

Metodología en la investigación
sobre discapacidad.
Introducción al uso de las ecuaciones
estructurales

VI Seminario Científico
SAID, 2008

MIGUEL ÁNGEL VERDUGO
MANUELA CRESPO
MARTA BADÍA
BENITO ARIAS
(Coordinadores)

Publicaciones del INICO
Colección Actas
Salamanca, 2008

ISBN: 978-84-691-5852-4

Depósito Legal: S. 1.436-2008

Imprime KADMOS

Salamanca, 2008

ÍNDICE

PRESENTACIÓN	9
PRÓLOGO	11
1. <i>Introducción a la metodología SEM: Concepto y propósitos fundamentales.</i> BEGO- ÑA ORGAZ BAZ	13
2. <i>Modelos de medida y análisis factorial confirmatorio.</i> RAMÓN FERNÁNDEZ PULIDO	29
3. <i>Introducción a los modelos de ecuaciones estructurales.</i> MIGUEL RUIZ DÍAZ	43
4. <i>Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS y SAS.</i> BENITO ARIAS MARTÍNEZ	75
5. <i>Meta-análisis de la investigación.</i> JULIO SÁNCHEZ MECA	121
6. <i>Metodologías cualitativas: Características, procesos y aplicaciones.</i> M ^a TERESA AN- GUERA ARGILAGA	141
7. <i>Informática y análisis cualitativo.</i> M ^a CRUZ SÁNCHEZ GÓMEZ	127
8. <i>La ética en la investigación sobre discapacidad.</i> M ^a Teresa ANGUERA ARGILAGA	177
ANEXO. Programas SAID	193

META-ANÁLISIS DE LA INVESTIGACIÓN*

JULIO SÁNCHEZ MECA
*Departamento de Psicología Básica y Metodología
Facultad de Psicología (Universidad de Murcia)*

INTRODUCCIÓN

Durante la segunda mitad del siglo pasado se ha producido una explosión de información que ha provocado la necesidad de desarrollar métodos objetivos y sistemáticos para la acumulación del conocimiento científico en cualquier ámbito de investigación. El meta-análisis ha surgido como una metodología capaz de integrar cuantitativamente los resultados de las investigaciones sobre un determinado tema para poder establecer qué es lo que la evidencia empírica, hasta ese momento, ha demostrado. La revisión cuantitativa de la investigación se ha convertido en una tarea imprescindible entre el quehacer científico del pasado y del futuro, para orientar y dirigir nuevas investigaciones. Pero el meta-análisis es más que un mero método de revisión de la investigación. El meta-análisis comporta un nuevo modo de entender el significado y el análisis de los datos, con su énfasis en el tamaño del efecto frente a las pruebas de significación; es un nuevo enfoque que afecta a la interpretación habitual de qué entendemos por descubrimiento científico (Glass, McGaw y Smith, 1981; Schmidt, 1992, 1996; Schmidt y Hunter, 1995).

La realización de un meta-análisis es un proceso de investigación que requiere el cumplimiento de las normas propias del método científico: objetividad, sistematización y replicabilidad. Ésta es la principal diferencia con respecto a las revisiones tradicionales de la investigación. Frente a las revisiones narrativas, también denominadas revisiones cualitativas o subjetivas, el meta-análisis propugna el mismo rigor científico que se exige en las investigaciones primarias. En un meta-análisis no se prejuzgan los resultados de las investigaciones, sino que se cuantifican y se analizan conjuntamente. Así mismo, un meta-análisis tiene que garantizar su replicabilidad por otros investigadores, por lo que todas las decisiones en el proceso de revisión cuantitativa tienen que hacerse explícitas (Rosenthal, 1995).

La característica que mejor identifica al meta-análisis es el uso de los métodos estadísticos para integrar cuantitativamente los resultados de los estudios. Para ello, se requiere definir una medida de los resultados que sea homogénea a lo largo de todos los estudios. El índice conocido como "tamaño del efecto" (*effect size*) representa la magnitud en que se manifiesta el fenómeno en cuestión en cada estudio empírico (Glass et al., 1981). Hoy día existen numerosos textos que exponen cómo llevar a cabo estudios meta-analíticos, cómo

* Trabajo financiado con un Proyecto del Fondo de Investigación Sanitaria en la convocatoria de Evaluación de Tecnologías Sanitarias (Expediente N° PI07/90384)

calcular el tamaño del efecto de cada estudio y cómo aplicar técnicas de análisis estadístico específicamente diseñadas para ser aplicadas en meta-análisis, y todo ello dirigido a diferentes ámbitos científicos, tales como las Ciencias Sociales y del Comportamiento (Botella y Gambara, 2002; Cooper, 1998; Cooper y Hedges, 1994; Glass et al., 1981; Gómez, 1987; Hedges y Olkin, 1985; Hunter y Schmidt, 2004; Lipsey y Wilson, 2001; Petticrew y Roberts, 2006; Rosenthal, 1991; Schulze, 2004; Wolf, 1986) y las Ciencias de la Salud (Davey Smith, 2001; Eddy, Hasselblad y Shachter, 1995; Egger, Smith y Altman, 2001; Martín, Tobías y Seoane, 2006; Sutton, Abrams, Jones et al., 2000; Whitehead, 2002).

FASES DE UN META-ANÁLISIS

A lo largo del proceso de realización de un meta-análisis se suceden varias etapas, cada una de las cuales pretende aportar objetividad y sistematización al proceso de revisión. Las principales fases de un meta-análisis pueden resumirse en (Lipsey y Wilson, 2001; Marín-Martínez, Sánchez-Meca, Huedo y Fernández-Guzmán, 2007; Roberts, Kuncel, Viechtbauer y Bogg, 2007; Sánchez-Meca, 1999; Sánchez-Meca y Ato, 1989): (a) formulación del problema, (b) búsqueda de la literatura, (c) codificación de los estudios, (d) análisis estadístico y (e) publicación del estudio.

1º Formulación del problema. Como en cualquier investigación, el primer paso que debe dar el meta-analista es definir el objeto del meta-análisis, que generalmente será estudiar la magnitud y sentido de la relación entre dos (o más) variables o conceptos. En esta fase se precisa definir los conceptos no sólo de forma teórica, sino también las variables que se aceptarán como operativizaciones de ellas. Las definiciones operativas de los conceptos constituyen un importante criterio para la inclusión de los estudios en el meta-análisis. Así mismo, hay que distinguir entre las variables fundamentales de la investigación (aquéllas cuya relación se pretende estudiar) y variables potencialmente moderadoras de tal relación. En esta fase deben también plantearse las hipótesis que se pretenden poner a prueba.

2º Búsqueda de la literatura. Una vez formulado el problema de investigación, el paso siguiente consiste en realizar una búsqueda de la literatura lo más completa y exhaustiva posible. Para ello, es preciso especificar los criterios que deben cumplir los estudios empíricos para que sean incluidos en el meta-análisis, tales como el rango temporal en el que se realizaron o publicaron o el tipo de información estadística que deben aportar para poder realizar posteriormente los cálculos del tamaño del efecto.

En esta fase se recomienda utilizar tanto procedimientos formales como informales de búsqueda. Dentro de los procedimientos formales destacan las búsquedas por computador, que permiten acceder con gran economía de tiempo y recursos a las bases de datos informatizadas mediante el uso de palabras clave y otros descriptores. Pero estos procedimientos formales deben completarse con el uso de otros sistemas de búsqueda no sistemáticos que permiten localizar la literatura "fugitiva", es decir, trabajos no publicados o de muy reducida difusión, pero que pueden aportar datos relevantes al meta-análisis.

Por muy completa que sea la búsqueda de los estudios siempre cabe la posibilidad de que no seamos capaces de recoger todos los existentes, en especial aquéllos que no han sido publicados, con la consiguiente amenaza que el sesgo de publicación contra los resultados nulos puede ejercer en el meta-análisis. Siempre es recomendable, pues, realizar un estudio de la tolerancia a los resultados nulos para poder desechar el sesgo de publicación como una posible amenaza contra la validez de las conclusiones del meta-análisis (Begg, 1994; Burdett, Stewart y Tierney, 2003; Rosenthal, 1991; Rothstein, Sutton y Borenstein, 2005; Sutton, Duval, Tweedie et al., 2000).

3° *Codificación de las variables.* Por regla general, la realización de un meta-análisis se justifica por la existencia de resultados contradictorios en un determinado problema de investigación. Para determinar las razones de tales inconsistencias cada estudio seleccionado para el meta-análisis es revisado en profundidad y sometido a un protocolo de registro de aquellas variables que, en teoría, podrían estar afectando a la heterogeneidad encontrada en los resultados de los estudios. Este protocolo incluye variables moderadoras de tres tipos: sustantivas, metodológicas y extrínsecas.

- (a) Las variables *sustantivas* son aquéllas propias del objeto de investigación e incluyen aspectos tales como las características demográficas de las muestras de sujetos sometidas a estudio, cómo se operativizaron las variables, el contexto social, cultural, geográfico, económico, etc. en el que se hizo la investigación, etc.
- (b) Por variables *metodológicas* se entienden aquéllas que se refieren al método y diseño de la investigación y, por tanto, se repiten en todos los meta-análisis. Son metodológicas características tales como el tamaño de la muestra, la mortalidad experimental, el tipo de diseño o la calidad del diseño de investigación.
- (c) Las variables *extrínsecas* se caracterizan por ser externas al propio desarrollo de la investigación, aunque en ocasiones pueden afectar a los resultados de los estudios y poner en evidencia la existencia de deficiencias, sesgos o artefactos en un campo de estudio. Son ejemplos de este tipo de características, el estatus de publicación del estudio (publicado versus no publicado), la filiación de los autores del estudio o el género de los autores.

Además del registro de las variables moderadoras, es preciso definir y calcular el tamaño del efecto, que es un índice cuantitativo que resume la magnitud de la relación encontrada en cada estudio. Dependiendo de que el tipo de diseño habitual en los estudios integrados en el meta-análisis sea experimental o correlacional, se distinguen dos familias de índices del tamaño del efecto: las familias *d* y *r*, la primera para los estudios con asignación de los sujetos a grupos o condiciones experimentales, y la segunda para diseños de tipo correlacional (Cooper, 1998; Cortina y Nouri, 2000; Grissom y Kim, 2005; Lipsey y Wilson, 2001; Rosenthal, 1991; Rosenthal, Rosnow y Rubin, 2000).

La principal amenaza contra la validez del proceso de registro de las variables moderadoras es la falta de fiabilidad en el proceso de codificación. Es por ello que se recomienda la realización de un estudio de la fiabilidad de este proceso, para lo cual dos (o más) investigadores deben codificar de forma independiente los mismos estudios y, posteriormente, analizar la fiabilidad mediante el cálculo de índices de acuerdo inter-codificadores.

4° *Análisis estadístico.* Una vez que se dispone de una estimación del tamaño del efecto de cada estudio y de las variables moderadoras registradas en cada uno de ellos, el análisis estadístico de estos datos implica tomar el tamaño del efecto como la variable dependiente y las variables moderadoras como potenciales factores explicativos de aquél. Las preguntas a las que puede responder un meta-análisis son:

- (a) ¿Cuál es la magnitud del efecto medio?
- (b) ¿Es estadísticamente significativo el tamaño del efecto medio?
- (c) ¿Es el tamaño del efecto medio representativo de todos los estudios individuales? O lo que es lo mismo, ¿existe homogeneidad en torno al tamaño del efecto medio?
- (d) Si no existe homogeneidad, ¿qué características de los estudios pueden estar moderando el tamaño del efecto? ¿Puede proponerse un modelo explicativo?

5º Publicación del estudio. Para lograr el carácter replicable del meta-análisis, es preciso que la publicación del mismo siga las mismas normas que la publicación de investigaciones primarias (Botella y Gambara, 2006; Cooper, 1998; Rosenthal, 1995). La presentación de un meta-análisis debe, pues, incluir una introducción en la que se especifique el objetivo del estudio, las definiciones conceptuales y operativas de las variables y las hipótesis que se pretenden probar. En la sección método se incluye el proceso de búsqueda de los estudios incluidos en el meta-análisis, la codificación de las variables y de los tamaños del efecto y la especificación de las técnicas estadísticas que se han aplicado. En la sección de resultados se presentan y discuten los resultados de los estadísticos y contrastes aplicados, finalizando con las conclusiones, que deben incorporar directrices para la investigación futura. En la publicación debe ofrecerse al lector todas las facilidades para disponer de toda la base de datos meta-analítica, con objeto de posibilitar su replicación por otros investigadores.

ÍNDICES DEL TAMAÑO DEL EFECTO

Para poder integrar los resultados de múltiples estudios o investigaciones primarias acerca de una misma temática, es preciso que vengan expresados en una métrica común que los haga directamente comparables. En Ciencias Sociales es muy frecuente que, para la medición de un mismo concepto, se empleen muy diversas definiciones operativas, con diferentes indicadores y escalas de medida. Desde el meta-análisis se ha propuesto el tamaño del efecto como el indicador idóneo para representar el resultado de una investigación, de forma que sea comparable para diferentes escalas y operacionalizaciones de las variables implicadas.¹ Son múltiples los índices del tamaño del efecto que se pueden utilizar para resumir cuantitativamente la magnitud de la relación entre las variables. En meta-análisis estos índices suelen agruparse en dos familias: Las familias *d* y *r*.

• **La familia *d*.** Cuando los diseños de los estudios implican asignación de sujetos a diferentes grupos o condiciones experimentales (generalmente, grupos experimental y control), el índice del tamaño del efecto más adecuado es la “diferencia media tipificada”, que se define como la diferencia entre las medias de los grupos experimental y control dividida por la desviación típica intra-grupo (Lipsey y Wilson, 2001; Shadish y Haddock, 1994):

$$d = c(m) \frac{\bar{y}_E - \bar{y}_C}{S} ,$$

donde \bar{y}_E y \bar{y}_C son las medias de los grupos experimental y control, respectivamente; S es la desviación típica intra-grupo, que se obtiene mediante:²

$$S = \sqrt{\frac{(n_E - 1)S_E^2 + (n_C - 1)S_C^2}{n_E + n_C - 2}} ,$$

1 Aunque la corriente meta-analítica preconizada por Rosenthal proponía como resultado de los estudios el nivel crítico de probabilidad, p , obtenido en la prueba de significación, actualmente no se utiliza este índice debido a su dependencia del tamaño de la muestra.

2 Glass et al. (1981) propusieron dividir la diferencia entre las medias por la desviación típica del grupo de control, S_C (cuando exista), en lugar de dividir por S , debido a que, en ocasiones, la aplicación de un tratamiento puede alterar la variabilidad natural de los sujetos en la variable dependiente, como resultado de un efecto interactivo sujeto x tratamiento. Pero siempre que se cumpla el supuesto de homogeneidad de varianzas, es más eficiente utilizar S (Hedges y Olkin, 1985).

siendo n_E y n_C los tamaños muestrales de los grupos experimental y control, respectivamente, y s_E^2 y s_C^2 las respectivas varianzas. El factor $c(m)$ corrige el ligero sesgo positivo para muestras pequeñas mediante:

$$c(m) = 1 - \frac{3}{4(n_E + n_C) - 9} .$$

En meta-análisis, cada estimación del tamaño del efecto se pondera por la inversa de su varianza muestral, con objeto de permitir que los estudios más precisos (que suelen ser los basados en tamaños muestrales mayores) ejerzan un mayor peso específico en los cálculos meta-analíticos. Por ello es fundamental disponer de un estimador eficiente de la varianza debida a error de muestreo (o varianza intra-estudio). Para el índice d , ésta viene dada por:

$$V(d) = \frac{n_E + n_C}{n_E n_C} + \frac{d^2}{2(n_E + n_C)} .$$

En el caso de que todos los estudios del meta-análisis hayan operativizado la variable dependiente con la misma escala, test o prueba, no sería necesario estandarizar la diferencia de medias dividiendo por la desviación típica conjunta, sino que la simple diferencia entre las dos medias sería un índice del tamaño del efecto apropiado (Bond, Wiitala y Richard, 2003; Rosenberg, Adams y Gurevitch, 2000):

$$D = \bar{y}_E - \bar{y}_C ,$$

siendo su varianza intra-estudio:

$$V(D) = \frac{(n_E - 1)S_E^2 + (n_C - 1)S_C^2}{n_E + n_C - 2} .$$

Cuando la variable dependiente es dicotómica, el resultado de cada estudio queda resumido en una tabla de contingencia 2 x 2. La literatura médica ha estudiado en profundidad estos casos, recomendando la aplicación de alguno de los tres siguientes índices del tamaño del efecto (Deeks y Altman, 2001; Fleiss, 1994; Haddock, Rindskopf y Shadish, 1998; Sánchez-Meca y Marín-Martínez, 2000, 2001): La diferencia de riesgos, dr (o de proporciones), la razón de riesgos, rr (o de proporciones) y el odds ratio, or (o razón de posibilidades, o de productos cruzados). Siendo n_E y n_C los tamaños muestrales de los dos grupos, O_{1E} y O_{1C} las frecuencias de éxito en ambos grupos, y O_{2E} y O_{2C} las frecuencias de fracaso respectivas, la diferencia de riesgos se obtiene mediante:

$$dr = p_E - p_C ,$$

donde $p_E = O_{1E} / n_E$ y $p_C = O_{1C} / n_C$, y la varianza intra-estudio viene dada por:

$$V(dr) = \frac{p_E(1-p_E)}{n_E} + \frac{p_C(1-p_C)}{n_C} .$$

La razón de riesgos y su varianza intra-estudio, esta última en términos del logaritmo de la razón de riesgos, se obtienen mediante:

$$rr = \frac{p_E}{p_C} .$$

$$V(Lrr) = \frac{1-p_E}{n_E p_E} + \frac{1-p_C}{n_C p_C} .$$

Y el odds ratio, junto con su varianza intra-estudios, también en función del logaritmo del odds ratio, se obtienen mediante:

$$or = \frac{p_E(1-p_C)}{p_C(1-p_E)} = \frac{O_{1E}O_{2C}}{O_{1C}O_{2E}}$$

$$V(Lor) = \frac{1}{O_{1E}} + \frac{1}{O_{1C}} + \frac{1}{O_{2E}} + \frac{1}{O_{2C}} .$$

Por otra parte, se dan situaciones en las que la variable de respuesta es continua en naturaleza, pero ha sido dicotomizada, de forma que la única información disponible en el estudio es una tabla de contingencia 2×2 del tipo justamente comentado. Lo más aconsejable en estos casos es estimar la diferencia media estandarizada aplicando alguna de las funciones de transformación de variable dicotomizada a variable continua. De las diversas propuestas, las que mejores propiedades poseen son una basada en la función probit, d_{Probit} (Glass et al., 1981), y otras dos basadas en la función logística propuestas por Cox, d_{Cox} (Haddock et al., 1998) y por Hasselblad y Hedges (1995), d_{HH} , respectivamente (Sánchez-Meca, Marín-Martínez y Chacón-Moscó, 2003). El índice d_{Probit} y su varianza intra-estudio se obtienen mediante:

$$d_{\text{Probit}} = (z_E - z_C)$$

$$V(d_{\text{Probit}}) = \left[\frac{2\pi p_E(1-p_E)e^{-z_E^2}}{n_E} + \frac{2\pi p_C(1-p_C)e^{-z_C^2}}{n_C} \right],$$

siendo z_E y z_C la inversa de la función de distribución de la curva normal para p_E y p_C , respectivamente

$$[z_E = \Phi^{-1}(p_E); z_C = \Phi^{-1}(p_C)].$$

El índice d_{Cox} y su varianza intra-estudio se obtienen mediante:

$$d_{\text{Cox}} = Lor / 1.65$$

$$V(d_{\text{Cox}}) = 0.367 \left[\frac{1}{O_{1E}} + \frac{1}{O_{2E}} + \frac{1}{O_{1C}} + \frac{1}{O_{2C}} \right]$$

siendo Lor el logaritmo del odds ratio. Y el índice d_{HH} y su varianza intra-estudio vienen dados por:

$$d_{\text{HH}} = Lor \frac{\sqrt{3}}{\pi},$$

$$V(d_{\text{HH}}) = \frac{3}{\pi^2} \left[\frac{1}{O_{1E}} + \frac{1}{O_{2E}} + \frac{1}{O_{1C}} + \frac{1}{O_{2C}} \right].$$

Todos los índices hasta aquí comentados tienen en común que son aplicables a diseños de dos grupos independientes. Si el diseño es de un solo grupo con pretest y postest, el índice más apropiado para reflejar el efecto es la diferencia entre las puntuaciones de cambio estandarizada, d_{Change} :

$$d_{\text{Change}} = \frac{\bar{y}_{\text{Pre}} - \bar{y}_{\text{Post}}}{S},$$

donde S es una estimación de la desviación estándar, para la cual se han propuesto diferentes opciones. En concreto, Gibbons, Hedeker y Davis (1993) proponen utilizar la desviación estándar de las puntuaciones de cambio entre el pretest y el posttest (índice d_2). Dunlap, Cortina, Vaslow y Burke (1996) y Becker (1988) consideran más apropiado utilizar la desviación estándar del pretest (índice d_1) como un estimador no contaminado por el efecto del tratamiento:

$$d_1 = \frac{\bar{y}_{\text{Pre}} - \bar{y}_{\text{Post}}}{S_{\text{Pre}}}$$

$$d_2 = \frac{\bar{y}_{\text{Pre}} - \bar{y}_{\text{Post}}}{S_D}$$

Las varianzas intra-estudio de los índices d_1 y d_2 son, respectivamente:

$$V(d_1) = \frac{2(1-r)}{n} \left(\frac{n-1}{n-3} \right) \left(1 + \frac{n}{2(1-r)} d_1^2 \right) - \frac{d_1^2}{[c(n-1)]^2}$$

$$V(d_2) = \frac{n-1}{n(n-3)} (1 + n d_2^2) - \frac{d_2^2}{[c(n-1)]^2},$$

siendo n el tamaño muestral, r el coeficiente de correlación de Pearson entre las puntuaciones del pretest y el posttest y $c(m)$ el factor de corrección del sesgo para muestras pequeñas, que se obtiene mediante:

$$c(n-1) = 1 - \frac{3}{4(n-1)-1}$$

Otra situación común en los diseños con asignación a grupos se da cuando los dos grupos (experimental y control) se registran en el pretest y en el posttest. En estas condiciones, el índice de la familia d más apropiado debería comparar el cambio que se produce en un grupo con el cambio que se produce en el otro. Pero, dado que existen al menos dos posibles índices para definir el cambio estandarizado del pretest al posttest en cada grupo (índices d_1 y d_2), según la estimación de la desviación estándar utilizada, también cabe proponer dos índices de la diferencia entre cambios medios estandarizados (Becker, 1988; Morris, 2000; Morris y DeShon, 2002). Siendo d_{1E} y d_{1C} las diferencias de cambio estandarizadas, tomando las desviaciones estándar del pretest, para los grupos experimental y control respectivamente, el primer índice de la diferencia entre las puntuaciones de cambio lo representaremos por Δ_1 , y viene definido por:

$$\Delta_1 = d_{1E} - d_{1C} = \frac{\bar{y}_{\text{Pre}}^E - \bar{y}_{\text{Post}}^E}{S_{\text{Pre}}^E} - \frac{\bar{y}_{\text{Pre}}^C - \bar{y}_{\text{Post}}^C}{S_{\text{Pre}}^C}$$

con varianza intra-estudio

$$V(\Delta_1) = \frac{n_E + n_C}{n_E n_C} \left(\frac{N-2}{N-4} \right) \left(1 + \frac{n_E n_C}{n_E + n_C} \Delta_1^2 \right) - \frac{\Delta_1^2}{[c(N-2)]^2}$$

y siendo el factor de corrección $c(N-2)$:

$$c(N-2) = c(m) = 1 - \frac{3}{4(n_E + n_C) - 9}$$

Del mismo modo, siendo d_{2E} y d_{2C} las diferencias de cambio estandarizadas, tomando las desviaciones estándar de las puntuaciones de cambio, para los grupos experimental y control respectivamente, el segundo índice de la diferencia entre las puntuaciones de cambio lo representaremos por Δ_2 , y viene definido por:

$$\Delta_2 = d_{2E} - d_{2C} = \frac{y_{\text{Pre}}^{-E} - y_{\text{Post}}^{-E}}{S_D^E} - \frac{y_{\text{Pre}}^{-C} - y_{\text{Post}}^{-C}}{S_D^C}$$

con varianza intra-estudio

$$V(\Delta_2) = \frac{n_E + n_C}{2(1-r)n_E n_C} \left(\frac{N-2}{N-4} \right) \left(1 + 2(1-r) \frac{n_E n_C}{n_E + n_C} \Delta_2^2 \right) - \frac{\Delta_2^2}{[c(N-2)]^2} .$$

• **La familia r .** Cuando los estudios han aplicado un diseño correlacional en el que han medido y/o registrado variables, el índice del tamaño del efecto más adecuado es el coeficiente de correlación de Pearson, r , o alguna extensión de éste cuando la variable no es cuantitativa (correlación de Spearman, correlación biserial-puntual, coeficiente *phi*, etc.). Se recomienda utilizar la transformación a Z de Fisher de los coeficientes de correlación antes de integrarlos en el meta-análisis, con objeto de homogeneizar sus varianzas (Hedges y Olkin, 1985; Rosenthal, 1991, 1994):

$$Z_r = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) .$$

En este caso, la varianza intra-estudio viene dada por:

$$V(Z_r) = \frac{1}{N-3} .$$

Cuando resulte oportuno, existen fórmulas algebraicas que permiten transformar un índice de la familia d a uno de la familia r , y viceversa. En concreto, una diferencia de medias estandarizada, d , puede transformarse a coeficiente de correlación mediante (Rosenthal, 1991; Sánchez Meca y Ato, 1989):³

$$r = \frac{d}{\sqrt{d^2 + 4}} .$$

MODELOS ESTADÍSTICOS ENMATA-ANÁLISIS

A partir de las estimaciones del tamaño del efecto obtenidas en los estudios primarios, es posible responder a las preguntas que se mencionaron en el epígrafe 5.1 referente a las fases de un meta-análisis: ¿Cuál es la magnitud del efecto medio? ¿Cuál es el intervalo de confianza en torno a ésta? ¿Es significativa la magnitud del efecto medio? ¿Existe heterogeneidad entre las estimaciones del tamaño del efecto? ¿Pueden encontrarse variables moderadoras de tal heterogeneidad? A lo largo de la corta historia del meta-análisis se han propuesto diferentes modelos estadísticos para abordar los análisis. Pueden consultarse Chalmers, Hedges y Olkin (2002), Field (2001, 2003a, 2003b, 2005), Schmidt (2008), Schmidt y Hunter (1999), Schmidt, Oh y Hayes (en prensa), Viechtbauer (2007a) y algunos de nuestros trabajos (Ma-

³ A excepción de los índices diferencia de riesgos, dr , razón de riesgos, rr , y odds ratio, or , esta fórmula de transformación puede aplicarse al resto de índices de la familia d .

rín-Martínez y Sánchez-Meca, 1998; Sánchez-Meca y Marín-Martínez, 1997, 1998a, 1998b, 2008; Sánchez-Meca, Marín-Martínez y Huedo, 2006) para una descripción más detallada de las similitudes y diferencias entre ellos.

Existe en la actualidad un claro consenso hacia el reconocimiento de aquellos modelos meta-analíticos que en sus análisis estadísticos ponderan cada tamaño del efecto en función de su precisión. Aun así, cabe distinguir entre dos modelos estadísticos específicamente diseñados para su aplicación sobre bases de datos meta-analíticas: Los *modelos de efectos fijos* y de *efectos aleatorios* (Erez, Bloom y Wells, 1996; Hedges, 1994; Hedges y Olkin, 1985; Hedges y Vevea, 1998; Hunter y Schmidt, 2000; Overton, 1998; Raudenbush, 1994; Schmidt, 2008; Sutton y Higgins, 2008).

Los modelos de efectos fijos y aleatorios difieren en la concepción de la población de estudios de partida. En el **modelo de efectos fijos** se asume que los estudios incluidos en el meta-análisis están estimando a un mismo, y único, tamaño del efecto paramétrico, θ , por lo que la única variabilidad asumida es la debida a error de muestreo aleatorio o varianza intra-estudio, $V(T_i)$, es decir, al hecho de que los estudios utilizan muestras de sujetos diferentes. Siendo T_i el *i*ésimo tamaño del efecto⁴ de un conjunto de k tamaños del efecto independientes que están estimando a un mismo efecto poblacional, θ , el modelo matemático de efectos fijos se formula como:

$$T_i = \theta + \varepsilon_i.$$

En el **modelo de efectos aleatorios** se asume que los estudios estiman a una distribución de tamaños del efecto paramétricos en la población, que sigue una ley normal [$\theta_i \sim N(\mu_\theta; \sigma_\theta^2)$] por lo que, además de la variabilidad debida al error de muestreo o intra-estudio, $V(T_i)$, hay que contemplar también la variabilidad inter-estudios, σ_θ^2 . En consecuencia, el modelo matemático incorpora dos términos de error, uno debido a la variabilidad intra-estudio, ε_i , que coincide con el del modelo de efectos fijos, y otro que refleja la variabilidad inter-estudios, ξ_i :

$$T_i = \mu_\theta + \xi_i + \varepsilon_i.$$

Mientras que en el modelo de efectos fijos el factor de ponderación de cada índice T_i viene determinado exclusivamente por la varianza intra-estudio, $w_i = 1/V(T_i)$, en el de efectos aleatorios el factor de ponderación lo está en función de la varianza intra-estudio y la varianza inter-estudios, $w_i^* = 1/[V(T_i) + \sigma_\theta^2]$. Las consecuencias de asumir uno u otro modelo afectan al grado de generalización de los resultados del meta-análisis. En el modelo de efectos fijos, la generalización se limita a la población de estudios de características similares a los incluidos en el meta-análisis. En el modelo de efectos aleatorios, por el contrario, los resultados pueden generalizarse a una población mayor de posibles estudios.

La varianza intra-estudio de cada tamaño del efecto se estima con las ecuaciones tratadas en el epígrafe anterior, de forma que cada estudio tiene su propia varianza intra-estudio, que estará fundamentalmente en función del tamaño muestral: a menor tamaño muestral, mayor varianza intra-estudio y, por tanto, menor precisión, por lo que el peso específico que el estudio ejerce en los cálculos meta-analíticos será menor.

La varianza inter-estudios tiene que estimarse a partir de las propias estimaciones del tamaño del efecto incluidas en el meta-análisis. El procedimiento habitualmente utilizado es el propuesto por DerSimonian y Laird (1986), basado en el método de los momentos:

4 Se entiende que T_i puede representar a cualquiera de los índices del tamaño del efecto tratados en el epígrafe anterior.

$$\hat{\sigma}_{\theta}^2 = \begin{cases} \frac{Q - (k - 1)}{c} & (\text{para } Q \geq k - 1) \\ 0 & (\text{para } Q < k - 1) \end{cases},$$

donde c se obtiene mediante:

$$c = \sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i},$$

siendo $w_i = 1/V(T_i)$ el factor de ponderación en función únicamente de la varianza intra-estudio. Y Q es el estadístico de heterogeneidad, que se obtiene mediante:

$$Q = \sum_{i=1}^k w_i (T_i - \bar{T})^2,$$

donde \bar{T} es el tamaño del efecto medio ponderado por la inversa de la varianza intra-estudio:

$$\bar{T} = \frac{\sum_{i=1}^k w_i T_i}{\sum_{i=1}^k w_i}.$$

No obstante, se han propuesto otros estimadores de la varianza inter-estudios que pueden funcionar mejor que el aquí descrito. Pueden consultarse a este respecto los trabajos de Viechtbauer (2005, 2007b, 2007c) y de Sánchez-Meca y Marín-Martínez (2008).

Desde el modelo de efectos fijos, el tamaño del efecto medio, \bar{T} , se estima según acabamos de exponer; mientras que asumiendo un modelo de efectos aleatorios, los factores de ponderación incluyen no sólo varianza intra-estudio, sino también una estimación de la varianza inter-estudios:

$$\bar{T}^* = \frac{\sum_{i=1}^k w_i^* T_i}{\sum_{i=1}^k w_i^*}.$$

Puede, así mismo, obtenerse un intervalo de confianza en torno al tamaño del efecto medio en ambos modelos, asumiendo una distribución normal:

$$\text{Modelo de efectos fijos: } P\left(\bar{T} - |z_{\alpha/2}| \sqrt{V(\bar{T})} \leq \theta \leq \bar{T} + |z_{\alpha/2}| \sqrt{V(\bar{T})}\right) = 1 - \alpha$$

$$\text{Modelo de efectos aleatorios: } P\left(\bar{T}^* - |z_{\alpha/2}| \sqrt{V(\bar{T}^*)} \leq \theta \leq \bar{T}^* + |z_{\alpha/2}| \sqrt{V(\bar{T}^*)}\right) = 1 - \alpha$$

Para cada caso, la varianza del tamaño del efecto medio viene dada por:

$$V(\bar{T}) = 1 / \sum_{i=1}^k w_i$$

$$V(\bar{T}^*) = 1 / \sum_{i=1}^k w_i^*.$$

En el modelo de efectos fijos, la hipótesis de homogeneidad entre los tamaños del efecto, se contrasta planteando ésta como hipótesis nula: $H_0: \theta_1 = \theta_2 = \dots = \theta_i = \dots = \theta_k$, y aplicando el estadístico Q arriba descrito, bajo cuya hipótesis se distribuye según χ^2 de Pearson con $k - 1$ grados de libertad.

En el modelo de efectos aleatorios, al constituir un modelo de componentes de varianza, la principal hipótesis a contrastar es la de que no existe variabilidad inter-estudios: $H_0: \hat{\sigma}_\theta^2 = 0$, que se contrasta con el mismo estadístico de heterogeneidad, Q .

Si existe heterogeneidad entre los tamaños del efecto estimados a través de los estudios, el paso siguiente deberá ser poner a prueba el influjo, o la asociación, de variables moderadoras sobre los tamaños del efecto, con objeto de poder explicar tal heterogeneidad. Cuando la variable moderadora es cualitativa el análisis estadístico es el equivalente a un modelo de ANOVA, pero estimando los parámetros por mínimos cuadrados ponderados; mientras que si la variable moderadora es continua, se aplica un modelo de regresión, nuevamente por el método de mínimos cuadrados ponderados. En general, el modelo de análisis cuando ponemos a prueba el influjo conjunto de un grupo de variables moderadoras (cualitativas y/o continuas), es un modelo lineal en el que el vector de coeficientes de regresión, β , se estima mediante:

$$\beta = (X'WX)^{-1}X'T$$

siendo X la matriz de diseño, que contiene las variables moderadoras, T el vector de tamaños del efecto, que actúa como variable dependiente, y W es una matriz de ponderación diagonal $n \times n$, que contiene los factores de ponderación para cada tamaño del efecto. Esta matriz de ponderaciones se obtiene como la inversa de la matriz diagonal de varianzas de los tamaños del efecto estimados, $W = \Sigma^{-1}$. Dependiendo de que el modelo asumido sea de efectos fijos o de efectos aleatorios, los elementos de la diagonal de W se obtienen en función de la varianza intra-estudio, para el primer modelo [$w_i = 1/V(T_i)$], o bien en función de la suma de la varianza intra-estudio y de una estimación de la varianza inter-estudios, para el modelo de efectos aleatorios [$w_i^* = 1/(V(T_i) + \hat{\sigma}_\theta^2)$]. En este último caso, la varianza inter-estudios tiene que estimarse teniendo en cuenta la matriz de diseño mediante (Rosenberg et al., 2000):

$$\hat{\sigma}_\theta^2 = \begin{cases} \frac{Q_E - (k - h)}{trW - tr[WX(X'WX)^{-1}X'W]} & (\text{para } Q_E \geq (k - h)) \\ 0 & (\text{para } Q_E < (k - h)) \end{cases}$$

donde k es el número de estudios (o tamaños del efecto), h es el número de parámetros del modelo, tr es la traza de una matriz y Q_E es la suma de cuadrados de error por mínimos cuadrados ponderados, que se obtiene mediante:

$$Q_E = T'WT$$

A partir del vector de coeficientes de regresión y de la matriz de varianzas y covarianzas estimadas de los coeficientes de regresión del modelo, S_β , es posible poner a prueba el influjo de variables moderadoras. Así, el contraste de la hipótesis nula de ausencia de efecto de las variables moderadoras, $H_0: \beta = 0$, se efectúa mediante la suma de cuadrados ponderada del modelo, Q_R :

$$Q_R = \beta' S_\beta^{-1} \beta$$

que bajo dicha hipótesis se distribuye según χ^2 de Pearson con h grados de libertad.

Así mismo, es posible poner a prueba la especificación del modelo contrastando la hipótesis nula de una correcta especificación, $H_0: \delta = X\beta$, siendo δ el vector de tamaños del

efecto poblacional, mediante la suma de cuadrados de error ponderada, Q_E , que bajo dicha hipótesis se distribuye según χ^2 de Pearson con $k - h$ grados de libertad.

El modo de operar analíticamente en un meta-análisis es similar para los dos modelos asumibles: efectos fijos o efectos aleatorios. Pero la capacidad de generalización de los resultados es mayor en efectos aleatorios que en efectos fijos. Además, se considera actualmente más plausible el modelo de efectos aleatorios, ya que es más difícil imaginar que en la realidad toda una población de estudios realizados sobre un mismo problema estén estimando exactamente a un mismo efecto poblacional. De hecho, puede considerarse el modelo de efectos fijos como un caso particular del modelo, más general, de efectos aleatorios, cuando no existe varianza inter-estudios. Pero ante una base de datos meta-analítica concreta, nunca puede saberse con certeza si el modelo más apropiado es uno u otro, ya que desconocemos los parámetros poblacionales.

Una estrategia que suele recomendarse consiste en iniciar los análisis poniendo a prueba la hipótesis de homogeneidad de los tamaños del efecto mediante la aplicación de la prueba Q . Si ésta resulta significativa, ello sería indicativo de que existe heterogeneidad entre los tamaños del efecto, por lo que debería asumirse un modelo de efectos aleatorios en los análisis subsiguientes. Si, por el contrario, el estadístico Q no alcanza la significación estadística, podría adoptarse el modelo de efectos fijos. Esta estrategia sería factible si no fuera porque la prueba Q de heterogeneidad se muestra poco potente cuando el número de estudios, k , no es elevado (Harwell, 1997; Sánchez-Meca y Marín-Martínez, 1997). Por tanto, un resultado no significativo de la prueba Q no es garantía suficiente para asumir un modelo de efectos fijos, pero un resultado significativo de dicha prueba sí lo es para asumir con fiabilidad un modelo de efectos aleatorios.

Otra opción, compatible en general con la acabada de exponer, se basa en la estimación de la varianza inter-estudios. Dado que el estimador de dicha varianza se trunca en el valor 0 para evitar estimaciones negativas de la varianza inter-estudios, se puede asumir un modelo de efectos fijos cuando se dé tal situación, y adoptar un modelo de efectos aleatorios en caso contrario. Por último, se ha propuesto recientemente un nuevo índice, denominado índice I^2 , para estimar el grado de heterogeneidad de los tamaños del efecto en un meta-análisis que, aunque está en función del estadístico Q , no se deja afectar por el número de estudios, k , y puede interpretarse como un porcentaje de heterogeneidad exhibido por los tamaños del efecto (Higgins y Thompson, 2002; Huedo-Medina, Sánchez-Meca, Marín-Martínez y Bote-lla, 2006).

VENTAJAS Y LIMITACIONES DEL META-ANÁLISIS

No cabe duda de que el meta-análisis se ha convertido en una herramienta metodológica de gran utilidad para ayudar a acumular el conocimiento científico en un determinado campo de investigación de una forma objetiva, sistemática y rigurosa. Entre sus ventajas cabe destacar las siguientes (Cooper, 1998; Hedges y Olkin, 1985; Rosenthal, 1991; Sánchez-Meca, 2003; Sánchez-Meca y Ato, 1989):

- (a) *Eficiencia*. El meta-análisis tiene una mayor capacidad para tratar grandes cantidades de información que las revisiones narrativas, gracias a sus posibilidades de cuantificar y de codificar objetivamente las variables implicadas en los resultados de los estudios.
- (b) *Rigor científico*. El meta-análisis cumple con las normas de rigor científico que se exigen en las investigaciones primarias, posibilitando la replicación del proceso. Se

trata, en definitiva, de exigir en el proceso de revisión de la investigación las mismas normas de rigor científico que se exigen en la realización de estudios primarios.

- (c) *Detección de efectos pequeños.* Al centrarse en la magnitud de los efectos en lugar de en los resultados de las pruebas de significación estadística, el meta-análisis tiene una mayor capacidad para detectar efectos pequeños que, sin embargo, pueden tener relevancia práctica o real. Las revisiones narrativas tienen más dificultades para detectar estos efectos que, precisamente, son los más habituales en las ciencias empíricas en general y en las Ciencias Sociales y del Comportamiento, en particular.
- (d) *Potencia estadística.* Al acumular los tamaños muestrales de los diferentes estudios integrados, los procedimientos de análisis estadístico propios del meta-análisis tienen mayor potencia estadística para detectar los efectos y las relaciones entre las variables implicadas.
- (e) *Énfasis en el tamaño del efecto.* Otra de las grandes aportaciones del meta-análisis al proceso de investigación es el gran énfasis que pone en el tamaño del efecto, relegando las pruebas de significación estadística a un segundo plano. Este énfasis ha contribuido a modificar las recomendaciones que actualmente se están proponiendo contra el uso abusivo de las pruebas de significación.
- (f) *Aprovechamiento de resultados contradictorios.* El meta-análisis dispone actualmente de procedimientos estadísticos que permiten detectar variables moderadoras responsables de los resultados heterogéneos y contradictorios que pueden encontrarse en un determinado campo de investigación.
- (g) *Seguridad.* Al basarse en una metodología sistemática, objetiva y rigurosa, las conclusiones a las que se llegan con los meta-análisis son más fiables y seguras que las alcanzadas en revisiones cualitativas o narrativas de la investigación.

No obstante, el meta-análisis tiene también limitaciones que es preciso conocer para aplicarlo correctamente. Éstas son las principales junto con las recomendaciones para tratar de paliar sus efectos adversos:

(a) *Propensión a cometer errores.* Datos e información defectuosa en los estudios primarios pueden afectar a la calidad de los datos meta-analíticos y, en consecuencia, a la fiabilidad de sus conclusiones. El meta-análisis está, pues, limitado por las propias deficiencias de los estudios primarios. Si no es posible obtener datos exactos de sus resultados, es preferible no hacer el meta-análisis.

(b) *Heterogeneidad.* Estudios muy heterogéneos entre sí no deberían ser integrados, ya que los resultados globales serían poco informativos. Es por ello que se recomienda actualmente realizar estudios meta-analíticos con unos objetivos claramente focalizados, para garantizar que las conclusiones tendrán relevancia práctica. Precisamente, una de las críticas más severas que ha recibido el meta-análisis es el denominado “problema de las manzanas y las naranjas” (‘the apples and oranges problem’), dirigido por Eysenck (1978, 1994) contra el ‘macro’ meta-análisis realizado originalmente por Smith y Glass (1977) sobre la eficacia de la psicoterapia. No exento de razón, estas críticas contra la mezcla de estudios muy heterogéneos, ha provocado que actualmente se lleven a cabo meta-análisis de tamaño mucho más reducido, con estudios cuyos diseños sean similares (por ejemplo, todos los estudios deben tener dos grupos, uno de control y otro experimental). En el límite de la homogeneización se encuentra la línea de investigación meta-analítica médica, que recomienda la integración mediante meta-análisis únicamente de ensayos controlados aleatorizados (Whitehead, 2002).

(c) *Calidad de los datos*. La mezcla de estudios de buena calidad metodológica con estudios de baja calidad puede dar lugar a estimaciones de los efectos sesgadas. Se recomienda codificar como una variable moderadora más la calidad del diseño del estudio y analizar su posible relación con los tamaños del efecto; o bien, fijar desde el principio normas estrictas de calidad metodológica que deben cumplir los estudios para ser incluidos en el meta-análisis. Existe un buen número de escalas para valorar la calidad metodológica de los estudios primarios, buena parte de las cuales se han desarrollado en la literatura médica (Jüni, Altman y Egger, 2001; Moher, Jones y Lepage, 2001; Tritchler, 1999; Wortman, 1994).

(d) *Representatividad*. Por muy exhaustiva que sea la búsqueda de la literatura, nunca será posible localizar todos los estudios seleccionables. Si, además, no se dispone de estudios no publicados, el sesgo de publicación puede ser una seria amenaza contra la validez de los resultados del meta-análisis. La inclusión de estudios no publicados es muy recomendable y, en cualquier caso, realizar un análisis del sesgo de publicación mediante el cálculo de índices de tolerancia a los resultados nulos (fail-safe N), para valorar la robustez del meta-análisis frente a esta amenaza. El problema del sesgo de publicación y sus efectos sobre los resultados de los meta-análisis es una línea de trabajo actualmente muy activa (cf. por ejemplo, Rothstein, Sutton y Borenstein, 2005).

(e) *Dependencia*. La inclusión de más de un índice del tamaño del efecto calculado sobre la misma muestra de sujetos atenta contra el supuesto de independencia de los datos que asumen las técnicas meta-analíticas de análisis estadístico. Este problema surge cuando un mismo estudio primario presenta resultados sobre diferentes variables dependientes. Si se calcula un tamaño del efecto para cada una de ellas y se incorporan todas las estimaciones al meta-análisis, se incurre en un problema de dependencia que afecta a la validez de la conclusión estadística. Se han propuesto varias soluciones para resolver este problema. Una de ellas es obtener un promedio de los tamaños del efecto correspondientes a un mismo estudio (Gleser y Olkin, 1994; Hedges y Olkin, 1985; Marín-Martínez y Sánchez-Meca, 1999; Rosenthal y Rubin, 1986). Otras soluciones implican, bien realizar meta-análisis diferentes para cada variable dependiente, bien modelar la estructura correlacional entre ellas (Becker, 2000; Kalaian y Raudenbush, 1996; Raudenbush, Becker y Kalaian, 1988).

Como conclusión, podemos decir que el meta-análisis es hoy día una metodología de investigación que se ha convertido en un elemento indispensable como puente entre la investigación pasada y la futura: Aunque se basa en la integración cuantitativa de las evidencias acumuladas sobre un determinado tópico, su principal función es orientar la investigación futura, detectando vacíos conceptuales y/o metodológicos, denunciando deficiencias en la literatura y revelando relaciones no anticipadas. Su gran ventaja está en la “suma de esfuerzos” y en su empeño por explicar las desviaciones que se observan entre los resultados de los estudios: “Estas pequeñas desviaciones, que por sí mismas pueden no llegar a ser significativas, cuando se combinan pueden revelar importantes conexiones científicas desde las que se puede extraer nuevas evidencias científicas” (Garfield, 1991, p. 5).

El meta-análisis no debe entenderse como un mero método para hacer revisiones de literatura, sino como un modo de afrontar el análisis de los datos que implica cambios en la forma de entender el progreso de una ciencia. Su énfasis en el tamaño del efecto ha ejercido una importante influencia a favor de las críticas contra la excesiva atención prestada a las pruebas de significación a lo largo de la historia de la psicología científica. Al mismo tiempo, el meta-análisis ha contribuido a relativizar el papel que los estudios primarios juegan en la demostración de hechos científicos y en la importancia de acumular de forma sistemática la investigación para el desarrollo de teorías cada vez más sólidas. En palabras de Schmidt (1992, p. 1179):

“Los datos nos llegan encriptados, y para entender su significado primero tenemos que descifrar el código. Para hacer esto se necesita del meta-análisis. Por tanto, los estudios individuales deben ser considerados como meros datos puntuales que contribuirán a un meta-análisis futuro. De esta forma, el estatus y el valor científico del estudio individual se ve necesariamente reducido”.

Siempre que se utilice con buen juicio y siendo consciente de sus limitaciones, el meta-análisis puede aportar una importante contribución a la acumulación del conocimiento científico y, por ende, al progreso de la ciencia.

REFERENCIAS BIBLIOGRÁFICAS

- Becker, B.J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical & Statistical Psychology*, 41, 257-278.
- Becker, B.J. (2000). Multivariate meta-analysis. En H.E.A. Tinsley y S. S.D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 499-525). San Diego, CA: Academic Press.
- Begg, C.B. (1994). Publication bias. En H. Cooper y L.V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 399-409). Nueva York: Russell Sage Foundation.
- Bond, C.F.Jr., Wiitala, W.L. y Richard, F.D. (2003). Meta-analysis of raw mean differences. *Psychological Methods*, 8, 406-418.
- Botella, J. y Gambara, H. (2002). *¿Qué es el meta-análisis?* Madrid: Siglo XXI.
- Botella, J. y Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical & Health Psychology*, 6, 425-440.
- Burdett, S., Stewart, L.A. y Tierney, J.F. (2003). Publication bias and meta-analyses. *International Journal of Technology Assessment in Health Care*, 19, 129-134.
- Chalmers, T.C., Hedges, L.V. y Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions*, 25, 12-37.
- Cooper, H. (1998). *Integrating research: A guide for literature reviews* (3ª ed.). Newbury Park, CA: Sage.
- Cooper, H. y Hedges, L.V. (Eds.) (1994). *The handbook of research synthesis*. Nueva York: Russell Sage.
- Cortina, J.M. y Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks, CA: Sage.
- Davey Smith (Ed.) (2001). *Clinical Meta-analysis*. Nueva York: Wiley.
- Deeks, J.J. y Altman, D.G. (2001). Effect measures for meta-analysis of trials with binary outcomes. En M. Egger, G.D. Smith y D.G. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (2ª ed.) (pp. 313-335). Londres: BMJ Books.
- DerSimonian, R. y Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177-188.
- Dunlap, W.P., Cortina, J.M., Vaslow, J.B. y Burke, M.J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1, 170-177.
- Eddy, D.M., Hasselblad, V. y Shachter, R. (1995). *Meta-analysis by the confidence profile methods: The statistical synthesis of evidence*. Boston, MA: Academic Press.
- Egger, M., Smith, G.D. y Altman, D.G. (Eds.) (2001). *Systematic reviews in health care: Meta-analysis in context* (2ª ed.). Londres: BMJ Books.

- Erez, A., Bloom, M.C. y Wells, M.T. (1996). Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology*, 49, 275-306.
- Eysenck, H.J. (1978). An exercise in mega-silliness. *American Psychologist*, 33, 517-517.
- Eysenck, H.J. (1994). Meta-analysis and its problems. *British Medical Journal*, 309, 789-792.
- Field, A.P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161-180.
- Field, A.P. (2003). Can meta-analysis be trusted? *The Psychologist*, 16, 642-645.
- Field, A.P. (2003). The problems of using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, 2, 77-96.
- Field, A.P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods*, 10, 444-467.
- Fleiss, J.L. (1994). Measures of effect size for categorical data. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). Nueva York: Russell Sage Foundation.
- Garfield, E. (1991). Meta-analysis and the metamorphosis of the scientific literature review. *Current Contents: Social & Behavioral Sciences*, 43, 5-8.
- Gibbons, R.D., Hedeker, D.R. y Davis, J.M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics*, 18, 271-279.
- Glass, G.V., McGaw, B. y Smith, M.L. (1981). *Meta-analysis in social research*. Newbury Park, CA: Sage.
- Gleser, L.J. y Olkin, I. (1994). Stochastically dependent effect sizes. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 339-355). Nueva York: Russell Sage Foundation.
- Gómez, J. (1987). *Meta-análisis*. Barcelona: PPU.
- Grissom, R.J. y Kim, J.J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Erlbaum.
- Haddock, C.K., Rindskopf, D. y Shadish, W.R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, 3, 339-353.
- Harwell, M. (1997). An empirical study of Hedges's homogeneity test. *Psychological Methods*, 2, 219-231.
- Hasselblad, V. y Hedges, L.V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, 117, 167-178.
- Hedges, L.V. (1994). Fixed effects models. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285-299). Nueva York: Russell Sage Foundation.
- Hedges, L.V. y Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L.V. y Vevea, J.L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486-504.
- Higgins, J.P.T. y Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558.
- Huedo-Medina, T.B., Sanchez-Meca, J., Marin-Martinez, F. y Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, 11, 193-206.
- Hunter, J.E. y Schmidt, F.L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection & Assessment*, 8, 275-292.

- Hunter, J.E. y Schmidt, F.L. (2004). *Methods of meta-analysis: Correcting error and bias in research synthesis* (2ª ed.). Sage.
- Jüni, P., Altman, D.G. y Egger, M. (2001). Assessing the quality of randomised controlled trials. En M. Egger, G.D. Smith y D.G. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (2ª ed.) (pp. 87-108). Londres: BMJ Books.
- Kalaian, H.A. y Raudenbush, S.W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological methods*, 1, 227-235.
- Lipsey, M.W. y Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Marín Martínez, F. y Sánchez Meca, J. (1998). Testing for dichotomous moderators in meta-analysis. *Journal of Experimental Education*, 67, 69-81.
- Marín Martínez, F. y Sánchez Meca, J. (1999). Averaging dependent effect sizes in meta-analysis: A cautionary note about procedures. *Spanish Journal of Psychology*, 2, 32-38.
- Marín-Martínez, F., Sánchez-Meca, J., Huedo, T.B. y Fernández-Guzmán, I. (2007). *Meta-análisis: ¿Dónde estamos y hacia dónde vamos?* En A. Borges y P. Prieto (Eds.), *Psicología y ciencias afines en los albores del siglo XXI (Homenaje al profesor Alfonso Sánchez Bruno)* (pp. 87-102). Grupo Editorial Universitario.
- Martín, J.L.R., Tobías, A. y Seoane, T. (Coords.) (2006). *Revisiones Sistemáticas en las Ciencias de la Vida*. Toledo: FISCAM.
- Moher, D., Jones, A. y Lepage, L. for the CONSORT Group (2001). Use of the CONSORT statement and quality of reports for randomized trials: A comparative before-and-after evaluation. *Journal of the American Medical Association*, 285, 1992-1995.
- Morris, S.B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical & Statistical Psychology*, 53, 17-29.
- Morris, S.B. y DeShon, R.P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7, 105-125.
- Overton, R.C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3, 354-379.
- Petticrew, M. y Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A practical guide*. Malden, MA: Blackwell.
- Raudenbush, S.V. (1994). Random effects models. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321). Nueva York: Russell Sage Foundation.
- Raudenbush, S.W., Becker, B.J. y Kalaian, H.A. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, 103, 111-120.
- Roberts, B.W., Kuncel, N.R., Viechtbauer, W. y Bogg, T. (2007). Meta-analysis in personality psychology: A primer. En R.W. Robins, R.C. Fraley y R.F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 652-672). Nueva York: Guilford Press.
- Rosenberg, M.S., Adams, D.C. y Gurevitch, J. (2000). *MetaWin: Statistical software for meta-analysis* (Vers. 2.0). Sunderland, MA: Sinauer Assoc.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (ed. Rev.). Newbury Park, CA: Sage.
- Rosenthal, R. (1994). Parametric measures of effect size. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). Nueva York: Russell Sage Foundation.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118, 183-192.
- Rosenthal, R., Rosnow, R.L. y Rubin, D.B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, UK: Cambridge University Press.

- Rosenthal, R. y Rubin, D.B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 99, 400-406.
- Rothstein, H.R., Sutton, A.J. y Borenstein, M. (Eds.) (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. Chichester, UK: Wiley.
- Sánchez-Meca, J. (1999). Meta-análisis para la investigación científica. En F.J. Sarabia (Coord.), *Metodología para la Investigación en Marketing y Dirección de Empresa* (pp. 173-201). Madrid: Pirámide.
- Sánchez-Meca, J. (2003). La revisión del estado de la cuestión: El meta-análisis. En C. Camisón, M.J. Oltra y M.L. Flor (Eds.), *Enfoques, problemas y métodos de investigación en economía y dirección de empresas. Tomo I* (pp. 101-110). Castellón: ACEDE / Fundación Universitat Jaume I – Empresa.
- Sánchez-Meca, J. y Ato, M. (1989). Meta-análisis: Una alternativa metodológica a las revisiones tradicionales de la investigación. En J. Arnau y H. Carpintero (Eds.), *Tratado de psicología general I: Historia, teoría y método* (pp. 617-669). Madrid: Alhambra.
- Sánchez-Meca, J. y Marín-Martínez, F. (1997). Homogeneity tests in meta-analysis: A Monte Carlo comparison of statistical power and Type I error. *Quality & Quantity*, 31, 385-399.
- Sánchez-Meca, J. y Marín-Martínez, F. (1998a). Testing continuous moderators in meta-analysis: A comparison of procedures. *British Journal of Mathematical & Statistical Psychology*, 51, 311-326.
- Sánchez-Meca, J. y Marín-Martínez, F. (1998b). Weighting by inverse-variance or by sample size in meta-analysis: A simulation study. *Educational & Psychological Measurement*, 58, 211-220.
- Sánchez-Meca, J. y Marín-Martínez, F. (2000). Testing the significance of a common risk difference in meta-analysis. *Computational Statistics & Data Analysis*, 33, 299-313.
- Sánchez-Meca, J. y Marín-Martínez, F. (2001). Meta-analysis of 2x2 tables: Estimating a common risk difference. *Educational & Psychological Measurement*, 61, 249-276.
- Sánchez-Meca, J. y Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, 13, 31-48.
- Sánchez-Meca, J., Marín-Martínez, F. y Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8, 448-467.
- Sánchez-Meca, J., Marín-Martínez, F. y Huedo, T. (2006). Modelo de efectos fijos y modelo de efectos aleatorios. En J.L.R. Martín, A. Tobías y T. Seoane (Coords.), *Revisiones Sistemáticas en Ciencias de la Vida* (pp. 189-204). Toledo: FISCAM.
- Schmidt, F.L. (1992). What do data really mean? *American Psychologist*, 47, 1173-1181.
- Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1, 115-129.
- Schmidt, F.L. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods*, 11, 96-113.
- Schmidt, F.L. y Hunter, J.E. (1995). The impact of data analysis method on cumulative knowledge: Statistical significance testing, confidence intervals, and meta-analysis. *Evaluation & the Health Professions*, 18, 408-427.
- Schmidt, F.L. y Hunter, J.E. (1999). Comparison of three meta-analysis methods revisited: An analysis of Johnson, Mullen, and Salas (1995). *Journal of Applied Psychology*, 84, 144-148.
- Schmidt, F.L., Oh, I.-S. y Hayes, T.L. (en prensa). Fixed versus random effects models in meta-analysis: Model properties and an empirical comparison of difference in results. *British Journal of Mathematical and Statistical Psychology*.
- Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Hogrefe & Huber Pub.

- Shadish, W.R. y Haddock, C.K. (1994). Combining estimates of effect size. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261-281). Nueva York: Russell Sage Foundation.
- Smith, M.L. y Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Sutton, A.J., Abrams, K.R., Jones, D.R., Sheldon, T.A. y Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester, UK: Wiley.
- Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R. y Jones, D.R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, 320, 1574-1577.
- Sutton, A.J. y Higgins, J.P.T. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27, 625-650.
- Titchler, D. (1999). Modelling study quality in meta-analysis. *Statistics in Medicine*, 18, 2135-2145.
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational & Behavioral Statistics*, 30, 261-293.
- Viechtbauer, W. (2007a). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Journal of Psychology*, 215, 104-121. Viechtbauer, W. (2007b). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, 26, 37-52.
- Viechtbauer, W. (2007c). Hypothesis tests for population heterogeneity in meta-analysis. *British Journal of Mathematical & Statistical Psychology*, 60, 29-60.
- Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*. Chichester, UK: Wiley.
- Wolf, F.M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage.
- Wortman, P.M. (1994). Judging research quality. En H. Cooper y L.V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 97-109). Nueva York: Russell Sage Foundation.