

Niveles de evidencia y grados de recomendación (I/II)

J. Primo

Hospital de Sagunto, Valencia.

Este trabajo se presentó en el Symposium "Gestión del conocimiento y su aplicación en la Enfermedad Inflamatoria Crónica Intestinal" celebrado el 24 de enero de 2003 en Valencia

El concepto de Medicina Basada en la Evidencia (MBE) fue desarrollado por un grupo de internistas y epidemiólogos clínicos liderados por Gordon Guyatt, de la Escuela de Medicina de la Universidad McMaster de Canadá ⁽¹⁾. En palabras de David Sackett, "la MBE es la utilización consciente, explícita y juiciosa de la mejor evidencia clínica disponible para tomar decisiones sobre el cuidado de los pacientes individuales" ⁽²⁾. En esencia, la MBE pretende aportar más ciencia al arte de la medicina, siendo su objetivo disponer de la mejor información científica disponible –la evidencia– para aplicarla a la práctica clínica ⁽³⁾.

En función del rigor científico del diseño de los estudios, pueden construirse escalas de clasificación jerárquica de la evidencia, a partir de las cuales se establecen recomendaciones respecto a la adopción de un determinado procedimiento médico o intervención sanitaria ^(4,5). Aunque hay diferentes escalas de gradación de la calidad de la evidencia científica, todas ellas son muy similares entre sí.

La primera de ellas fue formulada en 1979 por la *Canadian Task Force on the Periodic Health Examination* (www.ctfphc.org) para la evaluación de medidas preventivas ⁽⁶⁾ y adaptada en 1984 por la U.S. *Preventive Services Task Force* (USPSTF). En su tercera edición, publicada recientemente ⁽⁷⁾ y que puede consultarse en la web de la *Agency for Healthcare Research and Quality* (www.ahrq.gov/clincluspstfix.htm), se evalúa la calidad de la evidencia de una forma más elaborada que

no sólo tiene en cuenta el tipo de diseño de los estudios (Tabla I). Los grados de recomendación se establecen a partir de la calidad de la evidencia y del beneficio neto (beneficios menos perjuicios) de la medida evaluada (Tablas II y III). Además, en ella se realizan análisis de coste-efectividad ⁽⁸⁾.

Otra escala, la utilizada por la *Scottish Intercollegiate Guidelines Network* (SIGN), también ha sido revisada recientemente ⁽⁹⁾ (Tablas IV y V).

Para el simposio "Gestión del conocimiento y su aplicación en la enfermedad inflamatoria crónica intestinal", celebrado en Valencia el 24 de enero de 2003, se recomendó utilizar la propuesta por el *Centre for Evidence-Based Medicine* (CEBM) de Oxford, en la que se tienen en cuenta no sólo las intervenciones terapéuticas y preventivas (Tablas VI, VII y VIII), sino también las ligadas al diagnóstico, el pronóstico, los factores de riesgo y la evaluación económica. Está disponible en www.cebm.net/levels_of_evidence.asp y se mantiene permanentemente actualizada (última revisión: mayo 2001).

En nuestro país hay que destacar el esquema de gradación propuesto por la *Agència d'Avaluació de Tecnologia Mèdica* (AATM) de la Generalitat de Catalunya ⁽⁴⁾. Esta clasificación tiene en cuenta, además del diseño de los estudios, una valoración específica de su calidad (Tabla IX).

BIBLIOGRAFÍA

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992; 268: 2420-5.
2. Sackett DL, Rosenberg WMC, Guyatt JAM, Haynes RB, Richardson WS. Evidence based medicine: what is it and what it isn't. *BMJ* 1996; 312: 71-2.
3. Guerra Romero L. La medicina basada en la evidencia: un intento de acercar la ciencia al arte de la práctica clínica. *Med Clin (Barc)* 1996;107:377-82.
4. Jovell AJ, Navarro-Rubio MD. Evaluación de la evidencia científica. *Med Clin (Barc)* 1995;105:740-3.
5. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' Guides to the Medical Literature: IX. A method for grading health care recommendations. *JAMA* 1995;274:1800-4.
6. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J* 1979;121:1193-254.
7. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D, for the Methods Work Group, Third U.S. Preventive Services Task Force. Current methods of the U.S. Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001;20(3S):21-35.
8. Saha S, Hoerger TJ, Pignone MP, Teutsch SM, Helfand M, Mandelblatt JS, for the Cost Work Group of the Third U.S. Preventive Services Task Force. The art and science of incorporating cost effectiveness into evidence-based recommendations for clinical preventive services. *Am J Prev Med* 2001;20(3S):36-43.
9. Harbour R, Miller J, for the Scottish Intercollegiate Guidelines Network Grading Review Group. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323: 334-6.

Tabla I. Jerarquía de los estudios por el tipo de diseño (USPSTF) ⁽⁷⁾

Nivel de evidencia	Tipo de estudio
I	Al menos un ensayo clínico controlado y aleatorizado diseñado de forma apropiada.
II-1	Ensayos clínicos controlados bien diseñados, pero no aleatorizados.
II-2	Estudios de cohortes o de casos y controles bien diseñados, preferentemente multicéntricos.
II-3	Múltiples series comparadas en el tiempo, con o sin intervención, y resultados sorprendentes en experiencias no controladas.
III	Opiniones basadas en experiencias clínicas, estudios descriptivos, observaciones clínicas o informes de comités de expertos.

Tabla II. Establecimiento de las recomendaciones (USPSTF) ⁽⁷⁾

Calidad de la evidencia	Beneficio neto sustancial	Beneficio neto moderado	Beneficio neto pequeño	Beneficio neto nulo o negativo
Buena	A	B	C	D
Moderada	B	B	C	D
Mala	E	E	E	E

Tabla III. Significado de los grados de recomendación (USPSTF) ⁽⁷⁾

Grado de recomendación	Significado
A	Extremadamente recomendable (buena evidencia de que la medida es eficaz y los beneficios superan ampliamente a los perjuicios).
B	Recomendable (al menos moderada evidencia de que la medida es eficaz y los beneficios superan a los perjuicios).
C	Ni recomendable ni desaconsejable (al menos moderada evidencia de que la medida es eficaz, pero los beneficios son muy similares a los perjuicios y no puede justificarse una recomendación general).
D	Desaconsejable (al menos moderada evidencia de que la medida es ineficaz o de que los perjuicios superan a los beneficios).
I	Evidencia insuficiente, de mala calidad o contradictoria, y el balance entre beneficios y perjuicios no puede ser determinado.

Tabla IV. Niveles de evidencia (SIGN) ⁽⁹⁾

Nivel de evidencia	Tipo de estudio
1++	Meta-análisis de gran calidad, revisiones sistemáticas de ensayos clínicos aleatorizados o ensayos clínicos aleatorizados con muy bajo riesgo de sesgos.
1+	Meta-análisis bien realizados, revisiones sistemáticas de ensayos clínicos aleatorizados o ensayos clínicos aleatorizados con bajo riesgo de sesgos.
1-	Meta-análisis, revisiones sistemáticas de ensayos clínicos aleatorizados o ensayos clínicos aleatorizados con alto riesgo de sesgos.
2++	Revisiones sistemáticas de alta calidad de estudios de cohortes o de casos y controles, o Estudios de cohortes o de casos y controles de alta calidad, con muy bajo riesgo de confusión, sesgos o azar y una alta probabilidad de que la relación sea causal.
2+	Estudios de cohortes o de casos y controles bien realizados, con bajo riesgo de confusión, sesgos o azar y una moderada probabilidad de que la relación sea causal.
2-	Estudios de cohortes o de casos y controles con alto riesgo de confusión, sesgos o azar y una significativa probabilidad de que la relación no sea causal.
3	Estudios no analíticos (observaciones clínicas y series de casos).
4	Opiniones de expertos.

Tabla V. Grados de recomendación (SIGN) ⁽⁹⁾

Grado de recomendación	Nivel de evidencia
A	Al menos un meta-análisis, revisión sistemática o ensayo clínico aleatorizado calificado como 1++ y directamente aplicable a la población objeto, o Una revisión sistemática de ensayos clínicos aleatorizados o un cuerpo de evidencia consistente principalmente en estudios calificados como 1+ directamente aplicables a la población objeto y que demuestren globalmente consistencia de los resultados.
B	Un cuerpo de evidencia que incluya estudios calificados como 2++ directamente aplicables a la población objeto y que demuestren globalmente consistencia de los resultados, o Extrapolación de estudios calificados como 1++ o 1+.
C	Un cuerpo de evidencia que incluya estudios calificados como 2+ directamente aplicables a la población objeto y que demuestren globalmente consistencia de los resultados, o Extrapolación de estudios calificados como 2++.
D	Niveles de evidencia 3 o 4, o Extrapolación de estudios calificados como 2+.

Tabla VI. Niveles de evidencia (CEBM)

Nivel de evidencia	Tipo de estudio
1a	Revisión sistemática de ensayos clínicos aleatorizados, con homogeneidad.
1b	Ensayo clínico aleatorizado con intervalo de confianza estrecho.
1c	Práctica clínica ("todos o ninguno") (*)
2a	Revisión sistemática de estudios de cohortes, con homogeneidad.
2b	Estudio de cohortes o ensayo clínico aleatorizado de baja calidad (**)
2c	<i>Outcomes research</i> (***), estudios ecológicos.
3a	Revisión sistemática de estudios de casos y controles, con homogeneidad.
3b	Estudio de casos y controles.
4	Serie de casos o estudios de cohortes y de casos y controles de baja calidad (****)
5	Opinión de expertos sin valoración crítica explícita, o basados en la fisiología, <i>bench research</i> o <i>first principles</i> (*****)

Se debe añadir un signo menos (-) para indicar que el nivel de evidencia no es concluyente si:

- Ensayo clínico aleatorizado con intervalo de confianza amplio y no estadísticamente significativo.
- Revisión sistemática con heterogeneidad estadísticamente significativa.

(*) Cuando todos los pacientes mueren antes de que un determinado tratamiento esté disponible, y con él algunos pacientes sobreviven, o bien cuando algunos pacientes morían antes de su disponibilidad, y con él no muere ninguno.

(**) Por ejemplo, con seguimiento inferior al 80%.

(***) El término *outcomes research* hace referencia a estudios de cohortes de pacientes con el mismo diagnóstico en los que se relacionan los eventos que suceden con las medidas terapéuticas que reciben.

(****) Estudio de cohortes: sin clara definición de los grupos comparados y/o sin medición objetiva de las exposiciones y eventos (preferentemente ciega) y/o sin identificar o controlar adecuadamente variables de confusión conocidas y/o sin seguimiento completo y suficientemente prolongado. Estudio de casos y controles: sin clara definición de los grupos comparados y/o sin medición objetiva de las exposiciones y eventos (preferentemente ciega) y/o sin identificar o controlar adecuadamente variables de confusión conocidas.

(*****) El término *first principles* hace referencia a la adopción de determinada práctica clínica basada en principios fisiopatológicos.

Tabla VII. Grados de recomendación (CEBM)

Grado de recomendación	Nivel de evidencia
A	Estudios de nivel 1.
B	Estudios de nivel 2-3, o extrapolación de estudios de nivel 1.
C	Estudios de nivel 4, o extrapolación de estudios de nivel 2-3.
D	Estudios de nivel 5, o estudios no concluyentes de cualquier nivel.

La extrapolación se aplica cuando nuestro escenario clínico tiene diferencias importantes respecto a la situación original del estudio.

Tabla VIII. Significado de los grados de recomendación

Grado de recomendación	Significado
A	Extremadamente recomendable.
B	Recomendación favorable.
C	Recomendación favorable pero no concluyente.
D	Ni se recomienda ni se desaprueba.

Tabla IX. Niveles de calidad de la evidencia científica (AATM) ⁽⁴⁾

Nivel	Fuerza de la evidencia	Tipo de diseño	Condiciones de rigurosidad científica
I	Adecuada	Meta-análisis de ECA	Análisis de datos individuales de los pacientes Sin heterogeneidad Diferentes técnicas de análisis Meta-regresión Mega-análisis Calidad de los estudios
II	Adecuada	ECA de muestra grande	Evaluación del poder estadístico Multicéntrico Calidad del estudio
III	Buena a regular	ECA de muestra pequeña	Evaluación del poder estadístico Calidad del estudio
IV	Buena a regular	Ensayo prospectivo controlado no aleatorizado	Controles coincidentes en el tiempo Multicéntrico Calidad del estudio
V	Regular	Ensayo retrospectivo controlado no aleatorizado	Controles históricos Calidad del estudio
VI	Regular	Estudios de cohorte	Multicéntrico Apareamiento Calidad del estudio
VII	Regular	Estudios de casos y controles	Multicéntrico Calidad del estudio
VIII	Pobre	Series clínicas no controladas Estudios descriptivos: Vigilancia epidemiológica Encuestas Registros Bases de datos Comités de expertos Conferencias de consenso	Multicéntrico
IX	Pobre	Anécdotas o casos únicos	

(ECA: ensayo controlado aleatorizado)