

# NIVELES DE EVIDENCIA Y GRADOS DE RECOMENDACIÓN



Lilisbeth Perestelo Pérez  
Julio José Galiano García  
Ana M. Perdomo Hernández  
M. Elisa de Castro Peraza



“Lo que caracteriza al hombre de ciencia no es la posesión del conocimiento o de verdades irrefutables, sino la búsqueda desinteresada e incesante de la verdad”

*Karl Popper*

ULL



Servicio  
Canario de la Salud



Escuela Universitaria  
de Enfermería  
Mrs. Sra. de Candelaria



## NIVEL DE EVIDENCIA

- La validez de la evidencia científica depende de la calidad y cantidad de estudios realizados en un área de investigación.
- Los estudios cuantitativos, especialmente los ensayos controlados aleatorizados (ECA), son los que aportan la evidencia científica más sólida.
- La replicación de estudios con una metodología similar aumenta el poder de la evidencia científica.

ULL



## NIVEL DE EVIDENCIA

- Existen diversas clasificaciones que diferencian jerárquicamente los distintos estudios, en función de lo apropiado del diseño para responder a cada pregunta de investigación planteada → *niveles de evidencia*
- Los niveles de evidencia son un continuo:
  - *En un extremo la máxima calidad y en el otro extremo los estudios de calidad más débil*

ULL



# CLASIFICACIÓN DE LA EVIDENCIA

*Centre for Evidence-Based Medicine, Oxford*

Estudios sobre tratamiento, prevención, etiología y complicaciones

Grado de recomendación	Nivel de evidencia	Fuente
A	1 a	Revisión sistemática de ECA, con homogeneidad, o sea que incluya estudios con resultados comparables y en la misma dirección.
	1 b	ECA individual (con intervalos de confianza estrechos)
	1 c	Eficacia demostrada por la práctica clínica y no por la experimentación
B	2 a	Revisión sistemática de estudios de cohortes, con homogeneidad, o sea que incluya estudios con resultados comparables y en la misma dirección.
	2 b	Estudio de cohortes individual y ensayos clínicos aleatorios de baja calidad (< 80% de seguimiento)
	2 c	Investigación de resultados en salud
	3 a	Revisión sistemática de estudios de casos y controles, con homogeneidad, o sea que incluya estudios con resultados comparables y en la misma dirección.
C	3 b	Estudios de casos y controles individuales
	4	Serie de casos y estudios de cohortes y casos y controles de baja calidad.

\*Si tenemos un único estudio con IC amplios o una revisión sistemática con heterogeneidad estadísticamente significativa, se indica añadiendo el signo (-) al nivel de evidencia que corresponda y la recomendación que se deriva es una D

ULL



## NIVEL DE EVIDENCIA

- Las revisiones sistemáticas y metanálisis de estudios experimentales de alta calidad proporcionan la mejor evidencia científica.
- Los estudios experimentales (ensayos controlados aleatorizados) son los estudios primarios que aportan la evidencia científica más sólida.
- Los metanálisis y revisiones de estudios cuasi-experimentales y de resultados también aportan evidencia científica sólida.
- Los estudios correlacionales, descriptivos y cualitativos aportan conocimientos iniciales, que sirven de base para los estudios cuasi-experimentales, experimentales y de resultado.

Los niveles de evidencia ayudan a determinar la calidad y validez de la evidencia disponible para la práctica

ULL



## NIVEL DE EVIDENCIA

El tipo de diseño de los estudios de investigación hace que éstos tengan diferentes probabilidades de incurrir en sesgos o errores sistemáticos que limiten su validez interna.

El diseño y la existencia de sesgos permiten establecer una jerarquía y una clasificación de los artículos científicos.

### CLASIFICACIÓN SEGÚN EL NIVEL DE EVIDENCIA

- Muchos sistemas de clasificación
- Similares pero no idénticos
- No hay consenso internacional
- Algunos más utilizados
- Condiciona el grado de recomendación.

ULL



## NIVELES DE EVIDENCIA PARA LA PREGUNTA DE INVESTIGACIÓN<sup>1</sup>

		Tipo de estudios			
Estudios terapéuticos: Investigación de los resultados del tratamiento		Estudios de pronóstico: Investigación del efecto de una característica del paciente en el desenlace de una enfermedad	Estudios diagnósticos: Investigación de una prueba diagnóstica	Análisis económico de toma de decisiones: Desarrollo de un modelo económico de toma de decisiones	
Nivel I	<ul style="list-style-type: none"> <li>• Ensayo clínico controlado (ECC) de alta calidad con diferencia estadísticamente significativa o sin ésta pero con intervalos de confianza estrechos</li> <li>• Revisión sistemática<sup>2</sup> (metaanálisis) de ECC Nivel I (y homogeneidad de los resultados<sup>3</sup>)</li> </ul>	<ul style="list-style-type: none"> <li>• Estudio prospectivo de alta calidad<sup>4</sup> (todos los pacientes fueron enrolados en el mismo punto del curso de la enfermedad con <math>\geq 80\%</math> de seguimiento de los pacientes incluidos)</li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel I</li> </ul>	<ul style="list-style-type: none"> <li>• Prueba de criterios diagnósticos previamente desarrollados en pacientes consecutivos (con aplicación universal de un «estándar ideal» de referencia)</li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel I</li> </ul>	<ul style="list-style-type: none"> <li>• Costos y alternativas sensibles; valores obtenidos de muchos estudios; con análisis múltiple de sensibilidad</li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel I</li> </ul>	
Nivel II	<ul style="list-style-type: none"> <li>• ECC de menor calidad (Ej. <math>&lt; 80\%</math> seguimiento, sin cegamiento, o aleatorización inadecuada)</li> <li>• Estudio prospectivo<sup>4</sup> comparativo<sup>5</sup></li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel II o de estudios Nivel I con resultados inconsistentes</li> </ul>	<ul style="list-style-type: none"> <li>• Estudio retrospectivo<sup>6</sup></li> <li>• Controles no tratados de un ECC</li> <li>• Estudio prospectivo de menor calidad (Ej. pacientes enrolados en diferentes puntos del curso de la enfermedad o <math>&lt; 80\%</math> de seguimiento)</li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel II</li> </ul>	<ul style="list-style-type: none"> <li>• Desarrollo de criterios diagnósticos en pacientes consecutivos (con aplicación universal de un «estándar ideal» de referencia)</li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel II</li> </ul>	<ul style="list-style-type: none"> <li>• Costos y alternativas sensibles; valores obtenidos de un número limitado de estudios; con análisis múltiple de sensibilidad</li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel II</li> </ul>	
Nivel III	<ul style="list-style-type: none"> <li>• Estudio de casos y controles<sup>7</sup></li> <li>• Estudio retrospectivo<sup>6</sup> y comparativo<sup>5</sup></li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel III</li> </ul>	<ul style="list-style-type: none"> <li>• Estudio de casos y controles<sup>7</sup></li> </ul>	<ul style="list-style-type: none"> <li>• Estudio de pacientes no consecutivos; sin aplicación consistente de un «estándar ideal» de referencia</li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel III</li> </ul>	<ul style="list-style-type: none"> <li>• Análisis basado en alternativas y costos limitados; estimados deficientes</li> <li>• Revisión sistemática<sup>2</sup> de estudios Nivel III</li> </ul>	
Nivel IV	Series de casos <sup>8</sup>	Series de casos	<ul style="list-style-type: none"> <li>• Estudio de casos y controles</li> <li>• Estándar de referencia deficiente</li> </ul>	<ul style="list-style-type: none"> <li>• Análisis sin análisis de sensibilidad</li> </ul>	
Nivel V	Opinión del experto	Opinión del experto	Opinión del experto	Opinión del experto	

1. Una evaluación completa de estudios individuales requiere de una valoración crítica de todos los aspectos del diseño del estudio.

2. Una combinación de resultados de dos o más estudios previos.

3. Estudios que hayan brindado resultados consistentes.



## Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence

Question	Step 1 (Level 1*)	Step 2 (Level 2*)	Step 3 (Level 3*)	Step 4 (Level 4*)	Step 5 (Level 5)
<b>How common is the problem?</b>	Local and current random sample surveys (or censuses)	Systematic review of surveys that allow matching to local circumstances**	Local non-random sample**	Case-series**	n/a
<b>Is this diagnostic or monitoring test accurate?</b> (Diagnosis)	Systematic review of cross sectional studies with consistently applied reference standard and blinding	Individual cross sectional studies with consistently applied reference standard and blinding	Non-consecutive studies, or studies without consistently applied reference standards**	Case-control studies, or "poor or non-independent reference standard**	Mechanism-based reasoning
<b>What will happen if we do not add a therapy?</b> (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*	Case-series or case-control studies, or poor quality prognostic cohort study	n/a
<b>Does this intervention help?</b> (Treatment Benefits)	Systematic review of randomized trials or <i>n</i> -of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study**	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning
<b>COMMON harms?</b> (Treatment Harms)	Systematic review of nested case-control studies, <i>n</i> -of-1 trial with the patient you are raising the question about, or observational study with dramatic effect	or (exceptionally) observational study with dramatic effect	study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**	or historically controlled studies**	reasoning
<b>What are the RARE harms?</b> (Treatment Harms)	Systematic review of randomized trials or <i>n</i> -of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect			
<b>Is this (early detection) test worthwhile?</b> (Screening)	Systematic review of randomized trials	Randomized trial	Non-randomized controlled cohort/follow-up study**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning

\* Level may be graded down on the basis of study quality, imprecision, indirectness (study PICO does not match questions PICO), because of inconsistency between studies, or because the absolute effect size is very small; Level may be graded up if there is a large or very large effect size.

\*\* As always, a systematic review is generally better than an individual study.

### How to cite the Levels of Evidence Table

OCEBM Levels of Evidence Working Group\*. "The Oxford 2011 Levels of Evidence".

Oxford Centre for Evidence-Based Medicine. <http://www.cebm.net/index.aspx?o=5653>

\* OCEBM Table of Evidence Working Group = Jeremy Howick, Iain Chalmers (James Lind Library), Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard and Mary Hodgkinson

# NIVELES DE EVIDENCIA Y GRADOS DE RECOMENDACIÓN

- Los niveles de evidencia y los grados de recomendación permiten categorizar los estudios según su mayor o menor riesgo de presentar sesgos dado un determinado diseño.
- Tienen en cuenta aspectos como:
  - La calidad y la cantidad de estudios
  - La consistencia de los resultados
  - Los riesgos asociados
  - ...

ULL



# GRADOS DE RECOMENDACIÓN

- Un enfoque sistemático para calificar la fuerza de las recomendaciones puede minimizar los sesgos y ayudar a interpretar la evidencia científica.
- El sistema GRADE permite calificar la calidad de la evidencia y la fuerza de las recomendaciones.
- La evidencia científica o las recomendaciones pueden ser de diferente calidad.
- Las recomendaciones para realizar o no una intervención deberían estar basadas en el balance entre los beneficios por un lado y los riesgos, los inconvenientes y, potencialmente, los costes, por el otro.
- Si los beneficios superan a los riesgos e inconvenientes, los profesionales recomendarán el tratamiento a los pacientes.
- La incertidumbre asociada con el balance entre los beneficios, riesgos e inconvenientes, será la que determine la fuerza de las recomendaciones.

ULL





## FAQ

[Learn more](#)

[FAQ !\[\]\(23d9fc146e83b5c3013cfa32c784f8d5\_img.jpg\)](#)

[Organizations](#)

[Downloads](#)

[Courses](#)

[About us](#)

### Preguntas más frecuentes

[expand all](#) [collapse all](#) [only allow one expanded element](#)

#### ▼ ¿Qué es el grupo de trabajo GRADE?

El grupo de trabajo GRADE comenzó en el año 2000 como una colaboración informal de personas interesadas en abordar las deficiencias de los actuales sistemas de clasificación en la atención sanitaria. Nuestro objetivo es desarrollar un sistema común y razonable para calificar la calidad de la evidencia y la fuerza de las recomendaciones

#### ▼ ¿Por qué es importante evaluar las evidencias clínicas?

Las evidencias clínicas, o las recomendaciones que se basan en la evidencia, pueden ser de calidad diferente. Las fuentes de evidencia van desde los pequeños estudios de laboratorio o informes de series de casos a grandes estudios clínicos bien diseñados que han minimizado de manera importante los sesgos. Como la evidencia de calidad baja puede acabar en una recomendación que no responde a los intereses de los pacientes, es esencial conocer si una recomendación es fuerte (podemos confiar en la recomendación) o débil (no podemos confiar)

#### ▼ ¿Es algo novedoso calificar las evidencias y las recomendaciones?

Realmente no. Estos sistemas se han utilizado durante más de 25 años.

#### ▼ ¿Si existen muchos sistemas para calificar las evidencias y las recomendaciones, por qué necesitamos otro?

Porque existe la necesidad de disponer de un sistema único para evitar confusión. Este sistema debería evitar las deficiencias de los otros e incluir sus fortalezas. Algunos sistemas de clasificación están solamente basados en el tipo de diseño de estudio y no consideran explícitamente otros factores importantes que determinan la calidad de la evidencia. Algunos sistemas son excesivamente complejos. Un [análisis](#) de los actuales sistemas ha mostrado que éstas y otras deficiencias todavía no han sido abordadas adecuadamente por ninguno de ellos. [See: How GRADE compares to other systems](#)

#### ▼ De acuerdo. ¿Pero no deberían las personas implicadas en el desarrollo de los actuales sistemas trabajar con el grupo de trabajo GRADE?

Así es. De hecho, las personas que han desarrollado algunos de los sistemas de clasificación más utilizados han estado involucradas activamente en el desarrollo del sistema GRADE.

#### ▼ ¿Qué significa el acrónimo GRADE?

**TABLA  
1**

**Criterios GRADE para valorar la calidad de la evidencia**

Calidad de evidencia	Diseño de estudio	Disminuir si*	Aumentar si*
Alta	ECA	Importante (-1) o muy importante (-2) limitación de la calidad del estudio	Asociación fuerte, sin factores de confusión, consistente y directa (+1)**
Moderada		Inconsistencia importante (-1)	Asociación muy fuerte, sin amenazas importantes a la validez (no sesgos) y evidencia directa (+2)
Baja	Estudio observacional	Alguna (-1) o gran (-2)*** incertidumbre acerca de que la evidencia sea directa	
Muy baja	Cualquier otra evidencia	Datos escasos o imprecisos (-1) Alta probabilidad de sesgo de notificación (-1)	Gradiente dosis respuesta (+1) Todos los posibles factores confusores podrían haber reducido el efecto observado (+1)

\*1 = subir o bajar un nivel (por ejemplo, de alta a intermedia); 2 = subir o bajar dos niveles (por ejemplo, de alta a baja).

\*\*Un riesgo relativo estadísticamente significativo > 2 (< 0,5), basado en evidencias consistentes en 2 o más estudios observacionales, sin factores de confusión plausibles.

\*\*\*Un riesgo relativo estadísticamente significativo > 5 (< 0,2), basado en evidencia directa y sin amenazas importantes para la validez.

ULI



## NIVELES DE CALIDAD (GRADE)

### Calidad de la evidencia ALTA

Es muy poco probable que nuevos estudios modifiquen la confianza que tenemos en el resultado estimado

### Calidad de la evidencia MODERADA

Es probable que nuevos estudios tengan un gran impacto en la confianza que tenemos en el resultado estimado y éstos podrían modificar el resultado

### Calidad de la evidencia BAJA

Es muy probable que nuevos estudios tengan un gran impacto en la confianza que tenemos en el resultado estimado y modifiquen el resultado que tenemos

### Calidad de la evidencia MUY BAJA

Cualquier resultado estimado es incierto

ULL



## TABLA 2

### Decisiones acerca de la fuerza de las recomendaciones

#### Factores que pueden debilitar la fuerza de una recomendación

Calidad de la evidencia baja	Crear� una mayor incertidumbre sobre el tama�o (relativo) de los efectos (beneficios y da�os)
Incertidumbre acerca del balance entre beneficios y riesgos e inconvenientes	La incertidumbre sobre el riesgo basal, la prevalencia del problema o el estado de salud, que podr�an afectar el tama�o del efecto (absoluto)
Incertidumbre o diferencias en los valores	La incertidumbre sobre la importancia relativa de los beneficios e inconvenientes para los afectados, o diferencias en cuanto a lo importante para las distintas personas, lo que podr�a afectar al balance entre beneficios y riesgos e inconvenientes
Beneficios e inconvenientes netos marginales	Los beneficios esperados netos o los inconvenientes son peque�os (e inciertos)
Incertidumbre sobre si los beneficios netos valen el coste que suponen	La incertidumbre relacionada con la falta de informaci�n sobre el coste o si el gasto de los recursos est�a justificado para el beneficio esperado

ULL



Marzo Castillejo M, Monta o Barrientos A. El sistema GRADE para la toma de decisiones cl nicas y la elaboraci n de recomendaciones y gu as de pr ctica cl nica. Aten Primaria. 2007;39:457-60.

#### Recomendaciones

- Fuertes
- D biles



# GRADOS DE RECOMENDACIÓN

## Recomendación Fuerte

En base a la evidencia científica disponible, si los beneficios son claramente superiores a los riesgos e inconvenientes, o si contrariamente no son superiores.

## Recomendación Débil

En base a la evidencia científica disponible, si los beneficios, los riesgos y los inconvenientes mantienen un equilibrio estrecho, o existe incertidumbre sobre la magnitud de los beneficios y riesgos.

ULL





# FACTORES QUE INFLUYEN EN LAS RECOMENDACIONES

- La incertidumbre en las estimaciones del beneficio y riesgo, los inconvenientes y los costes.
- Importancia de la variable de resultado que el tratamiento previene.
- Magnitud del efecto del tratamiento.
- Precisión de la estimación del efecto del tratamiento.
- Riesgos asociados al tratamiento.
- Inconvenientes del tratamiento.
- Riesgo basal.
- Costes.
- Valores y preferencias de los pacientes y/o familiares.

ULL



# NIVELES DE EVIDENCIA Y GRADOS DE RECOMENDACIÓN



Lilisbeth Perestelo Pérez  
Julio José Galiano García  
Ana M. Perdomo Hernández  
M. Elisa de Castro Peraza