

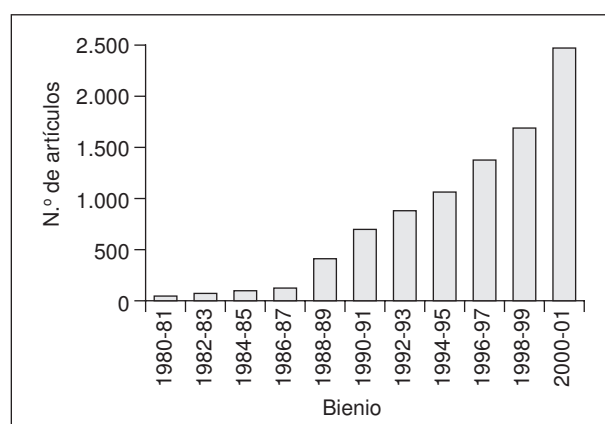
## Revisiones sistemáticas y metaanálisis

V. Abraira

Unidad de Bioestadística Clínica. Hospital Ramón y Cajal. Madrid.

En las clasificaciones de los tipos de estudios, las revisiones sistemáticas aparecen siempre como el diseño con mayor fuerza probatoria de la hipótesis en evaluación<sup>1</sup>, es decir, para contestar a una pregunta sobre la eficacia de un tratamiento, lo mejor sería buscar una revisión sistemática de ensayos clínicos que evalúen ese tratamiento, si la pregunta fuera sobre pronóstico, lo mejor sería una revisión sistemática de estudios de cohortes, etc. Aunque, obviamente, la frase anterior hay que matizarla, diciendo que lo mejor sería una *buen*a revisión sistemática porque, como en cualquier otro ámbito de la actividad humana, hay revisiones sistemáticas buenas y otras no tan buenas. Por ejemplo Silagy<sup>2</sup>, revisando 7 revistas de Atención Primaria en 1991, identificó 28 revisiones sistemáticas de las que, evaluadas usando 8 criterios estándar de rigor metodológico, sólo 7 (25%) tenían 8 puntos o más, de un máximo de 16.

Visto con la perspectiva actual parece muy natural la idea de responder a una pregunta revisando sistemáticamente y analizando toda la literatura sobre el tema, sin embargo los conceptos de revisión sistemática y metaanálisis son relativamente recientes. Para Altman la introducción y el continuo auge del metaanálisis es el cambio más prominente que se ha producido en el uso de la estadística en las revistas médicas en los últimos 10 años<sup>3</sup>. El término metaanálisis fue acuñado por Glass a mediados de los años 70 para describir la recopilación de información de varios estudios del mismo tipo<sup>4</sup> y el primer metaanálisis sobre un tema médico que figura en Medline se publicó en 1980 y se realizó sobre el tratamiento de la tartamudez<sup>5</sup>. Desde entonces, como se muestra en la figura 1, su aplicación en medicina ha aumentado espectacularmente, hasta el punto de que en 1988 Medline lo incluye como tipo de publicación estándar. En la actualidad, el término



**Figura 1.** Evolución temporal del número de artículos que figuran en Medline usando el término metaanálisis en el título o el resumen.

metaanálisis resulta algo ambiguo, pues aunque inicialmente se usaba para describir todo el proceso de revisar todos los estudios sobre un tema, desde su identificación hasta el análisis de los datos y su interpretación, con posterioridad se introdujo el término revisión sistemática para el proceso de identificar sistemáticamente y evaluar los artículos, con una metodología explícita y repetible, reservándose el término metaanálisis para la combinación numérica de los datos, aunque desgraciadamente la situación actual es que hay quien usa metaanálisis en este sentido restrictivo y quien lo sigue usando en sentido amplio para describir todo el proceso. A favor de la distinción, hay que destacar que no todas las revisiones sistemáticas incluyen metaanálisis, es decir, una combinación de los datos de los distintos estudios en un resultado global, bien porque los diseños, o la calidad de los estudios, son muy diferentes entre sí, o porque los resultados son muy heterogéneos.

La necesidad de las revisiones sistemáticas se sustenta en tres pilares: a) la inmanejable cantidad de información clínica producida: anualmente se publican más de dos millones de artículos en más de 20.000 revistas biomédicas<sup>6</sup>, que hace imprescindible un buen sistema de resumirlos; b) los estudios individuales dependen de sus característi-

Correspondencia:  
V. Abraira.  
Unidad de Bioestadística Clínica.  
Hospital Ramón y Cajal. Ctra. Colmenar km 9,100.  
28034 Madrid.  
Correo electrónico: victor.abraira@hrc.es

**Tabla 1. Pasos en la realización de una revisión sistemática**

Especificación de una pregunta a responder
Formulación de los criterios de elegibilidad de los estudios, que permitan identificar objetivamente aquéllos apropiados para resolver la pregunta
Producción de un protocolo en el que consten los criterios de selección de los estudios y los métodos que serán usados
Búsqueda rigurosa de todos los artículos relevantes (no sólo los que figuran en las bases de datos electrónicas)
Evaluación de si los artículos encontrados cumplen los criterios
Evaluación de la calidad de los artículos y su susceptibilidad a sesgos
Extracción de los datos que resumen tanto los resultados como el diseño
Combinación estadística de los datos (metaanálisis) si es apropiado y consideración de las diferencias entre artículos
Investigación de la robustez (estabilidad) de los resultados y análisis de sensibilidad
Interpretación de los resultados

Tomada de Altman<sup>3</sup>.

cas concretas y rara vez dan respuestas definitivas a las cuestiones clínicas<sup>7</sup>, las revisiones sistemáticas ayudan a establecer si los hallazgos de los estudios son consistentes y pueden ser generalizados y, en caso contrario, permiten explorar las razones de las inconsistencias, y c) las revisiones que incluyen metanálisis, al contar con mayor número de pacientes, tienen mayor precisión en sus estimaciones que los estudios primarios.

La revisión sistemática es en sí misma un diseño de investigación, observacional y retrospectivo, que sintetiza los resultados de múltiples investigaciones primarias. Es, por tanto, un tipo de diseño muy proclive a sesgos y debe planificarse cuidadosamente en un protocolo previo para intentar minimizarlos. Los pasos necesarios en la realización de una revisión se detallan en la tabla 1 y consisten básicamente en una revisión exhaustiva, objetiva y verificable de la investigación primaria que no se limite a las bases de datos electrónicas, sino que la complementa con búsquedas manuales en referencias bibliográficas, resúmenes de congresos, consultas con investigadores, registros de organismos evaluadores de investigación, etc.; con criterios explícitos de evaluación del diseño de los artículos y de su calidad y, si procede, combinación estadística de los datos usando métodos validados.

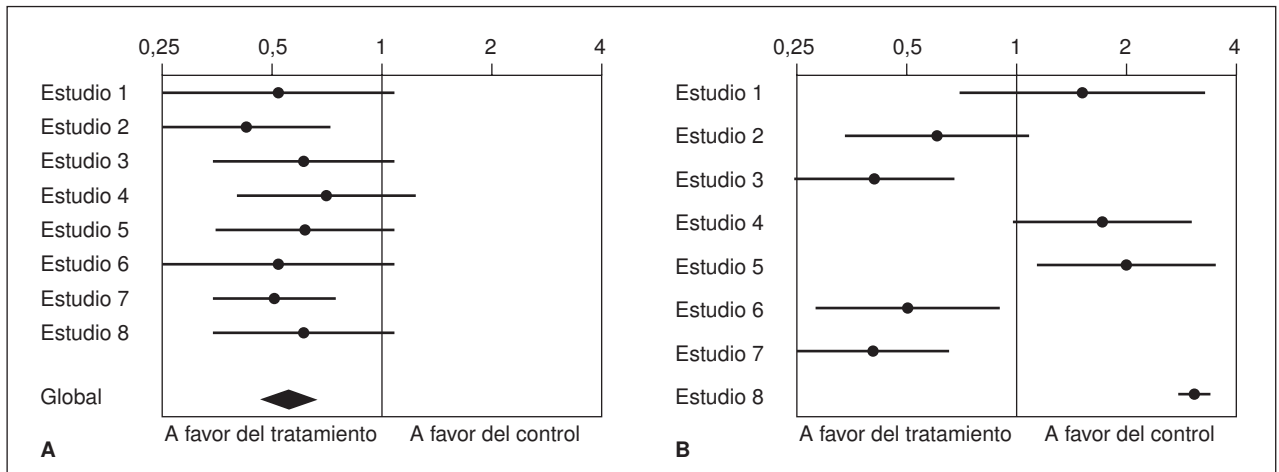
El método más simple de combinar los resultados sería dar como resultado global la media aritmética de los resultados de cada artículo, pero éste es un procedimiento que se presta a conclusiones erróneas porque los estudios con menor tamaño muestral son más propensos a dar por azar resultados más alejados del verdadero resultado. Los métodos usados en los metaanálisis usan medias ponderadas, de tal modo que los estudios con mayor tamaño muestral tengan mayor peso que los de menor tamaño. Hay básicamente dos métodos distintos, cuya diferencia estriba en cómo se trata la variabilidad entre estudios<sup>8</sup>. El modelo de efectos fijos considera que toda la variabilidad entre estudios es exclusivamente debida a la variación aleatoria producida por el muestreo, es decir, si los estudios primarios tuvieran un tamaño muestral infinito darían el mismo resultado. El modelo de efectos aleatorios conside-

## Puntos clave

- La introducción y el continuo auge de las revisiones sistemáticas y el metaanálisis es el cambio más prominente que se ha producido en el uso de la estadística en las revistas médicas en la última década.
- La revisión sistemática es en sí misma un diseño de investigación, es un diseño observacional y retrospectivo, que sintetiza los resultados de múltiples investigaciones primarias.
- Las revisiones sistemáticas son un método imprescindible para mantenerse al día, dada la inmanejable cantidad de información clínica producida actualmente.

ra que una parte de la variabilidad es debida a que, entre los diferentes estudios, hay diferencias subyacentes en el efecto. Aunque ambos métodos proporcionan estimaciones distintas del efecto global, en particular el modelo de efectos aleatorios da lugar a intervalos de confianza más anchos, la diferencia sólo es importante si los estudios son muy heterogéneos. En general, la decisión sobre qué método usar se suele hacer mediante un contraste de hipótesis de homogeneidad de los resultados, si son homogéneos se asume que el modelo de efectos fijos es el adecuado. La mayor limitación de este procedimiento es que se puede concluir que los resultados son homogéneos por falta de potencia estadística o, dicho de otro modo, no se debería ignorar la heterogeneidad simplemente aplicando una prueba estadística<sup>8</sup>. Otra cuestión a tener muy en cuenta es que una gran heterogeneidad puede indicar que el metaanálisis no es apropiado y no que haya que hacerlo usando el modelo de efectos aleatorios.

Los resultados de una revisión sistemática se suelen representar en una gráfica muy estandarizada y muy divulgada debido a que la Colaboración Cochrane la usa como logotipo. En la figura 2 se representan dos de estas gráficas que corresponden a dos revisiones sistemáticas hipotéticas de ensayos clínicos en los que se compara un tratamiento experimental (por ejemplo tratamiento con bloqueadores beta después de un infarto de miocardio) con otro control (por ejemplo placebo). En cada paciente el resultado se expresa con una variable binaria (mortalidad en los siguientes dos años). El resultado de cada ensayo se representa por uno de los índices habituales, por ejemplo el *odds ratio* (OR)<sup>9</sup>. El OR de cada ensayo se representa en la gráfica por un punto y su intervalo de confianza<sup>10</sup> al 95% con una línea horizontal, en escala logarítmica, para que el intervalo resulte centrado en la estimación puntual. Recuérdese<sup>9</sup> que si no hay efecto del tratamiento, el OR es 1; el convenio en las revisiones sis-



**Figura 2.** Representación gráfica del resultado de una revisión sistemática cuando el resultado en cada paciente se expresa con una variable binaria. Las distintas entradas en el eje Y representan los estudios individuales, para cada uno de ellos el punto muestra su resultado mediante alguno de los índices habituales, generalmente el *odds ratio*, y la línea horizontal su intervalo de confianza. En A se representa una situación en que los distintos estudios son homogéneos y por lo tanto tiene sentido realizar el metaanálisis, el intervalo de confianza del resultado global se suele representar mediante un rombo. Obsérvese como este intervalo es más estrecho que el de los estudios individuales, siendo esto uno de los objetivos del metaanálisis. En B una situación con resultados heterogéneos, en la que no estaría indicado el metaanálisis, lo pertinente sería explorar la razón de la heterogeneidad.

temáticas es construir el OR de modo que valores menores que 1, parte izquierda de la gráfica, correspondan a un efecto favorable del tratamiento. La gráfica contiene también una línea vertical de referencia en  $OR=1$ , valor de no efecto. Si se realiza metaanálisis, el OR combinado se representa por un rombo cuya diagonal horizontal representa su intervalo de confianza al 95%. En la figura 2A se observan 8 estudios con resultados homogéneos (todos encuentran efecto favorable del tratamiento experimental con OR comprendidos entre 0,4 y 0,7 y cada uno de ellos dentro de los intervalos de confianza de los otros), en seis de ellos, los intervalos cruzan la línea de no efecto, indicando que el efecto no es estadísticamente significativo. El rombo indica el efecto global ( $OR=0,55$ ) con un intervalo de confianza que establece que este efecto es claramente distinto de 1. La figura 2B muestra unos resultados muy heterogéneos, con los que no tendría sentido realizar metaanálisis, lo pertinente en este caso es explorar las razones de la heterogeneidad.

En estudios en que la variable respuesta es continua (por ejemplo presión arterial) el resultado en cada estudio es la diferencia de las medias entre ambos grupos, el resultado del metaanálisis es la media ponderada de las diferencias de medias y se construye una gráfica similar con las diferencias de medias y sus intervalos de confianza. En este caso la escala no es logarítmica y la línea de no efecto corresponde a diferencia de medias igual a 0.

Un elemento clave de las revisiones sistemáticas es la valoración de la calidad de los estudios. Aunque hay publicadas muchas escalas de valoración de la calidad metodológica de los distintos tipos de estudios, sobre todo de los ensayos clínicos, su comportamiento es muy poco concordante, algunas de ellas están incluso inversamente correlacionadas<sup>3</sup>, por lo tanto deben usarse con precaución. Un aspecto que necesita más desarrollo meto-

dológico es qué hacer luego con la información sobre la calidad, algunos autores usan las escalas de calidad como "pesos" en el metaanálisis, pero parece preferible examinar directamente la influencia en los resultados de distintos aspectos metodológicos mediante un análisis de sensibilidad.

Aunque la mayor parte de los metaanálisis publicados revisan ensayos clínicos, se empieza a extender su aplicación a otros tipos de estudios: estudios de pruebas diagnósticas, de pronóstico, de evaluación económica, incluso los propios metaanálisis (los denominados meta-metaanálisis) y en cada uno de ellos aparecen nuevos problemas metodológicos.

## BIBLIOGRAFÍA

- Phillips R, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, Dawes M. Levels of Evidence and Grades of Recommendations. Disponible en: <http://minerva.minervation.com/cebmdocs/levels.html> [consultado: 26/11/2002].
- Silagy CA. An analysis of review articles published in primary care journals. *Fam Pract* 1993;10:337-41.
- Altman DG. Statistics in medical journals: some recent trends. *Stat Med* 2000;19:3275-89.
- Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res* 1976;5:3-8.
- Andrews G, Guitart B, Howie P. Meta-analysis of the effects of stuttering treatment. *J Speech Hear Disord* 1980;45:287-307.
- Mulrow C, Cook D. Systematic Reviews. Synthesis of Best Evidence for Health Care Decisions. Philadelphia: American College of Physicians; 1998.
- Davidoff F, Case K, Fried PW. Evidence-Based Medicine: Why all the fuss? [Editorial]. *Ann Intern Med* 1995;122:727.
- Egger M, Smith GD, Phillips AN. Meta-analysis: Principles and procedures. *BMJ* 1997;315:1533-7.
- Abraira V. Medidas del efecto de un tratamiento (II): odds ratio y número necesario para tratar. *SEMERGEN* 2001;27:418-20.
- Abraira V. Estimación: intervalos de confianza. *SEMERGEN* 2002;28:84-5.