

METAANÁLISIS EN EPIDEMIOLOGÍA (SEGUNDA PARTE): MÉTODOS CUANTITATIVOS

Miguel Delgado Rodríguez* ** / María Sillero Arenas** / Ramón Gálvez Vargas*

*Cátedra de Medicina Preventiva, Facultad de Medicina / **Hospital «Virgen de las Nieves» Granada / ***Delegación Provincial de Salud y Consumo Jaén.

Resumen

El objetivo del presente trabajo es poner al alcance del lector interesado en el tema los principales métodos cuantitativos del metaanálisis que se han aplicado en el terreno sanitario, haciéndose especial hincapié en los más utilizados. Se ofrece también un pequeño resumen de los métodos más difundidos para determinar la presencia de un sesgo de publicación. Se discuten las ventajas e inconvenientes de cada procedimiento detallado.

Palabras clave: Metaanálisis. Métodos epidemiológicos. Revisiones.

METHA-ANALYSIS IN EPIDEMIOLOGY (SECOND PART): QUANTITATIVE METHODS

Summary

The main goal of this paper is to offer an introduction to the quantitative methods of meta-analysis used in epidemiology/public health, emphasizing on the most widespread used. We give also a brief reference of the methods to recognize publication bias. The limitations and advantages of the methods commented on are discussed.

Key words: Metha-analysis. Epidemiological methods. Reviews.

Introducción

El propósito que persigue la presente revisión es poner al alcance del lector interesado en el tema los métodos más frecuentes que se han utilizado y se utilizan en la combinación de resultados dentro del terreno sanitario. Los sanitarios, a diferencia de otros investigadores, generalmente analizan variables dicotómicas. Los tipos de pruebas dedicados a combinar resultados sobre este tipo de variables no se suelen encontrar en los textos más difundidos sobre metaanálisis. Así, el libro de Light y Pillémer¹, excelente para comprender lo que significa el metaanálisis y para iniciarse en él, se dedica, cuando trata los procedimientos estadísticos a emplear, a la llamada «magnitud del efecto» (*effect size*), que ha sido de rara utilización en la aplicación del metaanálisis al terreno epidemiológico. Otro ejemplo de ello lo constituye el texto de Hedges y Olkin², que se dedica exclusivamente a los métodos estadísticos sobre variables cuantitativas. Estos autores, centrados en su labor investigadora en el terreno de la Educación, focalizan los métodos que describen en el análisis de la «magnitud del efecto», no tratando en ningún momento los parámetros que en Salud Pública se manejan, tales como el riesgo relativo o el riesgo atribuible. A diferencia de los anteriores, el texto de Jenicek³, conocido autor por sus tres textos anteriores en el terreno de la epidemiología, aborda también los métodos de metaanálisis. Este autor, sin embargo, sólo hace referencia a uno de los procedimientos más utilizados en lo que podríamos llamar el metaanálisis epidemiológico: el método

Yusuf *et al.* (que los propios autores refieren como el método de Peto). De los demás métodos comentados por el autor no hemos podido encontrar en los últimos tres años ninguna aplicación en medicina. La revisión de Greenland⁴ sobre los métodos cuantitativos en la revisión de la literatura epidemiológica se centra prácticamente con exclusividad en el método de Woolf. Por todos estos hechos, se ha querido plasmar en esta revisión los métodos cuantitativos más difundidos y simplificar así la tarea para futuros investigadores que deseen iniciarse en el tema.

Métodos de combinación de los valores $p^{5,6}$

Sea p_i el valor de significación de una cola de estudio i y n en el número de estudios.

a) Estadístico de Fisher:

$$Z_1 = -2 \sum_{i=1}^n (1n p_i)$$

siendo Z_1 el valor de χ^2 con $2n$ grados de libertad.

b) Estadísticos de Edington:

b1) Cálculo de un valor p de significación cuando el número de estudios que se combinan es pequeño:

$$p = \left[\sum_{i=1}^n p_i \right]^n (n!)^{-1}$$

b2) Cuando el número de estudios a combinar es mayor, también puede calcularse un valor Z_2 a través del cálculo previo de un valor

Correspondencia: Miguel Delgado Rodríguez. Medicina Preventiva Facultad de Medicina. Avda. Madrid 11. 18012-GRANADA. Este artículo fue recibido el 6 de marzo de 1991 y fue aceptado tras revisión, el 11 de octubre de 1991.

Tabla 1. Estructura y notación de un estudio / *

		Enfermedad		Total
		Sí	No	
Exposición	Sí	a _i	b _i	n _{1i}
	No	c _i	d _i	n _{0i}
Total		m _{1i}	m _{0i}	N _i

*NOTA: Esta distribución puede servir para un estudio: a) de cohortes que valora proporción de incidencia, siendo n_{1i} y n_{0i} las cohortes expuestas y no expuestas; b) experimental, siendo n_{1i} y n_{0i} los grupos experimentales y control; o c) de casos y controles, siendo m_{1i} y m_{0i} los grupos de casos y controles.

ρ promedio:

$$p = \sum_{i=1}^n p_i / n$$

$$Z_2 = (0,5 - p) \sqrt{(12n)}$$

siendo Z₂ un valor Z que sigue una distribución normal.

Tanto la prueba Fisher como las dos de Edington utilizan un mínimo de información. En primer lugar, permiten la combinación de valores p pertenecientes a estudios que muestran efectos con sentidos opuestos. Por ejemplo, es posible que se alcance una significación del 5% (un valor Z₁ de 1,96) con un riesgo relativo de 0,5 o con un riesgo relativo de 2. Estos procedimientos combinan los dos estudios sin tener en cuenta el sentido de la asociación. El valor p es una medida de rareza, sin tener para nada en cuenta el sentido de la asociación global, con respecto a una hipótesis y no permite hablar o pensar de nada en términos de evidencia, sino sólo en contra de ella. Bajo la hipótesis nula, un riesgo relativo pequeño en un gran estudio puede ser tan raro como un riesgo relativo grande en uno pequeño, produciendo la misma Z₁ y el mismo valor p. El efecto absoluto (y su intervalo de confianza), una parte esencial de la evidencia, no se refleja en la Z₀ o en el valor p, de tal manera que en el metaanálisis se deben combinar los estimadores del efecto (el riesgo relativo u otras medidas) y los estadísticos de manera separada. Es necesario utilizar procedimientos que cuantifiquen la probabilidad con que se favorece una hipótesis u otra⁷. Otro defecto que tienen estos métodos es que no se puede estudiar la heterogeneidad existente entre los distintos estudios que se combinan.

Métodos de combinación de pruebas estadísticas^{5,6}

a) Combinación de valores t

$$[\sum_{i=1}^n t_i] [\sqrt{\sum_{i=1}^n \{ (g_i / (g_i - 2)) \}}]^{-1}$$

que sigue una distribución N (0,1). g_i indica grados de libertad.

b) Combinación de valores z

$$[\sum_{i=1}^n z_i] / \sqrt{n}$$

que sigue una distribución normal N (0,1). Tienen graves limitaciones, la más seria de las cuales es que cuanto mayor sea el número de estudios es posible que disminuya la significación

Tabla 2. Estructura y notación de un estudio de cohortes de densidad

		Enfermos	Población/Tiempo
		a _i	P _{1i}
Exposición	Sí	a _i	P _{1i}
	No	c _i	P _{0i}
Total		m _{1i}	P _i

estadística. También es posible ponderar la fórmula anterior por varios tipos de pesos, w (tamaño muestral, calidad del estudio, grados de libertad, etc.):

$$Z_3 = [\sum_{i=1}^n w_i z_i] / \sqrt{ \sum_{i=1}^n w_i^2 }$$

Este método fue propuesto por Mosteller y Bush en 1954. Los mismos comentarios que se han hecho con ocasión del apartado anterior son aplicables aquí. A pesar de sus limitaciones, aún se siguen utilizando. Un ejemplo de ello es el trabajo de Needleman y Gatsonis⁸ que estudiaron la relación entre la exposición a bajas concentraciones de plomo y el coeficiente intelectual de los niños.

Métodos basados en un modelo subyacente del efecto

Métodos para efectos dicotómicos (o binarios)

El principio del método consiste en la definición de una medida de asociación entre la exposición y el efecto, m_i, en cada estudio a combinar, de tal manera que si no hay asociación m_i sea igual a cero. Estos métodos se deben aplicar con precaución ya que se basan en condiciones asintóticas. Cuando los números de la tabla tetracórica son pequeños, los resultados pueden variar mucho dependiendo del método elegido; no se debe olvidar la asunción básica que subyace en estos métodos

$$\sqrt{(m^2 / \text{var}(m))} \approx \rightarrow N(0,1)$$

Si la medida que se elige es la diferencia de riesgos o tasas (el riesgo atribuible en los expuestos) no es necesaria ninguna transformación. Otro tipo de medidas necesitan una transformación para que se cumpla ese criterio (por ejemplo, la transformación logarítmica en el caso de riesgo relativo). Antes de comentar los métodos de combinación es necesario recordar los principales parámetros de cálculo epidemiológico, su varianza y su intervalo de confianza.

La tabla 1 es la distribución simple de un estudio de cohortes acumulado, de casos y controles o de prevalencia, mientras que la tabla 2 es la utilizada para los estudios de cohortes de densidad. Los estudios experimentales son similares a los estudios de cohortes. En la tabla 3 se ofrece el parámetro de asociación, su cálculo, su varianza y su intervalo de confianza para el 95%.

Métodos que se basan en el riesgo relativo (RR) y derivados

Método de Woolf⁹

Este método usa el inverso de la varianza como criterio de ponderación en el cálculo del parámetro resumen. Es uno de los

Tabla 3. Estimación de parámetros de asociación en diferentes diseños epidemiológicos

Modelo	Tipos de estudio		
	Cohortes acumulado	Cohortes de densidad	Casos y controles o de prevalencia
		<i>Medidas de fuerza de asociación</i>	
Parámetro	Riesgo relativo (RR)	Razón de densidades de incidencia (RDI)	Desigualdad relativa (OR)
Cálculo	$a/n_{11} / (c_i/n_{0i})$	$a_i/P_{1i} / (c_i/P_{0i})$	$a_d / (b_i c_i)$
Varianza (V)*	$1/a_i + 1/c_i - 1/n_{1i} - 1/n_{0i}$	$1/a_i + 1/c_i$	$1/a_i + 1/b_i + 1/c_i + 1/d_i$
IC 95%	$RR_i \cdot e^{(\pm 1.96 \sqrt{V_i})}$	$RDI_i \cdot e^{(\pm 1.96 \sqrt{V_i})}$	$OR_i \cdot e^{(\pm 1.96 \sqrt{V_i})}$
		<i>Medidas de carácter administrativo</i>	
Parámetro	Riesgo atribuible en expuestos (RAe)	Tasa atribuible en expuestos (TAE)	No calculable
Cálculo	$a/n_{11} - c_i/n_{0i} = R_{1i} - R_{0i}$	$a_i/P_{1i} - c_i/P_{0i} = DI_{1i} - DI_{0i}$	
Varianza	$R_{1i}(1-R_{1i})/n_{1i} + R_{0i}(1-R_{0i})/n_{0i}$	$DI_{1i}/n_{1i} + DI_{0i}/n_{0i}$	
IC 95%	$RAe_i \pm 1.96 \sqrt{V_i}$	$TAE_i \pm 1.96 \sqrt{V_i}$	

* Varianza del logaritmo neperiano del parámetro de fuerza.

procedimientos más utilizados y es el que recomienda Greenland⁴ sobre los restantes. Su recomendación parece estar basada en el hecho de que puede utilizarse en cualquier tipo de diseño, sea observacional o experimental. Este procedimiento lo hemos aplicado repetidas veces¹⁰⁻¹³ en el estudio de las hormonas exógenas y los tumores de mama y cuello uterino.

Estimación de parámetros resumen

Para la combinación de resultados, dado que el RR y derivados no siguen una distribución normal, se utilizará la transformación logarítmica de los mismos. El peso de cada RR será dado en función del inverso de su varianza:

$$1n RR_{pw} = \frac{\sum_{i=1}^n 1 / [v_i (1n RR_i)] \cdot 1n RR_i}{\sum_{i=1}^n 1 / [v_i (1n RR_i)]}$$

siendo RR_p el riesgo relativo ponderado. Para calcularlo será necesario realizar la transformación antilogarítmica.

La varianza del logaritmo natural del RR_p es igual al inverso de la suma de los pesos de cada RR considerado.

Esto es

$$V(1n RR_{pw}) = 1 / \left\{ \sum_{i=1}^n 1 / [v_i (1n RR_i)] \right\}$$

Una vez conocidos el RR_p y su varianza, se aplicará la fórmula siguiente para calcular su intervalo de confianza:

$$IC\ 95\% \text{ del } RR_{pw} = RR_{pw} \cdot EXP(\pm 1.96 \sqrt{V})$$

La significación global del parámetro se puede estimar dividiendo el logaritmo natural del RR_p por su error estándar¹⁴:

$$X_{asoc}^2 = 1n RR_{pw} / \sqrt{V(1n RR_{pw})}$$

que aproximadamente sigue una distribución normal bajo la hipótesis que se investiga.

El grado de homogeneidad se valora mediante el estadístico de Woolf^{9,14}

X_w^2 , o también llamada,

$$X_{hom}^2 = \sum_{i=1}^n 1 / [V_i (1n RR_i)] (1n RR_i - 1n RR_{pw})^2$$

que sigue una distribución X^2 con K-1 grados de libertad, siendo K el número de estudios que se combinan.

En epidemiología observacional, el parámetro que se utiliza más frecuentemente es el RR y parámetros similares (OR, RDI). Dado que los tres parámetros existentes de fuerza de la asociación ni significan lo mismo, ni sus cálculos son similares, el investigador debe precisar el tipo de parámetro a utilizar en primer lugar, especialmente cuando un estudio posibilita el cálculo u ofrece varios de ellos. De manera general, se prefiere el RR o la razón de densidades de incidencia antes que la desigualdad relativa. Es posible, y en algunos casos bastante frecuente, que existan varios informes de un mismo estudio. Los investigadores deben precisar qué informe utilizarán. Lo más recomendable es utilizar siempre el más reciente. La costumbre a la hora de ofrecer los resultados de un estudio es que los autores ofrezcan estimaciones crudas y ajustadas del valor de la fuerza de la asociación. Se prefiere en primer lugar las estimaciones ajustadas a las crudas, ya que las primeras intentan eliminar la influencia de los factores de confusión.

Extracción de la información de cada estudio

Para aplicar el método de Woolf es necesario sacar de cada estudio dos valores: el del RR o el de sus derivados y su varianza. Generalmente, los autores ofrecen el valor del RR. En caso contrario hay que calcularlo utilizando las fórmulas de la tabla 3. Éstas serían estimaciones crudas. Si los autores ofrecieran resultados en los que pudiera calcularse un RR ajustado por una o más variables debiera hacerse así. En este caso, las fórmulas de la tabla 3 no son las adecuadas. El lector puede consultar el texto de Kleinbaum *et al.*¹⁵ o el de Rothman¹⁶ para calcular cualquiera de los cuatro procedimientos clásicos de ajuste: el método de Woolf, la estandarización interna y externa, y el método de Mantel-Haenszel. En la mayoría de los casos, los autores ofrecen el RR o sus derivados, no así la varianza. Para su cálculo podemos distinguir varias situaciones:

a) Los autores ofrecen el intervalo de confianza del RR o derivados.

En este caso se aplicará la fórmula siguiente:

$$V(1n RR) = [(1n RR_s - 1n RR_i) / 3,92]^2$$

siendo RR_s y RR_i los límites superior e inferior del intervalo de confianza del RR. El valor 3,92 resulta de multiplicar el valor de la distribución normal para un error alfa del 5%, 1,96, por dos. Si los autores ofrecieran intervalos de confianza para otros niveles el valor anterior se cambiaría.

Si los autores ofrecieran un solo límite del intervalo de confianza del 95%, la varianza se calcula de manera similar.

$$V = [(1n RR_s - 1n RR) / 1,96]^2$$

$$V = [(1n RR_i - 1n RR) / 1,96]^2$$

b) Los autores no ofrecen el RR pero sí ofrecen los resultados de la regresión logística: la regresión logística es en la actualidad el procedimiento más empleado para ajustar por varios factores de confusión simultánea. Los coeficientes β se pueden transformar en la OR de manera sencilla: e^β . Si los autores ofrecen el error estándar del coeficiente β , la varianza del $1n(OR)$ será igual al cuadrado del error estándar. Si los autores no ofrecen el error estándar, pero sí el nivel de significación, se buscará en la tabla de distribución normal el valor de z para el nivel de significación alcanzado. El coeficiente β se dividirá por ese valor z y el resultado será el error estándar de β .

c) Los autores no ofrecen el intervalo de confianza pero sí el valor de la χ^2 : en este caso se puede utilizar una aproximación indirecta para conocer la varianza. Se puede utilizar el valor de la χ^2 para calcular el intervalo de confianza del 95% mediante el método de Miettinen¹⁷:

$$IC\ 95\% \text{ del } RR = RR^{\wedge} (1 \pm 1,96 / \sqrt{\chi^2})$$

siendo \wedge el signo potenciación.

Los límites calculados por este procedimiento se utilizarían para calcular la varianza por cualquiera de las fórmulas mencionadas en el apartado a). Esta aproximación no es adecuada cuando el tamaño muestral es pequeño (o el RR está alejado de la unidad).

d) Los autores no ofrecen ni intervalos de confianza, ni regresión logística, ni valor de la χ^2 : en estos casos hay que recurrir a la distribución cruda de los resultados, calculando la varianza por las fórmulas enumeradas en la tabla 3. Si los autores ofrecen un valor de RR ajustado, la varianza del mismo no es la de los datos crudos. Normalmente la varianza de datos ajustados es mayor que la de los crudos, por la introducción de varias variables en el análisis que disminuye la eficiencia. Esto introduce un error que, como el propio Greenland⁴ reconoce, no hay forma de evitar en muchas situaciones. En otras, sin embargo, es posible corregir este error. Greenland⁴ preconiza que a veces es posible calcular RR ajustados sin que los autores los ofrezcan. Cuando dos estudios son similares en su población participante y en la metodología empleada, y uno de ellos tenga estimaciones ajustadas, mientras que el otro no, es posible calcular el error —U— introducido por los factores de confusión en el cálculo del RR. Se estima U del modo siguiente:

$$U = RR_c / RR_a; \text{ y su varianza } V_u = V_a - V_c$$

siendo RR_c el riesgo relativo crudo; RR_a el ajustado; V_a la varianza del RR_a y V_c la del RR_c .

Este método es también aplicable cuando, dentro de un mismo estudio, existen estimaciones ajustadas y crudas simultáneamente para unos factores mientras que para otros sólo existen estimaciones crudas.

e) Los autores no ofrecen ni intervalos de confianza, ni regresión logística ni pruebas de χ^2 , ni la distribución completa de los datos crudos, pero sí se conocen el total de los grupos índice y de referencia: recientemente Trock *et al.*¹⁸ han propuesto un método de estimación de la varianza del $1n$ de la OR cuando no se conoce más que el total de casos (m_1) y el total de controles (m_0), es decir a, b, c , y d son desconocidos. Parten de dos asertos: 1) $OR = a' / c' / b' / d'$, donde $a' + c' = a + c = m_1$ y 2) $b' = d'$, donde $b' + d' = b + d = m_0$ (o $b' + 1 = d'$ si m_0 es un número impar). Bajo los asertos 1) y 2), la estimación de OR es igual a $a' d' / c' b'$ y la varianza del parámetro sería $1/a' + 1/b' + 1/c' + 1/d'$. En general, esta varianza infraestima el verdadero valor de la varianza, ya que la varianza alcanza su mínimo cuando $a=c$ y $b=d$. Puede ser una aproximación a tener en cuenta, especialmente si el número de estudios a los que es aplicable este procedimiento es pequeño con respecto al total de estudios que se combinan.

Método de Mantel-Haenszel^{15,16}

Planteamiento general

Este método es similar al caso anterior, pero mucho más simple. En este caso la ponderación se realiza por el tamaño muestral del estudio. Es un procedimiento que se recomienda para los estudios experimentales.

El RR global se estimará de modo diferente según el parámetro que se combine (razón de ventaja, OR, riesgo relativo, RR, o razón de sensibilidades de incidencia, RDI):

$$RR_{p_{MH}} = \frac{\sum_{i=1}^n a_i n_{0i} / N_i}{\sum_{i=1}^n c_i n_{1i} / N_i}$$

$$RDI_{p_{MH}} = \frac{\sum_{i=1}^n a_i P_{0i} / P_i}{\sum_{i=1}^n c_i P_{1i} / P_i}$$

$$OR_{p_{MH}} = \frac{\sum_{i=1}^n a_i d_i / N_i}{\sum_{i=1}^n c_i b_i / N_i}$$

En caso de que se combinen diferentes derivados del RR en un solo parámetro resumen, siguiendo el planteamiento de Duffy *et al.*¹⁹, que será detallado cuando se comente el método de Mantel-Haenszel para datos emparejados, se sumarán los numeradores por un lado y los denominadores por otro. Los primeros darán lugar al numerador del parámetro resumen final. Los segundos

constituirán el denominador total final.

La varianza del parámetro resumen será:

$$V(1nRRp_{MH}) = \frac{\sum_{i=1}^n p_i^2 V_i}{(\sum_{i=1}^n p_i)^2}$$

siendo p_i el peso asignado a cada estudio y V_i la varianza del $1nRRp_i$. El peso cambia según el tipo de diseño:

- a) Para los estudios de casos y controles sería $c_i b_i / N_i$
- b) Para los estudios de cohortes acumulados se estimaría como $c_i n_{11} / N_i$
- c) Para los estudios de cohortes de densidad valdría $c_i P_{11} / P_i$

Tras conocer el RRp_{MH} y la varianza del parámetro se puede estimar el intervalo de confianza del mismo y la χ^2 de asociación, utilizando las mismas ecuaciones descritas en el método de Woolf. Para la valoración de la heterogeneidad se pueden utilizar dos alternativas: la desarrollada en el método Woolf (sustituyendo el RRp_w por el RRp_{MH}) o la que se detalla en el apartado siguiente, donde se desarrolla el procedimiento de Duffy *et al.*

Método de Mantel-Haenszel para combinar datos emparejados y no emparejados

Se ha visto que el método de Mantel-Haenszel permite combinar resultados para datos no emparejados y obtener un RR resumen, tratando a cada estudio como un estrato. Cuando se aplica el procedimiento de Mantel-Haenszel a datos emparejados se realiza el mismo análisis que para datos no emparejados, pero cada pareja dentro de un estudio es tratada como un estrato. Esto es, cuando se combinan estudios emparejados, se estratifica por pares dentro de cada estudio, lo que supone que la estratificación por estudio queda hecha automáticamente. La tabla 4 muestra la estructura y notación de los estudios emparejados. La estimación combinada de una serie de n estudios emparejados es:

$$RRp = \frac{\sum_{i=1}^n \sum_{j=1}^k e_{ij} h_{ij} / 2}{\sum_{i=1}^n \sum_{j=1}^k f_{ij} g_{ij} / 2} = \frac{\sum_{i=1}^n z_{10i}}{\sum_{i=1}^n z_{01i}}$$

Se puede obtener una estimación resumen de datos emparejados y no emparejados del modo siguiente¹⁹:

$$RRp = \frac{\sum_{i=1}^n a_i d_i / N_i + 1/2 \sum_{i=1}^n z_{10i}}{\sum_{i=1}^n b_i c_i / N_i + 1/2 \sum_{i=1}^n z_{01i}}$$

La significación global de la asociación se puede calcular por el estadístico siguiente:

$$\chi^2 = \frac{(\sum_{i=1}^n a_i + 1/2 \sum_{i=1}^n z_{10i} - \sum_{i=1}^n n_{11} n_{01} / N_i - 1/2 \sum_{i=1}^n z_{01i})^2}{\sum_{i=1}^n m_{11} m_{01} n_{11} n_{01} / (N_i^3 - N_i^2) + 1/4 \sum_{i=1}^n (z_{10i} + z_{01i})}$$

con un grado de libertad.

Tabla 4. Estructura y notación de los estudios con datos emparejados

1. Estudios emparejados: sujetos pertenecientes al par j dentro del estudio i .

		Efecto		Total
		Sí	No	
Exposición	Sí	e_{ij}	F_{ij}	1
	No	g_{ij}	h_{ij}	1
Total		$e_{ij} + g_{ij}$	$f_{ij} + h_{ij}$	2

2. Estudios emparejados: estudio i de K estudios a combinar*

		No expuestos	
		Efecto	
		Sí	No
Expuestos	Sí efecto	z_{11i}	z_{10i}
	No efecto	z_{01i}	z_{00i}

$$* z_{11} = e_{ij} g_{ij}; z_{10} = e_{ij} h_{ij}; z_{01} = f_{ij} g_{ij}; z_{00} = f_{ij} h_{ij}$$

Este valor de la X^2 se puede utilizar para calcular el intervalo de confianza aplicando el procedimiento de Miettinen¹⁷.

$$IC \text{ del } RRp_{MH} = RRp_{MH} \wedge (1 \pm z_{1-\alpha} / \sqrt{X^2})$$

siendo \wedge el signo de potenciación.

Una alternativa para calcular el intervalo es utilizar su varianza. Para ello podemos recurrir a la estimación de la misma detallada para estudios no emparejados. En este caso habría que añadir que el peso de un estudio emparejado i es la Sumatoria de $f_{ij} g_{ij} / 2$ y su varianza $1/z_{10i} + 1/z_{01i}$.

La heterogeneidad del efecto se puede valorar mediante la X^2 siguiente:

$$X^2 = \sum_{i=1}^n \frac{(a_i - E(a_i | RRp))^2}{V(a_i | RRp)} + \sum_{i=1}^n \frac{(z_{10i} - E(z_{10i} | RRp))^2}{V(z_{10i} | RRp)}$$

El primer sumando representa los estudios no emparejados:

$$E(a_i | RRp) = A_i$$

$$V(a_i | RRp) = (1/A_i + 1/B_i + 1/C_i + 1/D_i)^{-1}$$

A_i es la solución positiva a la ecuación siguiente:

$$A_i = \frac{a_i - d_i - RRp(2a_i + b_i + c_i) \pm \sqrt{[(d_i - a_i - RRp(2a_i + b_i + c_i))^2 + 4(1 - RRp)n_{11}m_{10}RRp]}}{2(1 - RRp)}$$

Una vez calculado A_i es fácil calcular B_i , C_i y D_i , en base a los totales de filas y columnas.

El segundo sumando representa a los estudios emparejados:

$$E(z_{10i} | RRp) = [(z_{10i} + z_{01i})RRp] / (RRp + 1)$$

$$V(z_{10i} | RRp) = [(z_{10i} + z_{01i})RRp] / (RRp + 1)^2$$

Método de Peto

Este método es una adaptación del procedimiento anterior, el de Mantel-Haenszel, para combinar resultados de tablas de contingencia. Este método se encuentra descrito en detalle por Yusuf

et al.²⁰, siendo una derivación de un método propuesto con anterioridad por Peto et al.²¹. Al ser una modificación del procedimiento anterior, su uso es recomendado para la combinación de resultados de estudios experimentales, y no, por regla general, para agregar estudios observacionales.

Asuma, por razones de simplificación (para no emplear una anotación diferente), que la distribución de la tabla 1 es la de un estudio experimental, siendo los expuestos el grupo experimental y el no expuesto el control. Sea O_i el número observado de casos del fenómeno que se estudia en el grupo experimental (a en la tabla 1). El número esperado de casos en el grupo experimental se calcula tras asumir que, en la ausencia de efecto de la intervención que se valora, el número de casos que se esperan ocurrirían en la proporción que el tamaño muestral del grupo supone con respecto al total de la población de estudio. Esto es, $E_i = m_i n_i / N_i$. Bajo la hipótesis nula de ausencia de tratamiento, la diferencia $O_i - E_i$ debería variar alrededor de cero, con una varianza $V_i = E_i - [(N_i - n_i) / N_i] [N_i - m_i] / (N_i - 1)$

La significación global de la asociación se puede calcular mediante el estadístico

$$\frac{[\sum_{i=1}^n (O_i - E_i)]^2}{(\sum_{i=1}^n V_i)}$$

que tiene una distribución aproximada de X^2 con un grado de libertad.

El estimador resumen de la razón de ventaja con este método es

$$ORp_p = \exp \left\{ \frac{[\sum_{i=1}^n (O_i - E_i)]}{(\sum_{i=1}^n V_i)} \right\}$$

y un estimador aproximado de la varianza del logaritmo natural de la razón de ventaja ponderada (ORp) se calcula del modo siguiente:

$$V(1n ORp_p) = 1 / \sum_{i=1}^n V_i$$

Los autores también presentan una prueba de X^2 aproximada para comprobar la homogeneidad de las razones de ventaja; esto es, que los efectos observados varíen sólo aleatoriamente alrededor de un valor común. El estadístico

$$\frac{\sum_{i=1}^n [O_i - E_i] / V_i - [\sum_{i=1}^n (O_i - E_i)]^2 / \sum_{i=1}^n V_i}{\sum_{i=1}^n V_i}$$

tendría una distribución aproximada a la de X^2 con $K^* - 1$ grados de libertad, siendo K^* en número de varianzas que son distintas de cero.

Razón de verosimilitud (RV)

Este procedimiento ha sido propuesto recientemente. La RV es la razón de verosimilitudes de los datos observados bajo dos hipótesis simples. Es una comparación de lo bien que las dos hipótesis predicen los datos. La RV nos dice sólo lo que los datos afirman en la relación a las dos hipótesis dadas, no qué hipótesis elegir o lo que creer.

$$RV(H_1 \text{ vs } H_0) = \frac{\text{Prob}(\text{Datos} | H_1)}{\text{Prob}(\text{Datos} | H_0)} = \frac{f(\text{Datos} | H_1)}{f(\text{Datos} | H_0)}$$

según sea la variable discreta o continua. Es útil tanto para variables dicotómicas, como para efectos que se miden en una escala continua.

Una descripción detallada del método va más allá de los objetivos del presente trabajo, ya que no lo hemos visto aplicado, por el momento, a ninguna asociación. Los autores afirman que una de sus ventajas es que no se afecta por el problema de las comparaciones múltiples (que hace aconsejar por algunos autores que se disminuya el error alfa *a priori* asumido). Zucker y Yusuf²² no comparten las ventajas del método defendidas por Goodman⁷.

Métodos que se basan en el riesgo o la tasa atribuible en los expuestos (RAe)

El riesgo o la tasa atribuible en los expuestos se define como la diferencia entre las incidencias acumuladas o las densidades de incidencia de los expuestos y no expuestos. Cuando el objetivo es la diferencia de riesgos (incidencias acumuladas) el método recibe el nombre de DerSimonian y Laird²³. En realidad es la aplicación de Woolf al RAe. La aportación que realizan DerSimonian y Laird al procedimiento de Woolf es la incorporación de la variación entre los distintos estudios a la hora de estimar el intervalo de confianza del parámetro resumen final.

El RAe resumen será estimado como la media ponderada (por el inverso de la varianza) de los RAe individuales:

$$RAep = \frac{\sum_{i=1}^n 1 / [V_i (RAe_i)] \cdot RAe_i}{\sum_{i=1}^n 1 / [V_i (RAe_i)]}$$

siendo RAep el riesgo atribuible en los expuestos ponderado.

La varianza del RAep es igual al inverso de la suma de los pesos de cada riesgo atribuible considerado. Esto es

$$V(RAep) = 1 / \left\{ \sum_{i=1}^n 1 / [V_i (RAe_i)] \right\}$$

Una vez conocidos el RAep y su varianza, se aplicará la fórmula siguiente para calcular su intervalo de confianza:

$$IC \ 95\% \text{ del RAep} = RAep \cdot (\pm 1,96 \sqrt{V})$$

La heterogeneidad entre los diferentes RAe que se combinan se valora mediante el estadístico Q

$$Q = \sum_{i=1}^n 1 / V_i (RAe_i - RAep)^2$$

que sigue una aproximadamente distribución de X^2 con $K - 1$ grados de libertad (siendo K el número de estudios que se combinan).

Se puede incorporar explícitamente la variabilidad entre estudios (que puede producir variabilidad en los distintos efectos encontrados), por diferencias en los protocolos, periodos de seguimiento, selección de los pacientes. Se asume que en cada estudio se encuentra su propio RAe_i , siendo μ y τ^2 la media y varianza de los distintos RAe_i . La varianza entre estudios se puede estimar de manera sencilla por el método de momentos (MDM) como:

$$\tau^2 = \max \left\{ 0, \frac{Q - K + 1}{\sum_{i=1}^n 1/V_i - \left(\sum_{i=1}^n 1/V_i^2 / \sum_{i=1}^n 1/V_i \right)} \right\}$$

Dado este estimador de la variabilidad, el peso de cada investigación vendría dado por

$$w_i^* = 1 / (V_i + \tau^2)$$

Esto supondría que el RAe_i , medido como μ , se estimaría como

$$\mu = \frac{\sum_{i=1}^n (w_i^* RAe_i)}{\sum_{i=1}^n w_i^*}$$

La varianza y el intervalo de confianza de μ se estimarían como el caso anterior, cuando no se tuvo en cuenta la variabilidad entre estudios. En el presente caso se sustituiría $1/V_i$ por w_i^* .

Existen otros procedimientos basados en el RAe_i , pero no los hemos visto aplicados en epidemiología. Por ello no serán detallados. Entre ellos se encuentra el método de Cochran o la modificación de Mantel-Haenszel al de Cochran para muestras pequeñas⁶.

Variables continuas

Las variables de efecto continuas se utilizan más raramente en epidemiología, aunque en otros terrenos (educación, agricultura) son las más frecuentes. Una de las técnicas analíticas más empleadas es el análisis de la varianza (ponderado o no). En un análisis de varianza de dos vías, por ejemplo, una de las variables que se analizan es la variable cualitativa «número de estudio» y la restante es la «exposición» objeto de la investigación. En estos casos, el test de homogeneidad entre los distintos estudios que se combinan es el factor «número de estudio» y/o el análisis de la interacción estudio-exposición (un valor significativo en al menos una de las dos comparaciones anteriores reflejaría la existencia de diferencias significativas entre los resultados de las investigaciones que se combinan). Un ejemplo de estas aplicaciones lo constituye el trabajo de Dales *et al.*²⁴, en el que se analiza la sensibilidad, la especificidad y el valor global de la tomografía axial computarizada en la valoración del estadio de los pacientes afectados de cáncer de pulmón.

Para exposiciones que son variables continuas se pueden usar los modelos de regresión, en el que se relaciona el riesgo de enfermar con las dosis-tiempo recibida. Ejemplo de ello es el

modelo de regresión lineal usado por Gilbert *et al.*²⁵ al estudiar los efectos de la radiación en la salud.

$$R(t_1) = R(t_0) [1 + \beta z(t_1)]$$

siendo el riesgo a tiempo t función del riesgo basal a tiempo 0, la radiación media sufrida desde t_0 a t_1 , y el incremento de riesgo experimentado por esa radiación, β .

Gilbert *et al.*²⁵ no establecen ningún método de ponderación de los estudios que serán combinados. Longnecker *et al.*²⁶, en un metaanálisis de consumo de alcohol y cáncer de mama usan como criterio de ponderación el inverso de la varianza (método de Woolf) del RR estimado por cada autor en los estratos considerados.

La «magnitud del efecto» (*effect size*) no será detallada en el presente trabajo por una razón sencilla: Greenland⁴ argumenta su incapacidad para poder ser utilizada como parámetro resumen porque transforma los resultados de cada investigación en unidades estándar de la propia investigación. Es conocido que la desviación estándar de una variable es una función de la distribución marginal de la misma. Por ello, toda medida estandarizada en función de la desviación estándar proporcionan una valoración distorsionada de un determinado efecto, ya que no depende sólo de la magnitud del efecto, sino también de la distribución de la variable a nivel de la población^{4,27}.

El lector interesado encontrará en el texto de Hedges y Olkin² una valoración completa de los distintos procedimientos aplicados al estudio de variables continuas, incluyendo análisis multivariable.

Representaciones gráficas de las estimaciones del efecto

El *odd man out* es una aproximación gráfica al metaanálisis, útil para sintetizar la información procedente de cinco a diez estudios que proporcionan intervalos de confianza de un parámetro²⁸. El método consiste en la representación gráfica en escala logarítmica de los intervalos de confianza del 95% proporcionados por los estudios metodológicamente apropiados. Tras ello se identifican la serie de puntos que son comunes al menos para $N-1$ estudios. El método es claro y fácil de aplicar. Spector y Hochberg²⁹ han aplicado este procedimiento al estudio de la asociación entre los anticonceptivos orales y la artritis reumatoide. Produjo una estimación próxima al Mantel-Haenszel, aunque con un intervalo de confianza notablemente más estrecho. Otros autores, como Rifat³⁰, que lo ha utilizado para la valoración de las enfermedades tiroideas como factor de riesgo de la demencia de Alzheimer, lo encuentran muy adecuado.

Utilidad de los diversos procedimientos para variables dicotómicas

No existe un cuerpo de doctrina sobre las ventajas de un método sobre otro. No obstante, en la literatura se han establecido

algunas directrices, Fleiss¹⁴ compara tres procedimientos: el de Woolf, el de Mantel-Haenszel y el de Cornfield y Gart, especialmente los dos primeros. Este autor afirma que cuando el número de estudios que se combinan es grande es mejor utilizar el método de Mantel-Haenszel. Este autor no entra en disquisiciones sobre la utilidad de los procedimientos en los diferentes tipos de diseños.

El sesgo de confusión en los diseños observacionales es generalmente controlado en el análisis, aunque en los estudios de casos y controles la influencia de algunas variables distorsionantes puede ser eliminada en el diseño por el emparejamiento. Es difícil que el emparejamiento pueda suplir al análisis. De hecho, se ha constatado que el emparejamiento puede producir un descenso de la eficiencia estadística^{31,32}. La utilización del método de Woolf a los diseños observacionales (cohortes, casos y controles y estudios de corte) posibilita que los parámetros que se ponderan puedan estar ajustados por variables extrañas. Con ello se intenta conseguir que el sesgo de confusión sea menos probable. Sin embargo, la utilización del método de Mantel-Haenszel y derivados (procedimientos de Peto) obliga a utilizar la distribución original y cruda de los datos. Esto supone que si en el diseño no interviene un procedimiento que intente controlar el sesgo de confusión, es bastante probable que el resultado final pueda verse afectado por este tipo de error. Este hecho supone que estos métodos de combinación de resultados no se recomienden para los estudios observacionales. No sucede lo mismo con los diseños experimentales. En ellos existe un elemento que tiende a eliminar el sesgo de confusión: la aleatorización. Por ello, el método de Mantel-Haenszel y derivados sí son aplicables a este tipo de diseños. Esto no quiere decir que el método de Woolf no pueda ser utilizado en los estudios experimentales. De hecho, Greenland⁴ recomienda que éste sea el procedimiento de elección a aplicar en el metaanálisis. Existiría una razón para defender el método de Woolf frente al de Mantel-Haenszel en los estudios experimentales y es que no todos los ensayos clínicos que se realizan tienen el mismo grado de éxito en el control de la confusión mediante la aleatorización. Esto es un hecho dependiente del tamaño muestral: a mayor tamaño muestral, mayor éxito en la aleatorización^{33,34}. Sirva de ejemplo el metaanálisis realizado por Henry y O'Connell³⁵ para estudiar el efecto de los inhibidores de la fibrinólisis en la mortalidad por hemorragia digestiva alta. Los autores combinaron estudios que oscilaron entre 43 y 516 sujetos. Con 43 sujetos divididos en dos grupos es difícil que se garantice la homogeneidad completa de los grupos que se comparan, especialmente sin análisis multivariable (y aun con él en este caso). Igualmente sucedió con el análisis de O'Connor *et al.*³⁶ que estudiaron el ejercicio tras el infarto de miocardio, con tamaños muestrales oscilantes entre 54 y 1780. Por ello es frecuente que en los ensayos clínicos aleatorizados se utilicen también medidas ajustadas, como garantía de una mayor eliminación de la confusión residual existente tras la aleatorización. No obstante las premisas anteriores, el método más utilizado en la combinación de ensayos clínicos ha sido hasta el momento presente el de Mantel-Haenszel (y derivados)³⁷

La elección entre una medida basada en el RR y derivados o

en riesgo atribuible depende de los objetivos de la investigación. Las medidas basadas en el RR tienen como finalidad principal establecer la fuerza con que dos variables se encuentran relacionadas, mientras que la utilidad del RAe tiene una utilidad esencialmente administrativa^{37,38}. Independientemente de estos hechos, el método de DerSimonian y Laird se diferencia en esencia, aparte de adoptar un modelo aditivo, del método de Peto o del de Woolf en que en este procedimiento se tiene en cuenta la variabilidad entre los estudios que se combinan, mientras que en los demás sólo se toma en consideración la variabilidad intraestudio. Esto debería suponer que la estimación mediante este procedimiento produciría intervalos de confianza más amplios que con los restantes, siendo las estimaciones más conservadoras. Esto no es una ventaja desdeñable, ya que uno de los inconvenientes que se han señalado en el metaanálisis es la facilidad de obtener resultados estadísticamente significativos: tanto más fácil conforme el número de estudios que se combinan es mayor¹. A pesar de ello, Berlin *et al.*³⁷ han comprobado que no hay diferencias entre el método de DerSimonian-Laird y el de Peto cuando aplicaron ambos procedimientos a 21 metaanálisis que habían realizado con anterioridad.

Los métodos gráficos, como el *odd man out*, pueden presentar una utilidad interesante. Su principal limitación radica en que el número de estudios para que pueda ser aplicado es pequeño. Los autores no consideraron la posible aplicación del mismo cuando la población de estudios es superior a diez. Como antes se mencionó, sólo hemos podido localizar un estudio en la literatura que comparara este procedimiento con el más habitual Mantel-Haenszel. La estimación puntual de ambos métodos fue próxima. No así sucedió con los límites del intervalo de confianza, que fueron notablemente más estrechos en el *odd man out*. Esto, de comprobarse en futuras investigaciones, sería una desventaja de esta técnica: agravaría el inconveniente antes mencionado de la facilidad que puede tener el metaanálisis para encontrar resultados significativos. La experiencia con este tipo de técnicas es aún escasa.

Es necesario dejar claro un hecho: el metaanálisis es tan válido como lo son los diferentes estudios que son combinados. Es imposible inferir nada si los estudios incluidos en el análisis adolecen de graves deficiencias. En la literatura han sido mencionados algunos métodos para corregir errores en algunos estudios, en base a la información suministrada por otras investigaciones. Sirvan de ejemplo el método de Greenland⁴ para corregir por el sesgo de confusión o el desarrollado por Savitz y Barón³⁹ para corregir el error por la mala clasificación de los factores de confusión. Estas aproximaciones son, por regla general, de difícil aplicación en la práctica. Por ejemplo, en el análisis de los anticonceptivos orales y el cáncer de cuello uterino no pudo aplicar el método de Greenland para corregir por conducta sexual¹².

Incluso aunque todos los estudios que se incluyan en un metaanálisis sean absolutamente válidos, la inferencia puede presentar problemas (aunque no haya heterogeneidad). Los investigadores han de asegurarse de que los distintos estudios que se combinan de verdad pueden ser resumidos en un solo parámetro. Uno de estos problemas, por ejemplo, es el que se deriva de

combinar estudios con duraciones de observación muy diferentes. Se puede producir en cualquier estudio observacional (o experimental) si se cumplen dos condiciones: 1) el resultado es dependiente del tiempo; y 2) el número de individuos a riesgo cambia de manera acusada a lo largo del seguimiento. El cálculo en estos casos del número acumulado de casos observados y esperados tiende a infraestimar el RR. Una idea simple para superar esta dificultad es sólo incluir a cada punto y aquellos estudios que tenga seguimientos más allá de ese punto. Sin embargo, esto supone eliminar información y puede sesgar el resultado si los RR de los estudios no son iguales⁴⁰

Algunos autores, como Fleiss y Gross⁴¹, recomiendan el que se haga un análisis de sensibilidad, siempre y cuando sea posible. Consiste en analizar los resultados de las distintas investigaciones que son combinadas por más de un procedimiento, por ejemplo, métodos de Woolf y Peto, etc. Si los distintos análisis coinciden, esto reforzará la inferencia a realizar de los resultados.

Sesgo de publicación

En epidemiología es conocido que el sesgo de selección es difícilmente controlable (normalmente no) en el análisis. El sesgo de selección del metaanálisis es el sesgo de publicación, no en vano la mayoría de los metaanálisis realizados lo son sobre estudios publicados. Es necesario cuantificar de alguna manera la posible existencia de este sesgo que amenaza seriamente la inferencia. Una aproximación gráfica ha sido realizada por Vandembroucke⁴², que recomienda el uso de «funnel plots», representaciones en forma de embudo, en las que el logaritmo natural del RR se enfrenta al valor de su error estándar. La existencia de una figura simétrica alrededor del valor resumen del RR habla en favor de la ausencia de este error. Un ejemplo de su aplicación se puede encontrar en Tersmette *et al.*⁴³, cuando analizan si los pacientes con afecciones benignas gástricas que sufrieron gastrectomía tienen un riesgo mayor de cáncer de estómago. Otros autores lo han aplicado a otros parámetros de interés, no sólo al RR. Por ejemplo, Dales *et al.*²⁴, en su análisis de la utilidad de la tomografía axial computarizada en el estadiaje del cáncer de pulmón, extrapo-

laron la idea inicial de Vandembroucke al valor global de una prueba.

Dentro de las representaciones gráficas, quizás la más clásica sea la que enfrenta el RR y su intervalo de confianza en escala logarítmica al tamaño muestral del estudio, por extrapolación de las representaciones de la «magnitud del efecto»⁴. Se le puede hacer el mismo comentario que en el caso anterior. Personalmente creemos que este tipo de representaciones presenta la desventaja de que la figura puede distorsionarse cuando se ponderan estudios de casos y controles con estudios de cohortes, por las grandes diferencias de tamaño muestral existentes entre ellos. Por ejemplo, en el metaanálisis de anticonceptivos orales y cáncer de mama, se ponderaron estudios que oscilaron entre 198 sujetos y 1.090.814 participantes-año^{10,11}.

Otro método consiste en el cálculo del número adicional necesario de estudios (a los que han sido localizados) que hayan encontrado resultados *no* significativos que harían falta para conseguir que un resultado significativo ($p < 0,05$) encontrado en el metaanálisis se hiciera no significativo ($p > 0,05$). Rosenthal⁴⁴ nos ofrece una fórmula para la estimación

$$f = (Z_i)^2 / 2,706 - K$$

siendo f : el número adicional de estudios; Z_i , el valor absoluto de la raíz cuadrada del valor de la X^2 ; y K el número de estudios revisados.

Este procedimiento lo aplicaron, por ejemplo Ozminkowski *et al.*⁴⁵, que analizaron la supervivencia neonatal en su relación con el lugar de nacimiento. Encontraron que eran necesarios 307 estudios adicionales a sus 19 para rebatir sus conclusiones. Nosotros lo hemos utilizado también^{10,12,13}. Por ejemplo, en el análisis de la relación entre terapia hormonal sustitutiva tras menopausia natural y cáncer de mama, serían necesarios 70 estudios con resultados no significativos para hacer que la conclusión alcanzada fuera no significativa¹³. El problema que presenta este método deriva de su carácter elucubrativo: sólo asume estudios con resultados no significativos. Lo ideal sería disponer de un método que también asumiera estudios con resultados significativos pero de signo contrario al sentido global de la asociación.

Bibliografía

1. Light RJ, Pillemer DB. *Summing up. The science of reviewing research*. Cambridge (MA): Harvard University Press, 1984.
2. Hedges LV, Okin I. *Statistical methods for meta-analysis*. Orlando (FLA): Academic Press, 1985.
3. Jenicek M. *Meta-analyse en médecine*. Québec: Edisem, 1987.
4. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987; 9: 1-30.
5. DeMets DL. Methods for combining randomized clinical trials: strengths and limitations. *Stat Med* 1987; 6: 341-8.
6. Boissel J-P, Blanchard J, Panak E, Peyrieux J-C, Sacks H. Considerations for the meta-analysis of randomized clinical trials (Summary of a panel discussion). *Controlled Clin Trials* 1989; 10: 254-81.
7. Goodman SN. Meta-analysis and evidence. *Controlled Clin Trials* 1989; 10: 188-204.
8. Needleman HL, Gatsonis CA. Low-level exposure and the IQ of children. A meta-analysis of modern studies. *JAMA* 1990; 263: 673-8.

9. Woolf B. On estimating the relationship between blood group and disease. *Ann Human Genet*; 1955; 19: 251-3.
10. Delgado-Rodríguez M, Sillero-Arenas M, Rodríguez-Contreras R, López R, Gálvez-Vargas R. Oral contraceptives and breast cancer. A meta-analysis. *Rev Epidém Santé Publ* 1991; 39: 165-81.
11. Sillero Arenas M. *Metaanálisis de la relación entre los contraceptivos orales y los cánceres de mama y cuello uterino*. Tesis Doctoral. Granada: Universidad de Granada, 1989.
12. Delgado-Rodríguez M, Sillero-Arenas M, Rodríguez-Contreras R, Martín Moreno JM, Gálvez-Vargas R. Oral contraceptives and cancer of the cervix uteri. A meta-analysis. *Eur J Epidemiol* 1991 (enviado).
13. Sillero-Arenas M, Delgado-Rodríguez M, Rodríguez-Contreras R, Bueno-Cavanillas A, Gálvez-Vargas R. Hormone replacement therapy and breast cancer. A meta-analysis. *Obstet Gynecol* 1992; 79: 276-304.
14. Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. Nueva York: Wiley-Interscience, 1981.
15. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research*. Belmont (CA): Lifetime Learning Publications, 1982.

16. Rothman KJ. *Modern Epidemiology*. Boston: Little, Brown & Co., 1986.
17. Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol* 1976; 103: 226-35.
18. Trock B, Lanza E, Greenwald P. Dietary fiber, vegetables, and colon cancer: critical review and meta-analyses of the epidemiologic evidence. *JNCI* 1990; 82: 650-61.
19. Duffy SW, Rohan TE, Altman DG. A method for combining matched and unmatched binary data. *Am J Epidemiol* 1989; 130: 371-7.
20. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985; 335-71.
21. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observations of each patient. II. Analysis and examples. *Br J Cancer* 1977; 35: 1-39.
22. Zucker D, Yusuf S. The likelihood ratio versus the p value in meta-analysis: where is the evidence? *Controlled Clin Trials* 1989; 10: 205-8.
23. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7:177-88.
24. Dales RE, Stark RM, Raman S. Computed tomography to stage lung cancer. *Am Rev Resp Dis* 1990; 141: 1096-101.
25. Gilbert ES, Fry SA, Friggs LD, Volez GL, Cragle DL, Petersen GR. Methods for analyzing combined data from studies of workers exposed to low doses of radiation. *Am J Epidemiol* 1990; 131: 917-27.
26. Longnecker MP, Berlin JA, Orza MJ, Chalmers TC. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 1988; 260: 652-6.
27. Delgado Rodríguez M, Gómez Olmedo M, Fernández Sierra MA. Metaanálisis de la prevalencia de la úlcera péptica en relación al consumo de tabaco (Carta al Director). *Med Clin (Barc)* 1990; 95: 60.
28. Walker AM, Martín-Moreno JM, Rodríguez Artalejo F. Odd man out: a graphical approach to meta analysis. *Am J Public Health* 1988; 78: 961-6.
29. Spector TD, Hochberg MC. The protective effect of the oral contraceptive pill on rheumatoid arthritis: an overview of the analytical epidemiological studies using meta-analysis. *J Clin Epidemiol* 1990; 43: 1221-30.
30. Rifat SL. Graphic representations of effect estimates: an example from a meta-analytic review. *J Clin Epidemiol* 1990; 43: 1267-71.
31. Rodríguez-Coronado R, Delgado Rodríguez M, Sillero Arenas M, Lardelli Claret P, Gálvez Vargas R. Contraceptivos orales y cáncer de cuello uterino: análisis del poder de la asociación. *Rev San Hig Públ* 1991; 65: 25-38.
32. Delgado Rodríguez M, Sillero Arenas M, Rodríguez-Contreras R, Bueno Cavanillas A, Martín Moreno JM, Gálvez Vargas R. Anticonceptivos orales y cáncer de mama: análisis de la potencia estadística de asociación. *Gaceta Sanitaria* 1991; 5: 59-67.
33. Gálvez Vargas R, Delgado Rodríguez M. Epidemiología experimental. En: G Piédrola Gil et al. (Eds.): *Medicina preventiva y salud pública*, 8ª edición, Barcelona: Salvat, 1988: 125-34.
34. Rothman KJ. Epidemiologic methods in clinical trials. *Cancer* 1977; 39: 1771-5.
35. Henry DA, O'Connell DL. Effects of fibrinolytic inhibitors on mortality from upper gastrointestinal haemorrhage. *Br Med J* 1989; 298: 1142-6.
36. O'Connor GT, Buring JE, Yusuf S, Goldhaber SZ, Olmstead EM, Paffenbarger RS Jr, Hennekens CH. An overview of randomized trials of rehabilitation with exercise after myocardial infarction. *Circulation* 1989; 80: 234-44.
37. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989; 8: 141-51.
38. Gálvez Vargas R, Delgado Rodríguez M. Estudios de cohortes. En: G Piédrola Gil y cols (Eds.): *Medicina preventiva y salud pública*. 8ª edición, Barcelona: Salvat, 1988: 105-14.
39. Savitz DA, Baron AE. Estimating and correcting for confounder misclassification. *Am J Epidemiol* 1989; 129: 1062-71.
40. Abel UR, Edler L. A pitfall in the meta-analysis of hazard ratios. *Controlled Clin Trials* 1988; 9: 149-51.
41. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol* 1991; 44:127-39.
42. Vandenbroucke JP. Passive smoking and lung cancer: a publication bias? *Br Med J* 1988; 296: 391-2.
43. Tersmette AC, Offerhaus JA, Tersmette KWF, Giardello FM, Moore GW, Tytgat GMJ, Vandenbroucke JP. Meta-analysis of the risk of gastric stump cancer: detection of high risk patients subsets for stomach cancer after remote partial gastrectomy for benign conditions. *Cancer Res* 1990; 50: 6486-9.
44. Rosenthal R. *Meta-analytic procedures for social research*. Beverly Hills, California: Sage Publications, 1984:345.
45. Ozminkowski RJ, Wortman PM. Inborn / outborn status and neonatal survival: a meta-analysis of non-randomized studies. *Stat Med* 1988; 7: 1207-21.

