

Revisiones sistemáticas, paso a paso: una guía práctica sobre metodología e interpretación de resultados

Jose A Orozco

Área de Investigación en Medicina Interna, Servicio de Clínica Médica, Hospital Italiano de Buenos Aires, Argentina

Delfina Cirelli

Área de Investigación en Medicina Interna, Servicio de Clínica Médica, Hospital Italiano de Buenos Aires, Argentina

Las revisiones sistemáticas son estudios de investigación secundaria fundamentales para la medicina basada en la evidencia: identifican, evalúan y sintetizan de manera explícita y reproducible la evidencia disponible para responder una pregunta clínica concreta. Este artículo describe la arquitectura y el proceso metodológico de una revisión sistemática, brindando al lector una visión panorámica de sus etapas para facilitar la comprensión de sus componentes antes de abordar su evaluación crítica.

Introducción

Las revisiones sistemáticas (RS) son estudios que buscan, evalúan y resumen toda la evidencia disponible para responder una pregunta de investigación¹. Normalmente, parten de una pregunta en formato PICO (acrónimo de *Población, Intervención, Comparación y Outcome*), establecen criterios de inclusión y exclusión, y realizan una búsqueda bibliográfica exhaustiva en varias fuentes².

Si bien en este artículo nos centramos en las revisiones sistemáticas, también existen otros tipos de revisiones como las narrativas, que ofrecen información general sobre un tema. Estas no detallan cómo se buscaron o seleccionaron los estudios, ni incluyen un análisis crítico o una síntesis reproducible de los resultados. Son útiles para introducirse o actualizarse en una temática, pero no son confiables para la toma de decisiones clínicas³.

El propósito de este artículo es describir la arquitectura y el proceso metodológico de una revisión sistemática, brindando al lector una visión panorámica de sus etapas para facilitar la comprensión de sus componentes antes de abordar su evaluación crítica.

¿Por qué leer revisiones sistemáticas?

En la práctica clínica, cuando surge una pregunta sobre tratamiento, diagnóstico o pronóstico de una enfermedad, la fuente de información más útil suele ser una RS más que un estudio aislado. Un solo estudio puede estar sesgado, tener una muestra pequeña o no ser replicable. Por otro lado, buscar, leer y evaluar todos los estudios por cuenta propia requeriría semanas o meses, tiempo que rara vez tenemos disponible².

Para superar esta dificultad, la alternativa ideal es consultar una RS, diseño que sintetiza de forma rigurosa toda la evidencia relevante. Si esta además incluye un metanálisis (MA) —una técnica estadística que combina resultados para obtener un estimador único del efecto—, se aumenta la

precisión de las estimaciones de estudios individuales y se facilita la toma de decisiones clínicas. Pero no todos los datos pueden o deben resumirse mediante MA: si los resultados no son numéricamente compatibles o los estudios son clínicamente heterogéneos (diferencias en pacientes, intervenciones o desenlaces), la síntesis narrativa es la alternativa más adecuada⁴.

Al integrar datos de distintas poblaciones y contextos, las conclusiones de las RS con MA son en general más robustas y aplicables a los pacientes que atendemos en la práctica cotidiana.

Tipos de revisiones sistemáticas

Las RS se clasifican según el tipo de pregunta de investigación que nos planteamos¹:

- **Revisiones de intervenciones.** Reúnen estudios sobre los efectos de una intervención en los desenlaces de interés. Ejemplo: ¿El uso de aspirina en prevención primaria reduce eventos cardiovasculares en adultos sin enfermedad cardiovascular previa?
- **Revisiones de pruebas diagnósticas.** Sintetizan la evidencia sobre el rendimiento de pruebas diagnósticas frente a un estándar de referencia, informando sensibilidad, especificidad, y valores predictivos. Ejemplo: ¿El cuestionario PHQ-9 es una herramienta válida para detectar depresión en atención primaria?
- **Revisiones de factores pronósticos.** Analizan variables que influyen en la probabilidad de ocurrencia o progresión de un desenlace. Ejemplo: ¿Cuál es el riesgo de que la intolerancia a la glucosa progrese a diabetes en población seguida en atención primaria?

Etapas de una revisión sistemática

Para ilustrar cada una de las fases de manera práctica, tomaremos como hilo conductor una revisión sistemática sobre la eficacia y seguridad de la vacuna contra el virus herpes zóster, publicada por Bengolea et al. en la revista Medicina (Buenos Aires)⁵.

1. Definición de la pregunta de investigación

La pregunta guía los criterios de inclusión de los estudios que serán incorporados en la revisión, y el formato PICO es una herramienta práctica para estructurarla⁶.

- Población de interés (P): pacientes adultos sanos y pacientes adultos de alto riesgo (inmunodeprimidos, pacientes trasplantados, VIH/SIDA, pacientes oncológicos y aquellos tratados con corticoides de forma crónica)
- Intervención o exposición bajo estudio (I): vacuna recombinante de herpes zóster
- Comparador o alternativa (C): placebo
- Outcome o desenlace (O): eficacia (incidencia de infección por herpes zóster, incidencia de neuralgia post herpética, intensidad del dolor medida con escalas específicas) y seguridad (eventos adversos)

Algunos autores añaden un quinto elemento, el tipo de estudio (T), que especifica el diseño más adecuado para responder la pregunta, como por ejemplo, los ensayos clínicos para las preguntas de intervención¹. Una definición clara de estos elementos facilita la búsqueda y la selección de estudios.

2. Búsqueda y selección de artículos

Las RS buscan reunir todos los estudios relevantes que respondan a la pregunta de investigación. Para lograrlo, la búsqueda bibliográfica debe ser lo más completa posible y, además, quedar bien documentada: qué bases de datos se consultaron, en qué fechas y con qué estrategias. Habitualmente se utilizan fuentes como Medline, Embase o CENTRAL, pero también es importante

incluir la literatura gris. La estrategia de búsqueda se construye a partir de la pregunta de investigación, con una combinación de términos estandarizados para cada base de datos (como MeSH en Medline o Emtree en Embase) y términos de lenguaje libre cuando es necesario¹.

Una vez realizada la búsqueda, comienza el proceso de selección. Primero se hace un tamizaje inicial revisando títulos y resúmenes. Luego, los estudios potencialmente relevantes pasan a una segunda etapa, en la que se evalúa el texto completo. Este proceso debe realizarse por duplicado y de forma independiente. Cuando hay desacuerdos, éstos se resuelven por consenso o con la ayuda de un tercer revisor. En algunos casos, si falta información clave, se puede contactar a los autores del estudio para solicitarla².

3. Evaluación del riesgo de sesgo

No todos los estudios tienen la misma calidad y es de gran importancia tenerlo en cuenta. Evaluar el riesgo de sesgo significa analizar hasta qué punto los resultados de cada estudio pueden desviarse del efecto real de la intervención. Esto es fundamental porque las conclusiones de la revisión dependen directamente de los estudios incluidos: si la evidencia está sesgada, las conclusiones también pueden estarlo⁶. Al igual que en la selección de estudios y la extracción de datos, la evaluación del riesgo de sesgo debe realizarse por duplicado y de manera independiente, resolviendo las discrepancias por consenso o con la participación de un tercer revisor¹.

La herramienta utilizada para evaluar este riesgo debe adecuarse al tipo de estudio y a la pregunta de investigación. Por ejemplo, para preguntas sobre terapéutica suelen priorizarse los ensayos clínicos aleatorizados. En estos casos, es habitual utilizar herramientas como RoB 2⁷. Para estudios no aleatorizados de intervenciones, ROBINS-I⁸; para estudios de factores pronósticos, QUIPS⁹; para estudios de precisión diagnóstica, QUADAS-2¹⁰; y para modelos de predicción, PROBAST¹¹.

4. Extracción y síntesis de resultados

Cuando hablamos de **datos** nos referimos a toda información relevante para la revisión, por lo que resulta crucial que los autores planifiquen desde el inicio qué datos van a necesitar y cómo los van a recolectar¹.

La extracción se realiza mediante un formulario estandarizado, donde se registran las características de los estudios, la población, las intervenciones o exposiciones, los desenlaces y los resultados numéricos. Para esto pueden utilizarse planillas de cálculo o *software* específicos, lo que ayuda a que el proceso sea ordenado y reproducible. Al igual que en la selección, esta etapa también se realiza por duplicado, con resolución de discrepancias por consenso o mediante un tercer revisor².

Una vez recolectados los datos, se realiza la síntesis de resultados, que implica analizar y combinar la información de los estudios incluidos. La síntesis puede ser cualitativa (una descripción narrativa) o cuantitativa (un metanálisis), dependiendo de cuán comparables sean los estudios. Los resultados del MA suelen graficarse en un diagrama de bosque (o *forest plot* traducido al inglés)¹.

Presentación visual de resultados: el diagrama de bosque

El *forest plot* es una herramienta gráfica muy útil porque permite ver los resultados de todos los estudios incluidos y el efecto global de una intervención frente a su control.

En este tipo de gráficos, el resultado de cada estudio se representa con un cuadrado (que indica el estimador de efecto) y una línea horizontal (su intervalo de confianza, generalmente del 95%). Cuanto más grande es el cuadrado, mayor es el peso del estudio en el resultado final². También se incluye una línea vertical que representa la ausencia de efecto. El resultado combinado del MA suele representarse con un diamante en la parte inferior del gráfico, cuyos extremos muestran el

intervalo de confianza del efecto global^{1,2}.

El análisis de la amplitud de los intervalos de confianza permite interpretar la precisión de los resultados. En primer lugar, la precisión estadística, se valora por la amplitud del intervalo y por si este atraviesa o no la **línea de no efecto** (p. ej., el valor 1 para medidas relativas como la razón de riesgos, 0 en diferencias de medias). Un intervalo estrecho que no la atraviesa indica una estimación más precisa desde el punto de vista estadístico, mientras que si la atraviesa, el resultado es compatible tanto con beneficio como daño. En segundo lugar, la precisión clínica hace referencia a si nuestra decisión cambiaría según el verdadero efecto se encontrara en el extremo inferior o superior del intervalo de confianza. Si ambos extremos conducen a la misma conducta, la estimación puede considerarse clínicamente precisa, aun cuando el intervalo sea relativamente amplio¹².

Por otro lado, la heterogeneidad refleja cuán diferentes son los resultados entre los estudios¹³.

Así, el *forest plot* permite ver rápidamente qué resultados tuvo cada estudio, cuán precisos son estos resultados, cuán diferentes son entre sí y cuál es el efecto global. La [Figura 1](#) presenta un ejemplo de *forest plot* correspondiente al desenlace incidencia de herpes zóster en la RS de Bengolea et al⁵. Esta figura muestra los resultados para el subgrupo de pacientes sanos y para el subgrupo de alto riesgo, así como la estimación del efecto global para este desenlace.

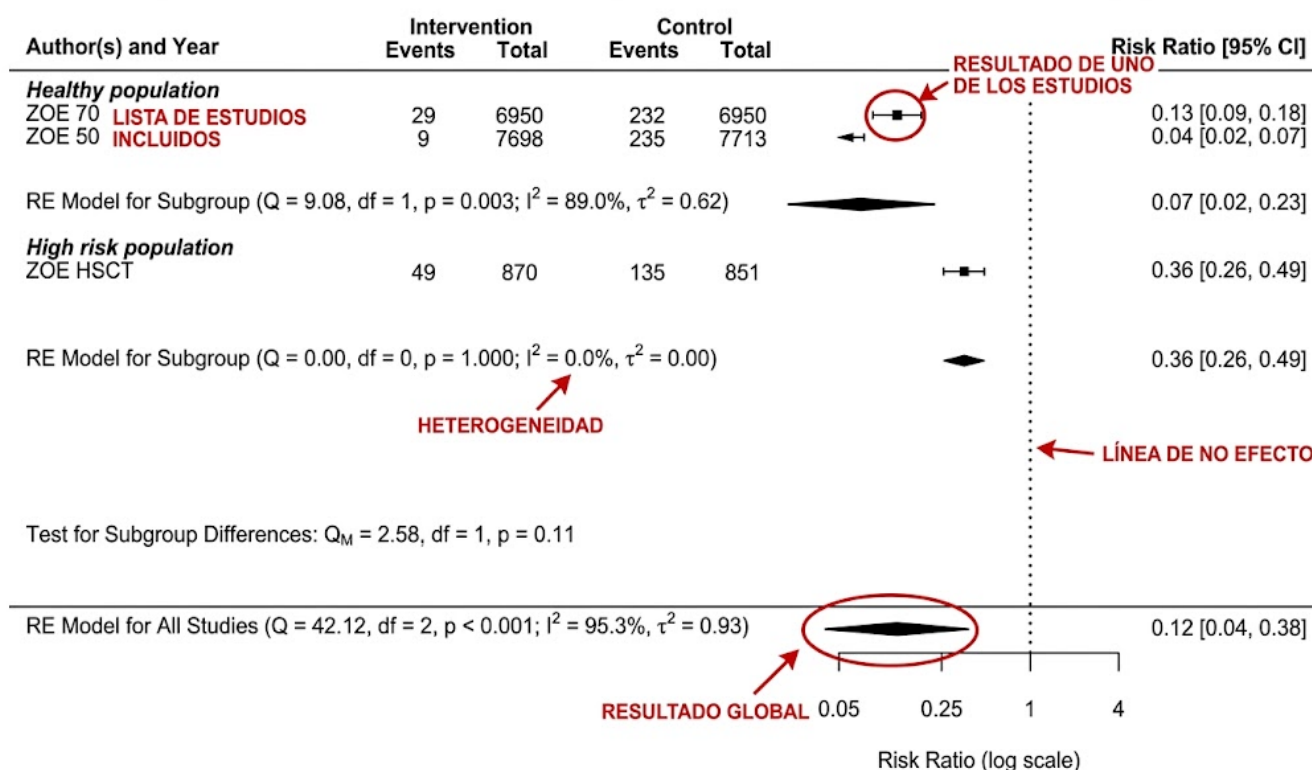


Figure 1. Forest plot que representa los efectos de la vacuna recombinante contra el herpes zóster sobre la incidencia de esta enfermedad en subgrupos de participantes sanos y de alto riesgo. Reproducido de la revisión sistemática de Bengolea et al., con autorización de los autores⁵. Las anotaciones en rojo son propias.

A continuación destacamos algunos aspectos del gráfico que son relevantes para su correcta interpretación. En primer lugar vemos que en el subgrupo de población sana el estudio “ZOE 70” estimó un riesgo relativo (RR) de 0,13 con un intervalo de confianza (IC) del 95% de 0,09 a 0,18. Este valor sugiere un beneficio de la vacuna (RR menor a 1), y dado que el intervalo no incluye el 1,

podemos afirmar que existe una diferencia significativa desde el punto de vista estadístico entre la vacuna y el placebo.

Al analizar el efecto combinado dentro de este subgrupo, observamos un RR de 0,07 con IC 95% 0,02 a 0,23. Si bien la dirección del efecto sigue favoreciendo a la intervención, el intervalo es más amplio, lo que indica menor precisión. Esto puede ocurrir cuando los estudios incluidos presentan resultados heterogéneos o con poca superposición entre sus intervalos de confianza^{1,13}. Finalmente, el resultado global muestra un RR de 0,12 con IC 95% 0,04 a 0,38. En este caso, la evidencia global sugiere un beneficio de la intervención.

En la parte inferior del gráfico también se informa la heterogeneidad. Un I^2 de 95% indica una heterogeneidad considerable, es decir una alta variabilidad entre los resultados de cada estudio individual. La heterogeneidad se puede apreciar también directamente en el gráfico, donde observamos que los intervalos de confianza de los estudios incluidos presentan escasa o nula superposición^{1,13}.

5. Evaluación de la certeza de la evidencia

Al llegar a los resultados, es importante preguntarse: ¿qué tan confiables son? La certeza de la evidencia refleja el grado de confianza que podemos tener en los resultados. Aunque una RS esté bien realizada, los estudios incluidos pueden tener limitaciones que afecten esa confianza. Por eso, es fundamental identificar y explicitar estas limitaciones al interpretar los resultados. El enfoque más utilizado para este propósito es el sistema GRADE ([Tabla 1](#))⁶.

Este sistema tiene en cuenta las siguientes características para determinar la certeza de la evidencia para cada desenlace en particular: el riesgo de sesgo, la inconsistencia, la evidencia indirecta, la imprecisión y el sesgo de publicación⁶.

Nivel de certeza	Definición
Alta	Existe alta confianza en que el efecto estimado es muy cercano al efecto real.
Moderada	Existe confianza moderada en el efecto estimado. Es posible que el efecto real sea diferente.
Baja	La confianza en el efecto estimado es limitada. El efecto real podría ser sustancialmente diferente.
Muy baja	La confianza en el efecto estimado es muy baja. Es muy probable que el efecto real sea sustancialmente diferente.

Table 1. Niveles de certeza de la evidencia propuestos por el sistema GRADE Fuente: elaboración propia a partir de Guyatt et al⁶.

6. Presentación y comunicación de hallazgos

En esta etapa se presentan los hallazgos de forma clara y útil para el lector. Algunos elementos clave que no deberían faltar son¹:

- Las medidas de efecto (en términos relativos) con sus intervalos de confianza.
- Los resultados en términos absolutos (número de eventos y participantes por grupo).
- La cantidad de estudios que aportan evidencia para cada desenlace.
- La certeza de la evidencia y las razones que la sustentan.

Una herramienta especialmente útil son las tablas de **resumen de hallazgos**, que organizan toda esta información de manera clara y accesible ([Figura 2](#))¹⁴.

Summary of Findings for la vacuna recombinante contra VZV Versus placebo for pacientes adultos sanos y pacientes adultos de alto riesgo

Outcome and follow-up	Patients (studies), N	Efecto relativo (95% CI)	Absolute effects (95% CI)			Certeza	Qué pasa
			placebo	la vacuna recombinante contra VZV	Diferencia		
Incidencia de herpes zóster en población sana	29311 (2 Experimentos controlados aleatorios [ECAs])	RR = 0.07 (0.02 a 0.23)	32 por 1000	2 por 1000 (1 a 7)	30 menos por 1000 (de 31 menos a 25 menos)	⊕⊕⊕⊖ Moderado ^a	La vacuna recombinante probablemente reduce la incidencia de herpes zóster en población sana.
Incidencia de herpes zóster en población de alto riesgo	1721 (1 ECA (experimento controlado aleatorizado))	RR = 0.36 (0.26 a 0.49)	159 por 1000	57 por 1000 (41 a 78)	102 menos por 1000 (de 117 menos a 81 menos)	⊕⊕⊕⊕ Alta	La vacuna recombinante reduce la incidencia de herpes zóster en población de alto riesgo.
Incidencia de herpes zóster en toda la población	31032 (3 Experimentos controlados aleatorios [ECAs])	RR = 0.12 (0.04 a 0.38)	39 por 1000	5 por 1000 (2 a 15)	34 menos por 1000 (de 37 menos a 24 menos)	⊕⊕⊕⊖ Moderado ^a	La vacuna recombinante probablemente reduce la incidencia de herpes zóster en toda la población.

CI: Intervalo de confianza ; RR: Razón de riesgo

a. Inconsistencia

Figure 2. Tabla de resumen de hallazgos que resume la evidencia sobre la efectividad de la vacuna recombinante contra herpes zóster sobre incidencia de esta enfermedad Fuente: elaboración propia con la herramienta GradePro GDT, a partir de la revisión de Bengolea et al.^{5,15}

Conclusiones

Las revisiones sistemáticas reúnen y analizan de forma rigurosa toda la evidencia disponible para responder una pregunta clínica específica, ofreciendo una visión más completa y confiable que los estudios individuales. Estas revisiones son fundamentales para la toma de decisiones informadas en salud, ya que sintetizan grandes volúmenes de información de manera crítica. Comprender este proceso metodológico estructurado y exigente permite al lector reconocer los elementos que garantizan la validez de sus conclusiones y tomar decisiones de salud mejor informadas.

Referencias

- Higgins JPT, Thomas J, Chandler J, et al., editores. Cochrane Handbook for Systematic Reviews of Interventions version 6.4 [Internet]. Cochrane; 2023 [cited 28 abr 2026]. Available from: <https://www.cochrane.org/handbook>
- Guyatt G, Rennie D, Meade MO, Cook DJ, editores. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 3rd ed. New York: McGraw-Hill Education; 2014. 697 p.
- Chambergo-Michilot D, Diaz-Barrera ME, Benites-Zapata VA. Revisiones de alcance, revisiones paraguas y síntesis enfocada en revisión de mapas: aspectos metodológicos y aplicaciones [Scoping reviews, umbrella reviews and focused mapping review synthesis: methodological aspects and applications]. Rev Peru Med Exp Salud Publica [Internet]. 2021 [cited 28 abr 2026];38(1):136-142. DOI: [10.17843/rpmesp.2021.381.6501](https://doi.org/10.17843/rpmesp.2021.381.6501) PMID: [34190906](https://pubmed.ncbi.nlm.nih.gov/34190906/)
- Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020 [cited 28 abr 2026];368:l6890. DOI: [10.1136/bmj.l6890](https://doi.org/10.1136/bmj.l6890) PMID: [31948937](https://pubmed.ncbi.nlm.nih.gov/31948937/)
- Bengolea A, Chamorro F, Ramos JT, et al. Effectiveness and safety of the recombinant herpes zoster vaccine in different population groups: a systematic review and meta-analysis. Medicina (B Aires) [Internet]. 2024 [cited 28 abr 2026];84(5):959-970. Available from: <https://www.medicinabuenosaires.com/revistas/vol84-24/n5/959.pdf> PMID: [39399936](https://pubmed.ncbi.nlm.nih.gov/39399936/)
- Guyatt G, Agoritsas T, Brignardello-Petersen R, et al. Core GRADE 1: overview of the Core GRADE approach. BMJ [Internet]. 2025 [cited 28 abr 2026];389:e081903.

DOI:[10.1136/bmj-2024-081903](https://doi.org/10.1136/bmj-2024-081903). PMID: [40262844](https://pubmed.ncbi.nlm.nih.gov/40262844/)

7. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ [Internet]. 2019 [cited 28 abr 2026];366:l4898. DOI: [10.1136/bmj.l4898](https://doi.org/10.1136/bmj.l4898). PMID: [31462531](https://pubmed.ncbi.nlm.nih.gov/31462531/)

8. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ [Internet]. 2016 [cited 28 abr 2026];355:i4919. DOI: [10.1136/bmj.i4919](https://doi.org/10.1136/bmj.i4919) PMID: [27733354](https://pubmed.ncbi.nlm.nih.gov/27733354/)

9. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. Ann Intern Med [Internet]. 2013 [cited 28 abr 2026];158(4):280-6. DOI: [10.7326/0003-4819-158-4-201302190-00009](https://doi.org/10.7326/0003-4819-158-4-201302190-00009) PMID: [23420236](https://pubmed.ncbi.nlm.nih.gov/23420236/)

10. Whiting PF, Rutjes AW, Westwood ME, et al.; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med [Internet]. 2011 [cited 28 abr 2026];155(8):529-36. DOI: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009) PMID: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)

11. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med [Internet]. 2019 [cited 28 abr 2026];170(1):W1-W33. DOI: [10.7326/M18-1377](https://doi.org/10.7326/M18-1377) PMID: [30596876](https://pubmed.ncbi.nlm.nih.gov/30596876/)

12. Guyatt G, Zeng L, Brignardello-Petersen R, et al. Core GRADE 2: choosing the target of certainty rating and assessing imprecision. BMJ [Internet]. 2025 [cited 28 abr 2026];389:e081904. DOI: [10.1136/bmj-2024-081904](https://doi.org/10.1136/bmj-2024-081904) PMID: [40300802](https://pubmed.ncbi.nlm.nih.gov/40300802/)

13. Guyatt G, Schandelmaier S, Brignardello-Petersen R, et al. Core GRADE 3: rating certainty of evidence-assessing inconsistency. BMJ [Internet]. 2025 [cited 28 abr 2026];389:e081905. DOI: [10.1136/bmj-2024-081905](https://doi.org/10.1136/bmj-2024-081905) PMID: [40328467](https://pubmed.ncbi.nlm.nih.gov/40328467/)

14. Guyatt G, Yao L, Murad MH, et al. Core GRADE 6: presenting the evidence in summary of findings tables. BMJ [Internet]. 2025 [cited 28 abr 2026];389:e083866. DOI: [10.1136/bmj-2024-083866](https://doi.org/10.1136/bmj-2024-083866) PMID: [40425239](https://pubmed.ncbi.nlm.nih.gov/40425239/)

15. GRADEpro GDT: GRADEpro Guideline Development Tool [software on the Internet]. Hamilton (ON): McMaster University & Evidence Prime; 2026 [cited 28 abr 2026]. Available from: <https://gdt.gradepro.org/app/>

Historial del artículo

Recibido 29/04/2026, Aceptado 05/08/2026 y Publicado 14/09/2026.

ORCID

Jose A. Orozco <https://orcid.org/0000-0001-7734-4133>

Delfina Cirelli <https://orcid.org/0000-0002-8143-8631>

Declaración de uso de inteligencia artificial

En la preparación de este manuscrito fue utilizada la herramienta ChatGPT para traducción de bibliografía del inglés al español y corrección gramatical del texto.

Fuentes de financiamiento / Conflicto de interés de los autores

Este artículo no contó con fuentes de financiamiento específicas. Los autores declararon no poseer conflictos de interés.

Declaración de revisión editorial

Artículo no evaluado por pares externos. Por sus características, este artículo fue evaluado por el Comité Editorial de Evidencia.

Licencia

[Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](#)

(CC BY-NC-SA 4.0)

Imagen de portada

Imagen de Mashiur Rahman bajo licencia de contenidos de Pixabay.