

Calidad de la evidencia y fuerza de la recomendación

Dra. Julia Colomer Revuelta y Dr. José Galbe Sánchez-Ventura

Grupo PrevInfad/PAPPS

Octubre 2005

Introducción

Lectura crítica

Discusión en el grupo de expertos

Definición de la calidad de la evidencia (CE)

Normas para asignar la calidad de la evidencia

Problemas para asignar la calidad de la evidencia

Definición de la fuerza de la recomendación (FR)

Normas para asignar la fuerza de la recomendación (FR)

Conclusiones

Anexos

Introducción

El proceso de elaboración de recomendaciones sobre actividades preventivas es complejo y comienza con la formulación de cuestiones o preguntas a examinar. A este primer paso le sigue la búsqueda bibliográfica realizada según la metodología de la medicina basada en pruebas y cuyo objetivo es encontrar la mejor evidencia posible para cada una de las cuestiones examinadas mediante un sistema de gradación y de jerarquización de la calidad de la evidencia y de la fuerza de la recomendación que permita a los destinatarios de las recomendaciones conocer el valor final de las mismas.

Existen muchos grupos de trabajo que elaboran recomendaciones, Guías de práctica clínica, documentos de consenso basados en pruebas etc y puede decirse que hay tantos sistemas de clasificación de la evidencia como grupos hay¹⁻⁵. El grupo PrevInfad utiliza en estos momentos, para elaborar sus recomendaciones, la metodología de la Canadian Preventive Services Task Force (CPSTF) y cuya última revisión se hizo en 2003.

Esta clasificación como muchas otras exige valorar por un parte la calidad del estudio que sustenta la recomendación mediante un sistema jerarquizado asignando un valor que se traduce en un número que utiliza la terminología de los números romanos (I, II, III) y en

segundo lugar clasifica la fuerza de la recomendación mediante una terminología alfanumérica (A, B, C, D, E, I) Este segundo componente de la valoración debe ponderar otras variables como lo factible de un determinada medida, aceptabilidad de la población, costes aspectos éticos, resultados finales en calidad de vida para los usuarios etc. Que determinan si finalmente una determinada actividad es o no recomendada y con qué grado de recomendación.

Lectura crítica

Una vez elaborada la pregunta a examinar y realizada la búsqueda bibliográfica ,se deben obtener los trabajos originales que sustentan la recomendación ya sea en formato electrónico o en papel. A ello sigue un proceso de lectura crítica que debe aplicar normas estrictas de lectura crítica de literatura científica como las que se pueden obtener de C.A.S.P.E No es el objeto de este manuscrito profundizar en la lectura crítica de la literatura científica.

Discusión en el grupo de expertos

Algunos grupos de trabajo tienen la posibilidad de que todos los trabajos sean revisados y leídos de forma independiente por dos o más revisores para luego establecer un sistema de resolución de controversias entre lectores. El grupo prevInfad somete a discusión a todo el grupo de trabajo de expertos todas las propuestas que se realizan aportando previamente un documento de trabajo dónde figuran claramente no sólo las recomendaciones sino las pruebas que sustentan cada una de las recomendaciones. En último lugar y dentro del contexto del Programa de Actividades Preventivas y de Promoción de la Salud (PAPPS-SemfyC) cada grupo de expertos expone sus actividades en una reunión plenaria dónde los expertos de otros grupos pueden establecer correcciones y aceptar o rechazar finalmente la propuesta de actividad a desarrollar.

Definición de la calidad de la evidencia

Es el nivel de confianza en que la estimación de un efecto es correcta.

Normas para asignar la calidad de la evidencia:

La CE se establece según la clasificación de la CPSTF⁵ como de grado I si las pruebas que la sustentan proceden de al menos un ensayo clínico aleatorizado controlado o de un estudio de meta análisis. La CE es de nivel II-1 si las pruebas que la sustentan proceden de Evidencia obtenida de al menos un ensayo clínico aleatorizado controlado o de un estudio de meta análisis y II-2 cuando proceden estudios analíticos de cohortes o de casos y controles, preferentemente hechos en más de un centro o quipo de investigación. El II-3 queda reservado para la Evidencia obtenida de series temporales, con o sin intervención. Los resultados sorprendentes en experiencias no controladas también pueden ser incluidos en este grupo de evidencia. El nivel III corresponde a las Opiniones de reconocidos expertos, basadas en la experiencia clínica, estudios descriptivos o informes de comités de expertos.

Para Asignar la CE es necesario en primer lugar atender a la calidad del estudio y que en su ejecución no haya sesgos. Es importante que tenga consistencia o similitud en las estimaciones del efecto en diferentes estudios. Valorando la homogeneidad o heterogeneidad y finalmente la similitud en las comparaciones en los diferentes estudios.

Serían motivos para subir grados⁶ a la CE si:

- Evidencia de asociación causal fuerte: +1 RR>2 (<0,5) en 2 ó más estudios sin sesgo de confusión
- Evidencia de asociación causal muy fuerte: +2 ó RR>5 (<0,2)
- Evidencia directa sin problemas de validez: +1
- Evidencia de gradación dosis-respuesta: + 1

Serían motivos para bajar grados a la CE si:

- Limitaciones en la calidad del estudio(QE) en -1 serio ó -2 muy serio
- Inconsistencia importante: -1 serio y -2 muy serio
- Inconsistencia parcial: -1 menor -2 mayor
- Datos imprecisos: -1
- Probabilidad de sesgo de publicación: -1
- Todos los factores de confusión plausibles actuarían disminuyendo el efecto

Problemas para asignar la calidad de la evidencia:

Este sistema de clasificar la CE ha sido fuertemente criticado por muchos autores ya que al trabajar específicamente con Actividades Preventivas (APR) muchas de ellas son clasificadas con un nivel bajo de calidad ya que el máximo rango lo ostentan los metaanálisis y los estudios aleatorizados y controlados y las pruebas procedentes de este tipo de estudios son difíciles de encontrar en APR. Es por esto que el grupo de trabajo del BMJ Health Services Research en 2004 opina que deben establecerse métodos de clasificación de la CE específicamente pensados para APR⁷⁻⁹.

Definición de la fuerza de la recomendación

Nivel de confianza en que si se hace una determinada intervención ésta será más beneficiosa que perjudicial.

Normas para asignar la fuerza de la recomendación

La asignación de la FR y siguiendo siempre a la CPSTF utiliza un sistema alfanumérico de cinco grados. La A que corresponde a: hay buenas pruebas para recomendar la actividad. B: Hay pruebas aceptables para recomendar la actividad. C: Las pruebas existente son conflictivas y no permite realizar una recomendación a favor o en contra de la actividad, sin embargo, otros factores pueden influenciar la toma de decisiones. D: Hay pruebas aceptables para pronunciarse en contra de la actividad. E: Hay buenas pruebas para pronunciarse en contra de la actividad. I: Hay insuficientes pruebas (en cantidad y/o calidad) para hacer una recomendación, sin embargo, otros factores pueden influenciar la toma de decisiones.

Establecer la CE para cada resultado importante, establecer qué resultados son críticos para tomar una decisión tener en cuenta la calidad global de la Evidencia para cada resultado importante y el balance entre beneficios y riesgo. Para ello se debe asignar siempre la calidad más baja entre las medidas de resultados críticas para tomar una decisión. Debe tomarse una decisión sobre qué medidas o resultados son críticos para realizar un juicio de valor y considerar siempre los posibles efectos adversos con un balance beneficio/riesgo favorable.

Otros aspectos serán asignar un valor implícito o explícito o un valor relativo a cada medida de resultado, especificar ámbitos específicos, grupos de pacientes, grupos de riesgo y tener en cuenta a la población desfavorecida.

Así como tener en cuenta factores específicos de cada ámbito que puedan modificar la aplicabilidad de una intervención, la incertidumbre sobre el riesgo basal y la magnitud del efecto de los resultados principales y los costes.

Problemas en la asignación de la fuerza de la recomendación:

Algunas instituciones y organizaciones que se dedican a la Atención primaria han desarrollado un sistema de clasificación que responde al acrónimo SORT¹⁰⁻¹² (Strength of Recommendation Taxonomy) y que intenta diferenciar entre resultados orientados al paciente y que le proporcionan una mejor calidad de vida (A), resultados orientados a la enfermedad o variables intermedias fisiopatológicas, histopatológicas etc, (C) o resultados parcialmente orientados al paciente (B). La calidad de la evidencia la clasifican en base al diseño del estudio valorando tres dimensiones: Calidad Cantidad y consistencia

Comentario final

No existe evidencia para utilizar una tabla con preferencia sobre otra por lo que mientras no existan pruebas sólidas sobre las ventajas de una clasificación respecto a otras el grupo Previnfad seguirá utilizando la de la CPSTF Con todo debe tenerse en cuenta que cuando se estudian cuestiones de cambios de estilo de vida o cambios sobre comportamientos las tablas tanto de la CPSTF como la de la USPSTF pueden no ser los más adecuados y en este caso podrían tener ventajas clasificaciones como la de S.O,R.T la de de la AHCPRR que figuran en los anexos. Sin embargo es muy importante insistir en mejorar la metodología para asignar de la forma más rigurosa posible la CE y la FR.

En el caso de recomendaciones de nivel C o I se propone considerar otros aspectos que harán mas pertinente la recomendación:

- Considerar todas las posibilidades de beneficios y daños asociados con cualquier maniobra preventiva, incluyendo el aumento de la calidad o duración de la vida, alivio de ansiedad o ahorro de dinero, costos, "etiquetado" y ansiedad, incluyendo la inducida por un diagnóstico precoz.
- Aumentar la implicación del paciente en la decisión.
- Minimizar daños y evitar "etiquetado" innecesario.
- Propugnar cambios mayores sólo si hay fuertes pruebas de su necesidad.
- Evitar intervenciones costosas sin claro beneficio.
- Centrarse en condiciones con una carga de enfermedad alta.
- Estar alerta hacia las necesidades especiales de grupos de riesgo.

Bibliografía

1. Levels of evidence in AFP. Am family physicians. 2003;68(12):2306.
2. Lam R, Kennedy S. Using metaanalysis to evaluate evidence: Practical trips and traps. Canadian Journal of Psychiatry. 2005;50(3):167-8.
3. Levels of Evidence and Grades of Recommendation Center Evidence Based Medicine CEBM. http://www.cebm.net/levels_of_evidence.asp
4. Carmen Castiñeira Pérez, Carmen Costa Ribas, Arturo Louro González. Grupo MBE Galicia. Lectura crítica de una Guía Clínica (GPC). Fisterra. Cómo leer una guía de práctica clínica <http://fisterra.com/guias2/fmc/leer.asp>
5. Primo J. Niveles de evidencia y grados de recomendación. Simposium de gestión del conocimiento y su aplicación en la enfermedad inflamatoria crónica intestinal. Organizado por GETECCU (Grupo Español de Trabajo en Enfermedad de Crohn y colitis ulcerosa). Valencia 24 de enero de 2003.
6. Anónimo. Niveles de Evidencia y Fuerza de la recomendación. Seminario metodológico de la Agencia de Evaluación de tecnologías Sanitarias de Andalucía AETSA. 3 de noviembre de 2004, <http://www.juntadeandalucia.es/salud/orgdep/aetsa/pdf/Seminario%20GRADE2.pdf>
7. Kroke A, Boeing H, Rossnagel K, Willich SN. History of the concept of levels of evidence and their current status in relation to primary prevention through lifestyle interventions. Public Health Nutrition. 2003;7(2):279-84.
8. Atkins D et al. Grading Working Group Systems for grading the quality of evidence and strength of recommendation I: Critical appraisal of existing approaches. THE GRADA WORKING GROUP. BMC Health Services Research. 2004;4:38.
9. Atkins D et al. & Grading working Group. BMJ 2004;328(7454):1494-6.
10. Weiss B. SORT: Strength of Recommendation Taxonomy. Fam.Med. 2004;36(2):141-3.
11. Holger J, Schünemann MJ, Best D, Vist G, Oxman D & for the grading working Group. Letters, numbers, symbols and words: How to communicate grade of evidence and recommendations. CMAJ. 2003;169(7): 677-80.
12. Ebell M et al. Simplifying the language of evidence to improve patient care: Strength of Recommendation Taxonomy (SORT): A patient-centered approach to grading evidence in the medical literature. The Journal of Family Practice. February 2004. Vol. 53, No. 2

Anexo1

Documento de evaluación de escalas de evidencia^{9,11}:

1. En qué medida es la escala de evaluación aplicable a diferentes tipos de problemas como: Efectividad, Daño, Diagnóstico, pronóstico
 - Sí
 - No
 - No es seguro

2. En qué medida puede la escala ser usada con diferentes audiencias
 - Pequeña escala
 - Media escala
 - Gran escala

3. La EV es clara
 - Sí
 - Parcialmente
 - No

4. ¿Ocurre con mucha frecuencia que haya información necesaria no disponible?
 - A menudo
 - A veces
 - Casi nunca

5. ¿Con qué frecuencia deben tomarse decisiones subjetivas?
 - A menudo
 - A veces
 - Casi nunca

6. ¿Existen dimensiones de valoración que no están en el documento?
 - Sí
 - Parcialmente
 - No

7. ¿Hay dimensiones de valoración que no están incluidas y deberían estarlo?
 - Sí
 - Parcialmente

- No

8. ¿La forma de agregación de las dimensiones de valoración son claras y simples?

- Sí
- Parcialmente
- No

9. ¿La forma en la que las dimensiones de valoración fueron agregadas, son apropiadas?

- Sí
- Parcialmente
- No

10. ¿Son las categorías incluidas lo suficiente discriminativas para distintos niveles de CE y FR?

- Sí
- Parcialmente
- No

11. ¿Qué probabilidad hay de que la escala sea discriminativa entre altos y bajos grados de CE y FR?

- Sí
- Parcialmente
- No

12. ¿Es reproducible la evaluación?

- Probablemente no
- No es seguro
- Probablemente sí

Anexo 2

Niveles de la “evidencia” científica. Clasificación de la Canadian Task Force

<http://www.ctfphc.org/>

Calidad de la evidencia	Criterio
I	Obtenida de al menos un ensayo clínico aleatorizado controlado o de un meta-análisis
II-1	Obtenida de ensayos clínicos con grupo control sin distribución aleatoria.
II-2	Obtenida de estudios analíticos de cohortes o de casos y controles, preferentemente hechos en más de un centro o equipo de investigación
II-3	Obtenida por estudios comparativos de series temporales o de lugares diferentes, con o sin intervención. Los resultados sorprendentes en experiencias no controladas también pueden ser incluidos en este grupo de evidencia.
III	Opiniones de reconocidos expertos, basadas en la experiencia clínica, estudios descriptivos o informes de comités de expertos

Fuerza de la recomendación	Criterio
A	Hay buenas pruebas para recomendar la actividad
B	Hay pruebas aceptables para recomendar la actividad
C	Las pruebas existente son conflictivas y no permite realizar una recomendación a favor o en contra de la actividad, sin embargo, otros factores pueden influenciar la toma de decisiones.
D	Hay pruebas aceptables para pronunciarse en contra de la actividad
E	Hay buenas pruebas para pronunciarse en contra de la actividad
I	Hay insuficientes pruebas (en cantidad y/o calidad) para hacer una recomendación, sin embargo, otros factores pueden influenciar la toma de decisiones

Anexo 3

Strength of Recommendation Taxonomy (SORT)

Fuerza de la recomendación	Definición
A	Recomendación basada en metaanálisis de buena calidad Ensayos clínicos aleatorizados (diagnóstico) Reglas de decisión clínica validadas (diagnóstico) Cambios espectaculares en series clínicas
B	Rec. basada en evidencias inconsistentes o limitadas en resultados orientados al paciente como: RS o MA de baja calidad Ensayos Clínicos de baja calidad Estudios de cohorte(tratamiento) Cohorte retrospectiva (Pronóstico) Estudios Caso control
C	Resultados basados en consenso, opiniones de expertos o práctica clínica empírica Evidencia orientada a enfermedad, series de casos como: Guías de consenso Opiniones de expertos Práctica clínica Series de casos

Anexo 4

Clasificación de la USPTF

Niveles de evidencia	
Grado I	Evidencia obtenida a partir de al menos un ensayo clínico randomizado y controlado bien diseñado.
Grado II- 1	Evidencia obtenida a partir de ensayos clínicos no randomizados y bien diseñados.
Grado II- 2	Evidencia obtenida a partir de estudios de cohortes (2ª: prospectivo y 2b: retrospectivo).
Grado II- 3	Evidencia obtenida a partir de múltiples series comparadas en el tiempo, con o sin grupo control. Incluye resultados "dramáticos" producidos por experimentos no controlados (p. Ej. Penicilina en los años 40).
Grado III	Opiniones basadas en experiencias clínicas, estudios descriptivos o informes de comités de expertos.

En la siguiente tabla se muestra cuál es la relación entre la calidad de la evidencia y la fuerza de la recomendación de una determinada práctica o intervención según el U.S. Preventive Task Force:

FUERZA DE LA RECOMENDACIÓN	NIVEL DE LA EVIDENCIA
A: Existe adecuada evidencia científica para adoptar una práctica.	I, II- 1
B: Existe cierta evidencia para recomendar la práctica.	II- 1, II- 2
C: Hay insuficiente evidencia para recomendar o no recomendar la práctica.	III
D: Existe cierta evidencia para no recomendar la práctica.	II- 1, II- 2
E: Existe adecuada evidencia científica para no adoptar la práctica.	I, II- 1

Anexo 5

Clasificación de la US Agency for Health Care Policy and Research (AHCPR)

Nivel de evidencia	Tipo de estudio	Grado de recomendación
1	Estudios RCTs bien diseñados con 100 pacientes o más	A
2	Estudios RCTs bien diseñados con menos 100 pacientes	A
3	Estudios de cohortes bien diseñados	A
4	Estudios bien diseñados de caso-control	B
5	Estudios no controlados o mal diseñados	B
6	Evidencia conflictiva acerca del peso de la evidencia	B
7	Opinión de expertos	C