

## **MODELOS MULTINIVEL APLICADOS AL META-ANÁLISIS**

*María Castro*

El meta-análisis es una técnica de síntesis de resultados de investigación procedentes de un conjunto de investigaciones que abordan el mismo tema de estudio. A nadie se le escapa que esta síntesis de resultados aparece bloqueada por una variada cantidad de problemas como múltiples tipos de evidencias, diferentes diseños experimentales, distintos tipos de resultados, distintas medidas de la magnitud del efecto, sesgos en la validez interna y externa, o la existencia de evidencias indirectas o incompletas, por sólo señalar los más importantes. Los métodos meta-analíticos tradicionales aportan herramientas para resolver algunos de estos complejos problemas, aún así, los procedimientos tienen algunas limitaciones.

El enfoque meta-analítico basado en la aplicación de la estructura que aportan los *modelos multinivel con procedimientos de estimación bayesianos y de máxima verosimilitud* se plantea como una alternativa novedosa. La innovación en el procedimiento de síntesis se define en dos sentidos: por un lado, el modelo de análisis y, por otro, los estimadores utilizados (Castro, 1997).

Con relación al modelo, la organización de los datos meta-analíticos responde a una estructura jerárquica o anidada, ya que para realizar un meta-análisis se necesita que el tema de investigación se repita de estudio individual a estudio individual, estando asociado a cada uno de ellos características específicas. El objetivo que se pretende es obtener información útil procedente del conjunto de investigaciones. Se ve que hay dos niveles de análisis, la relación que se establece entre variables dentro de cada estudio individual y la relación que se establece a lo largo de todo el conjunto de estudios, niveles que es importante diferenciar para no mezclar distintos tipos de efectos. Los modelos multinivel aportan una estructura general que permite representar cada uno de estos niveles con un submodelo.

Con respecto a la estimación, la lógica bayesiana no es nueva, desde que Bayes en 1763 enunciara el famoso teorema se ha desarrollado toda una rama de la inferencia estadística. Sin embargo, esta lógica ha sido escasamente aplicada en el ámbito de las Ciencias Sociales (Pollard, 1986) aunque sí se encuentran aplicaciones específicas al meta-análisis dentro del ámbito de las Ciencias Médicas (Eddy, Hasselblad y Shachter, 1992).

Este nuevo planteamiento está orientado a la toma de decisiones sobre una base de evidencia, aunque éstas sean incompletas o imperfectas. Su objetivo es, por tanto, hacer la mejor síntesis posible con la evidencia utilizable. Ésta es una de las principales diferencias con los tradicionales métodos meta-analíticos. Otra de las diferencias más significativas es la transformación de la conceptualización del problema de la síntesis de la evidencia. Deja de ser un problema básicamente descriptivo para convertirse en un problema de inferencia estadística, puesto que se propone la extracción de conclusiones que van más allá de la mera descripción de los datos. Si bien ésta ha sido una de las principales finalidades del meta-análisis clásicamente concebido, las estrategias metodológicas aplicadas no han sustentado este objetivo. Ahora, el principal problema del meta-análisis es la descomposición e identificación de fuentes de varianza y se reconoce la estructura jerárquica de los datos. Por tanto, parece más lógico aplicar modelos estadísticos que reflejen esta estructura y permitan identificar distintas fuentes de varianza, como es el caso de los modelos multinivel (Castro y De la Orden, 1998).

De una manera muy sintética, el meta-análisis es un procedimiento que pretende estudiar la posibilidad de generalización de los resultados de investigación. Las estrategias empleadas hasta ahora (Glass, 1976; Hunter y Schmidt, 1990 o Hedges y Olkin, 1985) trataban de depurar los estimadores de la magnitud del efecto trabajando sobre factores aislados (falta de homogeneidad, error muestral, debilidad en la validez de constructo, etc.) o aportan modelos no del todo realistas que presuponen que toda la variabilidad entre resultados puede ser explicada por las características de los estudios.

La alternativa propuesta desde los modelos multinivel parte de un planteamiento radicalmente distinto, intenta analizar las posibilidades de generalizabilidad de un conjunto de resultados de investigación de forma conjunta. La cuestión estadística central en un meta-análisis es el estudio de la consistencia de los resultados entre distintas investigaciones. Si se obtienen efectos similares se pueden resumir en un índice conjunto de magnitud del efecto promedio. Normalmente, los resultados son inconsistentes y la magnitud del efecto varía de investigación a investigación. A este problema se le han ido dando distintas soluciones, por ejemplo Hunter y Schmidt (1990) estiman la media y desviación típica de la verdadera magnitud del efecto o Hedges y Olkin (1985) desarrollan pruebas de homogeneidad para proceder a la explicación de la variabilidad en función de variables moderadoras.

Sin embargo, el problema de la inconsistencia es tan fuerte que incluso cuando las distintas investigaciones comparten la misma magnitud del efecto, los resultados pueden variar por efecto del error muestral. De ahí que el principal foco de atención del meta-análisis sea **diferenciar los componentes de variación de las magnitudes**

**del efecto** (Bryk y Raundenbush, 1985). Es decir, definir qué parte se debe al error muestral y qué parte se debe verdaderamente a la inconsistencia de los efectos. Si se detecta inconsistencia, se puede formular un modelo que de cuenta de ella.

En definitiva, esta nueva orientación centrada en la identificación de componentes de varianza está estrechamente relacionada con la Teoría de la Generalizabilidad desarrollada por Cronbach (Cronbach, Rajaratnam y Gleser, 1963; Cronbach, Gleser, Nada y Rajaratnam, 1972; Martínez Arias, 1995), en tanto que pretende diferenciar qué varianza entre elementos —en este caso investigaciones primarias— se puede considerar varianza verdadera y cuáles son al error. El concepto de error de medida deja de considerarse de forma unitaria e indiferenciada para introducir el concepto estadístico de *muestreo de fuentes de variación múltiples*, que permite considerar la variabilidad de cada una de las características de la investigación primaria, lo que, a su vez, permite considerar distintas fuentes de error que pueden ser tratadas empleando los procedimientos del modelo lineal general del diseño de investigación. De la misma forma, el modelo de medidas paralelas se sustituye por el de *medidas aleatoriamente paralelas*, considerando que las distintas características de las investigaciones pueden entenderse como una muestra aleatoria de un universo más amplio, definido por la población de características posibles bajo las cuales se ha definido el marco del meta-análisis. Probablemente, el concepto más significativo para el meta-análisis sea la extensión de fiabilidad al más amplio de *generalización* o grado de generalidad, entendido como la inferencia estadística a un universo o población a partir de un estimador observado que se considera como una muestra concreta de dicho universo. No es otra cosa que la invarianza dentro de una población de características o la generalizabilidad de un resultado concreto a una población de características.

Interesa, por tanto, encontrar o desarrollar un modelo matemático que se adapte mejor a la situación particular de investigación que se tiene y que permita utilizar la evidencia acumulada sin incurrir en problemas que invaliden —aunque sea parcialmente— las conclusiones. Y esta es la solución que aportan los modelos multinivel (Castro y Gaviria (a), en prensa).

## **ADECUACIÓN DE LOS MODELOS JERÁRQUICOS LINEALES AL PROBLEMA META-ANALÍTICO**

Una de las situaciones de investigación en las que se pueden descubrir una estructura anidada de los datos es en los estudios meta-analíticos (Bryk y Raundenbush, 1992; Draper, 1995). Una investigación individual pretende descubrir una determinada relación entre una variable de tratamiento y los efectos del mismo. Por otra parte, si se cuenta con un conjunto de investigaciones sobre el mismo tema —es decir, investigaciones que estudian la eficacia del mismo conjunto de tratamientos— lo que se quiere estudiar es la constancia o invariabilidad de los efectos en el total de la muestra. Hay, por tanto, dos niveles distintos de relación tratamientos-efectos, la que se establece dentro de cada investigación individual y la que se establece a lo largo de todo el conjunto de investigaciones. Ahora, la relación entre variable de tratamiento y efectos está anidada en el conjunto de estudios.

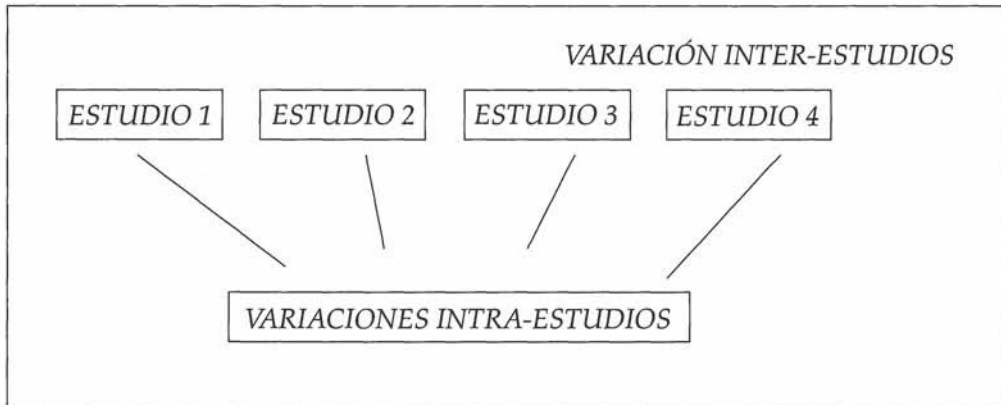


Figura 1. Fuentes de variación meta-analítica

Aparecen, como se ve, dos fuentes de variabilidad. En el primer caso, la variabilidad intra-estudio se debe a las características de la muestra. Sin embargo, además de la variabilidad muestral, entre las investigaciones primarias pueden existir diferencias en los diseños, en los instrumentos de recogida de información,..., en definitiva entre las peculiaridades y características propias de cada investigación. En este caso se puede considerar que se está trabajando con una muestra de investigaciones primarias extraídas de la población de investigaciones con características similares. Se perfila entonces otra fuente de variabilidad, la debida al muestreo entre la población de estudios posible. Como se ve en la Figura 1, se identifican claramente dos niveles, un nivel de estudios individuales y un nivel del conjunto de investigaciones. La relación tratamientos-efectos está afectada por distintos tipos de variación debida bien al muestreo entre sujetos bien al muestreo entre estudios, según el nivel de análisis (Castro, 1997).

Claramente, *los datos meta-analíticos tienen una estructura anidada, y es pertinente aplicar la lógica de los modelos multinivel*: por un lado, los efectos de los tratamientos están anidados en las investigaciones y, por otro, el tema sustantivo, objeto de investigación, es común al conjunto de estudios. Así, se necesitan modelos que estén preparados para dar cuenta de la variación en dos niveles distintos: dentro de cada investigación y entre el conjunto de investigaciones. En definitiva, se busca estudiar las fuentes de variación entre efectos de los tratamientos, dentro de cada investigación individual y dentro del conjunto de investigaciones. En el nivel de investigación individual, se evalúa la relación entre tratamiento y efectos en cada investigación concreta. Dentro del nivel del total de investigaciones, se evalúa la consistencia de esa relación a lo largo de todo el conjunto de estudios, estimando la inconsistencia de los resultados.

Las diferencias en la aplicación de los modelos multinivel a estudios meta-analíticos frente a otro tipo de estudios se observan principalmente en dos puntos (Bryk y Raundenbush, 1992): a) las puntuaciones brutas de cada investigación individual no suelen estar disponibles, teniendo normalmente los estadísticos publicados en los

informes, y b) que cada investigación suele emplear distintas medidas de los resultados, incluso en el caso de que se consideren medidas del mismo constructo, por eso siempre el primer objetivo del meta-análisis ha sido la búsqueda de una métrica común.

Desde el punto de vista de la *conceptualización*, los modelos multinivel aportan un entorno natural para expresar y comparar teorías sobre las relaciones estructurales entre variables en cada uno de los niveles definidos (Draper, 1995; Goldstein, 1995). Así, dentro de los estudios meta-analíticos, se pueden desagregar los determinantes procedentes del nivel del estudio individual, del nivel del conjunto de investigaciones y de la interrelación entre ellos. Por otra parte, resultan particularmente interesantes las implicaciones consideradas desde el punto de vista del *diseño y del análisis*. En el caso concreto del meta-análisis, al existir variación dentro de cada investigación individual y entre investigaciones parece clara la necesidad de una estrategia analítica que pueda representarla, lo que supone una formulación jerárquica de las fuentes de variación y capacidad para dar cuenta de ella. La supresión o no inclusión de alguno de los niveles puede llevar a obtener estimaciones sesgadas. Si las medidas de la magnitud del efecto tienen como base muestras moderadamente amplias ( $N \geq 30$ ), la distribución de los estadísticos será aproximadamente normal y se asume que con varianza conocida. Desde un punto de vista *estadístico*, el meta-análisis supone el estudio de un conjunto de estimaciones independientes, cada una de ellas normalmente distribuidas y con varianza conocida.

## MODELO LINEAL MIXTO APLICADO AL META-ANÁLISIS

Con el objeto de presentar una pequeña ilustración, se expone el meta-análisis de Goodman (1989), revisado por Draper y cols. (1993) y Draper (1995), con datos procedentes de seis investigaciones clínicas sobre los efectos de la aspirina en la mortalidad de pacientes que han sobrevivido a un ataque al corazón. De los seis estudios, cinco muestran que la aspirina reduce la tasa de mortalidad en estos pacientes. Sin embargo, el sexto estudio no sostiene esta conclusión, con la particularidad de que es el experimento con mayor número de sujetos. Las preguntas que, lógicamente, se plantean son: a) ¿por qué el sexto estudio arroja resultados diferentes?, b) si la respuesta a la anterior pregunta es incierta, qué hacer para reducir esa incertidumbre, y c) ¿qué terapia recomendar a los pacientes mientras se responde a la pregunta b)?

Para resolver este problema, Goodman define un modelo multinivel en el que, en el primer nivel, la variación intra-estudios es completamente aleatoria y se asume como conocida la variación inter-estudios. Al probar el ajuste del modelo, se observa que hay una gran cantidad de varianza sin explicar, siendo necesario formular un modelo de segundo nivel, donde se introduce una posible variable moderadora de los resultados para mejorar el ajuste del modelo. En este caso, la variable es el conocimiento previo sobre los efectos de la aspirina en las tasas de mortalidad de estos pacientes obtenido en el conjunto de investigaciones. El paso siguiente consiste en probar la existencia de varianza residual interestudios que este modelo de segundo nivel rechaza explicar.

En la literatura se emplean dos modelos para el tratamiento de la variabilidad de los resultados de investigación (Castro, 1997; Castro y De la Orden, 1998). Según Hedges y Olkin, uno de ellos entiende que toda la variabilidad entre los estudios se puede explicar a partir de las características de los estudios —modelo de efectos fijos—, que es el más desarrollado, y otro plantea que toda la variación es inexplicable —modelo de efectos aleatorios— debida, por tanto, al azar. Ambos modelos pueden parecer plausibles, sin embargo, no deja de ser ciertamente reduccionista e irreal considerar toda la variación posible totalmente aleatoria o totalmente determinada. Quizá sea más interesante pensar que una parte de la variación puede tener explicación a partir de las características de las investigaciones y otra parte puede deberse a la variación aleatoria dentro de la muestra de estudios, aparte de la variación debida al muestreo entre sujetos. Este planteamiento coincide con el de los modelos multinivel, que incorporan un modelo de efectos mixtos (Raundenbush y Bryk, 1985). En el ejemplo, el modelo propuesto es de efectos mixtos, ya que en el primer nivel reconoce que toda la variación tiene un origen aleatorio. Sin embargo, en el segundo nivel se introduce una variable de carácter fijo, propia del conjunto de estudios que contribuye a explicar la variabilidad restante. Este modelo supera al modelo de efectos aleatorios puesto que éste únicamente describe la variabilidad entre estudios sin intentar explicarla. Además, el modelo de efectos mixtos permite responder mejor a la tercera cuestión planteada, ya que emplea toda la información disponible en el conjunto de investigaciones, frente a una posible respuesta dada a partir de un modelo de efectos fijos, que no reconoce la posible variación entre investigaciones primarias, y donde toda la variación se puede atribuir a las distintas muestras de sujetos empleadas. Por otra parte, la respuesta a las dos primeras preguntas sólo es posible desde un modelo como el mixto, ya que es el único modelo que descompone y analiza distintas fuentes de variación.

En la propuesta de Raundenbush y Bryk (1985) para la aplicación de modelos de efectos mixtos al meta-análisis, dentro del nivel de estudios primarios, la variación se entiende como aleatoria, mientras que en el nivel del conjunto de investigaciones, la variación intra-estudios se trata como resultado, dependiente de las características de los estudios. Este modelo permite probar hipótesis sobre la influencia de las características en los resultados, encontrar estimadores mejorados de los efectos individuales (*empírico bayesianos*) y examinar la sensibilidad de todas las inferencias sustantivas contando con la imprecisión en la estimación de los componentes de varianza.

El modelo general plantea que la variabilidad entre magnitudes del efecto es completamente aleatoria. Una vez que se comprueba que hay una parte de la variabilidad que este modelo rechaza explicar, introduce el segundo nivel que expresa una variación explicable de forma fija a partir de las características conocidas de los estudios, más un componente aleatorio debido al muestreo entre investigaciones. La estimación se considera empírico-bayesiana puesto que considera la información procedente del conjunto de investigaciones para estimar los valores de la magnitud del efecto. Veámoslo con más detalle.

En el *primer nivel*, se formula un modelo separadamente para cada investigación individual, que estima los parámetros independientemente para cada unidad. Los

parámetros en este nivel varían aleatoriamente alrededor de una media central. El modelo explica la variabilidad intra-estudios como una función de las diferencias entre unidades. El modelo para este primer nivel queda formulado de la siguiente manera:

$$d_i = \delta_i + e_i \quad (1)$$

donde  $i = 1, 2, \dots, k$  son las distintas investigaciones primarias,  $\delta_i$  es el valor paramétrico de la magnitud del efecto y  $e_i$  es el término de error aleatorio debido al muestreo entre sujetos que se asume como independiente, normalmente distribuido y con varianza  $v_e$ ,  $e_i \sim N(0, v_e)$ .

En el *segundo nivel* se atiende a la variación inter-investigaciones y las variaciones inter-investigación son vistas como una función de las diferencias entre las características de investigaciones. Los parámetros de la magnitud del efecto ( $\delta_i$ ) varían como una función de las características conocidas de los estudios más un término de error aleatorio, procedente ahora del muestreo entre investigaciones. El modelo para este segundo nivel queda expresado de la siguiente forma:

$$\delta_i = \vec{W}_i \vec{\gamma} + u_i \quad (2)$$

donde:  $\vec{W}_i$  es un vector ( $q \times 1$ ) de constantes que representa las diferencias conocidas entre estudios,  $\vec{\gamma}$  es el vector ( $q \times 1$ ) de los parámetros inter-estudios y  $u_i$  es el error aleatorio debido al muestreo entre investigaciones, que se asume independiente, y que está normalmente distribuido, con media cero y varianza  $\tau^2$ ,  $u_i \approx N(0, \tau^2)$ .

La ecuación 2 para este segundo nivel se conoce dentro de la literatura sobre estadística bayesiana como la distribución *a priori* de  $\delta_i$ . La razón para esta denominación es que intenta estimar los parámetros de esta distribución con los datos que ya se conocen procedentes de un conjunto de investigaciones, el procedimiento se denomina «empírico bayesiano» (Bryk y Raundebush, 1985; Berger, 1985). Este estimador ha demostrado su mejor comportamiento estadístico frente al estimador tradicional de la magnitud del efecto (Castro, 1997; Castro y Gaviria (b), en prensa).

El modelo combinado de los dos niveles definidos es:

$$d_i = \vec{W}_i \vec{\gamma} + u_i + e_i \quad (3)$$

donde se asume que  $u_i$  y  $e_i$  son independientes y la distribución marginal de  $d_i$  es:  $d_i \approx N(\vec{W}_i \vec{\gamma}, v_e + \tau^2)$ . Entonces, el modelo multinivel aplicado al meta-análisis es equivalente a un modelo mixto con efecto fijos,  $\vec{\gamma}$ , y efectos aleatorios,  $u_i$ .

La aportación de este modelo está centrada en dos puntos:

- en primer lugar, el reconocimiento de la doble fuente de variabilidad aleatoria, que además tiene la ventaja de ser un modelo más generalizable que los modelos clásicos, a pesar de tener una varianza de error menor que el modelo de efectos aleatorios,  $y$ ,
- en segundo lugar, la consideración de la existencia de una distribución de parámetros,  $\delta_i$ , rompiendo la lógica clásica que prevé un único valor paramétrico central para cualquier estimación de la magnitud del efecto individual.

## REFERENCIAS BIBLIOGRÁFICAS

- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chance. Reeditado en *Biometrika*, 45, 293-315.
- Berger, J.O. (1985). *Statistical decision theory and bayesian analysis*. New York: Springer-Verlang.
- Bryk, A.S. y Raudenbush, S.W. (1992). *Hierarchical linear models: applications and data analysis methods*. London: Sage.
- Castro, M. (1997). Meta-análisis. Aportaciones metodológicas a la síntesis cuantitativa de la evidencia. Un estudio de simulación Montecarlo. Madrid: Universidad Complutense de Madrid.
- Castro, M. y De la Orden, A. (1998). Análisis y comparación de las características de los modelos de síntesis meta-analítica: aportaciones de los modelos jerárquicos lineales. *Bordón* 50 (2), pp. 119-134.
- Castro, M. y Gaviria, J.L. (a) (en prensa). Meta-análisis y modelos jerárquicos lineales: estudio analítico de las aportaciones de los índices de magnitud del efecto individual clásico y empírico-bayesiano. *Revista de Educación*.
- Castro, M. y Gaviria, J.L. (b) (en prensa). Application of herarchical lienar models to meta-analysis: study of the monte carlo simulation on the functioning of traditional and empirical-bayes effect size. *Quality and Quantity*.
- Cronbach, L.J., Gleser, G.C., Nada, H. y Rajartanam, N. (1972). *The dependability of behavioral measurement*. New York: John Wiley.
- Cronbach, L.J., Rajartanam, N. y Gleser, G.C. (1963). Theory of generalizability. A liberation of reability theory. *Britihs Journal of Statistical and Mathematical Psychology*, 16, 137-173.
- Draper, D. (1995). Inference and hierarchical modeling in social sciences. *Journal of Educational and Behavioral Statistics*, 20 (2), 115-149.
- Draper, D., Gaver, D.P., Goel, P.K., Greenhouse, J.B., Hedges, L.V., Morris, C.N., Tucker, J.R. y Waternaux, C. (1993). *Combinig information: statistical issues and opportunities for research*. Alexandria: American Statistical Associaton.
- Eddy, D.M.; Hasselblad, V. y Schacher, R. (1992). *Meta-analysis by de confidence profile method. The statistical synthesis of evidence*. London: Academic Press.
- Goldstein, H. (1995). Hierarchical data modeling in the social sciences. *Journal of Educational and Behavioral Statistics*, 20 (2), 201-205.
- Goodman, S.N. (1989). Meta-analysis and evidence. *Controled Clinical Trials*, 10, 188-204.
- Hedges, L.V. y Olkin, I. (1985). *Stadistical Methods for meta-analysis*. Orlando: Academic Press.
- Hunter, J.E. y Schmidt, F.L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. London: Sage.
- Martínez Arias, R. (1995). *Psicometría: teoría de los test psicológicos y educativos*. Madrid: Síntesis.
- Pollard, W.E. (1986). *Bayesian statistics for evaluation research. An introduction*. Beverly Hills: Sage.
- Raudenbush, S.W. y Bryk, A.S. (1985). Empirical bayes meta-analysis. *Journal of Educational Statistics*, 10 (2), 75-98.