



Medicina de Familia SEMERGEN

www.elsevier.es/semergen



FORMACIÓN CONTINUADA - METODOLOGÍA Y TÉCNICAS

Revisiones sistemáticas en cinco pasos: IV. Cómo sintetizar los resultados



K. Saeed Khan^{a,b}, A. Bueno-Cavanillas^{a,b,*} y J. Zamora^{b,c,d}

^a *Departamento de Medicina Preventiva y Salud Pública. Universidad de Granada, Granada, España*

^b *CIBER de Epidemiología y Salud Pública (CIBERESP-Spain)*

^c *Unidad de Bioestadística Clínica, Hospital Ramón y Cajal, Madrid, España*

^d *Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, Reino Unido*

Recibido el 16 de febrero de 2022; aceptado el 24 de febrero de 2022

Disponible en Internet el 20 de mayo de 2022

PALABRAS CLAVE

Revisión sistemática;
Evaluación de
calidad;
Síntesis de
resultados;
Metaanálisis;
Heterogeneidad

Resumen En este artículo, el cuarto de una serie de cinco sobre revisiones sistemáticas, se muestra cómo recopilar y resumir los resultados de los estudios incluidos en una revisión. La síntesis de los datos extraídos consiste en tabular las características y la calidad de los estudios incluidos, y resumir los efectos y la magnitud de las asociaciones encontradas. Opcionalmente pueden utilizarse métodos estadísticos para explorar la heterogeneidad y sus fuentes. La magnitud de la heterogeneidad puede plantear la duda de si es viable realizar un metaanálisis general o es preferible hacerlo en subgrupos. Por último, se proporcionan recursos para evaluar las razones clínicas y metodológicas de la heterogeneidad.

© 2022 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

KEYWORDS

Systematic review;
Quality assessment;
Evidence summary;
Meta-analysis;
Heterogeneity

Systematic reviews in five steps: IV. Summarizing the evidence

Abstract This fourth in a series of five articles on systematic reviews demonstrates how to compile and summarize the results of the studies included in a review. The synthesis of the extracted data consists of tabulating the characteristics, and the quality of the included studies, with the effects or the magnitude of the associations found in them. Statistical methods can be used to explore the differences between studies and the reasons for the inconsistencies. The magnitude of this heterogeneity influences whether it is feasible to perform an overall or subgroup meta-analysis. Finally, information is provided both to assessment of clinical and methodological reasons for heterogeneity.

© 2022 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Published by Elsevier España, S.L.U. All rights reserved.

* Autor para correspondencia.

Correo electrónico: abueno@ugr.es (A. Bueno-Cavanillas).

<https://doi.org/10.1016/j.semerg.2022.02.006>

1138-3593/© 2022 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introducción

Este artículo es el cuarto de una serie de cinco sobre revisiones sistemáticas. En este paso se estudiará cómo sintetizar los resultados obtenidos a partir de los procesos descritos en los artículos previos a este¹⁻³.

Para que sea posible sintetizar los resultados de los diferentes estudios incluidos en una revisión hace falta presentarlos con claridad, de forma que se puedan analizar rigurosamente. Por una parte, hay que evaluar si los efectos observados de las intervenciones son coherentes entre sí y, en caso de que no lo sean, explorar el por qué. Por otra parte, determinar si es o no adecuado hacer una combinación estadística de los efectos individuales, es decir, realizar un metaanálisis. Este artículo se centra en cómo sintetizar el contenido de una revisión sistemática sobre los efectos de una intervención (o de una exposición a un determinado factor) medidos en desenlaces de tipo binario. En la [tabla 1](#) se incluye un breve glosario en el que se definen los términos específicos tratados en el artículo.

Describir los datos de los estudios seleccionados

Para comenzar, se recoge y presenta la información sobre las características de los estudios incluidos: participantes, intervenciones o exposiciones, desenlaces, diseño del estudio y resultados (efectos estimados)^{4,5}. Para que esta información se pueda valorar e interpretar con claridad se recomienda utilizar figuras y tabulaciones simples. De esta manera se facilita la comprensión del conjunto de la evidencia recogida y se evitan errores de interpretación. La naturaleza y la complejidad de la tabla dependerá de la cantidad de estudios incluidos y de la información que proporcionan.

A la hora de decidir la estructura de la tabla, es preciso tener en cuenta las cuestiones que se consideraron importantes cuando se formuló la pregunta PICO (Pacientes, Intervención/Comparación y Resultados [Outcomes]) de la revisión, así como aquellas variables que inicialmente se pensó que podrían producir una variación en los efectos. Por ejemplo, se podría tabular la información en filas de acuerdo con una característica de los participantes, y en ellas proporcionar de forma resumida la información relativa a las intervenciones, desenlaces y efectos de cada estudio ([fig. 1](#)).

En este paso, también se deberían computar los efectos observados en cada estudio, junto con sus intervalos de confianza⁶⁻⁸. Esto ayudará a observar la dirección (el beneficio o el perjuicio) y la magnitud del efecto (el grado en el que el riesgo aumenta o disminuye) en los estudios incluidos. La [figura 2](#) presenta una manera gráfica de evaluar la dirección y la magnitud del efecto a través de un diagrama *forest plot*⁹. Este tipo de figuras son útiles para realizar juicios cualitativos sobre los efectos de las intervenciones. En ocasiones, es posible sacar conclusiones sobre la efectividad de una intervención a partir de la observación cualitativa de estos resultados, sin necesidad de un análisis estadístico. En cambio, cuando debido al tamaño reducido de las muestras los efectos no son del todo precisos, lo más provechoso será realizar un metaanálisis. Al combinar estadísticamente los resultados individuales de cada estudio se aumentará la precisión del resultado global. Pero, antes de todo, es necesario evaluar la heterogeneidad que se observa entre los

resultados de los estudios individuales para determinar si es sensato realizar este metaanálisis o no.

Es importante recordar que el análisis de heterogeneidad debe planificarse *a priori* y tiene que estar reflejado en el protocolo de la revisión que se registra prospectivamente. Si durante el proceso de revisión se detectara que faltan datos o información relevante en los artículos publicados, la mejor opción será contactar con los autores para recabar dicha información antes de seguir con el proceso.

Investigar las diferencias en los resultados de los estudios

A veces los estudios difieren respecto a las características fundamentales de los participantes, las intervenciones o exposiciones evaluadas, o en la definición del desenlace o la forma de medirlo (heterogeneidad clínica). Otras veces, los estudios varían en los diseños y en la calidad metodológica (heterogeneidad metodológica). El análisis de la heterogeneidad consiste en el análisis de la variación de los efectos entre los estudios y en la exploración de sus razones. Las tablas elaboradas previamente son un buen punto de partida para explorar la heterogeneidad. También un examen visual del diagrama *forest plot* puede dar alguna pista. Generalmente, si las estimaciones puntuales del efecto se sitúan en el mismo lado de la «línea de no efecto», el valor nulo, se puede considerar que los efectos de las intervenciones tienen la misma dirección, aumentando o disminuyendo la probabilidad del desenlace. Por el contrario, cuando las estimaciones puntuales se sitúan a ambos lados de la línea de no efecto indican la posibilidad de que exista heterogeneidad, la intervención puede producir tanto un aumento como una disminución del riesgo, dependiendo de los estudios. Adicionalmente, si los intervalos de confianza se solapan entre sí puede significar que las diferencias observadas en las estimaciones puntuales del efecto se deben al azar, y sugiere que la heterogeneidad es poco importante. En la [figura 2](#) se muestra una representación gráfica de la heterogeneidad.

Las pruebas estadísticas de heterogeneidad examinan si la variabilidad observada entre los efectos individuales es compatible con lo que se esperaría que ocurriera únicamente por azar. Debido al escaso número de estudios que generalmente se incluyen en una revisión, la evaluación de la heterogeneidad suele estar afectada por falta de potencia estadística. Se recomienda utilizar un valor de $p < 0,1$ (en lugar del habitual 0,05). Dadas las limitaciones de la evaluación formal de la heterogeneidad con una prueba estadística, para cuantificar la cantidad de heterogeneidad que supera a la esperada por azar se recomienda usar el estadístico $I^2 = (Q - [k-1])/Q$. El I^2 oscila entre el 0 y el 100%; 25% representaría una heterogeneidad baja y un 75% una heterogeneidad alta. Al utilizar I^2 se evita depender del número de estudios. No obstante, este estadístico también depende del tamaño muestral de cada estudio de forma que un metaanálisis de estudios con muchos participantes podría inflar el valor de la I^2 .

Una aproximación sensata a la hora de evaluar la heterogeneidad consistiría en examinar tanto las evaluaciones no estadísticas en el diagrama *forest plot* como el I^2 , sin depender únicamente del valor p de la prueba estadística. En caso de que se sospeche la existencia de heterogeneidad, aunque no esté confirmada estadísticamente, es

Tabla 1 Glosario sobre términos del cuarto paso en una revisión sistemática

Término	Definición
Metaanálisis	Se trata de una técnica estadística que combina los resultados de un conjunto de estudios sobre un mismo tema con el objetivo de producir un resultado promedio ponderado.
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 en el desenlace de los estudios.
Efecto	Es la medida de la asociación observada entre las intervenciones o exposiciones y los desenlaces, un dato estadístico utilizado para sintetizar la solidez de la asociación observada. El efecto tiene un punto de estimación, así como un intervalo de confianza.
Efecto resumen	Describe el efecto generado al reunir los efectos individuales en un metaanálisis.
Tamaño del efecto	Cuanto mayor sea el tamaño del efecto, mayor será la fuerza de la asociación entre exposición y desenlace.
Intervalo de confianza (IC)	En inglés, <i>confidence interval</i> (CI). Se trata del rango en que se espera que el valor «verdadero» de una medida (el resultado de una intervención) esté comprendido, con un determinado grado de certeza. Normalmente, se utilizan intervalos de confianza del 95%.
I^2	Es un dato estadístico utilizado para evaluar la heterogeneidad surgida durante el proceso de síntesis de los resultados. Oscila entre 0 y 100%. Proporciona el porcentaje de la variación total de los estudios debido a la heterogeneidad.
Estimación puntual	En inglés, <i>point estimate</i> . Se trata del valor observado del efecto de una intervención entre los sujetos incluidos en la muestra del estudio.
Valor de p	En inglés, <i>p value</i> . Es la probabilidad, dada una hipótesis nula, de que los efectos observados, o efectos extremos, en un estudio pueden haber ocurrido por azar (error aleatorio). Normalmente, un valor de p menor a 5% ($p < 0,05$) se considera estadísticamente significativo.

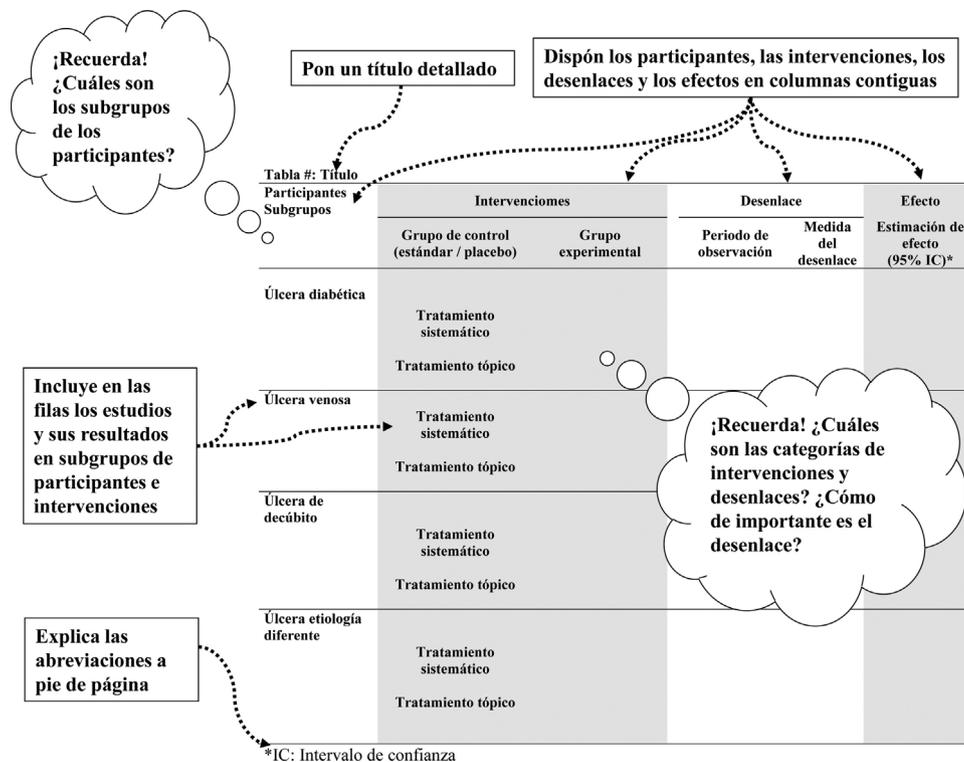


Figura 1 Cómo preparar una tabla.

recomendable buscar una explicación para ella. Cuando se detecta un exceso de heterogeneidad y, no se identifican fuentes que la puedan explicar, suele ser preferible evitar la estimación de un resultado combinado para los estudios incluidos en la revisión.

Metaanálisis de los resultados observados en los estudios

Empecemos recordando que una revisión sistemática no tiene por qué incluir siempre un metaanálisis. No

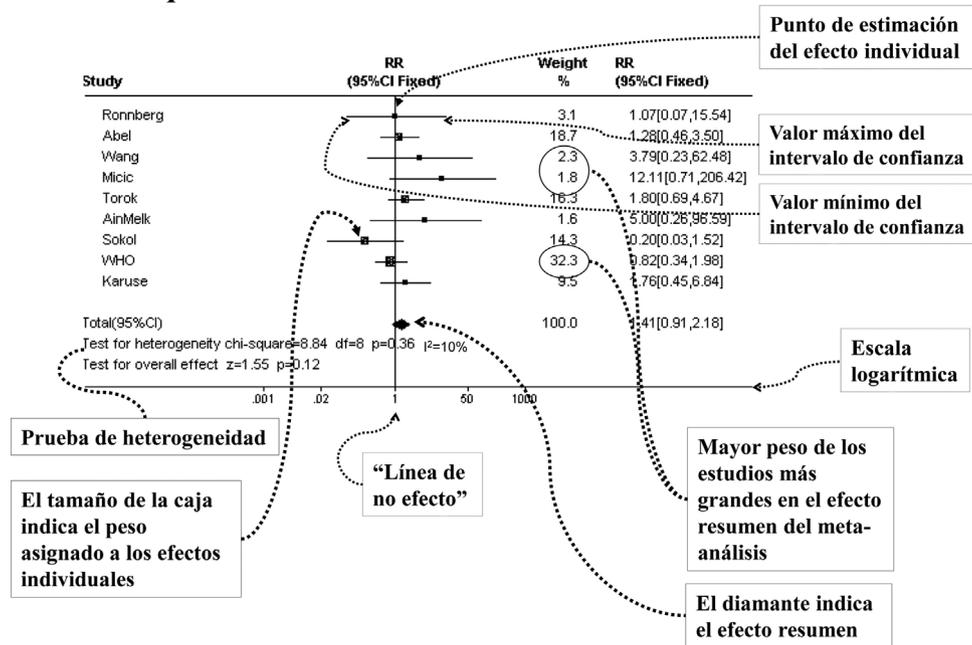
Forest plots

Figura 2 Cómo interpretar un forest plot.

siempre es factible realizar un metaanálisis, por ejemplo, no lo es cuando hay grandes disparidades entre los estudios en cuanto a los participantes, las intervenciones o exposiciones, el diseño o la calidad metodológica. De hecho, solo se debería efectuar si los estudios comparten características clínicas, calidad metodológica, y sus efectos son homogéneos.

En pocas palabras, un metaanálisis agrupa los efectos observados en los estudios individuales con el fin de estimar un efecto promedio, ponderado, de todos los estudios. Cada estudio se pondera según alguna medida relativa a su relevancia. Por ejemplo, se puede emplear un método que dé más importancia a los estudios más grandes (mayor tamaño muestral) con una estimación del efecto más precisa, y menos peso a estudios pequeños, con una estimación del efecto más imprecisa. Esto se consigue al asignar un peso determinado a cada estudio, en una proporción inversa al error de estimación del efecto, que en última instancia depende tanto del tamaño muestral como del número de eventos observados en el estudio.

Es esencial asegurar la robustez del efecto resumen calculado empleando un modelo de análisis estadístico adecuado. Existen fundamentalmente dos modelos: modelo de «efectos fijos» y modelo de «efectos aleatorios». El primero asume que solo existe un efecto subyacente «verdadero» y que los estudios incluidos son una muestra aleatoria de los estudios que estiman este efecto medio «fijo». El segundo modelo considera que no existe un único efecto subyacente, sino que hay varios efectos dependiendo de las características de los estudios, los participantes, o las diferencias en las intervenciones (o exposiciones). El modelo de efectos aleatorios calcula un efecto en el que además de considerar la precisión, se pondera cada estudio en función de la variación aleatoria que se asume que existe entre ellos. Esto hace que se pueda magnificar el sesgo de publicación (ver en el

último artículo de esta serie), puesto que a la hora de estimar el efecto resumen, se pueden ponderar a los estudios pequeños con un peso proporcionalmente mayor que en un modelo de efectos fijos.

Cuando existe heterogeneidad, un modelo de efectos aleatorios proporciona intervalos de confianza para el efecto global más amplios, y por tanto más conservadores. El modelo de efectos fijos puede generar una precisión indebida del efecto global si existe una heterogeneidad sin explicación entre estudios. En la práctica se usan ambos a modo de análisis de sensibilidad, pero en caso de elegir uno, la elección debe realizarse *a priori*.

Heterogeneidad clínica

Las diferencias respecto a los participantes, las intervenciones o exposiciones y los desenlaces pueden ayudar a interpretar la relevancia clínica de los resultados. Como se sugirió en el primer artículo de esta serie, es recomendable identificar las posibles fuentes de variación en los efectos durante la formulación de la pregunta. Para hacer la tarea más fácil, se pueden clasificar los estudios en subgrupos, de esta manera es posible examinar las diferencias en los efectos en cada uno de dichos subgrupos. Lo más apropiado es efectuar esta clasificación en el primer paso de la revisión sistemática y, en todo caso, evitar realizar análisis de subgrupos que no estaban planificados con anterioridad.

Heterogeneidad metodológica

También hace falta valorar si las diferencias de diseño y calidad entre los estudios incluidos en la revisión pueden explicar la variación de los efectos observados. Se espera que se hayan descartado los estudios con diseños deficientes

y que se hayan mantenido desde el principio solo aquéllos que presenten un mínimo de calidad. Asimismo, lo ideal sería haber evaluado la calidad de los estudios concienzudamente, como se hizo en el tercer paso de esta serie de artículos, con el fin de descubrir si la calidad varía entre los estudios. En este paso, la atención se pone en valorar si la calidad del estudio interfiere en la estimación de los efectos como parte del proceso de exploración de la heterogeneidad.

Como se ha visto anteriormente, la clasificación en subgrupos puede ser muy útil. En el caso de la heterogeneidad metodológica, los estudios deben tabularse en subgrupos de acuerdo con su diseño. Es importante tener en cuenta que un metaanálisis es plausible únicamente si incluye estudios con el mismo diseño y las inferencias se basan en los efectos observados en estudios con diseños de calidad superior.

Por último, para reducir el número de subgrupos, así como el riesgo de conclusiones espurias, se puede considerar la posibilidad de clasificar los estudios en subgrupos de calidad alta y baja, con base en su cumplimiento con los elementos de calidad.

Metarregresión

Se trata de una técnica para explorar la heterogeneidad que ha ganado muchos adeptos en los últimos años. En pocas palabras, la metarregresión utiliza un modelo de regresión lineal multivariable para examinar la influencia de las características y de la calidad de los estudios en el tamaño de los efectos observados individualmente. No obstante, su principal inconveniente es el riesgo de «sobreajuste» (*overfitting*), que tiene lugar cuando se incluyen en el modelo pocos estudios y un gran número de variables de ajuste. Para evitarlo, es frecuente encontrar análisis univariantes de regresión en lugar de multivariantes.

Metaanálisis cuando la heterogeneidad no tiene explicación

En algunas revisiones no se consigue explicar la heterogeneidad. Puede ser sencillamente porque el número de estudios no es suficiente para realizar un análisis que detecte el porqué de las diferencias en los efectos. En este caso cabe cuestionar la viabilidad de un metaanálisis. Es tentador atribuir la heterogeneidad a la variabilidad fortuita de los efectos y llevar a cabo un metaanálisis con un modelo de efectos aleatorios. No obstante, la utilización de este modelo puede producir sesgos en la estimación del efecto resumen. La interpretación del efecto resumen debe ser muy cautelosa, la heterogeneidad compromete la solidez de la evidencia recogida en la revisión. Por lo tanto, hay que examinar si el efecto general ponderado y los efectos de los estudios individuales de alta calidad son congruentes o no.

Conclusiones

A la hora de sintetizar los resultados es fundamental empezar por valorar los efectos de cada uno de los estudios

incluidos en la revisión. Resumir las características de los estudios en una tabla, dividida en subgrupos, puede resultar muy útil para explorar la heterogeneidad clínica y metodológica. Cuando las características de los componentes de la pregunta PICO no sean excesivamente dispares entre los estudios, se podrá realizar un metaanálisis para generar un efecto promedio ponderado a partir de los efectos individuales. El próximo y último paso de esta serie se centrará en la interpretación de los resultados.

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.

Khalid Saeed 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, España.

Agradecimientos

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

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 (In Press).
2. Khan KS, Bueno-Cavanillas A, Zamora J. Revisiones sistemáticas en cinco pasos: II. Cómo identificar los estudios relevantes. SEMERGEN. 2022 (In Press).
3. Khan KS, Bueno-Cavanillas A, Zamora J. Revisiones sistemáticas en cinco pasos: III. Cómo evaluar la calidad de los estudios. SEMERGEN. 2022 (In Press).
4. Khan KS, Kunz R, Kleijnen J, Antes G. Systematic Reviews to Support Evidence-Based Medicine. 2nd ed. CRC Press; 2011.
5. 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/10.1258/JRSM.96.3.118>.
6. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. John Wiley & Sons; 2019.
7. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Systematic Reviews of Effectiveness. En: Aromataris E, Munn Z, editores. JBI Manual for Evidence Synthesis. 2020. p. 71–88, <http://dx.doi.org/10.46658/JBIMES-20-04>.
8. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. New York, USA: Centre for Reviews and Dissemination; 2009. p. 1–108 [consultado 15 Mar 2022] <https://www.york.ac.uk/media/crd/Systematic.Reviews.pdf>
9. Khan KS, Daya S, Jadad AR. The Importance of Quality of Primary Studies in Producing Unbiased Systematic Reviews. Arch Intern Med. 1996;156:661–6, <http://dx.doi.org/10.1001/archinte.1996.00440060089011>.