

*Departamento de Medicina Social, Universidad de Bristol, Bristol BS8 2PR
Matthias Egger, lector en medicina social y epidemiología
George Davey Smith, profesor de epidemiología clínica*

*Departamento de Atención Primaria y Ciencias Poblacionales, Escuela de Medicina del Hospital
Royal Free, Londres NW3 2PF*

Andrew N Phillips, profesor de epidemiología y bioestadística Correspondencia: Dr. Matthias

*Egger, m.egger@bristol.ac.uk
BMJ1997; 315: 1533-7.*

El meta-análisis es un procedimiento estadístico que integra los resultados de diversos estudios independientes que se consideran "combinables". Un meta-análisis bien realizado permite una evaluación más objetiva de la evidencia que la que se obtiene con una revisión narrativa tradicional, proporciona una estimación más precisa del efecto del tratamiento y puede explicar la heterogeneidad entre los resultados de los estudios individuales. Por otra parte, un meta-análisis mal realizado puede estar sesgado debido a la exclusión de estudios relevantes o a la inclusión de estudios inadecuados. En general, los análisis erróneos pueden evitarse si se siguen una serie de principios básicos. Estos principios se discuten a lo largo de este artículo, junto con las etapas prácticas en la realización de un meta-análisis.

El estudio observacional de la evidencia

El meta-análisis debería ser considerado como un estudio observacional de la evidencia. Las etapas implicadas son similares a las que se emprenden en cualquier tipo de investigación: formulación del problema que se quiere abordar, recogida y análisis de los datos, e informe de resultados. Los investigadores deben escribir previamente un protocolo detallado de la investigación en el que se establezcan claramente los objetivos, las hipótesis que se quieren evaluar, los subgrupos de interés y los métodos y criterios propuestos para identificar y seleccionar los estudios relevantes y para extraer y analizar la información.

Igual que en el caso de los criterios de inclusión y exclusión para pacientes individuales en los estudios clínicos, se deben definir los criterios de selección de los datos a incluir. Estos criterios se relacionan con la calidad de los ensayos y con la "combinabilidad" de los tratamientos, los pacientes, los resultados y los períodos de seguimiento. La calidad y el diseño de un estudio pueden influir en sus resultados. De forma ideal, los investigadores deberían incluir exclusivamente ensayos controlados con una aleatorización adecuada de los pacientes y que presenten los resultados de todos los pacientes incluidos según el principio de intención de tratar, con una evaluación objetiva, preferiblemente ciega, de los resultados. Sin embargo, evaluar la calidad de un estudio puede ser un proceso subjetivo, dado que la información publicada suele ser inadecuada para este propósito. Por tanto, es preferible definir sólo criterios de inclusión básicos y realizar un análisis de sensibilidad minucioso (ver más adelante).

Se debe establecer claramente la estrategia para identificar los estudios relevantes. En particular, se debe decidir si la búsqueda se va a extender a estudios no publicados, ya que sus resultados pueden diferir de forma sistemática de los ensayos publicados. Como se discutirá en artículos posteriores, un meta-análisis que se limite a la evidencia publicada puede producir resultados distorsionados, ya que existe un sesgo de publicación. Las bases de datos electrónicas son útiles para localizar los estudios publicados, pero si sólo se utilizan estas bases se puede perder una parte sustancial de los estudios relevantes. En un intento de identificar todos los ensayos controlados publicados, la Colaboración Cochrane ha emprendido una búsqueda manual extensa en las revistas médicas publicadas en inglés y en otras muchas lenguas". El Registro Cochrane de Ensayos Controlados es probablemente la mejor fuente electrónica de ensayos; sin embargo, también deben escrutarse los índices de citas y las bibliografías incluidas en los artículos de revisión, en monografías, y en los estudios localizados.

Se necesita un formulario estandarizado de registro para la recogida de datos. Es útil si se utilizan dos observadores independientes para extraer los datos, ya que se evitan errores. En esta etapa se puede evaluar la calidad de los estudios, utilizando alguna de las diferentes escalas diseñadas específicamente para este propósito'. Además, se obtienen puntuaciones más consistentes si los observadores no conocen los nombres de los autores, sus instituciones, los nombres de las revistas, las fuentes de financiación, o los agradecimientos'. Esto se puede conseguir, bien fotocopiando los artículos sin los títulos y enmascarando el nombre de la revista y otras características identificativas con un rotulador, bien escaneando el texto mediante ordenador y preparando formatos estandarizados

Puntos esenciales

El meta-análisis debería ser cuidadosamente planificado, al igual que se hace con cualquier otro tipo de proyecto de investigación, con un protocolo detallado escrito con antelación

La definición a priori de los criterios de selección de los estudios que se van a incluir y la búsqueda exhaustiva y completa de tales estudios son elementos cruciales para un meta-análisis de alta calidad

La presentación gráfica de los resultados de los estudios individuales en una escala común es un paso intermedio importante, que permite examinar visualmente el grado de heterogeneidad que existe entre los estudios
Existen diferentes métodos estadísticos para combinar los datos, pero no existe un único método "correcto"

Un análisis de sensibilidad exhaustivo es esencial para evaluar la robustez de las estimaciones combinadas, en función de las diferentes asunciones y criterios de inclusión.

La medida estandarizada de los resultados

Los resultados de los estudios individuales tienen que ser expresados en un formato estándar para que se puedan realizar comparaciones entre los estudios. Si la medida del resultado es una variable continua - por ejemplo, la presión arterial - se utiliza la diferencia de las medias entre los grupos de tratamiento y control. Sin embargo, la magnitud de la diferencia está influida por el nivel de la variable subyacente en la población. Por ejemplo, es más probable que un fármaco antihipertensivo tenga un mayor efecto absoluto en yacientes con hipertensión severa que en pacientes con hipertensión leve. Por lo tanto, las diferencias se presentan a menudo como unidades de desviación estándar. En el caso de variables de resultado binarias - por ejemplo, enfermedad / no enfermedad, o vivo / muerto - se calculan habitualmente odds ratios o riesgos relativos (ver cuadro). El odds ratio viene propiedades matemáticas convenientes, que facilitan la combinación de los datos y a evaluación de la significación estadística de los resultados combinados. Las medidas absolutas, tales como la reducción del riesgo absoluto o el número de pacientes que es necesario tratar para prevenir un suceso son más útiles cuando se aplican los resultados a la práctica clínica (ver más adelante).

Odds ratio o riesgo relativo?

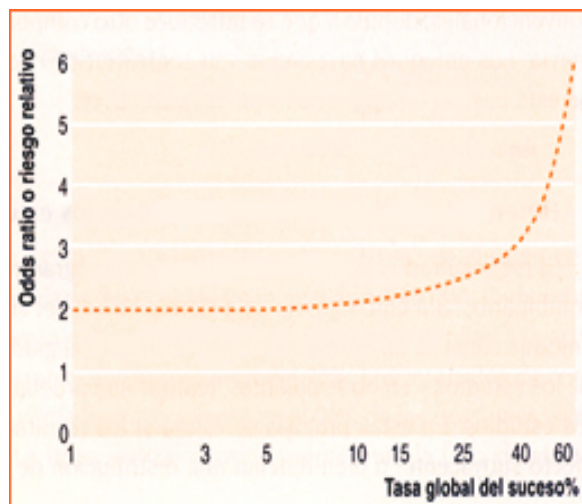
Ods y odsratio

El odds es el número de pacientes que cumplen los criterios de un determinado suceso dividido por el número de pacientes que no los cumplen. Por ejemplo, el odds de diarrea durante el tratamiento con un antibiótico en un grupo de 10 pacientes puede ser 4/6 (cuatro con diarrea dividido por seis sin diarrea, 0,66); en el grupo control, el odds puede ser 1/9 (0,11) (un corredor de apuestas se referiría a esta situación como 9 a 1). El odds ratio del grupo de tratamiento con respecto al grupo control sería 6 (0,66/0,11).

Riesgo y riesgo relativo

El riesgo es el número de pacientes que cumplen los criterios de un determinado suceso dividido por el número total de pacientes. En el ejemplo anterior, los riesgos serían 4/10 en el grupo de tratamiento y 1/10 en el grupo control, dando un riesgo relativo, o razón de riesgos, de 4 (0,4/0,1).

El odds ratio será similar al riesgo relativo si el suceso es y relativamente infrecuente, por ejemplo en menos del 20%. Si el resultado es más corriente (como en el ejemplo de la diarrea) entonces el odds ratio sobrestimará de forma considerable el riesgo relativo.



Métodos estadísticos para calcular el efecto combinado

El último paso consiste en calcular el efecto global mediante la combinación de los datos. Una media aritmética simple de los resultados procedentes de todos los estudios produciría resultados erróneos. Los resultados de estudios pequeños están más sujetos al azar y por tanto se les debería otorgar un peso menor. Los métodos utilizados en meta-análisis usan una media ponderada de los resultados, dando mayor influencia a los ensayos grandes que a los pequeños. Las técnicas estadísticas que se utilizan para esto se pueden clasificar, a grandes rasgos, en dos modelos, dependiendo de cómo se trate la variabilidad de los resultados entre los estudios. El modelo de "efectos fijos" considera que esta variabilidad inter-estudio se debe exclusivamente al azar, una asunción a menudo difícilmente justificable". Por tanto, si todos los estudios fueran infinitamente grandes, tendrían los mismos resultados. El modelo de "efectos aleatorios" asume un efecto subyacente diferente para cada estudio y lo considera como una fuente adicional de variabilidad, lo que conduce a intervalos de confianza más amplios que el modelo de efectos fijos. Se asume que los efectos subyacentes se distribuyen al azar y la estimación del efecto combinado consiste en estimar el punto central de esta distribución. Aunque ninguno de los dos métodos puede considerarse "correcto", sólo se observará una diferencia sustancial en el cálculo del efecto combinado entre ambos métodos si los estudios son marcadamente heterogéneos".

Meta-análisis Bayesiano

Algunos estadísticos piensan que otros enfoques estadísticos son más adecuados que los dos modelos expuestos anteriormente. Una aproximación utiliza el teorema de Bayes, así denominado en honor de un clérigo inglés del siglo XVIII. Los estadísticos bayesianos expresan su creencia acerca del tamaño de un efecto mediante una distribución de probabilidades previa al conocimiento de los datos, y actualizan esa creencia calculando una distribución de probabilidades posterior, que tiene en cuenta los datos observados. Los modelos bayesianos pueden utilizarse tanto bajo las asunciones de efectos fijos como las de efectos aleatorios. El intervalo de confianza (o más correctamente, según la terminología bayesiana, el intervalo de credibilidad del 95%, que cubre el 95% de la distribución de probabilidades posterior) a menudo será más amplio que el que se deriva de los modelos convencionales, debido a que se introduce otro componente de variabilidad, la distribución previa. Los enfoques bayesianos son controvertidos porque la distribución previa a menudo está basada en opiniones y juicios subjetivos.

Heterogeneidad entre los resultados de los estudios

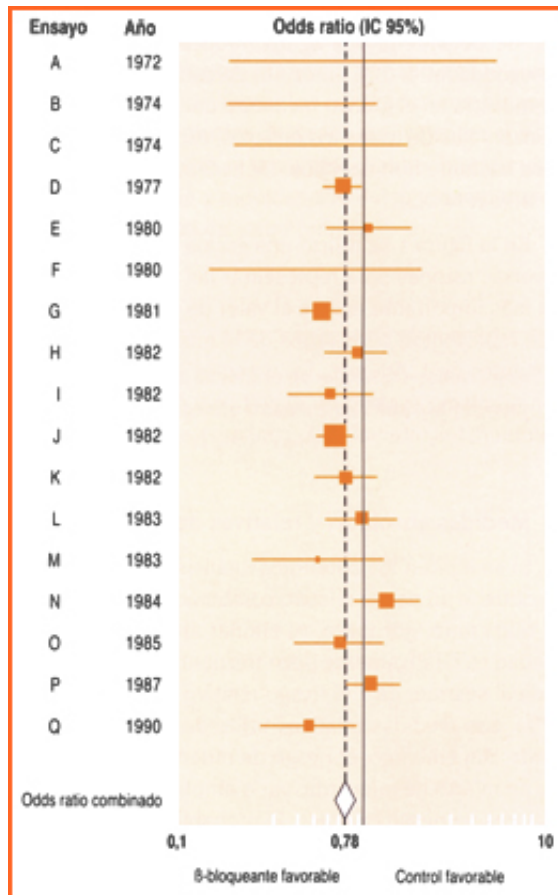
Si los resultados de los estudios difieren en gran medida puede que no sea adecuado combinarlos. Sin embargo, no está claro cómo saber si es apropiado o no combinarlos. Un enfoque consiste en examinar estadísticamente el grado de semejanza entre los resultados de los estudios - en otras palabras, realizar una prueba estadística de heterogeneidad entre los estudios. En estas pruebas se valora si los resultados de un estudio reflejan un único efecto subyacente, o bien reflejan una distribución de efectos. Si la prueba muestra result-

tados homogéneos, entonces se asume que las diferencias entre los estudios son consecuencia de la variación del muestreo y es adecuado utilizar un modelo de efectos fijos. Sin embargo, si la prueba muestra que existe una heterogeneidad significativa entre los resultados de los estudios, entonces resulta más apropiado utilizar un modelo de efectos aleatorios. Una limitación importante de este enfoque es la falta de potencia de las pruebas estadísticas de heterogeneidad, ya que a menudo no rechazan la hipótesis nula de resultados homogéneos, incluso cuando existen diferencias sustanciales entre los estudios. Aunque no existe una solución estadística para este tema, Odds ratio (IC 95%) heterogeneidad entre los resultados de los estudios no debe considerarse sólo un problema para el propio meta-análisis, ya que también proporciona una oportunidad de examinar por qué los efectos del tratamiento difieren según las circunstancias. Después de aplicar la prueba estadística, la heterogeneidad debería intentar explicarse.

Figura 1. Mortalidad total en los ensayos con (3-bloqueantes para la prevención secundaria después de un infarto de miocardio. El cuadrado negro y la línea horizontal corresponden al odds ratio y al intervalo de confianza del 95% de cada ensayo. El tamaño de cada cuadrado negro 1984 1 negro refleja el peso de cada ensayo. El diamante representa el odds ratio combinado y el intervalo de confianza del 95%, que muestra una reducción del 22% en la odds de muerte (las referencias se pueden conseguir de los autores)

Presentación gráfica

Una manera útil de presentar los resultados de cada ensayo y sus intervalos de confianza consiste en hacerlo de forma gráfica. La figura 1 representa un meta-análisis de ensayos con β-bloqueantes en la prevención secundaria después de un infarto de miocardio. Cada estudio se representa por un cuadrado negro y una línea horizontal, que corresponden a la estimación puntual del odds ratio y a su intervalo de confianza del 95%. Los intervalos de confianza del 95% contendrían el efecto verdadero subyacente en el 95% de las ocasiones, si el estudio se repitiera una y otra vez. La línea vertical continua representa la falta de efecto



del tratamiento (odds ratio = 1). Si el intervalo de confianza incluye el 1, entonces la diferencia entre el efecto en el grupo experimental y el grupo control no es significativa a niveles convencionales ($P > 0,05$). El área del cuadrado negro refleja el peso del estudio en el meta-análisis. El intervalo de confianza de todos los estudios, excepto dos, cruza la línea de no efecto, lo que indica que los efectos estimados fueron no significativos ($P > 0,05$).

El diamante representa el odds ratio combinado, que se ha calculado utilizando un modelo de efectos fijos, junto con su intervalo de confianza del 95%. El odds ratio combinado muestra que los f3-bloqueantes orales, tomados a los pocos días o a las pocas semanas después de la fase aguda, reducen la mortalidad subsiguiente en una estimación del 22% (odds ratio, 0,78; intervalo de confianza del 95%, 0,71 a 0,87). El odds ratio combinado se muestra en el gráfico mediante una línea vertical discontinua. Esta línea cruza las líneas horizontales de todos los ensayos, excepto la de uno (N). Esto indica un conjunto de estudios bastante homogéneos. De hecho, el test de heterogeneidad da un valor de P no significativo ($P = 0,2$).

En la figura 1 se utilizó una escala logarítmica para representar los odds ratios. Existen diversas razones para representar las medidas de cociente (ratios) en escala logarítmica. La más importante es que el valor de un odds ratio y su recíproco - por ejemplo, 0,5 y 2 -, que representan odds ratios de la misma magnitud pero de dirección opuesta, serán equidistantes de 1. Además, los estudios con odds ratios por encima y por debajo de 1 utilizarán el mismo espacio en el gráfico y por tanto parecerá que tienen la misma importancia. También, los intervalos de confianza serán simétricos alrededor de la estimación puntual.

Medidas absolutas y relativas del efecto

Si se repiten los anteriores análisis utilizando el riesgo relativo en lugar del odds ratio, se obtiene un riesgo relativo combinado de 0,8 (intervalo de confianza del 95%, 0,73 a 0,88). El odds ratio, por tanto, es similar al riesgo relativo, tal como se espera cuando la enfermedad es relativamente poco frecuente (ver cuadro). La reducción del riesgo relativo, obtenida al sustraer de 1 el riesgo relativo y expresar el resultado en porcentaje, es 20% (12% a 27%). Las medidas relativas utilizadas en este análisis ignoran el riesgo absoluto subyacente. Sin embargo, el riesgo de muerte entre los pacientes que sobreviven a la fase aguda de un infarto de miocardio varía ampliamente. Por ejemplo, entre pacientes con tres o más factores de riesgo cardiaco, la probabilidad de muerte a los dos años del alta varía entre un 24% y un 60%. Por el contrario, la mortalidad a los 2 años en pacientes sin factores de riesgo es menor del 3%. La reducción del riesgo absoluto o la diferencia de riesgos refleja tanto el riesgo subyacente sin tratamiento como la reducción del riesgo asociada al tratamiento. El recíproco de la diferencia de riesgos es el "número que es necesario tratar" ("number needed to treat"), el número de pacientes que se necesitan tratar para prevenir un suceso.

Para un riesgo base) del 1 % al año, la diferencia del riesgo absoluto muestra que se previenen dos muertes por cada 1000 pacientes tratados (ver tabla). Esto equivale a tratar 500 pacientes (1 [1 0,002] cada año para prevenir una muerte. Por el contrario, si el riesgo es mayor del 10%, se necesitan tratar menos de 50 pacientes para prevenir una muerte.

Muchos clínicos probablemente decidan no tratar a los pacientes con muy bajo riesgo, dado el gran número de pacientes que se tiene que exponer a los efectos adversos de los -bloqueantes para prevenir una muerte. Valorar el número de pacientes que es necesario tratar en función del riesgo estimado del paciente sin tratamiento y de la reducción del riesgo relativo por el tratamiento es una ayuda en la toma de decisiones sobre pacientes individuales. Se ha publicado recientemente un nomograma que facilita el cálculo, a pie de cama, del "número que es necesario tratar".

El meta-análisis de medidas absolutas del efecto, tales como la diferencia de riesgos, puede ser útil para ilustrar el rango de efectos absolutos en los diferentes estudios. Sin embargo, la diferencia de riesgos combinada (y el "número que es necesario tratar" correspondiente) estará determinada principalmente por el número y tamaño de los ensayos en pacientes de bajo, intermedio o alto riesgo. Por tanto, los resultados combinados serán aplicables exclusivamente a los pacientes con niveles de riesgo equivalentes al riesgo promedio de los ensayos incluidos. Por esto, tiene más sentido habitualmente usar medidas relativas del efecto para resumir la evidencia y medidas absolutas para aplicar la evidencia a una situación concreta, tanto clínica como de salud pública.

-bloqueantes en la prevención secundaria después de un infarto de miocardio-reducciones del riesgo absoluto y número de pacientes que es necesario tratar en un año para prevenir una muerte, según diferentes niveles de mortalidad en el grupo control

RIESGO DE MORTALIDAD EN UN AÑO ENTRE CONTROLES (%)	REDUCCIÓN DEL RIESGO ABSOLUTO	NÚMERO QUE ES NECESARIO TRATAR
1	0,002	500
3	0,006	167
5	0,01	100
10	0,02	50
20	0,04	25
30	0,06	17
40	0,08	13
50	0,1	10

Los cálculos asumen una reducción constante del riesgo relativo del 20%.

Análisis de sensibilidad

Con frecuencia existirán opiniones divergentes acerca de cuál es el método correcto para llevar a cabo un meta-análisis en concreto. Por tanto, siempre se debe examinar la robustez de los hallazgos mediante un análisis de sensibilidad exhaustivo que considere diferentes asunciones. Esto se ilustra en la figura 2 para el meta-análisis de los β -bloqueantes después de un infarto de miocardio. En primer lugar, se calculó el efecto global mediante diferentes métodos estadísticos, utilizando tanto modelos de efectos fijos como de efectos aleatorios. La figura muestra que las estimaciones combinadas son virtualmente idénticas y que los intervalos de confianza son sólo ligeramente más amplios con el modelo de efectos aleatorios. Esto se explica por la relativamente poca variabilidad entre los ensayos de este meta-análisis.

En segundo lugar, la calidad metodológica de los ensayos fue evaluada en función de cómo fueron asignados los pacientes al grupo de tratamiento activo o al de control, de cómo se evaluó el resultado y de cómo se analizaron los datos. Se dio un máximo de nueve puntos si la asignación del paciente fue verdaderamente aleatoria, si la evaluación del estado vital fue independiente del grupo de tratamiento y si los datos de todos los pacientes inicialmente incluidos fueron analizados de acuerdo con el principio de intención de tratar. La figura 2 muestra que los tres estudios de baja calidad (< 7 puntos) mostraron un efecto más beneficioso que los ensayos de alta calidad. Sin embargo, la exclusión de esos tres estudios deja prácticamente inalterados el efecto combinado y los intervalos de confianza.

En tercer lugar, los resultados significativos tienen mayor probabilidad de ser publicados que los hallazgos no significativos y esto puede distorsionar los resultados del meta-análisis. La presencia de este sesgo de publicación puede detectarse mediante la estratificación del análisis según el tamaño del estudio - los efectos más pequeños pueden ser significativos en los estudios más grandes. Si existe un sesgo de publicación, se espera que, entre los estudios publicados, los más grandes presentarán los efectos más pequeños. La figura 2 confirma esta situación, ya que los ensayos más pequeños (50 muertes o menos) muestran el efecto más grande. Sin embargo, la exclusión de los estudios más pequeños tiene poco efecto sobre la estimación combinada.

En último lugar, dos estudios (1 y N, ver figura 1) fueron interrumpidos antes de lo previsto debido a los resultados procedentes de análisis intermedios. Las estimaciones de los efectos de los tratamientos en los estudios que se interrumpen prematuramente suelen estar sesgadas, con valores que se alejan del valor nulo. Esto implica la introducción de sesgos en los meta-análisis que incluyen estos ensayos. Sin embargo, la exclusión de estos ensayos sólo afecta de forma marginal la estimación combinada.

Por tanto, los análisis de sensibilidad muestran que los resultados de este meta-análisis son robustos, independientemente de la elección del método estadístico, de la exclusión de ensayos de menor calidad o de la de los estudios que finalizan antes de tiempo. También sugieren que es improbable que el sesgo de publicación distorsione los hallazgos.

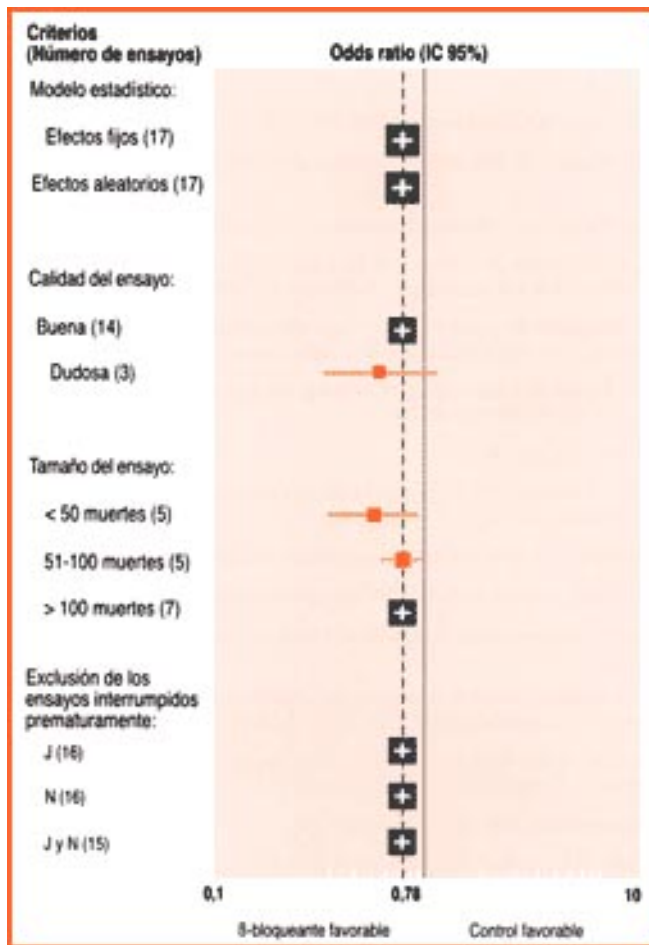


Figura 2 Análisis de sensibilidad del meta-análisis de los β -bloqueantes en la prevención secundaria después de un infarto de miocardio (ver texto para explicación)

Conclusiones

El meta-análisis debe ser visto como la estructuración del proceso a través del cual se realiza una revisión de la investigación existente. La exhaustividad de la revisión y el grado de "combinabilidad", que deben ser considerados en cualquier revisión, se abordan de forma explícita. ¿Fue razonable combinar los ensayos individuales que componen un meta-análisis? ¿Es el resultado robusto en relación con los cambios en las asunciones? La conclusión alcanzada, ¿tiene sentido desde el punto de vista clínico y fisiopatológico? Finalmente, ¿ha contribuido el análisis al proceso de toma de decisiones lógicas en el manejo de los pacientes? Estos temas se explorarán con más detalle en los siguientes artículos de esta serie.

El Departamento de Medicina Social de la Universidad de Bristol y el Departamento de Atención Primaria y de Ciencias Poblacionales de la Escuela de Medicina del Hospital Royal Free de Londres son parte de la colaboración sobre Investigación de Servicios Sanitarios del Consejo de Investigación Médica. Financiación: La Fundación Nacional Suiza de la Ciencia apoyó a ME.

Referencias

1. Huque MF. Experiences with meta-analysis in NDA submissions. Proceedings of the Biopharmaceutical Section of the American Statistical Association 1988;28-33.
2. Egger M, Davey Smith G. Meta-analysis: potentials and promise. *BMI* 1997;315:1371-4.
3. Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *BMI* 1997;315:629-34.
4. Sacks H, Chalmers TC, Smith HI. Randomized versus Historical controls for clinical trials. *Ann N Y Acad Sci* 1982;472:233-40.
5. Schulz KF, Chalmers I, Hayes RI, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
6. Prendiville W, Elbourne D, Chalmers I. The effects of routine oxytocic administration in the management of the third stage of labour: an overview of the evidence from controlled trials. *Br J Obstet Gynaecol* 1988;95:3-16.
7. Begg CB, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
8. Greenhalgh T. The Medicine database. *BMI* 1997;315:180-3.
9. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: a comparison of Medicine searching with a perinatal clinical trials database. *Controlled Clinical Trials* 1985;6:306-317.
10. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMI* 1994;309:1286-91.
11. Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *BMI* 1992;305:786-8.
12. The Cochrane Controlled Trials Register. I: Cochrane Library. CD ROM and online. Cochrane Collaboration Issue 1). Oxford: Update Software, 1997.
13. Moher D, Jadad AR, Nichol G, Penman M, Tugwell, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995;16:62-73.
14. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DIM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials* 1996; 17:1-12.
15. Chalmers TC. Problems induced by meta-analyses. *Stat Med* 1991;10:971-80.
16. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;347:363-6.
17. Laupacis A, Sackett DL, Roberts RS. An Assessment of clinically useful measures of the consequences of treatment. *New Engl J Med*. 1988;318:1728-33.
18. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining even rates from clinical trials. *Stat Med* 1989;8:141-51.
19. Yusuf S, Petp R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985;17:335-71.
20. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;7:177-88.
21. Carlin IB. Meta-analysis for 2x2 tables: a bayesian approach. *Stat Med* 1992;11:141-58.
22. Lilford RJ, Braunholtz D. The statistical basis of public policy: a paradigm shift is overdue. *BMI* 1996;313:603-7.
23. Eddy DM, Hasselblad S, Shachter R. Meta-analysis by the confidence profile method. The statistical synthesis of evidence. Boston: Academic Press, 1992.
24. Bailey KR. Inter-study differences: how should they influence the interpretation and analysis of results? *Stat Med* 1987;6:351-8.
25. Gallbraith R. A note on graphical presentation of estimated odd ratios from several clinical trials. *Stat Med*. 1988;7:889-94.
26. Multicenter Postinfarction Research Group. Risk stratification and survival after myocardial infarction. *New Engl J Med* 1983;309:331-6.
27. Chatellier G, Zapletal E, Lemaire D, Menard I, Degoulet P. The number needed to

- treat: a clinically useful nomogram in its proper context. *BMI* 1996;312:426-9.
28. Easterbrook PJ, Berlin IA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72. 29. Green S, Fleming TR, Emerson S. Effects on overviews of early stopping rules for clinical trials. *Stat Med* 1987;6:361-7. 30. Oxman AD. Checklist for review articles. *BMI* 1994;309:648-51.