

Revisiones sistemáticas y metanálisis (II)

Pértega Díaz, S.¹; Pita Fernández, S.²

1- Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Juan Canalejo. 2- Centro de Salud de Cambre.

CAD. ATEN. PRIMARIA 2005; 12: 166-171

INTRODUCCIÓN

En un trabajo anterior se han presentado los conceptos de revisión sistemática y metaanálisis en el proceso de síntesis de los resultados alcanzados por diversos estudios en relación a un tema determinado¹. Dicha publicación se centró fundamentalmente en la exposición de las limitaciones y las etapas de una revisión sistemática, con una breve descripción de las técnicas estadísticas habitualmente utilizadas durante la etapa del metaanálisis. En el presente trabajo, se tratará de describir con mayor extensión los métodos estadísticos disponibles para la combinación de resultados en este tipo de estudios.

EL ANÁLISIS DE LA HETEROGENEIDAD

Antes de optar por alguno de los distintos métodos estadísticos que permiten combinar los resultados individuales de cada estudio para obtener un estimador combinado del efecto, habrá que determinar:

- El tipo de respuesta a estudiar:** si la respuesta es binaria o dicotómica la medida de efecto utilizada será la diferencia de proporciones, el riesgo relativo o el valor del odds ratio. Si la respuesta es una variable numérica, el efecto se medirá mediante la diferencia de medias en los grupos de interés. Puesto que esta diferencia será probablemente mayor cuando las medias sean mayores, en lugar de usar las diferencias absolutas se suelen utilizar las diferencias estandarizadas. En estudios no aleatorizados, en los que no se tenga control sobre los posibles factores de confusión, será aconsejable utilizar como medidas del efecto los correspondientes coeficientes de regresión.
En este punto, no debe olvidarse además que los valores de significación (valores de la p), que no informan del sentido ni de la magnitud de la asociación, o los estadísticos de contraste, que tienen en cuenta el sentido del efecto pero no su magnitud, no son medidas apropiadas para utilizar en la fase del metaanálisis.
- La heterogeneidad entre estudios:** Se trata de analizar hasta qué punto los resultados de los diferentes estudios pueden combinarse en una única medida. Diferencias

en el diseño del estudio, las características de la población, etc. pueden llevar a resultados muy diferentes y comprometer los resultados del metaanálisis.

La evaluación del grado de heterogeneidad puede llevarse a cabo mediante pruebas estadísticas, siendo la más utilizada la prueba Q de Der Simonian y Laird². Dicha prueba se basa en calcular una suma ponderada de las diferencias entre el efecto $\hat{\theta}_i$ determinado en cada uno de los k estudios (odds ratio, riesgo relativo, diferencia de medias, etc.) y el promedio global:

$$Q = \sum_{i=1}^k w_i (\hat{\theta}_i - \bar{\theta})^2$$

con $\bar{\theta} = \frac{1}{k} \sum_{i=1}^k \hat{\theta}_i$. De esta forma, si los estudios son homogéneos, el estadístico Q sigue aproximadamente una distribución χ^2 con $k-1$ grados de libertad. El valor obtenido para dicho estadístico en cada caso concreto se confronta con la distribución teórica correspondiente, obteniéndose así un valor de significación que permite rechazar ($p < 0,05$) o aceptar ($p > 0,05$) la hipótesis de homogeneidad. No obstante, se trata de una prueba con escasa potencia estadística, por lo que un resultado no significativo suele ser insuficiente para concluir que no existe heterogeneidad entre los estudios, y conviene explorar esta posibilidad con otros métodos, fundamentalmente de tipo gráfico, como son el gráfico de Galbraith o el gráfico de L'Abbé^{3,4}.

Por un lado, en el gráfico de Galbraith³ se representa la precisión de cada estudio (el inverso del error estándar de la estimación del efecto) frente al efecto estandarizado (i.e., la estimación del efecto dividida entre su error estándar). Se representa también la línea de regresión ajustada a estos puntos y una banda de confianza, de modo que todos los puntos deberían situarse dentro de dicha banda. Los puntos fuera de esos márgenes de confianza son los que mayor variabilidad aportan al análisis. Además, aquellos estudios con un mayor peso en el metaanálisis serán los de mayor precisión y podrán identificarse, por lo tanto, a la derecha del gráfico. El gráfico de L'Abbé⁴ es otra herramienta útil en el caso de trabajar con una respuesta binaria (por ejemplo, respuesta a

un nuevo tratamiento frente a otro estándar). En él se representa la proporción de eventos en el grupo control frente a la proporción de eventos en el grupo de tratamiento. Cada uno de los puntos en el gráfico representa así el riesgo relativo correspondiente a los diferentes estudios, de modo que la diagonal que divide el gráfico en dos secciones dejará a uno de los lados los estudios favorables al grupo de tratamiento y al otro los favorables al grupo control. La presencia de puntos dispersos, que no se sitúen de forma paralela a dicha diagonal, indicará posible heterogeneidad.

Para ilustrar lo anterior, consideraremos un hipotético ejemplo en el que se desea realizar un metaanálisis de 10 ensayos clínicos que tratan de evaluar la eficacia de un nuevo fármaco para el tratamiento de una determinada enfermedad. En todos los estudios los pacientes son aleatorizados para recibir el fármaco experimental (Grupo de tratamiento) o bien el tratamiento habitual (Grupo control), contabilizándose en cada grupo el número de pacientes que se han recuperado de la enfermedad. La variable respuesta es por lo tanto la curación, y la medida de efecto el riesgo relativo ($\theta = RR$). Los datos utilizados para este ejemplo se muestran en la Tabla 1.

TABLA 1

Metaanálisis de diez ensayos clínicos que analizan la eficacia de un nuevo fármaco en el tratamiento de una determinada enfermedad.

Estudio	Grupo de tratamiento		Grupo control		RR	95% IC
	Enfermos	Sanos	Enfermos	Sanos		
1	80	20	82	18	1,11	0,63-1,97
2	40	40	60	30	1,50	1,04-2,16
3	67	40	80	25	1,57	1,03-2,39
4	20	34	22	32	1,06	0,79-1,44
5	387	107	350	100	0,97	0,77-1,24
6	765	222	830	150	1,47	1,22-1,77
7	75	30	80	25	1,20	0,76-1,89
8	703	345	765	240	1,38	1,20-1,59
9	255	120	290	77	1,53	1,19-1,95
10	111	45	109	42	1,04	0,73-1,48

La prueba de Der Simonian y Laird no revela, con un nivel de confianza del 95%, evidencia estadística de heterogeneidad ($Q=14,401$; $p=0,109$). Sin embargo, los gráficos de Galbraith y de L'Abbé sugieren cierto grado de heterogeneidad, con uno de los estudios fuera de las bandas de confianza en el primero (aquel que proporciona una menor esti-

mación del efecto) y con puntos que no se alinean en torno a una línea recta en el gráfico de L'Abbé (Figuras 1 y 2).

FIGURA 1

Gráfico de Galbraith para los datos del ejemplo en la Tabla 1.

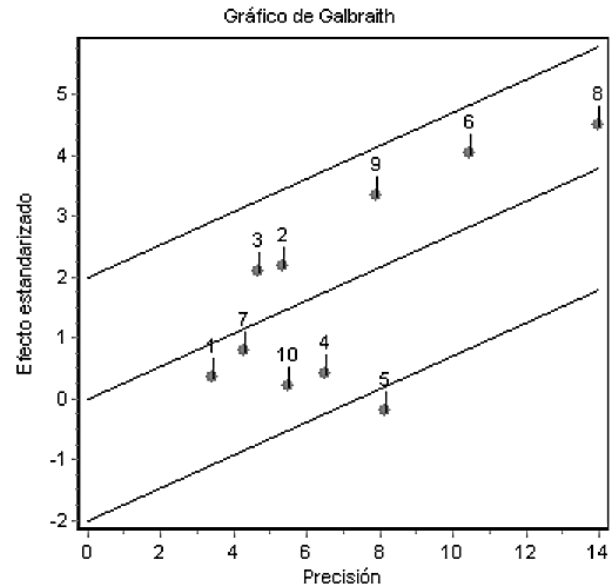
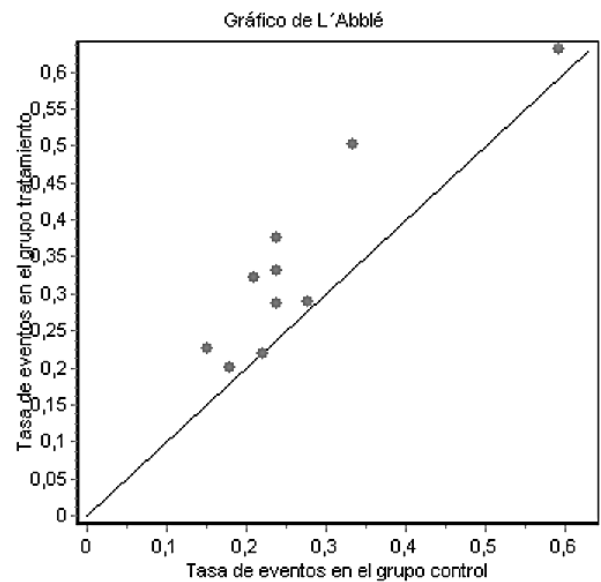


FIGURA 2

Gráfico de L'Abbé para los datos del ejemplo en la Tabla 1.



MÉTODOS ESTADÍSTICOS PARA LA COMBINACIÓN DE RESULTADOS

A pesar de las diferencias entre los distintos métodos disponibles para el metaanálisis, todos ellos siguen un esquema similar. En la mayoría de los casos, el estimador del efecto combinado se calcula como una media ponderada de los

estimadores de cada estudio, donde los pesos se asignan en base a la precisión de cada trabajo. De esta forma, los estudios con mayor variabilidad en la respuesta o con un tamaño muestral más reducido tendrán una contribución menor en el estimador global.

Fundamentalmente, los métodos estadísticos más utilizados en la práctica pueden clasificarse en dos grupos, según se tenga en cuenta o no la heterogeneidad entre estudios en el análisis⁵: los **modelos de efectos aleatorios** y los **modelos de efectos fijos**.

a) Modelo de efectos fijos.

En el modelo de efectos fijos se asume que no existe heterogeneidad entre los estudios incluidos en la revisión, de modo que todos ellos estiman el mismo efecto y las diferencias observadas se deben únicamente al azar.

Denotando una vez más por $\hat{\theta}_i, i = 1, \dots, k$ la medida de efecto (odds ratio, diferencia de medias, etc.) obtenida a partir de los datos del i-ésimo estudio, en el modelo de efectos fijos se asume que existe un efecto global fijo θ :

$$\hat{\theta}_i = \theta + \delta_i$$

siendo δ_i el error cometido al aproximar θ .

El efecto global puede estimarse como un promedio ponderado de los efectos individuales de cada estudio:

$$\hat{\theta} = \frac{\sum_{i=1}^k w_i \hat{\theta}_i}{\sum_{i=1}^k w_i}$$

donde los pesos w_i vienen dados como el inverso de la varianza de la estimación $\hat{\theta}_i$ correspondiente:

$$w_i = \frac{1}{Var(\hat{\theta}_i)}$$

La medida global del efecto así obtenida tendrá una varianza que viene dada por:

$$Var(\hat{\theta}) = \frac{1}{\sum_{i=1}^k w_i}$$

de modo que si se asume que $\hat{\theta}$ sigue una distribución normal podrá calcularse el intervalo de confianza correspondiente como $\hat{\theta} \pm 1,96 \times \sqrt{Var(\hat{\theta})}$.

b) Modelo de efectos aleatorios

Por el contrario, con un modelo de efectos aleatorios se asume que los estudios incluidos en la revisión constituyen

una muestra aleatoria de todos los estudios existentes. Ahora el efecto $\hat{\theta}_i$ de cada estudio se considera que tiene tres componentes:

$$\hat{\theta}_i = \theta + \theta_i + \delta_i$$

donde θ es el efecto común que interesa estimar, θ_i es el efecto a estimar en el i-ésimo estudio (en función de las características particulares del mismo) y δ_i el error cometido en la estimación.

Del mismo modo que en el modelo de efectos fijos, el efecto global se estima aquí como una media ponderada de los estimadores individuales, donde los pesos se calculan ahora como el inverso de la suma de la varianza del estudio individual más la varianza entre estudios:

$$\hat{\theta}^* = \frac{\sum_{i=1}^k w_i^* \hat{\theta}_i}{\sum_{i=1}^k w_i^*}; w_i^* = \frac{1}{Var(\hat{\theta}_i) + \tau^2}$$

La medida global del efecto así obtenida tendrá una varianza que viene dada por:

$$Var(\hat{\theta}^*) = \frac{1}{\sum_{i=1}^k w_i^*}$$

pudiendo calcularse el intervalo de confianza correspondiente como $\hat{\theta}^* \pm 1,96 \times \sqrt{Var(\hat{\theta}^*)}$.

Es frecuente encontrar trabajos en los que se presentan conjuntamente los resultados del metaanálisis tanto con el modelo de efectos fijos como con el modelo de efectos aleatorios. Mientras que algunos autores defienden la utilización del modelo de efectos aleatorios en todos los casos, otros hacen hincapié en sus posibles deficiencias, como el hecho de que es menos preciso, proporcionando intervalos de confianza más amplios que el modelo de efectos fijos. En general, debe tenerse en cuenta que el principal objetivo de un metaanálisis no será siempre el de obtener un estimador combinado del efecto. Cuando los resultados de los estudios revisados sean claramente heterogéneos el análisis e identificación de las causas de dicha heterogeneidad debe convertirse en nuestro principal objetivo. Si las discrepancias no son muy grandes el modelo de efectos aleatorios se convierte en la alternativa al modelo más sencillo con efectos fijos para combinar los resultados. En caso de una mayor variabilidad en los resultados la mejor opción será no realizar el metaanálisis, averiguar las causas de la heterogeneidad y realizar un análisis por subgrupos.

TABLA 2.

Resultados del metaanálisis para los datos de ejemplo de la Tabla 1.

Estudio	Tamaño muestral	RR	IC95% (RR)	Pesos		
				Modelo de efectos fijos	Modelo de efectos aleatorios	
4	108	1,0625	0,7861	1,4360	7,2094	9,5287
2	170	1,5000	1,0411	2,1612	4,9052	7,2288
1	200	1,1111	0,6263	1,9712	1,9908	3,4331
7	210	1,2000	0,7601	1,8945	3,1375	5,0711
3	212	1,5701	1,0306	2,3919	3,6921	5,7917
10	307	1,0371	0,7264	1,4806	5,1614	7,5103
9	742	1,5252	1,1910	1,9532	10,6925	12,2262
5	944	0,9747	0,7661	1,2401	11,2800	12,6111
6	1967	1,4695	1,2184	1,7723	18,6395	16,2978
8	2053	1,3785	1,1982	1,5860	33,2917	20,3012
Efectos fijos	6913	1,3119	1,2100	1,4224		
Efectos aleatorios	6913	1,2880	1,1504	1,4421		

PRESENTACIÓN DE RESULTADOS

Una vez realizados los cálculos anteriores, los resultados de un metaanálisis suelen representarse en una gráfica ("forest plot") en la que se muestra el efecto estimado en cada estudio junto con el valor obtenido combinando los resultados de todas las investigaciones, acompañados por sus respectivos intervalos de confianza. Además, suele representarse en la gráfica la línea vertical del valor correspondiente a la ausencia de efectos (RR=1 o Diferencia de medias=0). Podría resultar útil fijar también los límites de relevancia clíni-

ca para determinar si las diferencias, además de alcanzar significación estadística, son de una magnitud relevante. Recurriendo una vez más al ejemplo anterior, en la Tabla 2 y en la Figura 3 se muestran los resultados del metaanálisis utilizando tanto el modelo de efectos fijos como el modelo de efectos aleatorios. Todos salvo uno de los estudios muestran resultados homogéneos, con un efecto favorable del tratamiento experimental, y RR comprendidos entre 1,04 y 1,57. En los casos en los que los intervalos de confianza cruzan la línea de no efecto (RR=1) la diferencia en las

FIGURA 3

Resultados del metaanálisis para los datos del ejemplo de la Tabla 1.

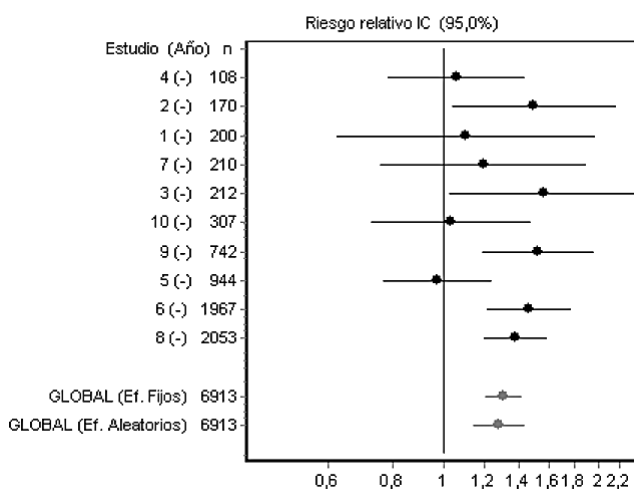


FIGURA 4

Análisis de sensibilidad para el metaanálisis de los datos del ejemplo en la Tabla 1.

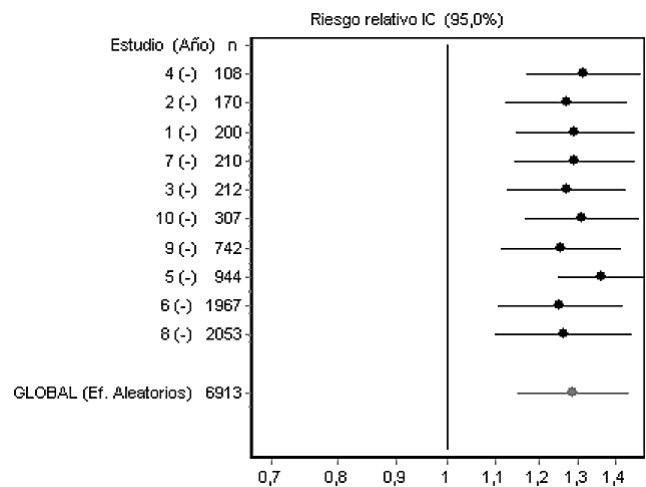
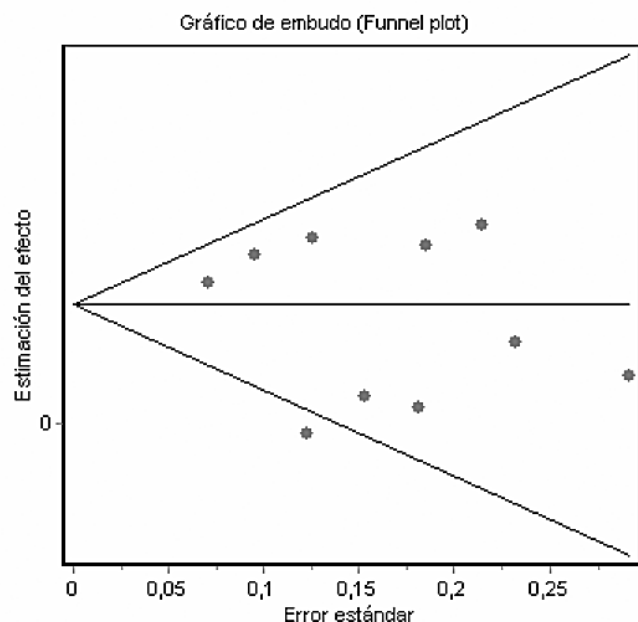


FIGURA 5

Funnel plot para los resultados del metaanálisis con los datos del ejemplo de la Tabla 1.



tasas de respuesta no ha resultado estadísticamente significativa. Con cualquiera de los dos métodos de análisis, los resultados permiten concluir que el nuevo tratamiento es significativamente mejor que el tratamiento estándar para conseguir la curación de los pacientes, obteniéndose medidas globales del efecto muy similares, de RR=1,21 con el modelo de efectos fijos y de RR=1,29 con el modelo de efectos aleatorios.

ANÁLISIS DE SENSIBILIDAD Y DEL SESGO DE SELECCIÓN

Después de realizar un metaanálisis, es recomendable estudiar la influencia de cada uno de los estudios en los resultados obtenidos. El análisis de sensibilidad consiste en replicar los resultados del metaanálisis excluyendo en cada paso uno de los estudios incluidos en la revisión. Si los resultados así obtenidos son similares, tanto en dirección como en magnitud del efecto y significación estadística indica que el análisis es robusto. Este mismo proceso podría repetirse eliminando a un mismo tiempo varios estudios (por ejemplo, aquellos de peor calidad metodológica, los no publicados, etc.) para determinar su posible influencia en los resultados. Junto con el análisis de sensibilidad, una vez obtenidos los resultados del metaanálisis se debe analizar la existencia de un posible sesgo de selección que pudiese poner en entredicho los resultados alcanzados. Entre los métodos

más utilizados para evaluar la existencia de este tipo de sesgos el más popular es el gráfico de embudo⁶ ("funnel plot"), el cual se basa en representar el tamaño muestral de cada trabajo frente al tamaño del efecto detectado. Lo normal sería que todos los estudios detectasen un efecto de magnitud similar, en torno a una línea horizontal, con mayor dispersión cuanto menor fuese el tamaño muestral. De esta forma, los puntos tenderían a distribuirse en forma de un embudo invertido. Si, por el contrario, existiese un sesgo de publicación, de los estudios de menor tamaño muestral solo se publicarían aquellos que encontrasen diferencias significativas entre los grupos, de forma que la nube de puntos aparecería deformada en uno de sus extremos. Existen otras técnicas estadísticas como la prueba de Begg (6) o de Egger (7), implementadas en la mayoría de los programas para la realización de metaanálisis, que permiten evaluar de una manera más objetiva la existencia de un posible sesgo de publicación.

Para los datos del ejemplo previo, los resultados del análisis de sensibilidad y el gráfico "funnel plot" se muestran en las Figuras 4 y 5, respectivamente. Tal y como se puede apreciar, ninguno de los trabajos parece modificar sustancialmente los resultados si es eliminado del metaanálisis. Así mismo, la Figura 5 no evidencia la existencia de sesgo de publicación, lo cual es refrendado por los resultados de las pruebas de Begg ($p=0,720$) y Egger ($p=0,316$).

En definitiva, las técnicas de metaanálisis constituyen así una herramienta sin excesiva complejidad estadística que permiten sintetizar los resultados de diferentes estudios en relación con un tema determinado. Su aplicación se ha visto facilitada en los últimos años gracias a la proliferación de programas informáticos que implementan este tipo de metodología, como es el caso del software EPIDAT (8). Sin embargo, su accesibilidad no debe potenciar su uso indiscriminado, obviando el hecho de que en ocasiones los diseños de los estudios que se incluyen en una revisión, su calidad metodológica o los resultados que estos alcanzan presentan un alto grado de heterogeneidad que desaconseja la realización del metaanálisis.

BIBLIOGRAFÍA

1. Pértega Díaz S, Pita Fernández S. Revisiones sistemáticas y metaanálisis. Cad Aten Primaria 2005; Vol. 12. Nº 2; 109-112.
2. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7: 177-188.
3. Galbraith R. A note on graphical presentation of estimated odds ratios from several clinical trials. Stat Med 1988; 7: 889-894.

4. L'Abbé, K, Detsky A, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987; 107: 224-233.
5. Molinero LM. Meta-análisis. [citado 23 Nov 2005]. Disponible en: <http://www.seh-lelha.org/pdf/metaanalisis.pdf>.
6. Begg CB, Mazumbar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; 50: 1088-1101.
7. Egger M, Smith GD, Schneider M, Zinder Ch. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629-634.
8. Epidat 3.0: Análisis epidemiológico de datos tabulados. [citado 23 Nov 2005]. Disponible en: <http://dxsp.ser-gas.es/default.asp>.