

Niveles de Evidencia Clínica y Grados de Recomendación

Levels of scientific evidence and degrees of recommendation

MELLA SOUSA, MARIO¹; ZAMORA NAVAS, PLÁCIDO²; MELLA LABORDE, MARIO³; BALLESTER ALFARO, JUAN JOSÉ⁴; UCEDA CARRASCOSA, PILAR⁵

¹ Editor Jefe Revista SATO

² Servicio Traumatología y Cirugía Ortopédica. Hospital U. Virgen de la Victoria, Málaga

³ Servicio de Cirugía General y del Aparato Digestivo. Hospital U. San Juan, Alicante.

⁴ Servicio Traumatología y Cirugía Ortopédica. Hospital U. Puerto Real, Cádiz.

⁵ Servicio Traumatología y Cirugía Ortopédica. Hospital U. Reina Sofía, Córdoba.

RESUMEN

Se presenta una puesta al día de los niveles de evidencia científica mencionando los diversos grados de recomendación para pacientes asintomáticos, indicando qué procedimientos son los más adecuados y cuáles debían ser evitados. De todos los sistemas descritos, se busca aquel que cumpla el principio de sencillez y utilidad. Elegimos para nuestro medio la valoración del Centre for Evidence-Based Medicine, Oxford (OCEBM). Esta clasificación tiene la ventaja que nos asegura el conocimiento más relativo a cada escenario, por su alto grado de especialización. Además tiene la prerrogativa de aclarar cómo afecta la falta de rigurosidad metodológica al diseño de los estudios, disminuyendo su valoración no sólo en la gradación de la evidencia, sino también en la fuerza de las recomendaciones.

ABSTRACT

An update of the levels of scientific evidence stating the varying degrees of recommendation for asymptomatic patients, indicating which procedures are most appropriate and what should be avoided all the systems described, we seek one that meets the principle of simplicity and utility. We chose for our setting the valuation of the Centre for Evidence-Based Medicine, Oxford (OCEBM). This classification has the advantage that assures us the knowledge on each scenario, its high degree of specialization. It also has the prerogative to clarify how it affects the lack of methodological rigor to the study design, reducing its assessment not only in the grading of the evidence, but also on the strength of recommendations.

Palabras clave: niveles de evidencia, grados de recomendación, clasificación de Oxford.

Key words: *evidence levels , recommendation grades, classification of Oxford.*

CONTACTO: Mario Mella Sousa.
mellasato@gmail.com

INTRODUCCIÓN

El uso de la evidencia en ciencias se atribuye a la medicina tradicional china, en los tiempos del emperador Qianlong, cuando ya se señalaba el método “kaozheng” que representa la “búsqueda de evidencia práctica”⁽¹⁾, pero su desarrollo, como se conoce en la actualidad, se originó con la creación en 1976 de la Canadian Task Force on Preventive Health Care (CTFPHC), quienes fueron los primeros en generar y organizar los niveles de evidencia y los grados de recomendación para pacientes asintomáticos, indicando qué procedimientos eran los más adecuados y cuáles debían ser evitados⁽²⁾. Esta metodología permitió tomar conciencia de la existencia de un orden jerárquico en la calidad de la evidencia entre los estudios científicos, donde lógicamente aquellos que presentan más sesgos, no debieran justificar acciones clínicas⁽³⁾. También en Canadá, desde 1992, un grupo de médicos internistas y epidemiólogos clínicos de la Facultad de Medicina de la Universidad de McMaster, definieron, sistematizaron y consolidaron el concepto de medicina basada en la evidencia (MBE) o “medicina basada en pruebas”, contribuyendo al cambio paradigmático para estudiar y ejercer las ciencias médicas, pues, entre otros tantos hechos, esta práctica se hacía cada vez más compleja, debido al incremento progresivo y abrumador de las publicaciones científicas que no se asociaban a calidad o a un aumento del tiempo necesario para leer y valorar de manera apropiada dicha información^(4,5).

El análisis constante de la evidencia disponible desde la perspectiva de los diferentes escenarios clínicos, permite establecer grados de recomendación para el ejercicio de procedimientos diagnósticos, terapéuticos, preventivos y económicos en salud; e indica la forma de valorar el conocimiento en función de etiología, daño, morbilidad y complicaciones; pronóstico, historia natural y curso clínico de una enfermedad o evento de interés. Estos han de actualizarse permanentemente en relación al avance del conocimiento, del desarrollo tecnológico y al estado del arte⁽⁶⁾.

Se estima que hasta la fecha se han descrito y propuesto alrededor de 100 sistemas distintos para valorar la evidencia⁽⁷⁾, lo que nos orienta acerca del valor que

le otorga la comunidad científica a esta forma de hacer medicina; pero que, por otra parte contradice el principio de sencillez y utilidad que inspiró al paradigma de la MBE. Esta situación puede confundir al clínico en la búsqueda de la mejor evidencia aplicable a su realidad, debido a que, a lo anterior se agrega la dificultad propia del lenguaje epidemiológico en que se expresa la información, siendo en ocasiones la comprensión de estos tópicos, sólo privilegio de algunos.

El objetivo de este artículo es entregar información referente a los niveles de evidencia para cada tipo de estudio según el escenario clínico o ámbito de la práctica clínica que corresponda, interpretándolos en un lenguaje clínico comprensible, especialmente para quienes se acercan a la jerarquización de la evidencia y sus recomendaciones, desde la práctica clínica.

CLASIFICACIÓN DE LA EVIDENCIA Y TIPOS DE ESTUDIO

El nivel o grado de evidencia clínica es un sistema jerarquizado, basado en las pruebas o estudios de investigación, que ayuda a los profesionales de la salud a valorar la fortaleza o solidez de la evidencia asociada a los resultados obtenidos de una estrategia terapéutica^(1,6).

No todos los conocimientos provenientes de los artículos científicos publicados, tienen el mismo impacto o valor sobre la toma de decisiones en materia de salud; por ello, se hizo necesario evaluar la calidad de la evidencia. Esto es, en términos simples, el análisis de la validez de los hallazgos en virtud de la calidad metodológica de las investigaciones que los soportan, garantizándonos por una parte un acercamiento a la veracidad científica; y por otra, a que esta verdad pueda traducirse en recomendaciones que a partir de la valoración crítica de los estudios, nos permitan aplicarlas a la problemática clínica o evento de interés que nos ocupe. Resulta importante señalar, en este momento, que no todos los diseños tienen el mismo poder para formular una recomendación; y que más aún, un mismo diseño de investigación puede tener un nivel de evidencia y grado de recomendación diferente según el escenario clínico o ámbito de la práctica clínica que corresponda.

Por lo anteriormente expuesto, nos parece que vale la

pena aclarar algunos conceptos que seguirán utilizándose de aquí en adelante:

Se define como diseño de investigación a los distintos tipos de estudios que con sus características metodológicas propias, permiten llevar a cabo una investigación clínica^(5,8).

Como escenario, ámbito o entorno, al ambiente en el que se desarrolla la situación clínica que se está evaluando; es decir: tratamiento, prevención, etiología, daño, pronóstico e historia natural, diagnóstico diferencial, prevalencia, estudios económicos y análisis de decisión.

Como niveles de evidencia, a herramientas, instrumentos y escalas que clasifican, jerarquizan y valoran la evidencia disponible, de forma tal que en base a su utilización se pueda emitir juicios de recomendación.

Como grados de recomendación a una forma de clasificación de la sugerencia de adoptar o no la adquisición o puesta en marcha de tecnologías sanitarias según el rigor científico de cada tipo de diseño⁽⁴⁾.

Dependiendo de los tipos de diseños de investigación clínica utilizados, podemos observar diferentes niveles o gradación de la calidad de estos según el escenario de investigación clínica a la que se refiera. Por una parte, llama la atención que las intervenciones terapéuticas sean las más abordadas desde esta concepción, situación que posiblemente no sea por azar, puesto que es un hecho que entre el 50% y el 60% de las publicaciones versan sobre el ámbito del tratamiento o los procedimientos terapéuticos^(9,10,11). No obstante ello, es deseable que todo el quehacer en salud sea analizado bajo la misma perspectiva. Lo anterior se relaciona al grado de avance

que los diversos grupos de expertos (autores de clasificaciones de la evidencia), han ido adquiriendo y que han adecuado, paulatinamente, estos fundamentos a los diferentes escenarios y diseños. Debemos advertir entonces, que sólo las clasificaciones propuestas por Sackett, National Institute for Health and Clinical Excellence (NICE) y el Oxford Centre for Evidence-Based Medicine (OCEBM), consideran otras áreas de la investigación clínica además de la concerniente a tratamiento. De esta forma, NICE adiciona evaluación de la evidencia para diagnóstico; mientras que Sackett considera cuatro grandes grupos temáticos: terapia, prevención, etiología y daño; pronóstico e historia natural; diagnóstico; y estudios económicos. OCEBM, que es una modificación “complementaria” de la clasificación de Sackett, incluye terapia, prevención, etiología y daño en un subgrupo; pronóstico e historia natural en otro; diagnóstico en el siguiente; diagnóstico diferencial y estudios de prevalencia en otro; y estudios económicos y análisis de decisión en otro.

A continuación se mencionan y describen las propuestas de jerarquización de la evidencia más utilizada en la actualidad (Tabla 1).

CANADIAN TASK FORCE ON PREVENTIVE HEALTH CARE (CTFPHC)^(2,12)

Esta propuesta de clasificación de la evidencia busca generar recomendaciones de una manera práctica, adoptando una posición binaria, “hágalo o no”, pero sólo en el ámbito de la prevención. Este grupo de estudio generó

Tabla 1: Propuestas de jerarquización de la evidencia analizadas

Propuesta	Tratamiento	Prevención	Etiología	Daño	Pronóstico	Diagnóstico	Prevalencia	Económicos
CTFPHC		X						
SACKETT	X	X	X	X	X	X		X
USPSTF						X		
OCEBM	X	X	X	X	X	X	X	X
SIGN	X							
NICE	X					X		

su primer informe en el año 1979, en el cual se divulgó el análisis de la evidencia hasta esa fecha para 78 enfermedades, lo que permitió ordenar y producir planes de salud para la población canadiense desde la perspectiva de “la prevención”. La metodología de este grupo hace énfasis en el tipo de diseño utilizado y la calidad de los estudios publicados, basándose finalmente en tres elementos claves:

- **1.** Un orden para los grados de recomendación, establecido por letras del abecedario donde, las letra A y B indican que existe evidencia para ejercer una acción (*se recomienda hacer*); D y E indican que no

debe llevarse a cabo una maniobra o acción determinada (*se recomienda no hacer*); la letra C, indica que la evidencia es “conflictiva”, o sea, que existe contradicción. Y la letra I que indica insuficiencia en calidad y cantidad de evidencia para establecer una recomendación (Tabla 2).

- **2.** Niveles de evidencia clasificados según diseño de estudio de I a III, disminuyendo en calidad según se acrecienta numéricamente. Para el número II se subdivide en números arábigos del 1 al 3 (Tabla 3).
- **3.** Niveles de evidencia clasificados según la validez

Tabla 2: Grados de recomendación para intervenciones de prevención (CTFPHC)

Grado de recomendación	Interpretación
A	Existe buena evidencia para recomendar la intervención clínica de prevención.
B	Existe moderada evidencia para recomendar la intervención clínica de prevención.
C	La evidencia disponible es conflictiva y no permite hacer recomendaciones a favor o en contra de la intervención clínica preventiva; sin embargo, otros factores podrían influenciar en la decisión.
D	Existe moderada evidencia para recomendar en contra de la intervención clínica de prevención.
E	Existe buena evidencia para recomendar en contra de la intervención clínica de prevención.
I	Existe evidencia insuficiente (en cantidad y en calidad) para hacer una recomendación; sin embargo, otros factores podrían influenciar en la decisión.

Tabla 3: Niveles de evidencia e interpretación de los tipos de estudio para intervenciones de prevención (CTFPHC)

Niveles de evidencia	Interpretación
I	Evidencia a partir de EC con asignación aleatoria.
II-1	Evidencia a partir de EC sin asignación aleatoria.
II-2	Evidencia a partir de estudios de cohortes y casos y controles, preferiblemente realizados por más de un centro o grupo de investigación.
II-3	Evidencia a partir de comparaciones en el tiempo o entre sitios, con o sin la intervención; podrían incluirse resultados espectaculares provenientes de estudios sin asignación aleatoria.
III	Opinión de expertos, basados en la experiencia clínica; estudios descriptivos o informes de comités de expertos.

interna o calidad metodológica del estudio, en buena, moderada e insuficiente (Tabla 4).

Como todo sistema de clasificación, este presenta algunas debilidades que se mencionan a continuación:

No abarca toda la dimensionalidad de la problemática de la “prevención”, respecto a las condiciones particulares de quienes son sujeto de la aplicación de medidas preventivas.

En su análisis no se incorpora la evaluación del aspecto financiero para la disponibilidad de las intervenciones preventivas.

La propuesta sólo se basó en población canadiense, por ende, el uso de las recomendaciones sólo puede ser extrapolada a poblaciones similares a ésta; por ello, esta propuesta tiene un inconveniente relacionado con su validez externa que ha de ser valorado al momento de pretender aplicarla en otros escenarios, pues de lo contrario se corre el riesgo de incurrir en esfuerzos económicos con resultados erráticos.

No contempla la relación del paciente, sus expectativas y su medio, para establecer las recomendaciones.

Puede inducir a errores al momento de valorar las recomendaciones para su puesta en práctica en los sistemas estatales de salud, esto quiere decir que una recomendación B puede ser menospreciada pudiendo tener un beneficio importante para la población.

No contempla otro tipo de áreas de investigación como tratamiento, etiología, daño, pronóstico, etc.

CLASIFICACIÓN DE LA EVIDENCIA SEGÚN SACKETT⁽¹³⁾

Esta sistematización propuesta por el epidemiólogo David L. Sackett (la que se emplea generalmente), jerarquiza la evidencia en niveles que van del 1 a 5; siendo el nivel 1 la “mejor evidencia” y el nivel 5 la “peor, la más mala o la menos buena”, según como se quiera leer (Tabla 5).

Ésta fue la primera propuesta que consideró otros escenarios clínicos o ámbitos de la práctica clínica diferentes de la prevención. Incorporó los análisis económicos, el diagnóstico y el pronóstico.

Hasta hoy, ha sido ampliamente utilizada por diferentes grupos científicos. A cada ámbito o escenario clínico le otorga el diseño de estudio más apropiado para la elaboración de las recomendaciones. Así, en el escenario de terapia, los diseños más puntuados corresponden a las revisiones sistemáticas (RS) de ensayos clínicos controlados con asignación aleatoria (EC); en escenarios de pronóstico, los estudios de cohortes; en escenarios de diagnóstico, los estudios de pruebas diagnósticas con estándar de referencia, etc.

Presenta desventajas que son comunes a todas las propuestas existentes, como que puede llevar a des-

Tabla 4: Validez interna e interpretación de los tipos de estudio para intervenciones de prevención (CTFPHC)

Niveles de evidencia	Interpretación
Buena	Un estudio (incluyendo la RS y el meta-análisis) que cumple los criterios específicos de estudio bien diseñado.
Moderada	Un estudio (incluyendo la RS y el meta-análisis) que no cumple (o no está claro que cumpla) al menos uno de los criterios específicos de estudio bien diseñado, aunque no tiene “defectos fatales”.
Insuficiente III	Un estudio (incluyendo la RS y el meta-análisis) que tiene en su diseño al menos un “defecto fatal” o no cumple (o no está claro que cumpla) al menos uno de los criterios específicos de estudio bien diseñado, aunque no presenta “errores fatales” o una acumulación de defectos menores que hagan que los resultados del estudio no permitan elaborar las recomendaciones.

preciar recomendaciones que se basan en evidencias inferiores a las de nivel 1; por ejemplo un estudio de nivel de evidencia 4, aún cuando se trate de un artículo cuyo aporte sea novedoso. No obstante, los autores son enfáticos en que lo que debe valorarse es «la mejor evidencia disponible actual»; pues «lo actual puede variar en el día a día»; y de este modo puede ocurrir que ante determinadas situaciones «la mejor evidencia disponible actual» sea una serie de casos y no un EC; y dos meses después, aparezca un estudio de cohorte prospectivo

que dará «la mejor evidencia disponible actual».

Esta clasificación fue pionera y ha servido de base para el desarrollo de clasificaciones más completas, como la propuesta por el OCEBM⁽¹⁴⁾.

U.S. PREVENTIVE SERVICES TASK FORCE (USPSTF)⁽¹⁵⁾

Este grupo de expertos jerarquizó y estableció la fuerza de sus recomendaciones a partir de la calidad de la evidencia y del beneficio neto; es decir, beneficios menos

Tabla 5: Clasificación de los niveles de evidencia según Sackett

Recomendación	Nivel	Terapia, prevención, etiología y daño	Pronóstico	Diagnóstico	Estudios económicos
A	1a	RS con homogeneidad y Meta-análisis de EC	RS con homogeneidad y Meta-análisis de estudios de cohortes concurrente	RS de estudios de diagnóstico nivel 1	RS de estudios económicos de nivel 1
	1b	EC individuales con intervalo de confianza estrecho	Estudio individual de cohorte concurrente con seguimiento superior al 80% de la cohorte	Comparación independiente y enmascarada de un espectro de pacientes consecutivos sometidos a la prueba diagnóstica y al estándar de referencia	Análisis que compara los desenlaces posibles, contra una medida de costos. Incluye un análisis de sensibilidad
B	2a	RS con homogeneidad de estudio de cohortes	RS de cohortes históricas	RS de estudios diagnósticos de nivel mayor a 1	RS de estudios económicos de nivel mayor a 1
	2b	Estudio de cohortes individual. EC de baja calidad	Estudio individual de cohortes históricas	Comparación independiente enmascarada de pacientes no consecutivos, sometidos a la prueba diagnóstica y al estándar de referencia	Comparación de un número limitado de desenlaces contra una medida de costo. Incluye análisis de sensibilidad
	3a	RS con homogeneidad de estudios de casos y controles			
	3b	Estudio de casos y controles individuales		Estudios no consecutivos o carentes de un estándar de referencia	Análisis sin una medida exacta de costo, pero incluye análisis de sensibilidad
C	4	Series de casos. Estudio de cohortes y casos y controles de mala calidad	Series de casos. Estudios de cohortes de mala calidad	Estudios de casos y controles sin la aplicación de un estándar de referencia	Estudio sin análisis de sensibilidad
D	5	Opinión de expertos sin evaluación crítica explícita, o basada en fisiología, o en investigación teórica	Opinión de expertos sin evaluación crítica explícita, o basada en fisiología, o en investigación teórica	Opinión de expertos sin evaluación crítica explícita, o basada en fisiología, o en investigación teórica	Opinión de expertos sin evaluación crítica explícita, o basada en teoría económica

Por homogeneidad se entiende que una RS que está libre de variaciones (heterogeneidad) en las direcciones o grados de resultados entre los estudios individuales.

perjuicios de la medida evaluada para su aplicación en “exámenes periódicos de salud”. Por otro lado, analizó el coste-efectividad de las intervenciones, por ende su aporte vino a complementar lo que había generado el grupo de CTFPHC^(2,12). La fuerza de las recomendaciones va desde la letra A hasta la E, otorgándose una A cuando existe buena evidencia que respalda la recomendación; y una E, que indica que existe buena evidencia que sustenta la recomendación de evitar la intervención (Tabla 6). La calidad de la evidencia es valorada en buena, justa o insuficiente en la fuerza de las recomendaciones y se basa en la consideración sistemática de tres criterios:

el concepto de “costo de sufrimiento de la condición estudiada”, las “características de la intervención” y la “efectividad de la intervención” demostrada en investigaciones clínicas publicadas; recibiendo un especial énfasis la efectividad de la intervención. En la revisión de los estudios clínicos, se utilizan criterios estrictos para la selección de la evidencia admisible y se otorga un acento especial en la calidad de los diseños de los estudios. Para jerarquizar la calidad de la evidencia, se otorga mayor peso a aquellos diseños que metodológicamente ofrecen menor riesgo de sesgos y errores aleatorios (Tabla 7). Es interesante observar que la correlación entre la fuerza de

Tabla 6: Recomendaciones a partir de calidad de evidencia para exámenes periódicos de salud (USPSTF)

Recomendación	Interpretación
A	La USPSTF recomienda claramente que los clínicos proporcionen la intervención a los pacientes que cumplan los criterios. La USPSTF ha encontrado buena evidencia de que la medida mejora de manera importante los resultados en salud y concluye que los beneficios superan ampliamente a los riesgos.
B	La USPSTF recomienda que los clínicos proporcionen la intervención a los pacientes. La USPSTF ha encontrado evidencia moderada de que la medida mejora de manera importante los resultados en salud y concluye que los beneficios superan a los riesgos.
C	La USPSTF no recomienda a favor o en contra de la intervención. La USPSTF ha encontrado al menos evidencia moderada de que la medida puede mejorar los resultados en salud, pero los beneficios son muy similares a los riesgos y no puede justificarse una recomendación general.
D	La USPSTF recomienda en contra que los clínicos proporcionen la intervención a los pacientes asintomáticos. La USPSTF ha encontrado al menos evidencia moderada de que la medida es ineficaz o que los riesgos superan a los beneficios.
I	La USPSTF concluye que la evidencia es insuficiente para recomendar a favor o en contra de la intervención. No existe evidencia de que la intervención es ineficaz, o de calidad insuficiente, o conflictiva y que el balance entre los riesgos y los beneficios no se puede determinar.

Tabla 7: Niveles de evidencia e interpretación de estos para exámenes periódicos de salud (USPSTF)

Evidencia	Interpretación
Buena	La evidencia incluye resultados consistentes a partir de estudios bien diseñados y realizados en poblaciones representativas que directamente evalúan efectos sobre resultados de salud.
Moderada	La evidencia es suficiente para determinar efectos sobre resultados de salud, pero la fuerza de la evidencia es limitada por el número, la calidad o la consistencia de los estudios individuales, la generalización a la práctica rutinaria, o la naturaleza indirecta de la evidencia sobre los resultados de salud.
Insuficiente	La evidencia es insuficiente para evaluar los efectos sobre los resultados de salud debido al número limitado o al poder de estudios, defectos importantes en su diseño o realización, inconsistencias en la secuencia de la evidencia, o falta de información sobre resultados de salud importantes.

Tabla 8: Clasificación de los niveles de evidencia de Oxford (OCEBM)

Grado de recomendación	Nivel de evidencia	Tratamiento, prevención, etiología y daño	Pronóstico e historia natural	Diagnóstico	Diagnóstico diferencial y estudios de prevalencia	Estudios económicos y análisis de decisión
A	1a	RS con homogeneidad de EC controlados con asignación aleatoria	RS de estudios de cohortes, con homogeneidad, o sea que incluya estudios con resultados comparables, en la misma dirección y validados en diferentes poblaciones	RS de estudios diagnósticos de nivel 1 (alta calidad), con homogeneidad, o sea que incluya estudios con resultados comparables y en la misma dirección y en diferentes centros clínicos	RS con homogeneidad de estudios de cohortes prospectivas	RS con homogeneidad de estudios económicos de nivel 1
	1b	EC individual con intervalo de confianza estrecho	Estudios de cohortes individuales con un seguimiento mayor de 80% de la cohorte y validados en una sola población	Estudios de cohortes que validen la calidad de una prueba específica, con estándar de referencia adecuado (independientes de la prueba) o a partir de algoritmos de estimación del pronóstico o de categorización del diagnóstico o probado en un centro clínico	Estudio de cohortes prospectiva con buen seguimiento	Análisis basado en costes o alternativas clínicamente sensibles; RS de la evidencia; e incluyendo análisis de la sensibilidad
	1c	Eficiencia demostrada por la práctica clínica. Considera cuando algunos pacientes mueren antes de ser evaluados	Resultados a partir de la efectividad y no de su eficacia demostrada a través de un estudio de cohortes. Series de casos todos o ninguno	Pruebas diagnósticas con especificidad tan alta que un resultado positivo confirma el diagnóstico y con sensibilidad tan alta que un resultado negativo descarta el diagnóstico	Series de casos todos o ninguno	Análisis absoluto en términos de mayor valor o peor valor
B	2a	RS de estudios de cohortes, con homogeneidad	RS de estudios de cohorte retrospectiva o de grupos controles no tratados en un EC, con homogeneidad	RS de estudios diagnósticos de nivel 2 (mediana calidad) con homogeneidad	RS (con homogeneidad de estudios 2b y mejores)	RS (con homogeneidad) de estudios económicos con nivel mayor a 2
	2b	Estudio de cohortes individual con seguimiento inferior a 80% (incluye EC de baja calidad)	Estudio de cohorte retrospectiva o seguimiento de controles no tratados en un EC, o GPC no validadas	Estudios exploratorios que, a través de una regresión logística, determinan factores significativos, y validados con estándar de referencia adecuado (independientes de la prueba)	Estudios de cohortes retrospectivas o de seguimiento insuficiente	Análisis basados en costes o alternativas clínicamente sensibles; limitado a revisión de la evidencia; e incluyendo un análisis de sensibilidad
	2c	Estudios ecológicos o de resultados en salud	Investigación de resultados en salud		Estudios ecológicos	Auditorías o estudios de resultados en salud
	3a	RS de estudios de casos y controles, con homogeneidad		RS con homogeneidad de estudios 3b y de mejor calidad	RS con homogeneidad de estudios 3b y mejores	RS con homogeneidad de estudios 3b y mejores
	3b	Estudios de casos y controles individuales		Comparación enmascarada y objetiva de un espectro de una cohorte de pacientes que podría normalmente ser examinado trastorno, pero el estándar de referencia no se aplica a todos los pacientes del estudio. Estudios no consecutivos o sin la aplicación de un estándar de referencia		Estudio no consecutivo de cohorte, o análisis muy limitado de la población basado en pocas alternativas o costes, estimaciones de datos de mala calidad, pero incluyendo análisis de la sensibilidad que incorporan variaciones clínicamente sensibles

Contiuación de Tabla 8

Grado de recomendación	Nivel de evidencia	Tratamiento, prevención, etiología y daño	Pronóstico e historia natural	Diagnóstico	Diagnóstico diferencial y estudios de prevalencia	Estudios económicos y análisis de decisión
C	4	Serie de casos, estudios de cohortes, y de casos y controles de baja calidad	Serie de casos y estudios de cohortes de pronóstico de poca calidad	Estudio de casos y controles, con escasos o sin estándares de referencia independiente	Series de casos o estándares de referencia obsoletos	Análisis sin análisis de sensibilidad
D	5	Opinión de expertos sin evaluación crítica explícita, ni basada en fisiología, ni en trabajo de investigación juicioso ni en "principios fundamentales"	Opinión de expertos sin evaluación crítica explícita, ni basada en fisiología, ni en trabajo de investigación juicioso ni en "principios fundamentales"	Opinión de expertos sin evaluación crítica explícita, ni basada en fisiología, ni en trabajo de investigación juicioso ni en "principios fundamentales"	Opinión de expertos sin evaluación crítica explícita, ni basada en fisiología, ni en trabajo de investigación juicioso ni en "principios fundamentales"	Opinión de expertos sin evaluación crítica o basado en teoría económica o en "principios fundamentales"

GPC: Guía de práctica clínica. Estudios con homogeneidad: se refiere a que incluya estudios con resultados comparables y en la misma dirección.

la recomendación y el nivel de evidencia no es exacta; por ejemplo, puede haber un buen nivel de evidencia que no prueba que una intervención es efectiva, como el caso de la mamografía en mujeres menores de 50 años; que recibe una recomendación «C». Por otra parte una recomendación tipo «A» fue otorgada al Papanicolaou para detección precoz de cáncer cérvico-uterino basado en el «costo de sufrimiento de la enfermedad» y un nivel II de evidencia sosteniendo dicha intervención⁽¹⁴⁾. Así, las recomendaciones permiten un cierto grado de flexibilidad y se amoldan de acuerdo al contexto imperante.

Esta clasificación tiene la ventaja de que contempla nuevos aspectos que condicionan la elaboración de recomendaciones ya no sólo a la calidad de la evidencia (siempre controvertido) sino también al beneficio neto y al "costo de sufrimiento de padecer el evento de interés en estudio", aproximándose a una perspectiva más global que emerge desde la posición de quien padece la condición de salud o evento de interés en estudio. Dentro de las desventajas podemos señalar que sólo se consideran estudios para diagnóstico y, aún resultan insuficientes los aspectos abordados para establecer

las recomendaciones; por ejemplo no se considera la disponibilidad de los recursos necesarios para llevar a cabo los objetivos o metas señaladas. de cobertura de salud estatal.

CENTRE FOR EVIDENCE-BASED MEDICINE, OXFORD (OCEBM)⁽¹⁴⁾

Esta propuesta se caracteriza por valorar la evidencia según el área temática o escenario clínico y el tipo de estudio que involucra al problema clínico en cuestión. Lo anterior es una innovación y es complementaria a lo expuesto por las otras iniciativas. Esta, tiene la ventaja que gradúa la evidencia de acuerdo al mejor diseño para cada escenario clínico, otorgándole intencionalidad, agregando las RS en los distintos ámbitos. Así por ejemplo, al tratarse de un escenario clínico relacionado con pronóstico de un evento de interés, la evidencia será valorada a partir de una RS de estudios de cohortes con homogeneidad, o en su defecto de estudios de cohortes individuales con un seguimiento superior al 80% de la cohorte; en cambio, si el escenario se refiere a terapia o tratamiento, la evidencia se valorará principalmente a

partir de RS de EC, o en su defecto de EC individuales con intervalos de confianza estrechos.

Esta clasificación tiene la ventaja que nos asegura el conocimiento más relativo a cada escenario, por su alto grado de especialización. Además tiene la prerrogativa de aclarar cómo afecta la falta de rigurosidad metodológica al diseño de los estudios, disminuyendo su valoración no sólo en la gradación de la evidencia, sino que también en la fuerza de las recomendaciones.

Presenta algunos inconvenientes para su práctica habitual. Por una parte, vemos como en su estructura se presentan términos epidemiológicos poco amigables y con múltiples aclaraciones que hacen su lectura poco fluida y, que rápidamente pueden frustrar o desanimar a quien se aproxima a ella por primera vez. En su intento por abarcar todos los aspectos con la máxima exhaustividad, pierde la simpleza para hacerla aplicable (Tabla 8).

Tabla 9: Niveles de evidencia para estudio de tratamiento con análisis cuantitativo (SIGN)

Nivel de evidencia	Interpretación
1++	Meta-análisis de alta calidad, RS de EC o EC de alta calidad con muy poco riesgo de sesgo
1+	Meta-análisis bien realizados, RS de EC o EC bien realizados con pocos riesgo de sesgos
1-	Meta-análisis, RS de EC o EC con alto riesgo de sesgos
2++	RS de alta calidad de estudios de cohortes o de casos y controles. Estudios de cohortes o de casos y controles con riesgo muy bajo de sesgo y con alta probabilidad de establecer una relación causal
2+	Estudios de cohortes o de casos y controles bien realizados con bajo riesgo de sesgo y con una moderada probabilidad de establecer una relación causal
2-	Estudios de cohortes o de casos y controles con alto riesgo de sesgo y riesgo significativo de que la relación no sea causal
3	Estudios no analíticos, como informes de casos y series de casos
4	Opinión de expertos

Los estudios clasificados como 1- y 2- no deben usarse en el proceso de elaboración de recomendaciones por su alto potencial de sesgo.

Tabla 10: Grados de recomendación para estudios de tratamiento con análisis cuantitativo (SIGN)

Grados de recomendación	Interpretación
A	Al menos un meta-análisis, RS o EC clasificado como 1++ y directamente aplicable a la población diana de la guía; o un volumen de evidencia científica compuesto por estudios clasificados como 1+ y con gran consistencia entre ellos
B	Un volumen de evidencia científica compuesta por estudios clasificados como 2++, directamente aplicable a la población diana de la guía y que demuestran gran consistencia entre ellos; o evidencia científica extrapolada desde estudios clasificados como 1++ o 1+
C	Un volumen de evidencia científica compuesta por estudios clasificados como 2+ directamente aplicables a la población diana de la guía y que demuestran gran consistencia entre ellos; o evidencia científica extrapolada desde estudios clasificados como 2++
D	Evidencia científica de nivel 3 o 4; o evidencia científica extrapolada desde estudios clasificados como 2+

SCOTTISH INTERCOLLEGIATE GUIDELINES NETWORK (SIGN)^(16,17)

Esta propuesta, se origina también teniendo como foco de interés la temática del tratamiento. Se diferencia de las anteriores por su particular énfasis en el análisis cuantitativo que involucra a las RS y otorga importancia a la reducción del error sistemático. Se compone de niveles de evidencia y grados de recomendación según esos niveles (Tablas 9 y 10). Como fortaleza, es interesante el hecho que considera la calidad metodológica de los estudios que componen las RS, situación que es de interés dada la alta producción anual de revisiones. Como debilidad podemos señalar que no considera en la elaboración de las recomendaciones la realidad científica y tecnológica del momento, pues estas se crean con una rigidez que puede ser peligrosa en el sentido de quienes usan con ortodoxia las recomendaciones para la implementación de políticas de salud, los que pueden evitar invertir con argumento en la evidencia. Por otro lado, se basa de forma puntual en los aspectos metodológicos y de diseño, pero no así en la dimensión de la perspectiva del

padecer una enfermedad o considerar las implicaciones económicas de las medidas recomendadas.

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE (NICE)⁽¹⁸⁾

Esta iniciativa que nace del National Health Service del Reino Unido (NHS), abarca la temática de la terapia y el diagnóstico. Adapta la clasificación hecha por SIGN para terapia y utiliza la de la OCEBM para diagnóstico; de tal modo que se efectúa una valoración de la evidencia disponible con base en estas dos herramientas. Por lo tanto, queda patente la semejanza con las clasificaciones anteriores respecto a la importancia otorgada a las RS al momento de generar las recomendaciones (Tablas 11-14). Presenta el inconveniente que hace una relación tan directa entre calidad de la evidencia y grado de recomendación que puede generar confusión sobre esos constructos.

DISCUSIÓN

La evidencia ha tenido un desarrollo a lo largo del tiempo marcado por el interés de la comunidad cientí-

Tabla 11: Niveles de evidencia para estudios terapia (NICE)

Nivel de evidencia	Interpretación
1++	Meta-análisis de gran calidad, RS de EC con asignación aleatoria o EC con asignación aleatoria con muy bajo riesgo de sesgos
1+	Meta-análisis de gran calidad, RS de EC con asignación aleatoria o EC con asignación aleatoria con bajo riesgo de sesgos
1-	Meta-análisis de gran calidad, RS de EC con asignación aleatoria o EC con asignación aleatoria con alto riesgo de sesgos*
2++	RS de alta calidad de estudios de cohortes o de casos-controles, o estudios de cohortes o de casos-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-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 sesgo*
3	Estudios no analíticos, como informe de casos y series de casos
4	Opinión de expertos

*Los estudios con un nivel de evidencia “-” no deberían utilizarse como base para elaborar una recomendación. Adaptado de Scottish Intercollegiate Guidelines Network.

fica por ordenar y valorar de una manera exhaustiva el conocimiento, siendo uno de los objetivos primordiales el contestar preguntas clínicas surgidas del quehacer médico diario. En principio, la idea principal fue dar directrices

para que el conocimiento más válido sea conseguido de manera expedita por el clínico. Paradójicamente, el desarrollo de múltiples propuestas de clasificar la evidencia y formular recomendaciones y el lenguaje epidemiológico

Tabla 12: Grados de recomendación para estudios de tratamiento con análisis cuantitativo (SIGN)

Grados de recomendación	Interpretación
A	Al menos un meta-análisis o un EC con asignación aleatoria categorizados como 1++, que sea directamente aplicable a la población diana; o una RS o un EC con asignación aleatoria o un volumen de evidencia con estudios categorizados como 1+, que sea directamente aplicable a la población diana y demuestre consistencia de los resultados. Evidencia a partir de la apreciación de NICE
B	Un volumen de evidencia que incluya estudios calificados de 2++, que sean 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 volumen de evidencia que incluya estudios calificados de 2+, que sean directamente aplicables a la población objeto y que demuestren globalmente consistencia de los resultados, o extrapolación de estudios calificados como 2++
D	Evidencia nivel 3 o 4, o extrapolación de estudios calificados como 2+, o consenso formal

D (DPP): Un buen punto de práctica (BPP) es una recomendación para la mejor práctica basado en la experiencia del grupo que elabora la guía. IP: Recomendación a partir del manual para procedimientos de intervención de NICE.

Tabla 13: Niveles de evidencia para estudios diagnóstico (NICE)

Grados de recomendación	Interpretación
Ia	RS con homogeneidad* de estudios de nivel 1+
Ib	Estudios de nivel 1+
II	Estudios de nivel 2++ RS de estudios de nivel 2
III	Estudios de nivel 3&& RS de estudios de nivel 3
IV	Consenso, informes de comités de expertos u opiniones y/o experiencia clínica sin valoración crítica explícita; o en base a la psicología, difusión de la investigación o “principios básicos”

* **Homogeneidad** significa que no hay variaciones o estas son pequeñas en la dirección y grado de los resultados entre los estudios individuales que incluye la RS. + **Estudios de nivel 1** son aquellos que utilizan una comparación enmascarada de la prueba con un estándar de referencia validado, en una muestra de pacientes que refleja a la población a quien se aplicaría la prueba. ++ **Estudios nivel 2** son aquellos que presentan una sola de estas características: población reducida (la muestra no refleja las características de la población a la que se le va a aplicar la prueba; utilizan un estándar de referencia pobre (definido como aquel donde la “prueba” es incluida en la “referencia”, o aquel en que las “pruebas” afectan a la “referencia”); la comparación entre la prueba y la referencia no está enmascarada; o estudios de casos y controles. && **Estudios de nivel 3** son aquellos que presentan al menos dos o tres de las características señaladas anteriormente.

empleado, ha llevado consigo confusión e incompreensión por parte de los clínicos, quienes ven en esta gran variedad de opciones más conflictos que ayuda al desarrollo de la práctica profesional. En este artículo hemos presentado y desarrollado de forma somera algunas de las clasificaciones existentes, utilizando como criterio de selección, el nivel de utilización de cada cual, el que se asocia directamente con el grado de aceptación de ellas por parte de la comunidad científica.

El problema de la jerarquización de la evidencia también ha planteado desafíos a los comités editoriales de las revistas biomédicas respecto de cómo estandarizar la valoración de la evidencia, para la aceptación de trabajos científicos. Entendemos la necesidad de simplificar y aunar criterios, incluso sería deseable que las recomendaciones que nacen de la valoración de la evidencia tuvieran un carácter universal, pero el generalizar también trae consigo algunos inconvenientes. ¿Por qué hay políticas de salud que fracasan aún cuando han contemplado en su elaboración la evidencia disponible? Una de las tantas razones es que algunas recomendaciones basadas en la evidencia pueden no ser aplicables a nuestro medio. De esta forma, al momento de acercarnos a la evidencia es necesario considerar la existencia de características propias de nuestra región, tanto poblacionales, culturales, económicas, tecnológicas como ambientales; es decir, darle relevancia al concepto de validez externa y valorar nuestra realidad.

Tabla 14: Grados de recomendación para estudios diagnóstico (NICE)

Grados de recomendación	Interpretación
A (EPD)	Estudios con un nivel de evidencia Ia o Ib
B (EPD)	Estudios con un nivel de evidencia II
C (EPD)	Estudios con un nivel de evidencia III
D (EPD)	Estudios con un nivel de evidencia IV

* EPD: estudios de pruebas diagnósticas.

Por otro lado, la enseñanza de las ciencias médicas no incentiva la validación de instrumentos extranjeros para la valoración de la lectura crítica, ni menos la creación de los propios. Así, nos vemos obligados no sólo a traducir el lenguaje epidemiológico desde el inglés, sino que también a adaptar e implementar esas traducciones a las problemáticas sanitarias nacionales e incluso regionales.

Es cierto que el análisis de la evidencia incluye, la adaptación a diferentes realidades; pero, complicando un poco más la contingencia ¿qué repercusiones podemos tener en el ámbito legal? En la actualidad vamos observando cómo la *lex artis* ya no se basa en opiniones de expertos, cada día se hace más próxima a la MBE, lo cual implica la contemplación de recomendaciones y protocolos en el momento de fallos judiciales. Necesitamos establecer recomendaciones basadas en estudios propios. No podemos pretender que las recomendaciones establecidas por científicos con distinto grado de desarrollo tecnológico y de habilidades, puedan ser entendidas por quienes ejecutan las leyes sin el riesgo de creerlas como determinantes.

En nuestro entorno, nos preguntamos: ¿Cuál de las propuestas de clasificación de la evidencia es más adecuada para nosotros? Desde nuestro punto de vista, destacamos la clasificación ofrecida por OCEBM⁽¹⁴⁾ sobre las otras, por ser capaz de asignar una valoración más completa de la evidencia según cada tipo de escenario. Sin embargo, es absolutamente imprescindible entender que las recomendaciones, son consejos desde la más alta perspectiva científica, no importa qué propuesta usemos sino cómo las empleemos, qué juicios hagamos y cómo interrelacionamos la evidencia con los factores propios de nuestro entorno.

De todos modos no hay que olvidar que el ejercicio de la medicina y la teoría médica, siempre se han basado en la observación y en lo que hoy llamamos “evidencia”. Hace 2.500 años Hipócrates⁽¹⁹⁾ escribió: “[...] No es posible sacar provecho de lo que se cumple sólo de palabra, sino de lo que llega a la demostración de la realidad... Por ello es menester atenerse por completo a los hechos y ocuparse de ellos, y no con la mayor brevedad, si es que se ha de poseer la actitud holgada e infalible que titulamos, precisamente, hábito médico”.

BIBLIOGRAFÍA

1. Sackett DL, Rosenberg WMC, Gary JAM, Haynes RB, Richardson WS. Evidence based medicine: what is it and what it isn't. *BMJ* 1996;312:71-2.
2. Canadian Task Force on Preventive Health Care. New grades for recommendations from the Canadian Task Force on Preventive Health Care. *CMAJ* 2003; 169: 207-220
3. Upshur RE. Are all evidence-based practices alike? Problems in the ranking of evidence. *CMAJ* 2003; 169: 672-73
4. Evidence-based medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992; 268: 2420-25.
5. Manterola C, Zavando D. Como interpretar los "Niveles de Evidencia" en los diferentes escenarios clínicos. *Rev Chil Cir* 2009; vol 61, Nº 6: 582-95.
6. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Medicina Basada en la Evidencia. Como ejercer y enseñar la MBE.* Churchill Livingstone España. Madrid.1997.
7. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, et al. Systems to Rate the Strength of Scientific Evidence. Health Services/Technology Assessment Text, National Library of Medicine. AHRQ Publication No. 02-E016, 2002. Disponible en: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat1.chapter.70996>.
8. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Diseño de Investigaciones clínicas.* Traducción de M^a Jesus del Sol Jaquotot. 3^a ed. Lippincott Williams and Wilkins. Barcelona, 2008.
9. 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;274:1800-04.
10. Laupacis A, Sackett DL, Roberts RS. Assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318: 1728-33.
11. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1992; 102 Supl 305-11.
12. Task Force Revitalization Process (CTFPHC). Evidence-Based Clinical Prevention, Updated August 17, 2005. Disponible en: <http://www.ctfphc.org>.
13. Sackett DL, Wennberg JE. Choosing the best research design for each question. *BMJ* 1997; 315(7123): 1636-40.
14. Oxford Centre for Evidence-based Medicine (CEBM). Centre for Evidence Based Medicine - Levels of Evidence (March 2009). Disponible en: <http://www.cebm.net/index.aspx?o=1025>
15. Task Force Ratings. Guide to Clinical Preventive Services, Second Edition. Disponible: <http://www.odphp.osophs.dhhs.gov/pubs/guidecps/PDF/APPA.PDF>.
16. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323: 334-336
17. Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50, A guideline developer's handbook, revisited edition 2008. Disponible en: <http://www.sign.ac.uk/pdf/sign50.pdf>
18. National Institute for Health and Clinical Excellence (NICE). The guidelines manual 2009. Disponible en: http://www.nice.org.uk/media/5F2/44/The_guidelines_manual_2009.
19. *Tratados Hipocráticos. Preceptos (2).* Madrid: Editorial Gredos; 1990. Tomo I. p. 310-1.