



## FORMACIÓN CONTINUADA - METODOLOGÍA Y TÉCNICAS

### Revisiones sistemáticas en cinco pasos: V. Cómo interpretar la evidencia



K.S. Khan<sup>a,b</sup>, A. Bueno Cavanillas<sup>a,b,\*</sup> y J. Zamora<sup>b,c,d</sup>

<sup>a</sup> Departamento de Medicina Preventiva y Salud Pública, Universidad de Granada, Granada, España

<sup>b</sup> CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, España

<sup>c</sup> Unidad de Bioestadística Clínica, Hospital Ramón y Cajal, Madrid, España

<sup>d</sup> Institute of Metabolism and Systems Research, Universidad de Birmingham, Birmingham, Reino Unido

Recibido el 5 de julio de 2022; aceptado el 17 de septiembre de 2022

Disponible en Internet el 18 de noviembre de 2022

#### PALABRAS CLAVE

Aplicabilidad;  
Generalización;  
Validez externa;  
Interpretación;  
Utilidad

**Resumen** El último paso en una revisión sistemática es interpretar la evidencia de forma que pueda ser útil para sustentar la toma de resultados. Es importante presentar los resultados de forma claramente diferenciada, asignar el nivel de fuerza de la evidencia que respalda cada uno de ellos teniendo en cuenta factores como el diseño de los estudios, su calidad metodológica y el sesgo de publicación. También es necesario considerar las variaciones en la fuerza de la asociación. El objetivo es detectar en qué tipo de pacientes la intervención analizada será más o menos efectiva, tendrá mayor o menor impacto una exposición determinada, o será de mayor utilidad una prueba diagnóstica. En esta fase, para facilitar la interpretación de los resultados, se puede añadir una estimación de la magnitud de la asociación, bien global o bien estratificada en función de las características de los participantes. Toda esta información ayudará a la hora de formular recomendaciones prácticas con utilidad clínica.

© 2022 Publicado por Elsevier España, S.L.U. en nombre de Sociedad Española de Médicos de Atención Primaria (SEMERGEN).

#### KEYWORDS

Applicability;  
Generalizability;  
External validity;  
Interpretation;  
Utility

#### Systematic reviews in five steps: V. Interpreting the findings

**Abstract** The last step in a systematic review is the interpretation of the findings. The important findings need to be explicitly identified. A level of strength of evidence should be assigned to support each key finding, based on factors such as study design, methodological quality and risk of publication bias. Variations in the magnitude of associations observed also need to be explored. The aim of this analysis is to determine in which clinical groups the intervention is more or less effective, the impact of exposure is greater or lesser, or a diagnostic test is

\* Autor para correspondencia.

Correo electrónico: [\(A. Bueno Cavanillas\).](mailto:abueno@ugr.es)

more useful. At this stage, for better interpretation of the findings, the magnitude of the association can be estimated either globally or stratified according to the characteristics of the participants. All this is helpful in formulating recommendations for clinical practice and policy. © 2022 Published by Elsevier España, S.L.U. on behalf of Sociedad Española de Médicos de Atención Primaria (SEMERGEN).

## Introducción

El objetivo final de una revisión sistemática es influir en la toma de decisiones para fundamentar la práctica clínica en la mejor evidencia científica disponible. La culminación del proceso de elaboración de una revisión sistemática, que se ha recorrido a lo largo de una serie de cinco artículos que finaliza con este último, es utilizar la evidencia científica, evaluada y sintetizada durante la revisión, para hacer recomendaciones sólidas que permitan tomar decisiones. Se tiende a pensar que con una revisión casi finalizada el significado de los resultados es obvio. No obstante, es necesario realizarse preguntas como: ¿Hasta qué punto son fiables los resultados? ¿La evidencia es realmente sólida? ¿Cómo se pueden generar conclusiones y recomendaciones para la práctica clínica en el futuro? En este paso se explorarán las alternativas de mayor transparencia y facilidad para ser extrapoladas, para determinar el grado o la fuerza de la evidencia. La finalidad es reforzar las recomendaciones que permitirán poner en práctica los frutos de la investigación.

## Fuerza de la evidencia

La fuerza de la evidencia depende de las fortalezas y debilidades de los distintos tipos de estudios. En una revisión sistemática, la fuerza se refiere al grado de certeza sobre la estimación de la magnitud con la que el efecto estudiado se asocia con los desenlaces relevantes. Para llegar a esta fase es necesario haber cumplido con los puntos clave expuestos en los pasos anteriores<sup>1-4</sup>. En otras palabras, hay que cerciorarse de que la evidencia responde a los componentes de la pregunta formulada; se ha llevado a cabo una búsqueda bibliográfica exhaustiva; se ha considerado tanto el diseño como la calidad de los estudios incluidos en la revisión; se ha explorado la heterogeneidad existente entre los efectos observados por los estudios individuales incluidos en la revisión; y por último, se han analizado los intervalos de confianza para evaluar la precisión o la imprecisión de los efectos<sup>5-9</sup>. Para una mejor comprensión del desarrollo de esta fase, se puede consultar la tabla 1, donde se encuentran recogidos los términos de mayor complejidad.

Durante el proceso de asignar un nivel de fuerza dado a la evidencia, *a priori* se tiende a otorgar un nivel alto a aquella que procede de estudios con un diseño experimental. Sin embargo, el análisis de elementos como la relevancia de la evidencia en relación con la pregunta, el sesgo de publicación, la calidad metodológica o la heterogeneidad de los resultados puede aumentar o disminuir el nivel de fuerza

asignado. Este análisis requiere un razonamiento adecuado, experiencia clínica y metodológica.

En ocasiones, está justificado que se disminuya el nivel de fuerza después de haber asignado un nivel alto. Es necesario, no obstante, revisar el nivel de fuerza para afianzar el resultado obtenido. Para ilustrar esta situación se puede tomar como ejemplo el uso de antibióticos en una otitis media, ilustrado en la tabla 2. Para el desenlace crítico «perforación del tímpano» se estimó una *odds ratio* (OR) = 0,51, lo que sugiere que las probabilidades de perforación timpánica, cuando la otitis se trata con antibióticos, se reducen a la mitad. Sin embargo, el intervalo de confianza osciló entre 0,20 y 1,26, esto supone que el verdadero efecto sobre la probabilidad de perforación puede oscilar entre una reducción del 80% (OR = 0,20) y un aumento del 26% (OR = 1,26), o simplemente no modificarse (OR = 1,0). Lo que realmente indica este resultado es que el efecto de los antibióticos sobre el riesgo de perforación es considerablemente incierto. Esta imprecisión es un motivo totalmente justificado para bajar el nivel de fuerza de la evidencia en un par de escalones.

## Presentación de la evidencia para facilitar su interpretación

Para aumentar la transparencia se recomienda crear una tabla que recoja, de forma resumida, la evidencia obtenida para cada uno de los desenlaces y/o grupos de estudio de interés. La fuerza de la evidencia debe evaluarse de forma independiente para cada desenlace, ya que puede variar entre ellos, incluso cuando procede de los mismos estudios. En el ejemplo de la otitis media (tabla 2), para el desenlace «existencia de dolor en un periodo de dos a siete días», la fuerza de la evidencia es alta, los participantes, las intervenciones y los desenlaces se refieren directamente a la pregunta inicial y no hay ninguna limitación en la calidad metodológica de los estudios. Además, el intervalo de confianza para la estimación puntual del efecto es significativo (no contiene el valor nulo) y no se detecta ninguna indicación de que exista sesgo de publicación. No obstante, para el desenlace «perforación del tímpano», además de la imprecisión del resultado, la calidad de la evidencia está limitada porque, entre otros, está sustentada por un ensayo de grandes dimensiones que excluyó del análisis a todos los pacientes que abandonaron antes de que el desenlace pudiera identificarse.

**Tabla 1** Glosario sobre términos del quinto paso en una revisión sistemática

Término	Definición
Diferencia de riesgos	Con frecuencia denominado riesgo atribuible. Se trata de la diferencia en las tasas de eventos entre dos grupos en un estudio comparativo. La inversa de la diferencia de riesgos indica el número necesario a tratar (NNT)
Fuerza de la evidencia	Muestra hasta qué punto es posible confiar en que los resultados son correctos
Heterogeneidad	El grado de diferencia en la magnitud del efecto entre estudios individuales. Para realizar un juicio clínico sobre la heterogeneidad hay que observar las diferencias entre participantes, intervenciones o exposiciones y desenlaces considerados en cada uno de los estudios
Número necesario para dañar (NNH)	Número de pacientes que deben ser tratados para que un paciente adicional experimente un efecto adverso o una complicación. Se calcula de la misma manera que el NNT
Número necesario a tratar (NNT)	Número de pacientes que deben ser tratados (o expuestos) para conseguir un desenlace positivo. Es una medida clínicamente intuitiva del impacto de una intervención
Odds ratio (OR)	Es la razón de probabilidades de un evento en el grupo experimental con respecto al grupo de control. Una OR de 1 indica que no hay diferencias entre los grupos de comparación. Una OR < 1 indica que la intervención reduce la probabilidad de que ocurra el desenlace, una OR > 1 indica que la intervención aumenta la probabilidad de que ocurra el desenlace
Riesgo basal	En inglés, <i>baseline risk</i> , se refiere a la frecuencia de un desenlace en ausencia de intervención. Está relacionado con la frecuencia del evento o desenlace de interés entre los sujetos de estudio. En el caso de que lo que se investigue sea el efecto de un tratamiento, un pronóstico favorable se asocia a un riesgo basal bajo, mientras que un pronóstico desfavorable se relaciona con un riesgo basal alto de resultados no deseados. El riesgo basal es importante para determinar qué tipo de pacientes se beneficiará más de la intervención
Riesgo relativo (RR)	Es la razón entre el riesgo en el grupo experimental y el riesgo en el grupo de control. Un RR de 1 indica que no hay diferencias entre los grupos de comparación. Un RR < 1 indica que la intervención reduce el riesgo de que ocurra el desenlace, un RR > 1 indica que la intervención aumenta el riesgo de que ocurra el desenlace
Sesgo de publicación	Surge cuando la importancia de los resultados de un estudio afecta a la probabilidad de que sea publicado

**Tabla 2** Resumen de los resultados de una revisión sobre la eficacia de antibióticos en otitis media entre los niños atendidos en atención primaria para desenlaces importantes

Desenlace	Odds ratio [IC 95%]	Riesgo basal (riesgo sin tratamiento)	Riesgo bajo tratamiento <sup>a</sup> [IC 95%]	Diferencia de riesgos [IC 95%]	NNT o NNH [IC 95%]
<b>Dolor 2-7 días</b> 9 ensayos	0,57 [0,45-0,73]	Medio: 260 por 1000 1000 (26%)	167 por 1000 [137-204]	93 menos por 1000 tratados [56-123]	11 [8-18]
<b>Perforación (evaluada con otoscopia o con examen de secreción ótica con un seguimiento de 7 días)</b> 2 ensayos	0,51 [0,2-1,26]	Bajo: 17 por 1000 (1,7%)	9 por 1000 [3-21]	8 menos por 1000 tratados [14 menos - 4 más]	125 [NNT: 250 - NNH: 71]
		Medio: 160 por 1000 (16%)	89 por 1000 [37-194]	71 menos por 1000 tratados [123 menos - 34 más]	14 [NNT: 29 - NNH: 8]
<b>Efectos adversos (vómitos, diarrea, erupción cutánea)</b> 4 ensayos	1,94 [1,28-2,94]	Bajo: 10 por 1000 (1%)	19 por 1000 [13-29]	9 más por 1000 tratados [3-19]	111 [52-333]
		Alto: 300 por 1000 (30%)	454 por 1000 [354-558]	154 más por 1000 tratados [54-258]	6 [3-18]

NNH: número necesario para dañar; NNT: número necesario a tratar.

<sup>a</sup> El riesgo bajo tratamiento se basa en el riesgo sin tratamiento y la odds ratio de la intervención, y se calcula con GRADEpro, que está disponible de forma gratuita en [www.gradeworkinggroup.org](http://www.gradeworkinggroup.org)Fuente: Tabla adaptada a partir del quinto capítulo del libro *Systematic Reviews to Support Evidence-Based Medicine*<sup>5</sup>.

## Funnel plots

### Simétricos

- Los efectos en estudios más pequeños (círculos abiertos) están más dispersos que en los estudios más grandes (círculos negros), y aparecen a ambos lados de los efectos observados en los estudios más grandes.
- Los efectos adquieren forma de embudo.

### Asimétricos

- Parece que faltan estudios más pequeños con efectos menores en comparación con los observados en estudios más grandes.
- El embudo parece truncado.

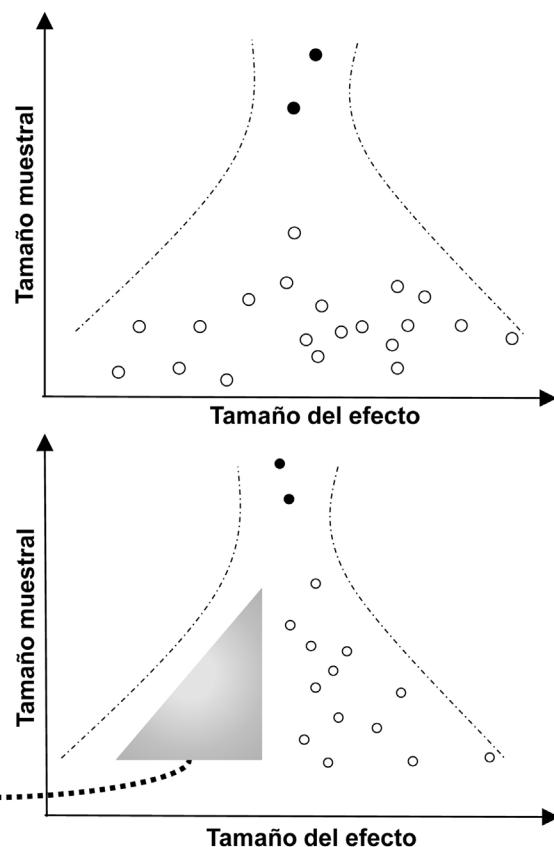


Figura 1 Ejemplo de un gráfico en chimenea o *funnel plot*.

## Aplicabilidad de la evidencia

A estas alturas de la revisión ya se sabrá si es posible o no tener suficiente confianza en sus resultados y se conoce el rango y magnitud de los beneficios esperados. Sin embargo, todavía se necesita un esfuerzo adicional para evaluar la aplicabilidad de la evidencia.

Las medidas relativas del efecto, como el riesgo relativo (RR) o la OR, son muy útiles para sintetizar la evidencia en un metaanálisis y evaluar la fuerza del efecto. Sin embargo, se necesita una estimación de la magnitud absoluta de los efectos, ajustada a los grupos específicos de participantes, para determinar si una intervención merece o no la pena. El efecto absoluto puede expresarse de modo fraccionario como la «diferencia de riesgos» (o riesgo atribuible). El inverso de la diferencia de riesgos convierte esta fracción en un número natural llamado «número necesario a tratar» (NNT), indica el número de sujetos que deben recibir la intervención para obtener un desenlace positivo. Cuando se refiere a los efectos adversos, el mismo cálculo pasa a llamarse «número necesario para dañar» (NNH, por sus siglas en inglés). Cuando el intervalo de confianza estimado para la diferencia de riesgos no alcanza la significación estadística, el NNT oscilaría entre un número positivo y otro negativo, que no tendría sentido y que se interpreta como NNH ([tabla 2](#)).

El NNT es absolutamente dependiente del riesgo inicial, por lo que permite individualizar el beneficio de las intervenciones. Cuanto más alto sea el NNT, mayor será el número de

pacientes que deberán someterse a tratamiento para conseguir un efecto beneficioso en uno de ellos. Por lo tanto, cuanto menor sea el NNT, más interesante será recomendar el tratamiento. Los NNT se pueden generar igualmente a partir de los efectos relativos sintéticos. La evidencia empírica sugiere que tanto el RR como la OR, obtenidos a partir de metaanálisis con modelos de efectos aleatorios, son bastante estables, independientemente del riesgo basal para el desenlace. Con esta condición se pueden utilizar para estimar el NNT, aunque se necesitará igualmente información sobre la condición clínica y el diagnóstico de los pacientes a tratar.

## Explorar el sesgo de publicación y otros sesgos

Uno de los métodos más simples, pero más frecuentemente utilizado, para detectar el sesgo de publicación es el gráfico en chimenea (*funnel plot*). Para que se pueda realizar e interpretar se necesita que la revisión incluya al menos 10 estudios. Como se aprecia en la [figura 1](#), se trata de una gráfica de dispersión (*scatter plot*) que recoge los efectos individuales de los estudios (eje x), representados en función de una medida de su precisión en el eje y (por ejemplo, el inverso del error estándar). Si la revisión incluye todos los estudios existentes, independientemente de su tamaño y precisión, la dispersión de los puntos que representan los datos formará una chimenea simétrica. Esto ocurre porque los estudios de menor tamaño son menos precisos y el margen de variación de los efectos estimados es más amplio. En

el supuesto de que se publicaran todos los estudios, tanto los que obtuvieron resultados positivos como los que proporcionaron resultados negativos, la gráfica sería simétrica. A medida que aumenta el tamaño y la precisión de los estudios, éstos van dando resultados menos variables a ambos lados del efecto promedio. Por lo tanto, la forma simétrica del gráfico significa que existe una baja probabilidad de que la revisión presente un sesgo de publicación u otros sesgos. Si el gráfico es asimétrico, podría ser porque faltan estudios pequeños en la base de la gráfica, estudios que son imprecisos y cuyos resultados debieron ser muy desfavorables o no significativos y que por tanto no se consideraron adecuados para publicarse; esto provoca un área menos poblada en la base del gráfico que denota la existencia de sesgos.

Debido a la variedad de razones que pueden provocar la asimetría del gráfico de chimenea (sesgo de publicación, de lenguaje, de citas, etc.), así como la dificultad para separar las unas de las otras, se ha acuñado el término «efecto de estudios pequeños» (*small-effect-study*) en lugar de sesgo de publicación. Sea cual sea la razón de la asimetría, la confianza en los resultados se verá limitada siempre que el gráfico aparezca truncado y asimétrico.

## Generar recomendaciones

Las recomendaciones deben transmitir un mensaje claro, elaborado a partir de revisiones de alta calidad, y ser tan sencillas de seguir en la práctica como sea posible. De esta forma, las conclusiones basadas en la evidencia alcanzarán la máxima utilidad para el conjunto de profesionales de la salud.

La credibilidad de una recomendación no solo depende de la fuerza de la evidencia de los estudios que la respaldan, intervienen otros cuatro factores relevantes, cuyo análisis deberá realizarse de forma explícita. En primer lugar, hay que prestar atención al equilibrio entre los efectos deseados y no deseados de una intervención. En segundo lugar, se debe pensar en la fuerza global de la evidencia para los diferentes desenlaces; cuanto mayor sea la fuerza de la evidencia, mayor será la probabilidad de que la recomendación derivada sea firme. En tercer lugar, hay que tener muy presentes las preferencias de los pacientes respecto a las intervenciones; por ejemplo, es más probable que una mujer joven, diagnosticada de cáncer de mama, se decante por la quimioterapia, aunque sea más agresiva, ya que esta opción le proporciona más probabilidades de prolongar sus años de vida. Por último, es necesario tener en cuenta el coste asociado a una recomendación; cuando el coste de una intervención sea muy elevado, la fuerza de la recomendación se verá atenuada.

## Conclusiones

Es evidente que cuanto más sólida sea la evidencia identificada en una revisión, más firme será el conocimiento aportado a la práctica clínica. Para optimizar la interpretación de la fuerza de la evidencia, lo ideal es presentar los resultados obtenidos de forma que se pueda evaluar separadamente la fuerza de la evidencia que avala cada desenlace. Para valorar la aplicabilidad de los resultados, las medidas

relativas del efecto deben complementarse con una estimación de la magnitud absoluta de los efectos, ajustada a los grupos específicos de participantes. Por otro lado, es imprescindible detectar sesgos de publicación o similares. Para ello, se recomienda la realización de un gráfico en chimenea; la asimetría en este gráfico sugiere que la confianza en los resultados podría estar limitada.

Si cada uno de los pasos de la revisión sistemática se sigue correctamente y las recomendaciones finales tienen en cuenta factores como el equilibrio entre los efectos deseados y no deseados, la fuerza total de la evidencia, los valores y preferencias de los pacientes o el coste de la intervención, los resultados se pueden traducir en recomendaciones firmes, verosímiles y dignas de confianza y, por lo tanto, contribuirán a establecer una base más sólida y fiable para futuras intervenciones.

## Financiación

Los autores no han recibido ningún tipo de financiación para la elaboración de este documento.

## Conflicto de intereses

Todos los autores declaran no tener ningún conflicto de intereses relacionado con este manuscrito.

## Agradecimientos

Agradecemos a Daniel Gavilán Cabello su colaboración en la traducción de este artículo.

Khalid S. Khan está contratado por la Universidad de Granada como investigador distinguido gracias a una ayuda del programa Beatriz Galindo del Ministerio de Ciencia, Innovación y Universidades de España.

## Bibliografía

1. Khan KS, Bueno Cavanillas A, Zamora J. Revisiones sistemáticas en cinco pasos: I. Cómo formular una pregunta para la que se pueda obtener una respuesta válida. Semergen. 2022;48:356-61, <http://dx.doi.org/10.1016/J.SEMERG.2021.12.005>.
2. Khan KS, Bueno-Cavanillas A, Zamora J. Revisiones sistemáticas en cinco pasos: II. Cómo identificar los estudios relevantes. Semergen. 2022;48:431-6, <http://dx.doi.org/10.1016/J.SEMERG.2021.12.006>.
3. Khan KS, Bueno-Cavanillas A, Zamora J. Revisiones sistemáticas en cinco pasos: III. Cómo evaluar la calidad de los estudios. Semergen. 2022, <http://dx.doi.org/10.1016/j.semerg.2022.05.001>.
4. Saeed Khan K, Bueno-Cavanillas A, Zamora J. Revisiones sistemáticas en cinco pasos: IV. Cómo sintetizar los resultados. Semergen. 2022, <http://dx.doi.org/10.1016/j.semerg.2022.02.006>.
5. Khan KS, Kunz R, Kleijnen J, Antes G. *Systematic Reviews to Support Evidence-Based Medicine*. 2nd ed. CRC Press; 2011.
6. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. J R Soc Med. 2003;96:118-21, <http://dx.doi.org/10.1177/014107680309600304>.

7. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. 2 nd ed. John Wiley & Sons; 2019.
8. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. En: Aromataris E, Munn Z, editores. *JBI Manual for Evidence Synthesis*. JBI; 2020. p. 71–88, <http://dx.doi.org/10.46658/JBIMES-20-04>.
9. Centre for Reviews and Dissemination. Core principles and methods for conducting a systematic review of health interventions. En: *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. University of York; 2009:1-108. [https://www.york.ac.uk/media/crd/Systematic\\_Reviews.pdf](https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf)