

Distribución del tamaño del efecto y del tamaño muestral en los meta-análisis dentro del ámbito psicológico

M. Rubio¹, J. Sánchez², F. Marín³, J. A. López⁴

¹Departamento de Psicología Básica y Metodología, Universidad de Murcia, maria.rubio1@um.es

²Departamento de Psicología Básica y Metodología, Universidad de Murcia, jsmeca@um.es

³Departamento de Psicología Básica y Metodología, Universidad de Murcia, fulmarin@um.es

⁴School of Social and Community Medicine, University of Bristol, ja.lopez-lopez@bristol.ac.uk

Introducción

El meta-análisis es una forma de revisión sistemática cuantitativa, donde los resultados de una serie de estudios empíricos sobre un mismo tópico de investigación son integrados estadísticamente. Con este objetivo, los resultados de los estudios son expresados en una métrica común a través de un índice del tamaño del efecto. La diferencia de medias tipificada es uno de los más utilizados en estudios donde dos o más grupos son comparados en una variable de resultado continua.

Los tres principales objetivos estadísticos en un meta-análisis son estimar el tamaño del efecto medio a través de los estudios primarios, estudiar la heterogeneidad de las estimaciones de los tamaños del efecto en torno al tamaño del efecto medio, y buscar variables moderadoras que pudieran explicar al menos parte de la heterogeneidad entre las estimaciones de los tamaños del efecto individuales. En las ciencias sociales, del comportamiento y de la salud, estos moderadores son las características diferenciales de los estudios, incluyendo por ejemplo el tipo de diseño o tipo de intervención.

Existen dos modelos estadísticos para llevar a cabo un meta-análisis: el modelo de efecto fijo y el modelo de efectos aleatorios. Bajo el modelo de efecto fijo, se asume que todos los estudios primarios en el meta-análisis estiman un tamaño del efecto medio paramétrico común, la única fuente de variabilidad entre los tamaños del efecto es debida a la selección aleatoria de los participantes en cada estudio (error de muestreo). Sin embargo, en el modelo de efectos aleatorios, se asume que cada estudio en el meta-análisis estima a una población diferente de tamaño del efecto. Estos estudios son seleccionados aleatoriamente de una población de estudios, donde los correspondientes tamaños del efecto paramétricos se distribuyen normalmente. Como consecuencia, en el modelo de efectos aleatorios, los tamaños del efecto presentan dos fuentes de variabilidad: la variabilidad inter-estudios y la variabilidad intra-estudio.

Actualmente, se considera el modelo de efectos aleatorios como más realista que el modelo de efecto fijo. Por tanto, en este estudio nos focalizaremos sobre las características metodológicas del modelo de efectos aleatorios en meta-análisis.

Hay cientos de simulaciones Monte Carlo que investigan el desarrollo de los métodos estadísticos en meta-análisis. Estos estudios Monte Carlo intentan simular las condiciones más realistas y representativas en un meta-análisis empírico. Además, varios autores han planteado en los últimos años diversas amenazas que cuestionarían la validez de los resultados obtenidos con la mayoría de las técnicas meta-analíticas. La primera de estas

amenazas es la posible violación del supuesto de normalidad en la distribución de los tamaños del efecto, en especial cuando se asume un modelo de efectos aleatorios [1,2]. La segunda consiste en que la posible existencia de una correlación entre tamaño muestral y tamaño del efecto constituye un problema adicional para las técnicas meta-analíticas más comúnmente utilizadas en las que se suelen aplicar métodos de ponderación que están en función del tamaño muestral [3]. En concreto, la existencia de una correlación negativa entre tamaño muestral y tamaño del efecto puede ser indicativa de un problema de sesgo de publicación.

Dado estas circunstancias, el objetivo principal del presente estudio fue explorar las características metodológicas de 50 meta-análisis publicados en el campo de la psicología clínica, basados en la efectividad de intervenciones psicológicas, y con la diferencia de medias tipificada como índice del tamaño del efecto.

Método

Los meta-análisis procedían de revistas con factor de impacto localizadas en el primer cuartil del 2011 con JCR (*Journal Citation Reports*) en el campo de la psicología clínica (*Clinical Psychology Review, Psychological Medicine, Journal of Consulting and Clinical Psychology, Depression and Anxiety, Health Psychology, Neuropsychology, Behaviour Research and Therapy, and Journal of Substance Abuse Treatment*). La búsqueda se llevó a cabo en Google Académico y limitada a meta-análisis publicados entre los años 2000 y 2012 con las palabras clave “*meta-analysis*” OR “*systematic review*” en el título.

En primer lugar, la lectura del título y del resumen de cada referencia nos permitió preseleccionar meta-análisis sobre la eficacia de programas, tratamientos e intervenciones psicológicas sobre trastornos psicológicos, educacionales y psicosociales. Otro criterio de inclusión era que los meta-análisis utilizarán índices de la familia d como tamaño del efecto: diferencia de medias tipificada, cambio medio tipificado y diferencia de cambios medios tipificados. Además, los meta-análisis debían reportar los tamaños del efecto individuales y los tamaños muestrales para los estudios primarios. Para asegurar que los meta-análisis seleccionados tuvieran suficientes datos para proporcionar resultados válidos, éstos tuvieron que estar formados por siete o más estudios y por al menos cinco sujetos por grupo.

Un total de 206 meta-análisis publicados se revisaron de los cuales finalmente se incluyeron 50 en el estudio. Estos meta-análisis resumen los resultados de 1285 estudios individuales.

Se creó una base de datos en SPSS y se codificaron los tamaños del efecto y tamaños muestrales de los estudios individuales de cada meta-análisis. Para los meta-análisis que incluyeron varios resultados, se seleccionó el resultado más relevantes desde el punto de vista clínico para el objetivo del meta-análisis. También se codificó el tipo de diseño y el tipo de índice de la familia d . Los diseños fueron clasificados como inter-grupos o intra-grupos y el tipo de d como diferencia de medias tipificada, cambio medio tipificado y diferencia de cambios medios tipificados.

Se utilizó R para realizar los análisis estadísticos, aplicando el paquete *metafor* [4]. Para cada meta-análisis se obtuvieron los siguientes índices meta-analíticos:

- El número de estudios
- El tamaño muestral
- El tamaño del efecto medio ponderado a partir de los tamaños del efecto individuales
- La varianza inter-estudios residual
- El estadístico Q para informar de la existencia o no de heterogeneidad de los tamaños del efecto en torno al efecto medio
- El índice I^2 para informar del grado (porcentaje) de heterogeneidad. Valores iguales al 25%, 50% y 75% se interpretan como heterogeneidad baja, media y alta, respectivamente [5].

Se llevaron a cabo análisis descriptivos sobre estos índices, concretamente se calculó la media, mediana, mínimo y máximo de las distribuciones.

En cada meta-análisis también se estudió el grado de cumplimiento del supuesto de normalidad de la distribución de tamaños del efecto. Para ello, se examinó la asimetría y la curtosis y la prueba de bondad de ajuste para la distribución normal, en concreto, se aplicó el test de Shapiro-Wilks para muestras pequeñas

Por último, se calculó la correlación de Pearson entre tamaño del efecto y tamaño muestral.

Resultados

Tabla 1. Análisis descriptivo de los cálculos meta-analíticos

	Min.	1 st . Qu.	Median	Mean	3 rd . Qu.	Max.
k	7	14	19	24.78	28.75	88
\bar{d}	.038	.291	.553	.584	.739	2.222
p_norm	.000	.004	.130	.221	.330	.852
$d_skewness$	-1.947	.004	.567	.534	1.068	3.677
$d_kurtosis$	-1.758	-.736	-.148	.768	1.149	12.220
N_median	9	29.25	40.5	43.98	59.72	87.5
$N_skewness$	-1.085	.829	1.350	1.324	1.746	3.487
$N_kurtosis$	-1.859	-.509	1.102	1.627	2.549	14.170
$r_{N,d}$	-.736	-.320	-.119	-.085	.112	.734
p_Q	.000	.000	.000	.078	.019	.981
I^2	0	43.94	62.87	56.19	74.96	93.61
τ^2	.000	.062	.112	.175	.171	1.024

Nota. Min. = mínimo; 1st. Qu. = primer cuartil; 3rd. Qu. = tercer cuartil; Max. = máximo; k = número de estudios; \bar{d} = tamaño del efecto medio; p_norm = p -valor asociado al test Shapiro-Wilks; $d_skewness$ = asimetría del índice d ; $d_kurtosis$ = curtosis del índice d ; N_median = mediana de los tamaños muestrales; $N_skewness$ = asimetría de los tamaños muestrales; $N_kurtosis$ = curtosis de los tamaños muestrales; $r_{N,d}$ = correlación entre tamaño del efecto y tamaño muestral; p_Q = p valor asociado al estadístico Q de heterogeneidad; I^2 = índice para cuantificar la cantidad de heterogeneidad (en %); τ^2 = varianza inter-estudios

En general, los resultados muestran la distribución de los tamaños muestrales, la distribución de los tamaños del efecto y su grado de ajuste a la distribución normal, el rango de valores de la varianza inter-estudios y el sentido y magnitud de la correlación entre tamaños muestrales y tamaños del efecto. La tabla 1 resume los estadísticos descriptivos de los índices meta-analíticos extraídos para cada meta-análisis.

En cuanto al número de estudios primarios, los valores oscilaron entre $k = 7$ y $k = 88$, siendo la media 24.78 y la mediana 19. El número mínimo de estudios coincide con el criterio de selección prefijado para que el meta-análisis pudiera ser incluido en la revisión sistemática.

Para calcular el tamaño del efecto medio, se tomaron los índices d en valor absoluto, ya que el signo positivo o negativo puede variar dependiendo del sentido de la relación entre los dos grupos, no siendo este criterio uniforme en todos los meta-análisis. El tamaño del efecto medio fue 0.584 y el primer y el tercer cuartil fueron 0.291 y 0.739, respectivamente. Estos resultados son similares al criterio de Cohen, donde valores iguales a 0.2, 0.5 y 0.8 reflejan tamaños del efecto de magnitud baja, media o alta, respectivamente. La forma de la distribución de los índices d fue examinada. El 38% de los test de Shapiro-Wilks de normalidad fueron estadísticamente significativos, siendo 0.221 el p valor asociado a dicho test de normalidad. Esto significa que un amplio número de los meta-análisis incluido no seguían una distribución normal. En relación a la asimetría de los tamaños del efecto, se puede observar como la media (0.534) y la mediana (0.567) son casi iguales, sin la existencia de valores extremos siendo el mínimo igual a -1.947 y el máximo 3.487. Por otro lado, hay una gran dispersión entre los valores de curtosis de los tamaños del efecto, el mínimo fue -1.758 y el máximo 12.22, con lo que la media y la mediana fueron más diferentes con valores 0.768 y -0.148, respectivamente.

Para examinar la forma de la distribución de los tamaños muestrales, se calcularon los índices de asimetría y curtosis. En relación a la asimetría, la media fue 1.324 y la mediana 1.350, sin valores extremos. Sin embargo, la curtosis reflejó una gran dispersión de valores (min. = -1.859 y máx. =14.170), por lo que la media y la mediana fueron muy discrepantes. Debido a la asimetría encontrada en la distribución de tamaños muestrales, los estadísticos descriptivos se calcularon para la mediana. La media fue 43.98, el mínimo 9 fue y el máximo 87.5.

El 70.73% de las correlaciones entre tamaño del efecto y tamaño muestral fueron negativas. Este resultado refleja la existencia de un amplio número de correlaciones negativas, siendo -0.085 el valor promedio. Además, se encontró un amplio rango de correlaciones con una correlación mínima igual a -0.736 y una máxima de 0.734. El 20% de las correlaciones fueron estadísticamente significativas, de las cuales 5 fueron positivas y 5 negativas.

El mínimo, el primer y tercer cuartil y la mediana de los p valores asociados a los estadísticos Q de heterogeneidad fueron menores de 0.05. Además, el 80% de las pruebas Q fueron estadísticamente significativas y el 86% de los índices I^2 fueron superiores al 25%. Esto indica la presencia de una gran variabilidad de los tamaños del efecto en la mayor parte de los meta-análisis.

Por último, en relación a la varianza inter-estudio, los percentiles 25, 50 y 75 fueron 0.062, 0.112 y 0.171, respectivamente.

Conclusiones

Los resultados de este estudio han mostrado que los meta-análisis reales presentan claras desviaciones del supuesto de normalidad para la distribución de tamaños del efecto, una gran variabilidad en la correlación entre tamaño del efecto y tamaño muestral y gran heterogeneidad en el número de estudios, distribución de los tamaños muestrales, tamaños del efecto y varianza inter-estudios.

Este estudio sirve para guiar estudios de simulación bajo condiciones realistas. Recomendamos que otros investigadores que realicen simulaciones Monte Carlo manipulen la asimetría y la cursosis de la distribución de los tamaños muestrales y tamaños del efecto, así como la correlación entre tamaño muestral y tamaño del efecto siguiendo los valores encontrados en esta revisión.

El grado en que los resultados de esta revisión sistemática de meta-análisis pueden generalizarse a otras disciplinas fuera del ámbito psicológico, dependerá de las similitudes de las características de los meta-análisis en esa disciplina.

Referencias

- [1] Borenstein, M., Hedges, L.V, Higgins, J.P.T. y Rothstein, H.R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1, 97-111.
- [2] Kontopantelis, E. y Reeves, D. (2010). Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. *Statistical Methods in Medical Research*, 0, 1-18.
- [3] Shuster, J. J. (2010). Empirical vs natural weighting in random effects meta-analysis. *Statistics in Medicine*, 29, 1259-1265.
- [4] Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, 36, 1-48.
- [5] Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F. y Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological methods*, 11(2), 193.