrecho | Estudios consistentes de nivel I |
| Scottish Intercollegiate Guidelines Network | I++ = MA de alta calidad, RS de ECA o ECA con muy baja posibilidad de sesgo I+ = MA bien diseñados, RS de ECA o ECA con baja posibilidad de sesgo I- = MA, RS de ECA o ECA con alta posibilidad de sesgo | Al menos un MA, RS o ECA calificado como I++ o directamente aplicable a la población diana; o RS de ECA o evidencia suficiente de estudios valorados como I+ directamente aplicables a la población diana y consistentes |

ECA: Ensayo controlado aleatorizado; RS: Revisión sistemática; MA: Metaanálisis

EL SISTEMA GRADE: UNA PROPUESTA DE HOMOGENEIZACIÓN

Por ello, desde el año 2000, un grupo internacional integrado por más de 60 expertos, en su mayoría metodólogos y clínicos, muchos de ellos procedentes de las organizaciones antes mencionadas, constituyeron la iniciativa GRADE (The Grading of Recommendations Assessment, Development and Evaluation: Clasificación de la evaluación, desarrollo y valoración de las recomendaciones)¹³. Su objetivo principal era desarrollar un método común y razonable para calificar la calidad de la evidencia y la fuerza de las recomendaciones. Desde entonces han mantenido más de 40 encuentros, con una sistemática clara:

- 1) Evaluar los diferentes sistemas disponibles.
- 2) Desarrollar un nuevo sistema de clasificación.
- 3) Diseminar el nuevo sistema a través de la comunidad científica y de sus publicaciones.

De manera sintética, el sistema GRADE de juicios secuenciales presenta algunos aspectos diferenciales que se pueden resumir en tres^{14,15}:

- Inicialmente categoriza las **variables** de resultado y su **importancia** relativa. La evaluación tradicional se centraba en el fármaco, de modo que el fármaco es el que tenía las necesidades y se buscaba a pacientes que las satisficieran. La evaluación GRADE se centra en el paciente, de modo que este es el que tiene las necesidades y se busca a los fármacos que las satisfacen. Se parte de lo que le importa al paciente y se gradúa esta importancia.
- Posteriormente evalúa la **calidad de la evidencia**, entendiendo por tal hasta qué punto nuestra confianza en la estimación de un efecto es adecuada para apoyar una recomendación. Se distinguen cuatro categorías (alta, moderada, baja, muy baja) para cada una de las variables de resultado, y después se analiza la calidad global de la evidencia.
- Finalmente gradúa la **fuerza de las recomendaciones** en dos categorías (recomendaciones fuertes o débiles, bien a favor o en contra), indicando hasta qué punto podemos confiar si poner en práctica la recomendación conllevará más beneficios que riesgos.

a) Clasificación de la importancia relativa de las variables de resultado

En la etapa inicial del sistema GRADE deben establecerse de forma explícita las variables de resultado que realmente tienen interés para los pacientes y clasificarlas según su importancia relativa, diferenciando aquellas decisivas de otras importantes aunque no críticas. Se recomienda utilizar la siguiente escala de nueve puntos:

- 1-3: variable de resultado no importante. No incluir en la tabla de evaluación de la calidad o de resultados. Estas variables de resultado no jugarán un papel importante en la formulación de las recomendaciones.

- 4-6: variable de resultado importante pero no clave para la toma de decisiones.
- 7-9: variable de resultado clave para la toma de decisiones.

b) Clasificación de la calidad de la evidencia en el sistema GRADE

La evaluación de la calidad se realiza para cada una de las variables de resultado seleccionadas. Al igual que con los sistemas iniciales de clasificación de la calidad de la evidencia, la estrategia GRADE empieza considerando el diseño del estudio y su adecuación para responder a cada tipo de pregunta.

En principio, los ensayos aleatorizados sin limitaciones importantes constituyen evidencias de alta calidad, mientras que los estudios observacionales sin especiales puntos fuertes o limitaciones importantes son considerados evidencias de baja calidad.

No obstante, existen determinados factores que pueden incidir directamente en nuestra decisión sobre la calidad de la evidencia, modificando la valoración inicial:

b.1) Factores que pueden disminuir la calidad de la evidencia:

- Limitaciones del estudio: ausencia de ocultación de la asignación; ausencia de enmascaramiento; falta de seguimiento de un número importante de participantes; falta de cumplimiento de un análisis por intención de tratar; interrupción del estudio antes de la fecha planificada; etc.
- Falta de coherencia de los resultados: similitud en las estimaciones del efecto entre los diferentes estudios (diferencias en la población, en la intervención, en las variables de resultado, etc.).
- Carácter indirecto de la evidencia: comparación de dos fármacos con placebo en lugar de directamente entre ellos, diferencias en la población, la intervención, la comparación o el resultado de interés medido, etc.
- Imprecisiones: muestras pequeñas, intervalos de confianza amplios, etc.
- Sesgo de publicación/notificación: no inclusión de todos los estudios o variables de interés.

b.2) Factores que pueden aumentar la calidad de la evidencia:

- Gran magnitud del efecto: una asociación muy fuerte (riesgo relativo [RR] > 5 o RR < 0,2).
- Factores de confusión verosímiles que reducirían el efecto demostrado: pacientes con peor pronóstico asignados al grupo experimental y que aun así obtienen mejores resultados que los controles, etc.
- Gradiente dosis-respuesta.

Una vez determinada la calidad de la evidencia para cada una de las variables consideradas, son las variables clínicas clave las que determinan la calidad de la evidencia global. En el caso de que la calidad sea diferente, la estrategia GRADE sugiere que

TABLA 2. Implicaciones de las recomendaciones

| | Pacientes/cuidadores | Clínicos | Gestores |
|--------|--|--|--|
| Fuerte | La inmensa mayoría de las personas estaría de acuerdo con la acción recomendada y únicamente una pequeña parte no lo estaría | La mayoría de los pacientes debería recibir la intervención recomendada | La recomendación puede ser adoptada como política sanitaria en la mayoría de las situaciones |
| Débil | La mayoría de las personas estaría de acuerdo con la acción recomendada pero un número importante de ellas no | Reconoce que diferentes opciones serán apropiadas para diferentes pacientes y que el médico tiene que ayudar a cada paciente a llegar a la decisión más consistente con sus valores y preferencias | Existe necesidad de un debate importante y la participación de los grupos de interés |

TABLA 3. Comparación de GRADE y otros sistemas (GRADE Working Group)

| Elemento | Otros sistemas | GRADE | Ventajas de la clasificación de GRADE* |
|--|--|--|---|
| Definiciones | Definiciones implícitas de calidad (nivel) de evidencia y fuerza de recomendación | Definiciones explícitas | Clarifica el significado de los niveles y qué debería tenerse en cuenta al realizar estos juicios |
| Juicios | Juicios implícitos en cuanto a qué variables de resultado son importantes, la calidad de la evidencia para cada resultado importante, la calidad de la evidencia global, el balance entre beneficios y riesgos y el valor del incremento de los beneficios | Juicios secuenciales, explícitos | Clarifica cada uno de estos juicios y reduce el riesgo de introducir errores o sesgos que pueden surgir cuando los juicios se realizan implícitamente |
| Componentes clave de la calidad de la evidencia | No son considerados para cada variable de resultado importante. Los juicios sobre la calidad de la evidencia están a menudo basados únicamente en el diseño de estudio | Consideración sistemática y explícita del diseño del estudio, la calidad del estudio, la consistencia y si la evidencia es directa o indirecta en los juicios sobre la calidad de la evidencia | Asegura que estos factores son considerados de manera apropiada |
| Otros factores que pueden afectar a la calidad de la evidencia | No son tenidos en cuenta de manera explícita | Consideración explícita de los datos imprecisos o escasos, sesgo de notificación, fuerza de la recomendación, evidencia sobre el gradiente dosis-respuesta y posibles factores de confusión | Asegura la consideración de otros factores |
| Calidad global de la evidencia | Basados implícitamente en la calidad de la evidencia sobre los beneficios | Basado en la calidad más baja de la evidencia para cualquiera de las variables de resultado que son críticas para la toma de decisiones | Disminuye la probabilidad de clasificar mal la calidad global de la evidencia cuando la evidencia para una variable de resultado crítica no está disponible |
| Importancia relativa de las variables de resultado | Considerada implícitamente | Juicios explícitos sobre qué variables de resultado son críticas, cuáles son importantes pero no críticas, y cuáles son poco importantes y pueden ser ignoradas | Al clasificar la calidad global de la evidencia y la fuerza de las recomendaciones, asegura la consideración apropiada de cada variable de resultado |
| Balance entre beneficios de salud y riesgos | No considerado explícitamente | Consideración explícita del compromiso entre beneficios importantes y riesgos, la calidad de la evidencia de estos, la aplicación de la evidencia en circunstancias específicas, y el conocimiento acerca del riesgo basal | Clarifica y mejora la transparencia de los juicios sobre los beneficios y riesgos |
| Si el incremento de beneficios en salud merece los costes | No considerado explícitamente | Consideración explícita tras la consideración inicial de si existen beneficios netos para la salud | Se asegura que los juicios sobre el valor neto de los beneficios para la salud sean transparentes |
| Resúmenes de la evidencia y los hallazgos | Presentación inconsistente | Tablas de síntesis de evidencia GRADE consistentes, incluyendo la evaluación de la calidad y un resumen de los hallazgos | Asegura que todos los miembros del grupo basan sus juicios en la misma información y que esta información está accesible a otros |
| Grado de utilización | Raramente usado por más de una organización y escasa, si alguna, evaluación empírica | Colaboración internacional a través de una amplia gama de organizaciones que las desarrollan y evalúan | Construido sobre la experiencia previa para conseguir un sistema que sea más razonable, fiable y ampliamente aplicable |

*La mayoría de los otros sistemas no incluye ninguna de estas ventajas, aunque algunos incorporan algunas de estas

los responsables de formular las recomendaciones tengan en cuenta la calidad más baja de todas las variables decisivas.

Finalmente, se distinguen cuatro categorías jerárquicas que clasifican con este sistema la evidencia disponible para formular las recomendaciones:

- Alta calidad: es muy improbable que las investigaciones adicionales modifiquen la certidumbre con respecto al cálculo del efecto.
- Calidad moderada: probablemente, las investigaciones adicionales tendrán una repercusión importante en la certidumbre con respecto al cálculo del efecto, y pueden modificarlo.
- Baja calidad: muy probablemente, las investigaciones adicionales tendrán una repercusión importante en la certidumbre con respecto al cálculo del efecto, y es posible que lo modifiquen.
- Muy baja calidad: cualquier cálculo del efecto es muy dudoso.

c) Graduación de la fuerza de las recomendaciones

La fuerza de la recomendación refleja el grado de confianza que tiene el panel que la formula en que los efectos deseables de la intervención superan a los no deseables (o viceversa). El sistema GRADE solo tiene dos categorías para designar la fuerza de las recomendaciones:

- **Fuerte:** Existe un alto grado de confianza en que los efectos deseables de la intervención superan a los no deseables (**recomendación fuerte a favor**) o viceversa (**recomendación fuerte en contra**).
- **Débil:** Probablemente los efectos deseables de la intervención superan a los no deseables (**recomendación débil a**

favor), o viceversa (**recomendación débil en contra**), pero existe menor grado de certeza.

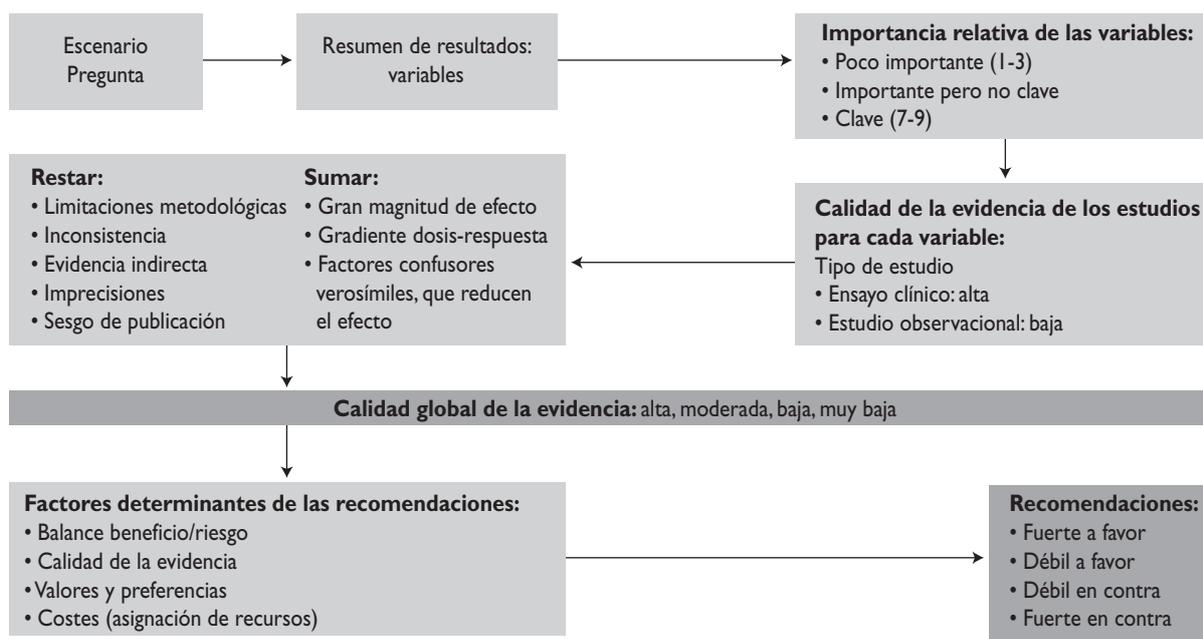
Cada tipo de recomendación conlleva una serie de implicaciones, tanto para los profesionales como para los pacientes y gestores, que pueden resumirse en la tabla 2.

Finalmente, resulta obligado conocer y tener en cuenta una serie de factores que van a determinar la fuerza de la recomendación, y que pueden resumirse en:

- Equilibrio entre efectos deseables y adversos: cuanto mayor es la diferencia entre los efectos deseables e indeseables, mayor es la probabilidad de que esté justificada una recomendación fuerte.
- Calidad de la evidencia: cuanto mayor es la calidad de la evidencia, mayor es la probabilidad de que se justifique una recomendación fuerte.
- Valores y preferencias: cuanto más varían los valores y preferencias, o mayor es la incertidumbre con respecto a ellos, mayor es la probabilidad de que esté justificada una recomendación débil.
- Costes (asignación de recursos): cuanto mayores son los costes de una intervención (es decir, mayores los recursos consumidos), menor es la probabilidad de que esté justificada una recomendación fuerte.

En la tabla 3 se expone una comparación entre el sistema GRADE y otros sistemas de clasificación de la calidad de la evidencia y fuerza de las recomendaciones, basada en los artículos que ha publicado el GRADE Working Group en British Medical Journal¹⁶⁻²⁰. En la figura 2 se expresa gráficamente la sistemática de cómo utilizar GRADE.

FIGURA 2. Propuesta de esquema de utilización de GRADE



BIBLIOGRAFÍA

1. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *CMAJ*. 1979;121:1193-254.
2. U.S. Preventive Services Task Force Ratings: Grade Definitions. Guide to Clinical Preventive Services, Third Edition: Periodic Updates, 2000-2003. [consultado: 10-VIII-2010]. Disponible en: <http://www.uspreventiveservicestaskforce.org/3rduspstf/ratings.htm>
3. Centre for Evidence-Based Medicine de Oxford. Levels of Evidence and Grades of Recommendation [consultado: 10-VIII-2010]. Disponible en: <http://www.cebm.net/index.aspx?o=1025>
4. Scottish Intercollegiate Guidelines Network. A guideline developer's handbook. Edinburg: SIGN; 2008; Publication nº 50 [consultado: 10-VIII-2010]. Disponible en: <http://www.sign.ac.uk/guidelines/fulltext/50/>
5. National Institute for Clinical Excellence (NICE): Guideline Development Methods. Information for National Collaborating Centers and Guideline Developers. London [consultado: 10-VIII-2010]. Disponible en: <http://www.nice.org.uk/>
6. 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 [consultado: 10-VIII-2010]. Disponible en: <http://archive.ahrq.gov/clinic/epcsums/strengthsum.htm>
7. West S, King V, Carey TS, Lohr KN, McKey N, Sutton SF, et al. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality. April 2002. [consultado: 10-VIII-2010]. Disponible en: <http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=erta47>
8. González de Dios J. Niveles de evidencia y fuerza de las recomendaciones: necesidad de homogeneización. *Espacioasma*. 2010;3: 24-8.
9. González de Dios J, Ochoa Sangrador C. De la evidencia a la recomendación: una tarea pendiente. *Med Clin (Barc)*. 2010 (en prensa).
10. 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.
11. 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: 1490-4.
12. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. and 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.
31. The Grading of Recommendations Assessment, Development and Evaluation (short GRADE) Working Group [consultado: 10-VIII-2010]. Disponible en: <http://www.gradeworkinggroup.org/>
14. Marzo Castillejo M, Montaña Barrientos A. El sistema GRADE para la toma de decisiones clínicas y la elaboración de recomendaciones y guías de práctica clínica. *Aten Primaria* 2007;39:457-60.
15. Marzo Castillejo M, Alonso-Coello P, Rotaecche del Campo R. ¿Cómo clasificar la calidad de la evidencia y la fuerza de las recomendaciones? *Aten Primaria*. 2006;37:5-7.
16. Guyatt GH, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924-6.
17. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008;336:995-8.
18. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336:1106-10.
19. Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A, et al; GRADE working group. Incorporating considerations of resources use into grading recommendations. *BMJ*. 2008;336:1170-3.
20. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al; GRADE Working Group. Going from evidence to recommendations. *BMJ*. 2008;336:1049-51.