

Hablemos de...

La revisión sistemática y metaanálisis

JOSÉ I. EMPARANZA E IRATXE URRETA

Unidad de Epidemiología Clínica. Red MBE G(03/090). Hospital Donostia. San Sebastián. Guipúzkoa. España.
jemparan@chdo.osakidetza.net; iratxeube@hotmail.com

Una revisión sistemática (RS) es un tipo de estudio que pretende recopilar toda la investigación sobre un tema determinado, evaluarla críticamente y obtener unas conclusiones que resuman el efecto de una intervención sanitaria.

No todas las revisiones son sistemáticas. Los métodos para revisar la bibliografía científica varían desde síntesis formales y cuantitativas hasta resúmenes subjetivos de datos observacionales.

Una RS es un ejercicio exhaustivo de búsqueda, evaluación y síntesis de la información. Intenta ser objetiva y cuantitativa usando números para sintetizar toda la evidencia disponible.

para una toma de decisiones clínica.

Con una RS no se pretende combinar los datos de una manera descoordinada. Por el contrario, se pretende mantener el efecto individual de cada estudio utilizando unos métodos estadísticos que den más peso en el cálculo del efecto combinado a aquellos estudios con tamaño muestral mayor.

Al proceso estadístico de combinar cuantitativamente los resultados ponderados de los estudios individuales (2 o más) se le llama metaanálisis. La bondad del metaanálisis depende, en primer término, de la bondad de la RS. Sin embargo, existen metaanálisis que no se hacen en el contexto de una RS. Su inconveniente es que el resultado puede estar totalmente sesgado y, por tanto, carecer de valor y credibilidad

Revisión sistemática

Justificación de las revisiones sistemáticas


Las razones para realizar una RS son de índole práctica¹. Estamos desbordados por una cantidad de información imposible de manejar, necesitamos RS para integrar eficientemente toda la información válida y proporcionar una base para tomar decisiones de manera racional. Precisamente éste es el propósito básico de la Colaboración Cochrane: preparar, mantener y disseminar RS de la efectividad de las intervenciones sanitarias².


Ventajas y desventajas


La revisión sistemática es un tipo de estudio que pretende revisar todo lo investigado sobre un tema específico.


Las RS tienen 2 ventajas fundamentales. La primera es que, al combinar la información de diversos estudios, permiten analizar la consistencia de los resultados. Muchos de los estudios, tomados individualmente, son demasiado pequeños incluso para detectar efectos modestos pero importantes, es decir, tienen un poder estadístico insuficiente. Combinando


Puntos clave

 Una revisión sistemática es un ejercicio exhaustivo de búsqueda, evaluación y síntesis de la información sobre un tema claramente definido.

 Una pregunta estructurada es el primer paso ineludible.

 Aseguraremos la exhaustividad de la búsqueda con una estrategia definida en todas las bases de datos pertinentes.

 Es necesario evaluar la calidad metodológica de los estudios incluidos en la revisión para asegurar el control de sesgos.

 El metaanálisis es el paso final; trata de resumir numéricamente toda la información de calidad.

La razón para realizar una revisión sistemática es de índole práctica, dada la cantidad de información

El primer paso es redactar una pregunta de modo estructurado.

La búsqueda de estudios que respondan a la pregunta planteada necesita una estrategia de búsqueda explícita y decidir qué bases de datos son

El tercer paso es la selección de los estudios tras la lectura del título y resumen para saber si responden a la pregunta planteada.

El cuarto paso es la extracción de los datos y la evaluación de la calidad metodológica por 2 de los investigadores.

los estudios que han intentado responder a la misma pregunta se consigue aumentar el tamaño muestral y, por tanto, se incrementa el poder estadístico (capacidad para descubrir efectos de modesta magnitud).

Una segunda ventaja importante es que un efecto similar en diferentes ámbitos y utilizando diseños distintos (p. ej., con criterios de inclusión y exclusión diferentes para los pacientes que entran en cada estudio) puede darnos una idea de cuán robustos y trasladables son los resultados de nuestra RS a otros ámbitos.

Si los resultados de los estudios individuales varían mucho entre uno y otro, deberemos examinar cuál es la fuente que está produciendo estas diferencias en los resultados entre los estudios.

Así pues, mientras que algunos dicen que hacer una RS es mezclar peras con manzanas, otros pueden opinar que esta característica es una ventaja porque aumenta lo que llamamos la "validez externa" o "generalizabilidad" (o capacidad de trasladar resultados entre ámbitos diferentes).

Una desventaja potencial es que, si se incluyen estudios con un diseño pobre, que no asegura el control de los hipotéticos sesgos, la RS producirá unos resultados combinados que no serán acordes con la realidad. Por ello es necesario evaluar la calidad metodológica de los estudios para incluir aquellos que acreditan una calidad suficiente.

En el caso particular de los ensayos clínicos, hay una serie de características de diseño que sabemos perturban notablemente los resultados obtenidos. Tales características son: una aleatorización incorrecta o sin ocultación de la secuencia de aleatorización, un enmascaramiento incorrecto y la pérdida de sujetos que lleve a evaluar una población final diferente de la aleatorizada.

Esquemáticamente, las ventajas de la RS son:

1. Es una técnica científica eficiente.
2. Aumenta la generalizabilidad en los resultados.
3. Aumenta la consistencia de los resultados.
4. Con ella aumentan el poder y la precisión de la estimación.
5. Permite hacer una evaluación exacta de las informaciones publicadas.

Etapas en la realización de una revisión sistemática

El primer paso es trasladar al papel la pregunta que ha

surgido con un determinado paciente³. Explicitar la pregunta no es fácil. Existen diferentes preguntas: sobre eficacia de una intervención terapéutica, exactitud diagnóstica, etiología, predicción, pronóstico, evaluación económica, etc. Idealmente la pregunta debe contener 4 componentes (PICO):

1. Una descripción lo más explícita y detallada posible del *paciente* que motiva esta pregunta.
2. Debemos detallar cuál es la *intervención* que queremos considerar, ya sea terapéutica, diagnóstica o de otra índole.
3. Describiremos adecuadamente la *comparación* con esta intervención, ya sea terapéutica, diagnóstica, etc.
4. Debemos ser cuidadosos al escoger la medida de *resultado* que utilizaremos para evaluar si una intervención es mejor que su comparación o no en este paciente.

El segundo paso consiste en realizar una búsqueda de estudios que respondan a la pregunta estructurada en los 4 componentes. Debemos saber que cada tipo de pregunta (sobre etiología, diagnóstico, tratamiento, pronóstico, etc.) requiere para su respuesta óptima, desde el punto de vista metodológico y práctico, un determinado diseño de estudio (descriptivo, de casos y controles, ensayo clínico, etc.). Es importante conocer qué bases de datos ofrecen mayor rendimiento para responder a qué tipo de preguntas. Por ejemplo, los ensayos clínicos publicados se pueden encontrar en la base de datos MEDLINE, pero se recogen también en 2 registros específicos: clinicaltrials.gov y controlled-trials.com. La Biblioteca Cochrane dispone asimismo de una base de datos (Central) de ensayos clínicos.

Igualmente es necesario definir la estrategia de búsqueda, es decir, qué términos vamos a utilizar para describir los estudios que queremos recuperar y cómo los vamos a combinar con los operadores booleanos (OR, AND, NOT). Para una búsqueda óptima es aconsejable consultar con una persona que tenga entrenamiento en documentación.

El tercer paso es la selección de estudios en función de su pertinencia con la pregunta. Habitualmente realiza la selección uno de los investigadores accediendo al título y al resumen de los artículos recuperados en la etapa de la búsqueda.

El cuarto paso es la extracción de datos. Es conveniente diseñar una hoja de recogida de datos, que debe ser consensuada con el equipo investigador, donde se recojan todas las características del estudio (autor, país, año, intervención, paciente, etc.). La extracción de datos debe ser

realizada por 2 investigadores para evitar errores y subjetividades en lo posible. En caso de que haya discrepancias, deben resolverse por discusión y consenso, o recurriendo a un tercer investigador que emita su juicio. De manera simultánea a la extracción de datos, 2 miembros

del equipo investigador evalúan la calidad de los estudios. Tras esta evaluación se seleccionan los de buena calidad. En caso de discrepancias, se procederá igual que en la extracción de datos.

Al concluir estos pasos tendremos una base de datos con los estudios que tienen suficiente calidad de entre los

seleccionados y sus características. Podremos entonces proceder al metaanálisis propiamente dicho si lo que queremos es obtener un estimador común para todos los estudios.

Metaanálisis

El metaanálisis es un proceso que tiene 2 etapas. La primera consiste en calcular las medidas de efecto para cada estudio y su intervalo de confianza (fig. 1). La segunda etapa es calcular el efecto global, resumen o combinado del tratamiento como una media ponderada de los efectos obtenidos en los estudios individuales. Las técnicas estadísticas a utilizar varían considerablemente de acuerdo con las medidas que se emplean para describir el efecto en cada estudio.

Básicamente podemos encontrar 2 tipos de medida del efecto o resultado (*outcome* en inglés). Los resultados pueden medirse mediante una variable binaria (p. ej., curados o no) o mediante una variable continua (colesterol plasmático, medida de pico flujo, número de leucocitos, etc.). En el caso de los resultados medidos mediante una variable binaria, podemos utilizar diferentes modos para expresar numéricamente el efecto en cada estudio. Existen medidas relativas como la *odds ratio* (OR) y el riesgo relativo (RR), y medidas absolutas como la reducción absoluta de riesgo (RAR) y el NNT (número de pacientes que es necesario tratar) (tabla 1).

Las medidas relativas expresan el efecto o resultado que se observa en un grupo en relación con el efecto en el otro grupo. Tienen propiedades matemáticas muy convenientes que facilitan el análisis, pero, por contra, su utilización puede ser engañosa. La reducción de la mortalidad de 2 por cada 10 millones a 1 por cada 10 millones supone una reducción relativa del 50%, indistinguible de la reducción relativa de 2 de cada 4 (control) a 1 de cada 4 (experimental). Sin embargo, mientras que en el primer caso necesitaremos tratar a 10 millones de personas para evitar una muerte, en el segundo por cada 4 que tratemos con el fármaco experimental evitaremos una muerte (NNT = 4). Las medidas absolutas del efecto, como NNT y RAR, se

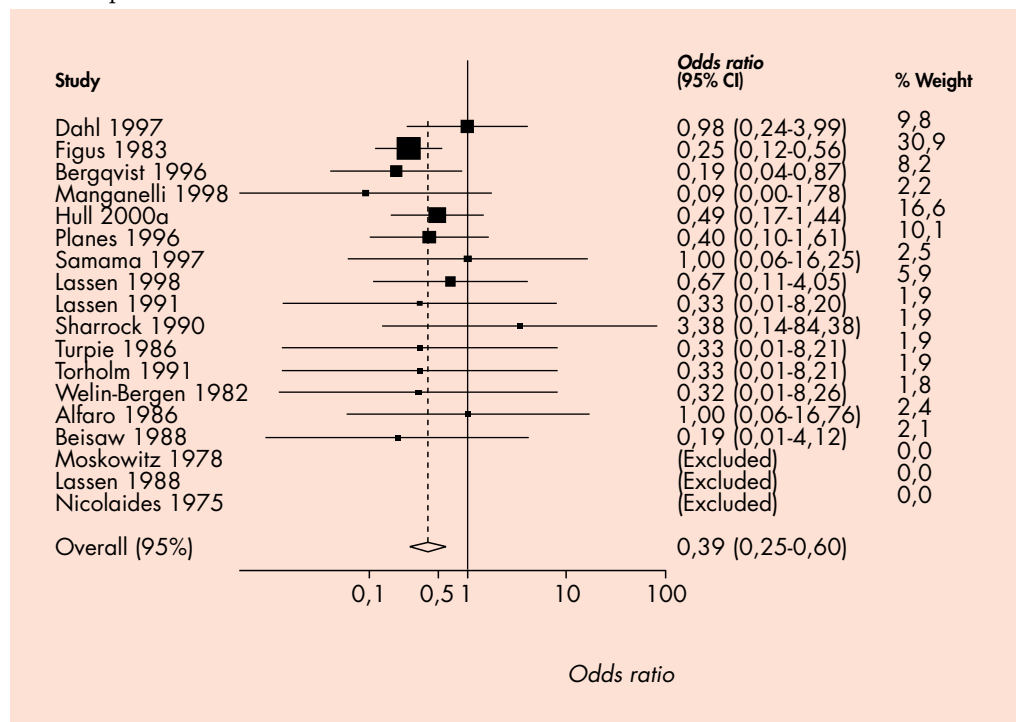


Figura 1. Representación gráfica de los resultados de cada estudio.

interpretan de manera más verosímil por el clínico⁴. Una vez elegida la medida del efecto que vamos a utilizar para el metaanálisis, procederemos a obtener un estimador único que resuma en un número el efecto real de la intervención para el conjunto de estudios analizados. Una amenaza importante a este estimador combinado o resumen de todos los estudios es la llamada heterogeneidad. Gráficamente podemos sospechar

Tabla 1. Estimadores del efecto para una variable binaria

	Muerte	No muerte	
Experimental	a	b	a + b
Control	c	d	c + d
	a + c	b + d	
RR = (a/a + b)/(c/c + d)			
OR = a · d/c · b			
RAR = (a/a + b) - (c/c + d)			
NNT = 1/RAR			

RR: riesgo relativo; OR: *odds ratio*; RAR: reducción absoluta de riesgo; NNT: número de pacientes que es necesario tratar.

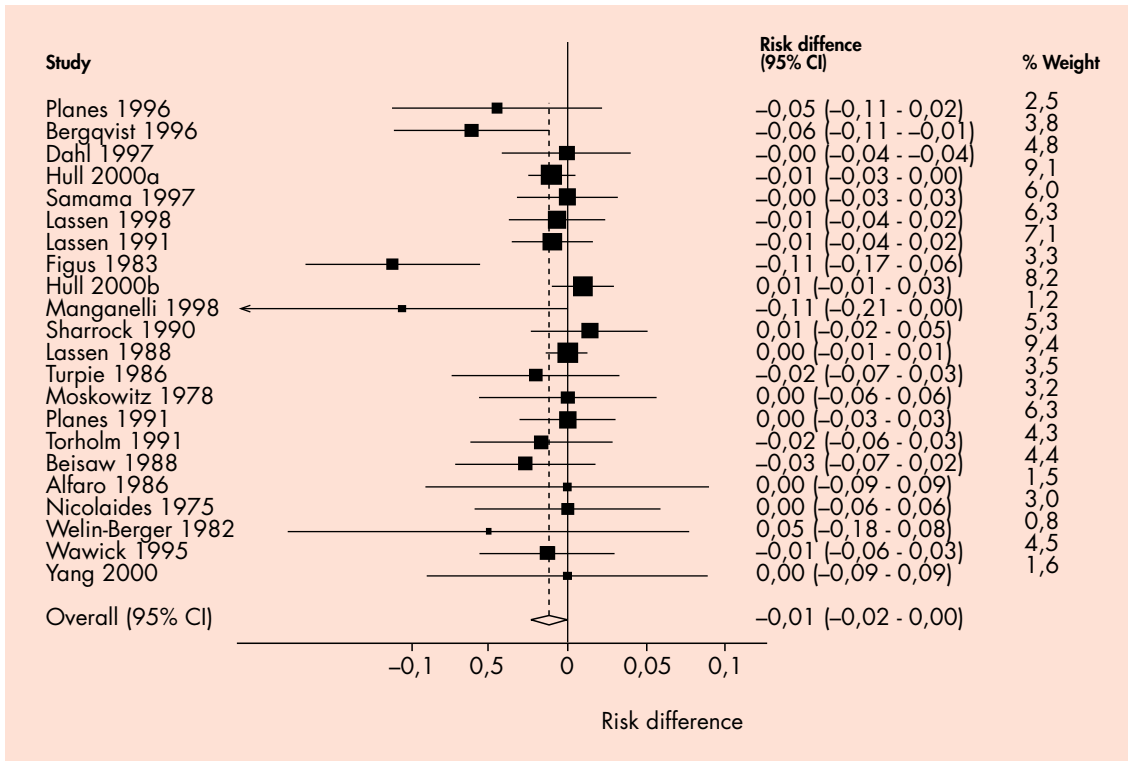


Figura 2.
 Representación gráfica que muestra heterogeneidad.

la presencia de heterogeneidad si el *forest plot* (representación gráfica de los estimadores de cada estudio y su intervalo de confianza) muestra líneas que no se superponen (fig. 2).

En efecto, cuando los estudios individuales recogidos en una RS difieren excesivamente entre sí (esto puede deberse a que los criterios de inclusión o de selección de paciente son distintos, a que los métodos de medición de un mismo fenómeno son diferentes o a cualquier otra razón de índole clínica o estadística), no parece razonable hacer un estimador único resumen del efecto. En su lugar, se puede recomendar hacer una descripción por subgrupos, uniendo en subgrupos los estudios que tienen más semejanza entre sí.

Disponemos de 2 métodos o modelos para obtener el estimador resumen del efecto en un conjunto de estudios. Son el modelo de efectos fijos y el de efectos aleatorios. El modelo de efectos fijos sólo incluye como fuente de variación la debida a la imprecisión de cada estudio, mientras que el de efectos aleatorios incluye 2 componentes de variación: a) el mismo que el de modelos fijos, la imprecisión en la estimación de cada estudio, y b) la variación de estudio a estudio. Por tanto, dado que la

variación que se incluye en el modelo de efectos aleatorios es mayor que la que se incluye en el modelo de efectos fijos, el intervalo de confianza alrededor del estimador resumen del metaanálisis obtenido mediante un modelo de efectos aleatorios es mayor, más amplio, que el obtenido mediante un modelo de efectos fijos.

Hay muchos investigadores y metaanalistas que opinan que el modelo que hay que utilizar es el de efectos aleatorios, mientras que otros opinan lo contrario. En cualquier caso, si existe algo de heterogeneidad no parece razonable (y en esto está casi todo el mundo de acuerdo) utilizar un modelo de efectos fijos.

Realmente, si existe heterogeneidad en el metaanálisis, deberemos investigar cuáles son los factores que están produciendo ese incremento de la variabilidad, y esto se puede analizar mediante modelos de metarregresión, cuya descripción está fuera del alcance de este artículo.

La última fase de un metaanálisis es analizar el sesgo de

El metaanálisis intenta obtener un estimador único que resuma el efecto real de la intervención.

Es más sencillo matemáticamente utilizar medidas relativas del efecto como el riesgo relativo y la *odds ratio*, pero se interpretan más fácilmente las absolutas, como la reducción absoluta de riesgo y el NNT (número de pacientes que es necesario tratar).

Una heterogeneidad excesiva entre los estudios desaconseja el cálculo de medidas resumen (o combinadas) de efecto.

Existen 2 modelos para estimar el efecto combinado: el modelo de efectos fijos y el de efectos aleatorios.

La última fase es analizar el sesgo de publicación básicamente mediante inspección visual.

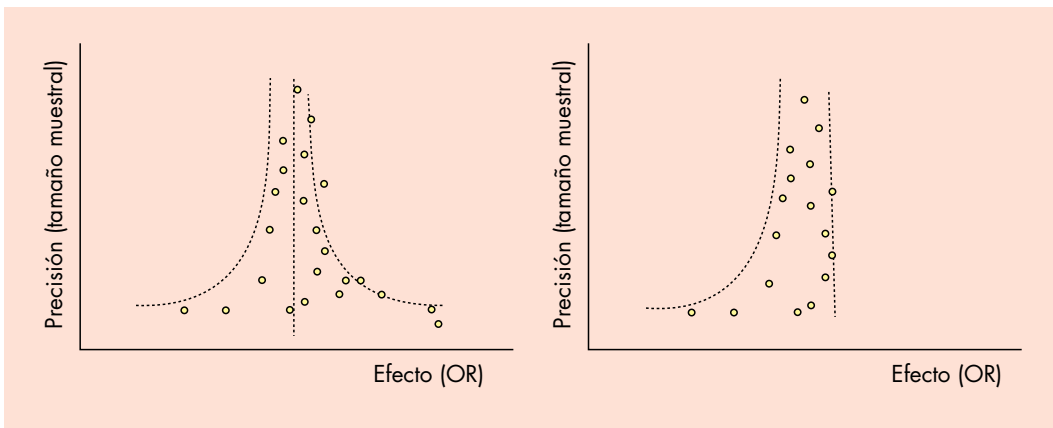


Figura 3. Ausencia (izqda.) y presencia (dcha.; asimetría) de sesgo de publicación. OR: odds ratio.

publicación. Éste se produce debido a la tendencia que tienen tanto los autores, los investigadores, como las revistas a publicar los artículos o estudios que dan resultados positivos, es decir, que muestran diferencias entre la intervención y el control. Para estudiar el sesgo de publicación existen algunas técnicas estadísticas, pero podemos hacernos una idea mediante la inspección visual de los estimadores individuales de cada estudio frente a la inversa de su variancia o, lo que es casi lo mismo, su tamaño muestral⁵. Observaremos una figura simétrica en ausencia de sesgo de publicación, como en la parte izquierda de la figura 3.

El programa CASPe (<http://www.redcaspe.org>), cuyo objetivo es enseñar habilidades en la búsqueda de información y en la lectura crítica de la bibliografía científica, ha diseñado unas plantillas con los puntos clave como ayuda para evaluar críticamente una RS.

Bibliografía



1. Egger M, Smith GD, Altman DG. Systematic reviews in health care. 2nd ed. London: BMJ Publishing Group; 2001.
2. Colaboración Cochrane. Disponible en: <http://www.cochrane.es/>
3. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Cómo formular preguntas clínicas que usted pueda responder. En: Sackett DL, Richardson WS, Rosenberg W, Haynes RB, editores. Medicina basada en la evidencia. Madrid: Churchill Livingstone; 1997.
4. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318:1728-33.
5. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629-34.