Psicológica (1996) 17, 467-482.

Estimadores del tamaño del efecto en meta-análisis: Un estudio Monte Carlo del sesgo y la eficiencia

Fulgencio Marín* y Julio Sánchez

Universidad de Murcia

La diferencia media tipificada es uno de los índices del tamaño del efecto más empleados para cuantificar e integrar los resultados de los estudios primarios en un meta-análisis. Mediante un estudio de simulación Monte Carlo, hemos comparado el sesgo y la eficiencia de cinco estimadores de la *diferencia media tipificada*: (a) la media aritmética a partir del índice $\hat{\Delta}$ de Glass (1976); (b) la media aritmética a partir del índice g de Hedges (Hedges, 1982); (c) la media aritmética a partir del índice corregido d de Hedges (Hedges y Olkin, 1985); (d) la media aritmética ponderada por la inversa de la varianza de g; y (e) la media aritmética ponderada por la inversa de la varianza de d (Hedges y Olkin, 1985). Partiendo de estudios primarios con un grupo experimental y un grupo de control, hemos manipulado el tamaño muestral (n se genera como una variable aleatoria uniforme), el tamaño del efecto paramétrico ($\delta = 0.2$, 0.5 y 0.8) y el número de estudios (k = 25, 50 y 100) en el meta-análisis. En las condiciones estudiadas, nuestros resultados muestran que el índice d se comporta como el menos sesgado. Sin embargo, se detectan diferencias en las alternativas de promediarlo a través de los estudios, destacando la mayor eficiencia en la media ponderada y el menor sesgo en el promedio sin ponderar.

Palabras clave: Tamaño del efecto, Monte Carlo, meta-análisis.

En los últimos 20 años la investigación en Ciencias del Comportamiento ha experimentado el desarrollo del meta-análisis, una nueva metodología para la revisión cuantitativa de la investigación sobre un determinado campo científico (Cooper, 1989; Glass, McGaw y Smith, 1981; Hedges y Olkin, 1985; Hunter y Schmidt, 1990; Rosenthal, 1991). En

---

* Parte de este trabajo se presentó al IV Symposium de Metodología de las Ciencias del Comportamiento, La Manga (Murcia), abril de 1995. Dirección de contacto: Dpto Psicología Básica y Metodología. Facultad de Psicología. Campus de Espinardo. Apdo 4021. 30080-Murcia. E-mail: fulmarin@fcu.um.es.

castellano pueden consultarse Bustamante y Delgado (1994), Gómez (1987), Marín (1996), Sánchez (1986), Sánchez y Ato (1989) y Vázquez (1990).

El meta-análisis permite sintetizar los resultados de un conjunto de estudios primarios sobre un determinado tópico de investigación, ofreciendo un índice medio de la relación entre las variables implicadas. Para ello, el meta-análisis aplica al proceso de revisión de la literatura las mismas normas de rigor científico que se exige a la realización de las investigaciones primarias. De esta forma, un meta-análisis se lleva a cabo siguiendo una serie de etapas similares a las de un estudio primario:

(1) *Formulación del problema.* Consiste en determinar el objetivo de la revisión que se pretende realizar.

(2) *Búsqueda de la literatura.* Una vez definido el problema, se lleva a cabo una búsqueda bibliográfica exhaustiva, utilizando métodos formales e informales, y tratando de recuperar tanto trabajos publicados como no publicados.

(3) *Codificación de las variables.* Si se sospecha que ciertas características de los estudios primarios pueden estar afectando a los resultados de los mismos, pueden codificarse con objeto de comprobar posteriormente su posible influencia moderadora.

(4) *Cálculo de un índice de los resultados de los estudios.* Para poder integrar los resultados de los estudios, es preciso definir un indicador aplicable a todos los estudios. El más común es el tamaño del efecto, que refleja la magnitud de la relación entre las variables implicadas. Según las características metodológicas del campo de investigación que se pretenda integrar, se utilizan distintos índices del tamaño del efecto, siendo los de uso más frecuente la diferencia media tipificada, para la investigación de corte experimental y cuasi-experimental, y el coeficiente de correlación de Pearson, para la investigación correlacional. No obstante, es posible transformar una diferencia media tipificada en coeficiente de correlación, y viceversa. De esta forma, cada estudio primario queda representado en el meta-análisis por un índice del tamaño del efecto.

(5) *Análisis e interpretación de los resultados.* El análisis de los índices del tamaño del efecto obtenidos en los estudios primarios debe comenzar con la obtención de estadísticos descriptivos, tales como la tendencia central y la variabilidad, acompañados de representaciones gráficas, seguidos de la comprobación de la homogeneidad de los tamaños del efecto en torno a su media y la verificación de la posible influencia de variables moderadoras sobre la distribución de los tamaños del efecto.

(6) *Publicación del estudio.* Finalmente, el meta-análisis debe publicarse siguiendo el esquema típico de cualquier investigación primaria: resumen, introducción, método, resultados y discusión.

El presente trabajo se centra en el estudio de uno de los indicadores del tamaño del efecto más frecuentemente utilizados en meta-análisis: la diferencia media tipificada. Diversos autores han propuesto diferentes estimadores de la diferencia media tipificada que pueden aplicarse a un mismo meta-análisis;

además, se han propuesto varios procedimientos para promediar un conjunto de diferencias medias tipificadas como resumen de un meta-análisis. Sin embargo, las propiedades de estos estadísticos apenas han sido investigadas. Si además tenemos en cuenta que los promedios que se pueden obtener con los diferentes procedimientos pueden diferir en mayor o menor medida, el estudio del comportamiento de estos estadísticos constituye una cuestión que requiere investigación. Este trabajo se enmarca en esta línea de investigación y trata de analizar, mediante simulación Monte Carlo, el sesgo y la eficiencia de varios estimadores de la diferencia media tipificada que se pueden obtener al promediar un conjunto de índices en un meta-análisis.

Se presentan en este estudio cinco estimadores de la diferencia media tipificada que se pueden aplicar para promediar los resultados de un meta-análisis; a continuación, se aplicarán dichos índices a un ejemplo; seguirá una presentación de la metodología seguida en el estudio de simulación y se finalizará con la interpretación y discusión de los resultados obtenidos.

1. La diferencia media tipificada.

En los diseños de investigación que implican la comparación de dos (o más) grupos, el índice del tamaño del efecto más usual es la *diferencia media tipificada*, que se define como la diferencia entre las medias de dos grupos (experimental y control) dividida por una estimación de la desviación típica. El modelo que subyace a este tipo de diseño asume la independencia y normalidad de las distribuciones de cada grupo en la variable dependiente. Es decir, siendo Y_j^E e Y_j^C las puntuaciones en la variable dependiente de dos poblaciones (experimental y control), suponemos que se distribuyen según una ley normal con medias μ^E y μ^C y varianza común σ^2:

$$Y_j^E \approx N(\mu^E, \sigma^2), \qquad j = 1,\ldots,N^E \qquad (1)$$

$$Y_j^C \approx N(\mu^C, \sigma^2), \qquad j = 1,\ldots,N^C \qquad (2)$$

A partir de estas dos distribuciones se define la diferencia media tipificada poblacional, δ, como:

$$\delta = \frac{\mu^E - \mu^C}{\sigma} \qquad (3)$$

Se han propuesto varios estimadores de δ. Glass, McGaw y Smith (1981) proponen el estimador $\hat{\Delta}$, según el cual la desviación típica, σ, se estima a partir del grupo de control:

$$\hat{\Delta} = \frac{\overline{y}^E - \overline{y}^C}{s^C} \qquad (4)$$

siendo $\overline{y}^E$ e $\overline{y}^C$ las medias muestrales de los grupos experimental y control, respectivamente; y s^C es la desviación típica insesgada de la muestra del grupo de control. La propuesta de Glass et al. (1981) es especialmente útil cuando el estudio implica un grupo que recibe alguna intervención y el otro actúa como grupo de control. En estas circunstancias, la aplicación de un tratamiento puede alterar la variabilidad natural de los sujetos en la variable dependiente, debido a un efecto interactivo sujeto-tratamiento que afectaría al grupo experimental, por lo que este grupo debe excluirse de la estimación de la desviación típica.

Por otra parte, Hedges (1981), partiendo del supuesto de homogeneidad de las varianzas, propone como estimador de δ, el índice *g* mediante

$$g = \frac{\bar{y}^E - \bar{y}^C}{s} \qquad (5)$$

donde *s* es la desviación típica intragrupo, que se obtiene por

$$s = \sqrt{\frac{(n^E - 1)(s^E)^2 + (n^C - 1)(s^C)^2}{n^E + n^C - 2}} \qquad (6)$$

Dado que la desviación típica del grupo de control, s^C, tiene mayor error muestral que la desviación típica intragrupo, *s*, el índice $\hat{\Delta}$ de Glass es menos eficiente que la *g* de Hedges.

Pero Hedges (1981) también demostró que $\hat{\Delta}$ y *g* son estimadores positivamente sesgados de δ, por lo que propuso un estimador insesgado, *d*, mediante:

$$d = J(m)g, \qquad (7)$$

donde *g* se obtiene por (5) y *J(m)* es una buena aproximación al factor de corrección del sesgo:

$$J(m) \approx 1 - \frac{3}{4m - 1} \qquad (8)$$

siendo

$$m = n^E + n^C - 2$$

Así pues, en la realización de un mismo meta-análisis diferentes investigadores pueden seleccionar distintos estimadores de la diferencia media tipificada para integrar los resultados de un conjunto de estudios, lo que puede dar lugar a ciertas discrepancias en las estimaciones del tamaño del efecto.

Uno de los objetivos del meta-análisis es obtener un índice global de la magnitud del efecto, promediando las estimaciones de la diferencia media tipificada del conjunto de estudios. Diferentes estimadores de δ darán lugar a diferentes índices globales. Pero también se pueden producir discrepancias debido a las diferentes formas de promediar tamaños del efecto. En efecto, Glass et al. (1981) proponen una simple media aritmética de los valores $\hat{\Delta}$.

Es decir, dado un conjunto de k estudios independientes, el valor medio de la diferencia media tipificada se obtendría mediante:

$$\bar{\Delta} = \frac{\sum \hat{\Delta}_i}{k} \qquad (9)$$

donde $\hat{\Delta}_i$ representa la diferencia media tipificada del *i*ésimo estudio, que se obtiene por (4) y k es el número de estudios.

Si en lugar de $\hat{\Delta}$ se aplica el índice g, o el índice insesgado, d, la media aritmética de cada uno de ellos se obtendría de igual forma mediante:

$$\bar{g} = \frac{\sum g_i}{k} \qquad (10)$$

$$\bar{d} = \frac{\sum d_i}{k} \qquad (11)$$

donde g_i y d_i se obtienen mediante (5) y (7), respectivamente, para el *i*ésimo estudio.

Sin embargo, como Hedges y Olkin (1985; véase también Shadish y Haddock, 1994) han demostrado, el modo más exacto de promediar un conjunto de k tamaños del efecto es calculando una media ponderada por la inversa de la varianza de cada estimación del tamaño del efecto. Para los índices g_i y d_i en el *i*ésimo estudio, estimamos sus respectivas varianzas según:

$$\mathrm{Var}(g_i) = \frac{n_i^E + n_i^C}{n_i^E n_i^C} + \frac{g_i^2}{2(n_i^E + n_i^C)} \qquad (12)$$

$$\mathrm{Var}(d_i) = \frac{n_i^E + n_i^C}{n_i^E n_i^C} + \frac{d_i^2}{2(n_i^E + n_i^C)} \qquad (13)$$

donde los valores g_i y d_i en ambas ecuaciones estiman el parámetro desconocido δ, que es el valor que permite obtener las varianzas exactas.

Así, las medias ponderadas de los índices g y d se calculan mediante:

$$\bar{g}_p = \frac{\sum w_i g_i}{\sum w_i} \qquad (14)$$

$$\bar{d}_p = \frac{\sum w_i d_i}{\sum w_i} \qquad (15)$$

donde w_i en las ecuaciones (14) y (15) representa la inversa de la varianza estimada de g_i y d_i, respectivamente; es decir, $1/Var(g_i)$ y $1/Var(d_i)$.

De las ecuaciones (12) y (13) se hace evidente que cuanto mayores son los tamaños muestrales de un estudio, menor será la varianza de la estimación y, en consecuencia, mayor será su peso específico en el cálculo de la media ponderada. Intuitivamente parece razonable que aquellos tamaños del efecto con menor varianza (y tamaños muestrales más elevados) aporten un mayor peso al cálculo del promedio. Además, Hedges y Olkin (1985) han demostrado que la ponderación por la inversa de la varianza maximiza la eficiencia del promedio.

Los estimadores $\overline{\Delta}$, $\overline{g}$ y $\overline{d}$ tienen errores típicos de estructura similar:

$$\hat{\sigma}_{\overline{\Delta}} = \frac{s_{\hat{\Delta}}}{\sqrt{k}} \qquad (16)$$

$$\hat{\sigma}_{\overline{g}} = \frac{s_g}{\sqrt{k}} \qquad (17)$$

$$\hat{\sigma}_{\overline{d}} = \frac{s_d}{\sqrt{k}} \qquad (18)$$

siendo k el número de estudios y $s_{\hat{\Delta}}$, s_g y s_d son las desviaciones típicas de los k tamaños del efecto promediados.

Por su parte, los estimadores $\overline{g}_p$ y $\overline{d}_p$, al ser medias ponderadas tienen como errores típicos (Hedges y Olkin, 1985):

$$\hat{\sigma}_{\overline{g}_p} = \frac{1}{\sum w_i} \qquad (19)$$

$$\hat{\sigma}_{\overline{d}_p} = \frac{1}{\sum w_i} \qquad (20)$$

donde w_i en (19) y (20) se obtiene como la inversa de (12) y (13), respectivamente.

Así pues, partiendo de la teoría estadística exacta, cabe esperar que los estimadores más eficientes de la diferencia media tipificada de un conjunto de k estudios independientes sean $\overline{d}_p$ y $\overline{g}_p$, ya que ponderan por la inversa de la varianza; les seguirán los estimadores $\overline{d}$ y $\overline{g}$, siendo $\overline{\Delta}$ el menos eficiente al utilizar sólo la desviación típica del grupo de control. Serán insesgados los estimadores $\overline{d}_p$ y $\overline{d}$ que corrigen el sesgo mediante (8). Sin embargo, dado que (8) es una aproximación al factor exacto de corrección del sesgo, y (13) una estimación aproximada de la varianza de d_i, los estimadores $\overline{d}$ y $\overline{d}_p$ pueden ver alterada su tendencia central y su eficiencia.

2. Un ejemplo.

La Tabla 1 presenta los resultados de un meta-análisis ficticio en el que un conjunto de $k = 6$ estudios con dos grupos (experimental vs. control) pretenden ser integrados para obtener un índice del tamaño del efecto medio. La Tabla 1 muestra los tamaños muestrales, n^E y n^C, las medias, $\overline{y}^E$ e $\overline{y}^C$, y las desviaciones típicas, s^E y s^C. Suponemos en este hipotético meta-análisis una situación frecuente: el grupo experimental sufre un ligero incremento de su variabilidad, respecto del grupo de control, debido a la exposición al tratamiento o fenómeno de interés. No obstante, se cumple el supuesto de homogeneidad de varianzas en todos los casos (*n.s.* = 1%).

Tabla 1. Datos de un meta-análisis ficticio.

Estudio	n^E n^C	$\bar{y}^E$ $\bar{y}^C$	s^E s^C	$\hat{\Delta}$	g Var(g)	d Var(d)
1	20	8.150	10.734	0.4600	0.3788	0.3712
	20	4.614	7.687		0.1018	0.1017
2	18	8.070	9.888	0.5326	0.4707	0.4603
	18	3.854	7.916		0.1142	0.1141
3	22	6.858	12.495	0.5895	0.4208	0.4132
	22	2.552	7.304		0.0929	0.0928
4	15	6.791	12.956	0.2079	0.1771	0.1721
	15	4.759	9.776		0.1339	0.1338
5	12	7.490	10.956	0.4109	0.3897	0.3763
	12	3.420	9.906		0.1698	0.1696
6	16	7.074	11.516	0.3252	0.2636	0.2569
	16	4.455	8.054		0.1261	0.1260

n^E (experimental) y n^C (control): tamaños muestrales de los grupos. $\bar{y}^E$ y $\bar{y}^C$: Medias de los grupos.
s^E y s^C: Desviaciones típicas insesgadas.
$\hat{\Delta}$: Estimador de la diferencia media tipificada (Glass).
g: Estimador sesgado de la diferencia media tipificada (Hedges).
d: Estimador insesgado de la diferencia media tipificada (Hedges).
Var(g) y Var(d): Varianzas de los estimadores g y d.

El primer cálculo que deberemos hacer es obtener la diferencia media tipificada en cada estudio. La Tabla 1 presenta los valores obtenidos por los estimadores de la diferencia media tipificada $\hat{\Delta}$, g y d, resultantes de aplicar las ecuaciones (4), (5) y (7), respectivamente; así como las varianzas de g y d, *Var(g)* y *Var(d)*, mediante las ecuaciones (12) y (13).

Una vez realizados los cálculos para cada estudio individual, procedemos a obtener las cinco estimaciones del promedio de la diferencia media tipificada del conjunto de estudios, es decir: $\overline{\Delta}$, $\overline{g}$, $\overline{d}$, $\overline{g}_p$ y $\overline{d}_p$. Para ello, aplicamos las ecuaciones (9), (10), (11), (14) y (15), respectivamente. La Tabla 2 presenta los resultados obtenidos. El promedio más alto ha sido obtenido por el estimador $\overline{\Delta}$(0.4210), como era de esperar debido a la menor variabilidad de los grupos de control. Los dos estimadores *g* sesgados, $\overline{g}$ y $\overline{g}_p$, presentan valores del orden de un 16% inferiores a $\overline{\Delta}$ (0.3501 y 0.3571, respectivamente); mientras que los dos estimadores *d* insesgados, $\overline{d}$ y $\overline{d}_p$, son un 18% inferiores al valor de $\overline{\Delta}$ (0.3417 y 0.3489, respectivamente). Así pues, en cuanto a la tendencia central el ejemplo pone de manifiesto que pueden obtenerse discrepancias considerables en función del estimador seleccionado.

En la Tabla 2 también se muestran los errores típicos de los cinco estimadores de la diferencia media tipificada, para lo cual se aplicaron las ecuaciones (16), (17), (18), (19) y (20). Nuevamente, el estimador menos eficiente fue $\overline{\Delta}$, con un error típico de 0.0569. Respecto de $\overline{g}$ y $\overline{d}$, $\overline{\Delta}$ presenta una eficiencia relativa de tan sólo el 60%; y menos aún, el 12%, respecto de $\overline{g}_p$ y $\overline{d}_p$.

Tabla 2. Medias y errores típicos de los estimadores.

	Estimador				
	$\overline{\Delta}$	$\overline{g}$	$\overline{d}$	$\overline{g}_p$	$\overline{d}_p$
Medias	0.4210	0.3501	0.3417	0.3571	0.3489
Errores típicos	0.0569	0.0445	0.0437	0.0197	0.0197

$\overline{\Delta}$: Media sin ponderar de los valores $\hat{\Delta}$. $\overline{g}$: Media sin ponderar de los valores *g*. $\overline{d}$: Media sin ponderar de los valores *d*. $\overline{g}_p$: Media ponderada de los valores *g*. $\overline{d}_p$: Media ponderada de los valores *d*.

SIMULACION

Los resultados del ejemplo ponen de manifiesto que las discrepancias entre los diversos estimadores de la diferencia media tipificada de un conjunto de estudios pueden no ser triviales, requiriendo un estudio del alcance y consecuencias que pueden tener en las investigaciones meta-analíticas. No son muchos los trabajos que han tratado esta cuestión. Hedges y Olkin (1985, pp. 114-117) estudiaron mediante simulación Monte Carlo el comportamiento del estimador $\overline{d}_p$, detectando un sesgo negativo debido a que el factor *J(m)* es una aproximación al verdadero factor de corrección del sesgo. Pero su estudio sólo contempló dos valores de *k*, 2 y 5, que no son muy habituales en los meta-análisis reales. Por otra parte, Thomas (1986) comparó la eficiencia de varios estimadores de la diferencia media tipificada, de los cuales en nuestro estudio sólo se incluye el cálculo del error típico del estimador $\overline{\Delta}$, $\hat{\sigma}_{\overline{\Delta}}$, según la ecuación (16). Así pues, no se ha comparado el sesgo y la eficiencia de los estimadores de uso más frecuente de la diferencia media tipificada para comprobar la posible influencia que puede ejercer sobre los resultados de un meta-análisis la elección del estimador del tamaño del efecto.

MÉTODO

Todo el estudio se programó en GAUSS (1992). Se definieron dos poblaciones, experimental y control, con distribución normal y varianzas homogéneas $[N(\mu^E,\sigma^2),\ N(\mu^C,\sigma^2)]$, a partir de las cuales se generaron aleatoriamente pares de muestras independientes de tamaño $n^E = n^C = n$. En concreto, se simularon estudios constituidos por un grupo experimental y un grupo control, a partir de un tamaño del efecto paramétrico, δ, definido según la ecuación (3).

Cada par de muestras generadas simula los datos de una investigación primaria. Para cada estudio se realizaron los siguientes cálculos (Hedges y Olkin, 1985):

1) Cálculo de las medias, $\bar{y}^E$ e $\bar{y}^C$, y las varianzas insesgadas, $(s^E)^2$ y $(s^C)^2$, de las dos muestras.

2) Cálculo del valor $\hat{\Delta}$, según la ecuación (4).

3) Cálculo del índice *g*, mediante la ecuación (5).

4) Cálculo del estimador insesgado *d*, aplicando la ecuación (7).

De esta forma, se obtuvieron para cada estudio simulado los tres estimadores de la diferencia media tipificada: $\hat{\Delta}$, *g* y *d*.

En un segundo paso se generaban una serie de *k* estudios que simulaban los datos de un meta-análisis. Para ello se manipularon los siguientes parámetros: (a) el tamaño muestral por estudio, $N = n^E + n^C$, donde $n^E = n^C$ se generó como una variable aleatoria uniforme entre los valores 20 y 40; (b) el número de estudios, *k*, con valores de 25, 50 y 100; (c) y el tamaño del efecto paramétrico del que procedían los *k* estudios, constante en cada meta-análisis, con los valores $\delta = 0.2$, 0.5 y 0.8. Siguiendo a Cohen (1988), estos tres valores de δ se corresponden con un efecto típicamente bajo, medio y alto, respectivamente, en Ciencias Sociales.

Se realizaron 500 réplicas de cada una de las 3 (número de estudios) x 3 (efecto paramétrico) = 9 condiciones del estudio, siendo cada réplica la simulación de un meta-análisis. Por tanto, se simularon 4500 meta-análisis. Sobre cada una de estas réplicas (o meta-análisis) se aplicaron los cinco procedimientos de integración que son objeto de estudio:

1) Cálculo de la media aritmética de los k valores $\hat{\Delta}_i$, g_i y d_i, aplicandc las ecuaciones (9), (10) y (11), respectivamente.

2) Cálculo de la media ponderada por la inversa de la varianza de los k valores g_i y d_i, según las ecuaciones (14) y (15), respectivamente.

Para una determinada condición, el sesgo de los cinco procedimientos se obtuvo promediando las diferencias entre el estimador del efecto medio y e valor paramétrico constante δ, a través de las 500 réplicas. Es decir, siendc $\hat{\theta}$ el estimador del efecto medio y θ el parámetro, calculamos para cada réplica la diferencia $\hat{\theta} - \theta$, y obtuvimos la media aritmética. Por su parte, la eficiencia se estimó calculando la varianza insesgada de las estimaciones er las 500 réplicas.

RESULTADOS Y DISCUSIÓN

La Tabla 3 muestra los valores del sesgo para los cinco estimadores er función de la magnitud del efecto paramétrico (δ) y del número de estudios (k) en un meta-análisis. A través de todas las condiciones, el estimador $\overline{\Delta}$ presenta el sesgo medio de mayor magnitud (+0.0140), seguido de los estimadores $\overline{d}_p$ (-0.0081), $\overline{g}$ (+0.0068), $\overline{g}_p$ (-0.0018) y $\overline{d}$ (-0.0001). Los estimadores $\overline{\Delta}$ y $\overline{g}$ presentan un sesgo sistemáticamente positivo en todas las condiciones, sobreestiman el parámetro, siendo el sesgo de $\overline{\Delta}$ sistemáticamente superior al de $\overline{g}$. Por su parte, los estimadores $\overline{d}_p$ y $\overline{g}_p$ presentan un sesgo sistemáticamente negativo, infraestiman el parámetro, con un mayor sesgo para $\overline{d}_p$ en todas las condiciones. El estimador $\overline{d}$ presenta las estimaciones menos sesgadas en gran parte de las condiciones, con valores de sesgo tanto positivos como negativos, que en valor absoluto suelen mostra una magnitud inferior a la del resto de estimadores.

Las Figuras 1 y 2 muestran el sesgo medio de los cinco estimadores, en función de la magnitud del efecto paramétrico y del número de estudios de meta-análisis, respectivamente. Se observa un incremento del sesgo en lo estimadores $\overline{\Delta}$, $\overline{g}$ y $\overline{d}_p$, conforme aumenta el efecto paramétrico, mientra que los estimadores $\overline{g}_p$ y $\overline{d}$ apenas se dejan afectar por δ (Figura 1). Por su parte, el número de estudios del meta-análisis tampoco parece afectar al sesgo de los cinco estimadores, que en los tres valores de k mantienen la misma ordenación: el mayor sesgo positivo para $\overline{\Delta}$, seguido de $\overline{g}$, el menor sesgo para $\overline{d}$ y un ligero sesgo negativo para $\overline{g}_p$ que se incrementa en el caso de $\overline{d}$ (Figura 2).

Tabla 3. Sesgo de los estimadores del tamaño del efecto.

		Tamaño del efecto				
δ	k	$\overline{\Delta}$	$\overline{g}$	$\overline{d}$	$\overline{g}_p$	$\overline{d}_p$
0.2	25	0.0043	0.0013	-0.0014	-0.0012	-0.0037
	50	0.0048	0.0021	-0.0006	-0.0009	-0.0035
	100	0.0042	0.0014	-0.0013	-0.0023	-0.0048
0.5	25	0.0161	0.0089	0.0019	-0.0004	-0.0067
	50	0.0126	0.0055	-0.0013	-0.0030	-0.0093
	100	0.0155	0.0083	0.0014	-0.0005	-0.0068
0.8	25	0.0211	0.0099	-0.0012	-0.0034	-0.0135
	50	0.0244	0.0125	0.0014	-0.0021	-0.0121
	100	0.0230	0.0115	0.0004	-0.0025	-0.0126

δ: Efecto paramétrico. k: Número de estudios.
$\overline{\Delta}$: Media sin ponderar de los valores $\hat{\Delta}$. $\overline{g}$: Media sin ponderar de los valores g. $\overline{d}$: Media sin ponderar de los valores d. $\overline{g}_p$: Media ponderada de los valores g. $\overline{d}_p$: Media ponderada de los valores d.

Así, la propuesta del índice insesgado d y de su ponderación por la inversa de la varianza (Hedges y Olkin, 1985) precisa de una serie de matices. En efecto, se consigue reducir el sesgo positivo, más pronunciado en el estimador $\overline{\Delta}$ (suponiendo varianzas homogéneas en las poblaciones) y de cierta magnitud en el estimador $\overline{g}$. Sin embargo, el sentido del sesgo se invierte con la media ponderada de índices d, lo que supone una infraestimación de δ. A este respecto, la simple media aritmética de índices d presenta el mejor comportamiento, con las estimaciones menos sesgadas. Aunque estadísticamente la ponderación por la inversa de la varianza minimiza el sesgo, el carácter aproximado de la ecuación (13), donde d estima el parámetro desconocido δ, provoca la infraestimación del parámetro. De hecho, el promedio ponderado $\overline{g}_p$, que parte del estimador positivamente sesgado g, también comporta un sesgo sistemáticamente negativo cuando se pondera por la inversa de la varianza estimada según la ecuación (12). Así, la imposibilidad de derivar en la práctica un estimador más ajustado de la inversa de la varianza aconseja la opción por la simple media aritmética de índices d, como el estimador menos sesgado.

En la Tabla 4 presentamos las varianzas de los cinco estimadores, como un indicador de su eficiencia, en función de la magnitud del efecto paramétrico (δ) y del número de estudios (k) en un meta-análisis. En todas las condiciones el estimador $\overline{d}_p$ resulta ser el más eficiente o con la menor varianza, seguido de los estimadores $\overline{g}_p$, $\overline{d}$, $\overline{g}$ y $\overline{\Delta}$, en ese orden. Por consiguiente, el criterio de ponderación por la inversa de la varianza, en primer lugar, y la corrección del sesgo de g, en segundo lugar, incrementan la

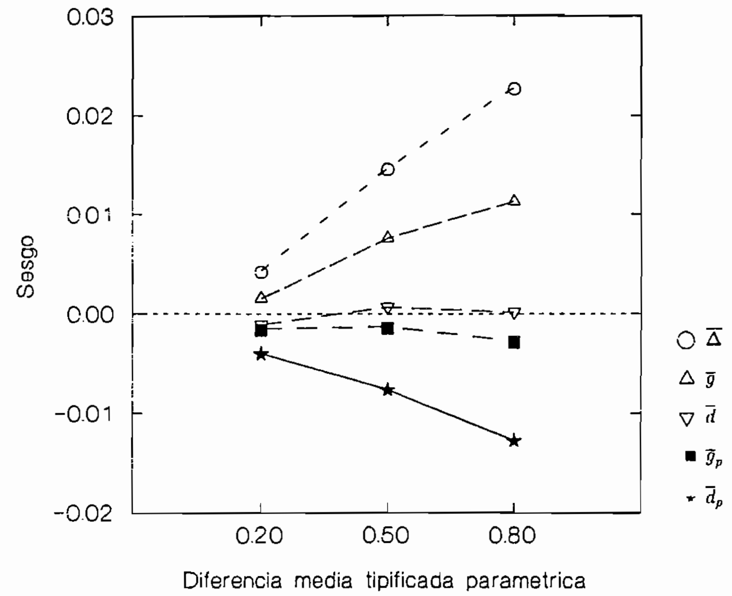

Figura 1. El sesgo en función de la magnitud del efecto paramétrico. (Nota: $\bar{\Delta}$: Media sin ponderar de los valores $\hat{\Delta}$. $\bar{g}$: Media sin ponderar de los valores *g*. $\bar{d}$: Media sin ponderar de los valores *d*. $\bar{g}_p$: Media ponderada de los valores *g*. $\bar{d}_p$: Media ponderada de los valores *d*.).

precisión del estimador de δ en un meta-análisis. Por su parte, la mayor variabilidad de $\bar{\Delta}$ es una consecuencia directa de la menor eficiencia del estimador Δ frente a *g* y *d*, bajo el supuesto de homogeneidad de las varianzas.

La Figura 3 muestra la eficiencia de los cinco estimadores en función del tamaño del efecto paramétrico. Mientras que el estimador $\bar{\Delta}$ disminuye en eficiencia conforme aumenta la magnitud de δ, los otros estimadores no reflejan una tendencia lineal clara. En la Figura 4 se ilustra la eficiencia en función del número de estudios del meta-análisis: en los cinco procedimientos el aumento de *k* se asocia con un incremento en la eficiencia (disminución de la varianza) de los estimadores. Asimismo, en ambas Figuras (3 y 4) puede apreciarse cómo se mantiene la ordenación en la eficiencia de los estimadores, independientemente de las condiciones manipuladas.

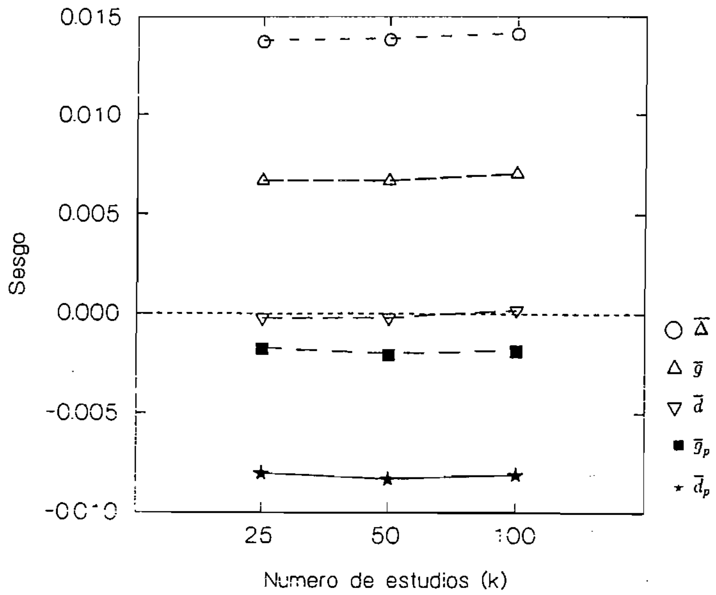

Figura 2. El sesgo en función del número de estudios. (Nota: $\overline{\Delta}$: Media sin ponderar de los valores $\hat{\Delta}$. $\overline{g}$: Media sin ponderar de los valores *g*. $\overline{d}$: Media sin ponderar de los valores *d*. $\overline{g}_p$: Media ponderada de los valores *g*. $\overline{d}_p$: Media ponderada de los valores *d*.).

Tabla 4. Eficiencia de los estimadores del tamaño del efecto.

		Tamaño del efecto				
δ	k	$\overline{\Delta}$	$\overline{g}$	$\overline{d}$	$\overline{g}_p$	$\overline{d}_p$
0.2	25	0.0555	0.0539	0.0531	0.0518	0.0511
	50	0.0409	0.0401	0.0396	0.0380	0.0375
	100	0.0269	0.0263	0.0259	0.0251	0.0248
0.5	25	0.0554	0.0533	0.0525	0.0518	0.0511
	50	0.0421	0.0405	0.0400	0.0389	0.0384
	100	0.0287	0.0278	0.0274	0.0266	0.0263
0.8	25	0.0588	0.0550	0.0542	0.0524	0.0517
	50	0.0400	0.0379	0.0374	0.0372	0.0367
	100	0.0303	0.0285	0.0280	0.0269	0.0265

δ: Efecto paramétrico. k: Número de estudios.
$\overline{\Delta}$: Media sin ponderar de los valores $\hat{\Delta}$. $\overline{g}$: Media sin ponderar de los valores g. $\overline{d}$: Media sin ponderar de los valores d. $\overline{g}_p$: Media ponderada de los valores g. $\overline{d}_p$: Media ponderada de los valores d.

En resumen, el estimador ideado por Hedges y Olkin (1985) como e
más apropiado para estimar la diferencia media tipificada paramétrica en ur
meta-análisis, $\bar{d}_p$, no está exento de problemas. Aun siendo el más eficiente
de los estimadores comparados, proporciona estimaciones negativamente
sesgadas, infraestimando el parámetro δ en mayor medida que la simple
media aritmética de índices *d* (estimador $\bar{d}$). Mientras que con el índice *c*
(ecuación 7) conseguimos las estimaciones menos sesgadas de δ, e
promedio ponderado por la inversa de la varianza de estos índices *d* a travé
de los *k* estudios de un meta-análisis queda sesgado debido a la inexactitud

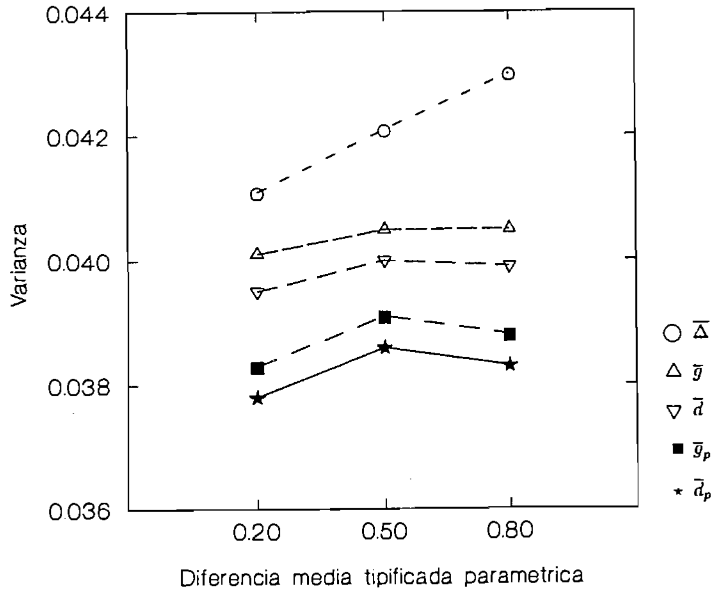

Figura 3. La eficiencia en función de la magnitud del efect paramétrico. (Nota: $\bar{\Delta}$: Media sin ponderar de los valores $\hat{\Delta}$. $\bar{g}$ Media sin ponderar de los valores *g*. $\bar{d}$: Media sin ponderar d los valores *d*. $\bar{g}_p$: Media ponderada de los valores *g*. $\bar{d}_p$: Medi ponderada de los valores *d*.).

del factor de ponderación, tal y como se calcula en la práctica (ecuación 13)
En consecuencia, la opción por un promedio sin ponderar de los índices *d* s
muestra como la más apropiada en cuanto al sesgo, con el inconveniente d
una menor eficiencia frente al promedio ponderado. Por su parte, l
estimación mediante índices $\hat{\Delta}$, en primer lugar, y mediante índices *g*, e
segundo lugar, provoca las estimaciones más sesgadas y menos eficientes

Estas consideraciones generales se mantienen en las diferentes condiciones de aplicación de un meta-análisis manipuladas en nuestro estudio de simulación.

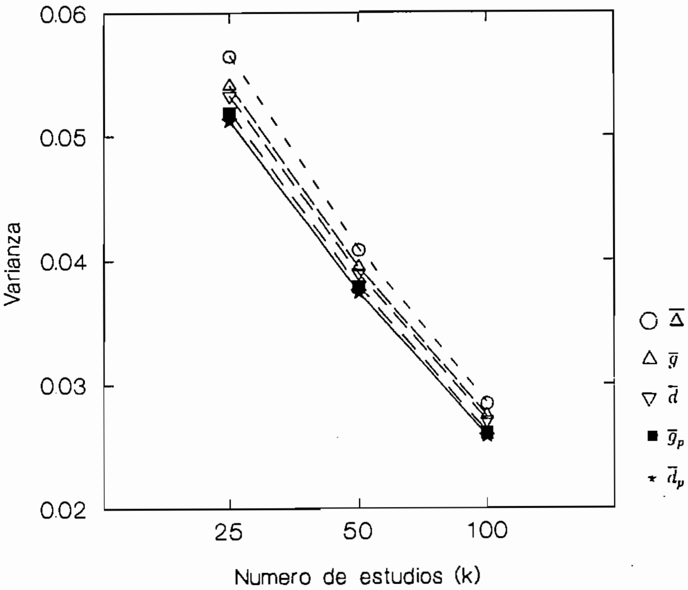

Figura 4. La eficiencia en función del número de estudios. (Nota: $\bar{\Delta}$: Media sin ponderar de los valores $\hat{\Delta}$. $\bar{g}$: Media sin ponderar de los valores *g*. $\bar{d}$: Media sin ponderar de los valores *d*. $\bar{g}_p$: Media ponderada de los valores *g*. $\bar{d}_p$: Media ponderada de los valores *d*.).

ABSTRACT

Standardized mean difference is one of the most applied effect size indexes, in order to quantifying and pooling the primary studies results in meta-analysis. With a Monte Carlo study, the bias and efficiency of five estimators of the *standardized mean difference* are compared: (a) the mean of Glass's $\hat{\Delta}$ indexes (Glass, 1976); (b) the mean of biased *g* indexes (Hedges, 1982); (c) the mean of unbiased *d* indexes (Hedges and Olkin, 1985); (d) the inverse-variance weighted mean of biased *g* indexes; and (e) the inverse-variance weighted mean of unbiased *d* indexes (Hedges and Olkin, 1985). We generate primary studies with two groups, experimental and control ones, manipulate sample size (*n* is generated as an uniform random variable), the population effect size (δ= 0.2, 0.5 y 0.8) and the number of studies (k = 25, 50, 100) in the meta-analysis. Through all the analyzed conditions, our findings show the lowest bias in the *d* index. However, there are discrepances in the averaging procedure of d index through the primary studies. So, we find the largest efficiency in the weighted mean d and the lowest bias in the average d without weighting.

Key words: Effect size, Monte Carlo, meta-analysis.

REFERENCIAS

Bustamante, F.J. y Delgado, J. (1994). Las corrientes del meta-análisis: Líneas de convergencia. *Psicológica, 15*, 255-274.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New York: Academic Press.

Cooper, H.M. (1989). *Integrating Research: A Guide for Literature Reviews* (2nd ed.). Beverly Hills, CA: Sage.

GAUSS (1992). *The GAUSS System* (Vers. 3.0). Washington: Aptech Systems, Inc.

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Research*, 5, 3-8.

Glass, G.V.; McGaw, B. y Smith, M.L. (1981). *Meta-analysis in Social Research.* Beverly Hills, CA: Sage.

Gómez, J. (1987). *Meta-análisis.* Barcelona: PPU.

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.

Hedges, L.V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics*, 7, 119-137.

Hedges, L.V. y Olkin, I. (1985). *Statistical Methods for Meta-analysis.* Orlando, FL: Academic Press.

Hunter, J.E. y Schmidt, F.L. (1990). *Methods of Meta-analysis: Correcting Error and Bias in Research Findings.* Beverly Hills, CA:Sage.

Marín, F. (1996). *Enfoques Meta-analíticos: Un Estudio Comparativo mediante Simulación Monte Carlo.* Tesis doctoral no publicada, Universidad de Murcia.

Rosenthal, R. (1991). *Meta-analytic Procedures for Social Research* (rev. ed.). Newbury Park, CA: Sage.

Sánchez, J. (1986). La revisión cuantitativa: Una alternativa a las revisiones tradicionales. *Anales de Psicología*, 3, 79-108.

Sánchez, 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.

Shadish, W.R. y Haddock, C.K. (1994). Combining estimates of effect size. En H.M. Cooper & L.V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 261-283). New York: Sage.

Thomas, H. (1986). Effect size standard errors for the non-normal non-identically distributed case, *Journal of Educational Statistics*, 11, 293-303.

Vázquez, C. (1990). Revisiones cuantitativas de la literatura: el meta-análisis. *Evaluación Psicológica*, 6, 261-288.

(Revisión aceptada: 26/6/96)