

ARTICULO METODOLÓGICO

El Meta-Análisis: La Integración de los Resultados Científicos

Guillermo Macbeth^{1*} - Nuria Cortada de Kohan* - Eugenia Razumiejczyk*

* Instituto de Investigaciones Psicológicas - Universidad del Salvador (IIPUS).

Resumen. Los investigadores pueden emprender varios tipos de análisis al revisar el estado del conocimiento científico. El meta-análisis es uno de ellos y se define como una revisión cuantitativa de la bibliografía científica. El propósito de este artículo es presentar un panorama de sus técnicas. Sus ventajas principales son la aproximación a una mayor objetividad en la revisión de los conocimientos acumulados, la uniformidad de las conclusiones, la condición cuantitativa de sus procedimientos y su replicabilidad. Mientras que el método de Winer emplea la prueba t para la combinación de estudios, el método de Stouffer propone una combinación que se realiza sobre puntajes Z estandarizados. Son múltiples las fuentes de error en meta-análisis. Se analiza el sesgo de publicación. Se concluye que el meta-análisis es una técnica de gran utilidad pero con importantes limitaciones. **Palabras clave:** meta-análisis, revisión bibliográfica, integración de resultados, tamaño del efecto, sesgo de publicación

Abstract: Researchers can perform different kinds of analysis when reviewing the state of the scientific knowledge. Meta-analysis is one of them and can be defined as a quantitative revision of the scientific bibliography concerning a specific research problem. The aim of this article is to provide a general account of its techniques. The main advantages of the meta-analysis are a better approach to objectivity, the uniformity of its conclusions, the quantitative condition of its procedures, and its replicability. Winer's and Stouffer's methods are reviewed and analyzed. The publication bias is presented as a source of error. It is concluded that meta-analysis is a useful but limited technique. **Key Words:** meta-analysis, bibliography revision, findings integration, effect size, publication bias

1. Introducción

Señala Glass (1976) que los investigadores pueden emprender tres tipos de análisis al revisar el estado del conocimiento científico que se ha acumulado sobre un problema de

¹ Por favor dirigir la correspondencia relacionada con este artículo a:
Guillermo Macbeth
Doctor en Psicología. - Becario postdoctoral del CONICET,
Instituto de Investigaciones Psicológicas, Universidad del Salvador (IIPUS).
Dirección: Marcelo T. de Alvear 1314 (C1058AAV), Buenos Aires, Argentina.
Teléfono: (+5411)-4813-3404
E-mail: guillermo.macbeth@mail.salvador.edu.ar

investigación: 1) El análisis primario: el investigador puede comunicar sus hallazgos empíricos en un artículo científico, de este modo, el análisis consiste en un conjunto de pruebas estadísticas y decisiones metodológicas que se llevan a cabo sobre la evidencia obtenida de manera directa; 2) El análisis secundario: implica un re-análisis sobre los datos aportados por artículos primarios con nuevas técnicas estadísticas que no se han aplicado en el artículo original; 3) El meta-análisis: su propósito es la integración de los resultados dispersos en múltiples estudios que ponen a prueba una misma hipótesis, es decir, se trata de un análisis conjunto de otros análisis realizados anteriormente. Glass (1976) define el meta-análisis como el análisis estadístico de un conjunto de estudios individuales que se efectúa con el fin de integrarlos mediante un estimador global, de manera que resulte posible obtener una conclusión sumaria. Por lo tanto, el meta-análisis puede ser definido, en sentido amplio, como una revisión cuantitativa de la bibliografía científica (Petticrew & Roberts, 2005).

Se ha demostrado que la revisión meta-analítica lleva, en algunos casos, a conclusiones inalcanzables para una revisión tradicional (Schmidt, 1992). Los resultados de estas revisiones cuantitativas poseen relevancia para la toma de decisiones en la administración de las políticas de salud, sobre todo para la prevención de enfermedades (Song, Sheldon, Sutton, Abrams & Jones, 2001). Los estudios meta-analíticos de Richard Peto, por ejemplo, han demostrado que se podrían haber salvado, sólo en Estados Unidos, de 10.000 a 20.000 vidas entre 1976 y 1985 si las políticas de prevención de infarto de miocardio hubiesen contemplado los resultados de sucesivos estudios meta-analíticos sobre el efecto beneficioso de la aspirina (Letón Molina & Pedromingo Marino, 2001; Peto, Yusuf, Lewis, Collins & Sleight, 1985).

Contar con la autorizada opinión de los especialistas en un área acotada de investigación, tal como ocurre en los estados del arte habitual o cualitativo puede, en ocasiones, desaprovechar información que sólo puede obtenerse mediante el análisis estadístico integrado de los estudios conocidos. Tradicionalmente, suelen sacarse conclusiones sumarias narrativas, condicionadas en distinta medida por las preferencias teóricas y metodológicas del revisor. La dificultad de las revisiones tradicionales se observa, además, en la acumulación de conocimientos especializados que deben poseerse como condición indispensable para poder realizarlas. Se debe ser un especialista para emprender una revisión cualitativa tradicional. En este sentido, resultan claramente preferibles los estados-del-arte preparados por uno o varios especialistas, a los preparados por un no

especialista (Letón Molina & Pedromingo Marino, 2001).

Se comprende la utilidad y eficacia de las revisiones bibliográficas meta-analíticas, cuya presencia en las publicaciones periódicas de investigación se hace cada vez más notoria (Hunter, Schmidt & Le, 2006; L'Abbé, Detsky & O'Rourke, 1987; Rodríguez & Maeda, 2006). Las ventajas principales que presenta este tipo de análisis son, por un lado, la aproximación a una mayor objetividad en la revisión y síntesis de los conocimientos acumulados y, por otro lado, la uniformidad de las conclusiones, la condición cuantitativa de sus procedimientos y su replicabilidad (Wolf, 1986). La réplica de un estudio meta-analítico resulta relativamente fácil de aprender en comparación con lo dificultoso que resulta adquirir pericia en un área específica de investigación, tarea que a veces requiere años de estudio.

Durante la década de 1930 coincidieron Ronald A. Fisher (1932) y Karl Pearson (1933) en señalar la importancia de construir procedimientos estadísticos que permitan analizar, de manera conjunta, los resultados de una colección de pruebas independientes de una misma hipótesis. Desde entonces se han ido construyendo numerosos procedimientos para lograr esta combinación (Hunter & Schmidt, 2004). Entre las más importantes técnicas que se han propuesto en la historia del meta-análisis se destacan el método de Fisher (1932), el de Winer (1971) y el atribuido a Stouffer, Suchman, De Vinney, Star & Williams (1949), tal como lo señala Wolf (1986). Estos procedimientos utilizan información obtenida de otros estudios ya publicados. La combinación se hace sobre la base de niveles de significación (*p* valores) y sobre estadísticos como la prueba *t* o puntajes *Z*, entre otros (Wolf, 1986). El método de Fisher, por su parte, presenta severas limitaciones que han sido analizadas en profundidad por varios especialistas (Fisher, 1932; Rosenthal & Rosnow, 1991; Wolf, 1986).

2. Métodos meta-analíticos

El método de Winer (1971) se emplea para la combinación de estudios cuyos datos se obtienen de la prueba *t* para medir diferencias entre medias. La distribución que se obtiene al aplicar esta prueba estadística supone la normalidad de la distribución que se expresa formalmente de la siguiente manera: $Z_c \approx N(0,1)$, donde Z_c significa combinación de pruebas *t*, el símbolo \approx significa distribución aproximada a, y $N(0,1)$ significa que la aproximación es una normal con media cero y varianza uno (como parámetros). Estos requisitos se cumplen cuando los grados de libertad son iguales o mayores a 10, por lo cual se recomienda no aplicar este método cuando se incluyen muy pocas pruebas en la combinación. La expresión

formal se presenta a continuación.

$$Z_c = \frac{\sum t}{\sqrt{\sum [gl/(gl-2)]}} \approx N(0,1)$$

El método consiste en obtener la Z_c que implica la combinación de varias pruebas t de los estudios conocidos en un único número que se obtiene de dividir la suma de las t , como se indica en el numerador, por un denominador que emplea los grados de libertad. El denominador es la raíz cuadrada de la suma (Σ) de los cocientes que se obtienen de dividir la cantidad de grados de libertad (gl) por los grados de libertad menos dos ($gl - 2$). Es decir, la prueba de meta-análisis propuesta por Winer (1971) consiste en obtener un cociente que surge de la ponderación de la suma de las pruebas t . La Z_c presenta una distribución aproximadamente normal, es decir, $\approx N(0,1)$ para 10 o más grados de libertad.

Por otra parte, Wolf (1986) ha expuesto el método de Stouffer (Stouffer et al., 1949), cuyo propósito es la combinación que se realiza sobre puntajes Z estandarizados que se obtienen de los p valores de las pruebas originales. Son estos puntajes Z los que se suman y luego dividen por la raíz cuadrada de N , siendo N la cantidad de estudios que se incluyen en la combinación. La expresión formal es la siguiente:

$$Z_c = \frac{\sum z}{\sqrt{N}} \sim N(0,1)$$

Los resultados en este método se comportan como una distribución normal, con media cero y varianza uno, $\sim N(0,1)$. En comparación con el método de Winer (1971), que analiza las respectivas t de los estudios combinados y cuya distribución es aproximadamente normal $\approx N(0,1)$, el método de Stouffer trabaja con una distribución estrictamente normal, esto es $\sim N(0,1)$, ocupándose de la combinación que se realiza sobre puntajes Z estandarizados.

Por lo tanto, el método de Stouffer et al. (1949) presenta mayores ventajas en tanto su cálculo es directo y no necesita, entre otras condiciones, transformaciones logarítmicas, como ocurre con el método de Fisher (1932), o correcciones para los grados de libertad, como lo exige el método de Winer (1971), que además resulta no aplicable para muestras muy pequeñas menores a 10 grados de libertad. No obstante, los resultados que se obtienen de aplicar los tres métodos son consistentes entre sí (Wolf, 1986). Sus diferencias pueden ser consideradas despreciables en la práctica. La opción por uno o por otro método deberá considerar las características de la información que ofrecen los estudios originales y la

facilidad con que puedan combinarse.

3. El tamaño del efecto

Cohen (1988) define el tamaño del efecto como el grado en que el fenómeno bajo estudio está presente en la población. Esta medida indica cuán fuerte es la falsedad de la hipótesis nula (que afirma la ausencia de efecto), sin implicar por ello afirmaciones de tipo causal (Wolf, 1986). Se refiere, pues, a la medida en que el efecto trasciende los límites de la muestra aleatoria estudiada de manera que los análisis estadísticos permitan al investigador sostener afirmaciones referidas a la población. El tamaño del efecto es uno de los componentes que posee mayor relevancia para el análisis del poder estadístico. El poder (o potencia) es una medida que indica con cuánta fuerza está el investigador evitando equivocarse con un error de tipo II, es decir, cuánto acierta al arriesgarse en la aceptación de un efecto que existe realmente, más allá de la decisión tomada. Siendo el poder definido como $1 - \beta$, y entendiendo que β es la probabilidad de cometer un error de tipo II, se puede afirmar que a mayor tamaño del efecto, mayor poder. Es por ello que el tamaño del efecto posee un gran interés para cualquier prueba de hipótesis, en general, y para el meta-análisis, en particular (Wolf, 1986).

Cohen (1988) propone obtener la medida d , que requiere estandarizar las diferencias observadas entre los grupos.

$$d = \frac{M_1 - M_2}{\sigma}$$

Se dividen las diferencias brutas entre los grupos (experimental y control, o pre-test y post-test) por el desvío estándar de las respectivas poblaciones (σ). La medida d de Cohen (1988) es el tamaño del efecto expresado en unidades de desvío estándar para la diferencia de medias entre los grupos, es decir, M_1 y M_2 . Esta diferencia se analiza mediante la prueba t . Nótese que el desvío estándar poblacional (σ) para estas diferencias no es generalmente conocido, por lo que se debe emplear en su lugar el desvío estándar (de) de las medias de los grupos estudiados, que es conocido. Como el diseño experimental procura que la variabilidad de los grupos experimental y control sea homogénea, resulta legítimo emplear el desvío estándar de cualquiera de los dos grupos. La variabilidad interna de la variable dependiente se comporta de una manera estable entre los grupos porque las condiciones de variabilidad se encuentran controladas (Ades, Lu & Higgins, 2005). El desplazamiento de la variable

dependiente ocurre, idealmente, en función de la exclusiva manipulación de la variable independiente, por lo que se supone cierta homogeneidad en la dispersión de las mediciones de cada grupo, más allá de los errores de muestreo (Hönekopp, Beck & Oswald, 2006). Así, resulta legítimo reemplazar la σ poblacional por el desvío estándar (*de*) de cualquiera de los dos grupos, el experimental o el control, el de pre-test o el de post-test.

$$d = \frac{M_1 - M_2}{de}$$

En esta expresión, la *d* es el tamaño del efecto para la diferencia estandarizada de la media de los dos grupos comparados en el experimento, es decir M_1 y M_2 (también podría expresarse como \bar{X}_1 y \bar{X}_2). El denominador *de* se refiere al desvío estándar homogéneo, es decir, a la variabilidad de cualquiera de los grupos conocidos, y no de la población desconocida (σ). Esta homogeneidad debe calcularse mediante algún test de homogeneidad de las varianzas, por ejemplo, el test de Levene (Letón Molina & Pedromingo Marino, 2001).

En ambos casos, la *d* adquirirá valores más altos cuánto más grande sea el numerador y más pequeño el denominador, es decir, cuando mayor sea la diferencia entre los grupos y menor sea la dispersión interna. Siendo la *d* una medida que trabaja en unidades de desvío estándar, más grande será la *d* cuánto más grande sea el tamaño del efecto, entendido como magnitud de la fuerza de la diferencia estandarizada que los experimentos lograron producir entre los grupos.

El promedio de las diferencias estandarizadas (d_{prom}) es un estimador global meta-analítico que se obtiene de sumar las diferencias estandarizadas entre los grupos de cada experimento y de dividirlos por la cantidad de experimentos estudiados. El énfasis en la consideración de la hipótesis de nulidad resulta evidente en los estudios meta-analíticos de Glass (Glass, McGaw & Smith, 1981; McGaw & Glass, 1980) que se basan en la combinación de medidas del tamaño del efecto y son los más empleados en las revisiones bibliográficas cuantitativas actuales. El propósito fundamental de este procedimiento es obtener una medida común que permita resumir los resultados de una colección de experimentos independientes que ponen a prueba una misma hipótesis. Para que tal resumen sea posible, es necesario que los experimentos considerados sean comparables, es decir, que presenten ciertas condiciones de similitud o semejanza que hagan homogéneos sus resultados. Este autor analiza conjuntamente los resultados de una colección de experimentos que comparan las mediciones de una misma variable en dos grupos diferentes, por ejemplo, en un

grupo experimental y en un grupo control, mediante la prueba *t*. Dando por cumplidas estas condiciones o corrigiendo los datos originales para que se cumplan (Hunter & Schmidt, 2004), es posible obtener una combinación de los tamaños del efecto de los respectivos experimentos mediante el cálculo directo del promedio. Llamamos d_{prom} a la diferencia estandarizada entre las medias, hallada en cada experimento.

$$d_{prom} = \frac{\sum d}{n}$$

Donde d_{prom} se refiere al promedio de las diferencias estandarizadas entre los grupos, $\sum d$ indica la suma de las diferencias en los experimentos estudiados, y n es la cantidad de experimentos incluidos en el meta-análisis.

La diferencia observada entre el grupo experimental y el grupo control (o entre el pre-test y el post-test) se calcula previamente para cada experimento. De esta manera, se realiza finalmente el cálculo del promedio de las diferencias estandarizadas de los experimentos. Ésta es la medida común que se necesita para que los resultados de experimentos diferentes puedan compararse en un mismo análisis.

En síntesis, el método de Glass consiste en promediar los tamaños del efecto observados en una colección de experimentos que ponen a prueba una misma hipótesis y cumplen ciertas condiciones de homogeneidad mediante el procedimiento de las diferencias estandarizadas entre grupos mediante pruebas *t*. Este autor también propone técnicas para otras pruebas de significación como la *r* (análisis de la correlación) y la *F* (análisis de la varianza), entre otras, e incluso para pruebas no paramétricas (Wolf, 1986). Todos los procedimientos se basan, en términos generales, en la misma combinación de puntajes transformados a una métrica común que los hace comparables.

Se han propuesto varios métodos gráficos para representar los procedimientos meta-analíticos, sobre todo para los que combinan variables de respuesta binaria (por ejemplo, muerte/no muerte; curación/enfermedad) y que trabajan con riesgos relativos o razones de *odds*, de uso corriente en la investigación biomédica (Letón Molina & Pedromingo Marino, 2001; Thompson, Turner & Warn, 2001). Asimismo, en ciencias sociales se han desarrollado técnicas para la corrección de sesgos en la información original de los estudios individuales que luego habrán de combinarse (Hunter & Schmidt, 2004).

4. El sesgo de publicación

Son múltiples las fuentes de error en meta-análisis. Muchas de ellas han sido estudiadas y resueltas satisfactoriamente, otras han sido detectadas pero no resueltas todavía, y otras tantas quizás no han sido detectadas aún (Hunter & Schmidt, 2004; Wolf, 1986).

El sesgo de publicación (Letón Molina & Pedromingo Marino, 2001) se debe a que no todas las investigaciones, más allá de su calidad científica, poseen la misma probabilidad de ser publicadas. Los estudios que obtuvieron resultados estadísticamente significativos logran más fácilmente la publicación que los estudios obligados a no rechazar la hipótesis nula. Se cuenta con muchos estudios publicados que presentan resultados significativos, no así con resultados desfavorables. De manera que al realizar la búsqueda bibliográfica para la compilación de investigaciones que luego serán meta-analizadas, se corre el riesgo de omitir resultados desfavorables para probar un determinado efecto. Los meta-análisis pueden ser, de este modo, propensos a generar conclusiones demasiado optimistas al no contar con los datos completos que reflejen la heterogeneidad del universo de estudios que han puesto a prueba una misma hipótesis (Sutton, Song, Gilbody & Abrams, 2000). En consecuencia, muchos meta-analistas incluyen en la combinación registros de investigaciones no publicadas en revistas científicas, por ejemplo, actas de congresos, archivos de tesis de grado y postgrado o informes de investigación de universidades u otros centros científicos.

Rosenthal (1979) ha propuesto que se calcule e informe en todo meta-análisis (que rechaza la hipótesis nula en su resultado combinado o global) la cantidad de estudios no significativos que deberían surgir para que la hipótesis nula no pueda ya rechazarse. Esta propuesta ha recibido de Cooper (1979) la etiqueta de *N segura* o *N a prueba de fallas* (*Fail Safe N*). La *N segura* es la cantidad de estudios adicionales que se necesitan combinar en un meta-análisis ya realizado para revertir la conclusión que afirma el rechazo de la hipótesis nula. De manera que cuanto mayor sea la *N segura*, más a salvo estará la hipótesis de investigación y más remota será la necesidad de resignarse a no rechazar la hipótesis de nulidad.

Cooper (1979) señala que el cálculo de la *N segura* exige determinar previamente un p valor de corte que se fija, tradicionalmente, en $p = 0,05$ ó en $p = 0,01$ para obtener resultados significativos o altamente significativos, respectivamente. La siguiente expresión muestra los cálculos necesarios para su cálculo con un p valor de 0,05.

$$N_{segura\ 0,05} = \left(\frac{\sum Z}{1,645} \right)^2 - N$$

Se calcula el cuadrado del cociente que tiene por numerador a la suma de los puntajes Z de los estudios combinados (usando el método de Stouffer) y por denominador al número 1,645 (que es el puntaje Z correspondiente a un valor de área bajo la curva normal para una $p = 0,05$, como riesgo unilateral). Finalmente se resta N , siendo N la cantidad de estudios incluidos en la combinación.

Del mismo modo, si se prefiere calcular una N_{segura} para un p valor de 0,01, debe emplearse el mismo procedimiento pero cambiando el denominador por 2,33, que es el puntaje Z que corresponde a un valor de área para ese p valor.

$$N_{segura\ 0,01} = \left(\frac{\sum Z}{2,33} \right)^2 - N$$

A mayor nivel de significación meta-analítica, mayor cantidad de N_{segura} se necesita para no revertir la conclusión que rechaza la hipótesis nula. A partir de estos cálculos se obtiene un número que indica cuántos estudios con resultados no significativos (es decir, que no rechazan la hipótesis nula) deberían agregarse a un meta-análisis ya realizado (que rechaza la hipótesis nula) para que éste deba dejar de rechazar la hipótesis nula (Wolf, 1986).

Letón Molina & Pedromingo Marino (2001) proponen otra manera de pensar la relación entre los estudios conocidos, con sus respectivos p valores, y los estudios desconocidos, con sus eventuales rechazos o no rechazos de la hipótesis nula. Estos autores sugieren un método para estimar la cantidad de estudios que se desconocen (por efecto del sesgo de publicación), dadas ciertas condiciones en los estudios conocidos (vinculadas al p valor y a la cantidad de estudios combinados).

Para aplicar este método se asume la hipótesis nula y se necesita conocer el p valor ordenado de los estudios combinados ($P_{(k)}$), junto con la cantidad de estudios publicados o conocidos (K) que se han incluido en la combinación. Los cálculos necesarios son los siguientes:

$$N = \text{int} \left(\frac{K \{1 - P(k)\}}{P(k)} \right)$$

La estimación de la cantidad de estudios desconocidos (N) se realiza sobre la base del cálculo en el que *int* indica que se tomará el número entero (*integer*) que resulte del cociente

que figura entre paréntesis, aún cuando arroje un número real, es decir, con decimales. La K se refiere a la cantidad de estudios conocidos que hemos incluido en la combinación meta-analítica, y la $P(k)$ indicia el p valor más alto que se ha registrado entre los estudios conocidos y combinados. Si mantenemos fija la cantidad de estudios conocidos (K), cuánto mayor sea el p valor máximo entre los estudios conocidos $P(k)$, menor será la cantidad de estudios que estimamos desconocer (N). Del mismo modo, cuánto menor sea el p valor máximo entre los estudios conocidos, mayor será la cantidad estimada de estudios desconocidos (N). Este cálculo parece estar pensado desde la precaución, es decir, desde la evitación de un optimismo ingenuo fundado en el desconocimiento de una colección de estudios desfavorables para la hipótesis de investigación, razón por la cual no han sido quizás publicados. Este método resulta, sin embargo, un tanto restringido por maximizar intuitivamente los efectos del sesgo de publicación y por no considerar los beneficios complementarios que ofrece el cálculo de la N_{segura} propuesta por Rosenthal (1979). La consideración de un único p valor (el máximo conocido), con la correspondiente omisión de toda una colección de p valores que aportan información valiosa y conocida, puede ser señalada como otra importante limitación de método de Letón Molina & Pedromingo Marino (2001) para la estimación de la cantidad de estudios desconocidos.

5. Conclusiones

El meta-análisis puede ser entendido como un conjunto de técnicas estadísticas que se emplea para la revisión de los conocimientos acumulados sobre un problema específico de investigación científica (Wolf, 1986). Su procedimiento general consiste en integrar una serie de resultados de estudios individuales, para lo cual deben primero transformarse las mediciones independientes a una métrica común.

Los métodos del meta-análisis propuestos por Glass et al. (1981) le conceden al manejo de la hipótesis de nulidad una gran importancia, en continuidad con los trabajos de Cohen (1988, 1990, 1994) sobre el tamaño del efecto y el poder o potencia de una prueba, entendido como $1-\beta$.

Frank L. Schmidt (1992) considera que el meta-análisis no es sólo una nueva técnica para confeccionar estados-del-arte cuantitativos, sino que es una nueva manera de entender el significado de los datos. Esta visión propone que la formulación de leyes científicas se desarrolle sólo mediante estudios meta-analíticos. En este sentido, algunos se refieren al

meta-análisis como una técnica para la generalización de la confiabilidad (Rodríguez & Maeda, 2006). Sostiene Schmidt que el mero análisis de los resultados obtenidos en investigaciones individuales corre el riesgo de sobreestimar el significado de la información. Tal proceder se concentra, sobre todo, en el error de tipo I y descuida el error de tipo II.

La cuestión permanece, sin embargo, como un problema metodológico aún abierto. Algunos autores sostienen que todavía deben estandarizarse los criterios de calidad para los estudios meta-analíticos, ya que no todos cumplen con las mismas exigencias (Sacks, Berrier, Reitman, Ancona-Berk & Chalmers, 1987). La combinación de estudios individuales de baja calidad sólo puede generar estudios meta-analíticos de baja calidad (Hunt & McKibbin, 1997; Lau, Ioannidis, & Schmid, 1997). Es por ello que Hunter & Schmidt (2004; Hunter et al., 2006) han centrado el desarrollo de sus técnicas meta-analíticas en la corrección de los sesgos y errores de los estudios individuales que se integran en la combinación. Sin embargo, muchas de estas limitaciones no son necesariamente analíticas sino que son, en ocasiones, conceptuales. La definición de los criterios de inclusión y exclusión que guían la búsqueda o revisión de los estudios que participarán o han participado de la combinación meta-analítica resulta siempre dificultosa (Counsell, 1997; Meade & Richardson, 1997). No obstante, esta circunstancia describe la situación de cualquier método de investigación, en tanto ofrece múltiples fuentes de error.

Agradecimientos: a Griselda y Raúl Krasuk, por la revisión de los cálculos. Al Doctor Alfredo López Alonso, por su apoyo a esta investigación.

Referencias

- Ades, A.E., Lu, G. & Higgins, J.P.T. (2005). The Interpretation of Random Effects. Meta-Analysis in Decision Models. *Medical Decision Making*, 25(6), 646-654.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences. Second Edition*. Hillsdale, N.J.: Lawrence Erlbaum.
- Cohen, J. (1990). Things I Have Learned (So Far). *American Psychologist*, 45(12), 1304-1312.
- Cohen, J. (1994). The Earth Is Round ($p < .05$). *American Psychologist*, 49(12), 997-1003.
- Cooper, H.M. (1979). Statistically Combining Independent Studies. A Meta-Analysis of Sex Differences in Conformity Research. *Journal of Personality and Social Psychology*, 37, 131-136.

- Counsell, C. (1997). Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews. *Annals of Internal Medicine*, 127(5), 380-387.
- Fisher, R.A. (1932). *Statistical Methods for Research Workers*. Londres: Oliver & Boyd.
- Glass, G. (1976). Primary, Secondary and Meta-Analysis of Research. *Educational Researcher*, 5, 3-8.
- Glass, G., McGaw, B. & Smith, M. (1981). *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage.
- Hönekopp, J., Becker, B.J. & Oswald, F.L. (2006). The Meaning and Suitability of Various Effect Sizes for Structured Rater × Ratee Designs. *Psychological Methods*, 11(1), 72-85.
- Hunt, D.L. & McKibbin, K.A (1997). Locating and Appraising Systematic Reviews. *Annals of Internal Medicine*, 126(7), 532-538.
- Hunter, J.E. & Schmidt, F.L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings (2nd Edition)*. Londres: Sage Publications.
- Hunter, J.E., Schmidt, F.L. & Le, H. (2006). Implications of Direct and Indirect Range Restriction for Meta-Analysis Methods and Findings. *Journal of Applied Psychology*, 91(3), 594-612.
- L 'Abbé, K.A. Detsky, A.S. & O'Rourke, K. (1987). Meta-Analysis in Clinical Research. *Annals of Internal Medicine*, 107(2), 224-233.
- Lau, J., Ioannidis, J.P.A., & Schmid, C.H. (1997). Quantitative Synthesis in Systematic Reviews. *Annals of Internal Medicine*, 127(9), 820-826.
- Letón Molina., E. & Pedromingo Marino, A. (2001). *Introducción al Análisis de Datos en Meta-Análisis*. Madrid: Ediciones Díaz de Santos.
- McGaw, B. & Glass, G. (1980). Choice of the Metric for Effect Size in Meta-Analysis. *American Educational Research Journal*, 17, 325-337.
- Meade, M.O. & Richardson W.S. (1997). Selecting and Appraising Studies for a Systematic Review. *Annals of Internal Medicine*, 127(7), 531-537.
- Pearson, K. (1933). On a Method of Determining Whether a Sample of Size n Supposed to Have Been Drawn From a Parent Population Having a Known Probability Integral Has Probably Been Drawn at Random. *Biometrika*, 25, 379-410.
- Petticrew, M. & Roberts, H. (2005). *Systematic Reviews in the Social Sciences. A Practical Guide*. London: Blackwell.

- Peto, R., Yusuf, S., Lewis, J., Collins, R. & Sleight, P. (1985). Beta Blockade During and After Myocardial Infarction. An Overview of the Randomized Trials. *Progress in Cardiovascular Diseases*, 27, 335-371.
- Rodriguez, M.C. & Maeda, Y. (2006). Meta-Analysis of Coefficient Alpha. *Psychological Methods*, 11(3), 306-322.
- Rosenthal, R. (1979). The File Drawer Problem and Tolerance for Null Hypothesis. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R.G. & Rosnow, R.L. (1991). *Essentials of Behavioral Research. Research Methods and Data Analysis (2nd Edition)*. Boston, Mass.: McGraw Hill.
- Sacks, H.S. Berrier, J., Reitman, D., Ancona-Berk, V.A. & Chalmers, T.C. (1987). Meta-Analysis and Randomized Controlled Trials. *The New England Journal of Medicine*, 316(8), 450-455.
- Schmidt, F.L. (1992). What Do Data Really Mean? Research Findings, Meta-Analysis and Cumulative Knowledge in Psychology. *American Psychologist*, 47(10), 1173-1181.
- Song, F., Sheldon, T.A., Sutton, A.J., Abrams, K.R. & Jones, D.R. (2001). Methods for Exploring Heterogeneity in Meta-Analysis. *Evaluation & The Health Professions*, 24(2), 126-151.
- Stouffer, S.A., Suchman, E.A., De Vinney, L.C., Star, S.A. & Williams, R.M. (1949). *The American Soldier. Adjustment During Army Life. Vol 1*, Princeton, N.J.: Princeton University Press.
- Sutton, A.J., Song, F., Gilbody, S.M. & Abrams, K.R. (2000). Modelling Publication Bias in Meta-Analysis: A Review. *Statistical Methods in Medical Research*, 9(5), 421-445.
- Thompson, S.G., Turner, R.M. & Warn, D.E. (2001). Multilevel Models for Meta-Analysis, and their Application to Absolute Risk Differences. *Statistical Methods in Medical Research*, 10(6), 375-392.
- Winer, B.J. (1971). *Statistical Principles in Experimental Design (2nd edition)*. N.Y.: McGraw-Hill.
- Wolf, F.M. (1986). *Meta-Analysis: Quantitative Methods for Research Synthesis*. Londres: Sage Publications.