# EL COSTE DE LA DICOTOMIZACIÓN EN META-ANÁLISIS*

# THE COST OF DICHOTOMIZING IN META-ANALYSIS

Fulgencio Marín Martínez, Julio Sánchez Meca y Tania Bibiana Huedo Medina

*Dpto Psicología Básica y Metodología*

*Universidad de Murcia*

14 de Noviembre, 2003

Dirección de contacto:
Fulgencio Marín Martínez
Dpto Psicología Básica y Metodología. Facultad de Psicología. Campus de Espinardo. Apdo 4021. 30100 Murcia.
E-mail: fulmarin@um.es
Tfno.: 968363471
Fax: 968364115

* Esta investigación ha sido financiada por el Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2000-2003 del Ministerio de Ciencia y Tecnología y por fondos FEDER (Proyecto Nº: SEC 2001-3821-C05-05).

# EL COSTE DE LA DICOTOMIZACIÓN EN META-ANÁLISIS


## Resumen

La dicotomización de variables dependientes continuas conlleva un sesgo negativo (infraestimación) en la estimación del tamaño del efecto. Recientemente se han propuesto nuevos índices del tamaño del efecto que corrigen tal sesgo. Sin embargo, cuando estos nuevos índices se calculan en un estudio primario, su variabilidad (incertidumbre) en torno al tamaño del efecto en la población es mayor que la de los índices tradicionales no corregidos. Sólo cuando el tamaño muestral en los estudios es muy elevado o cuando los estudios se incluyen en un meta-análisis y los tamaños del efecto se promedian, los índices insesgados resultan ser también más eficientes que los no corregidos. El objetivo de este trabajo fue explorar la influencia de diferentes condiciones meta-analíticas en el sesgo y eficiencia de dos índices del tamaño del efecto calculados sobre datos dicotomizados y promediados a través de los estudios en el meta-análisis: uno prácticamente insesgado ($\bar{d}_{Probit}$) y el otro sin corregir ( $\bar{d}_p$). Se presenta el mínimo número de estudios en el meta-análisis necesario para conseguir que el índice insesgado se comporte mejor que el no corregido, en función del tamaño muestral de los estudios y de la magnitud del tamaño del efecto en la población.


**Palabras-clave**: Dicotomización de variables continuas, tamaño del efecto, diferencia media tipifica, meta-análisis, tablas 2x2.


## Abstract

Dichotomizing continuous outcomes implies a negative bias (underestimation) in the estimation of an effect size. New effect-size indices have recently been proposed which correct such a bias. However, when these new indices are applied on a primary study, their variability (uncertainty) around the population effect size is larger than that of the traditional uncorrected indices. Only when the sample size in the studies is extremely large or when the studies are included in a meta-analysis and the effect sizes are averaged, the unbiased indices are also more efficient than the uncorrected ones. The purpose of this work was to explore the influence of different meta-analytic conditions in the bias and efficiency of two effect-size indices applied on dichotomized outcomes and averaged through the studies in the meta-analysis: one practically unbiased ($\bar{d}_{Probit}$) and the other uncorrected ( $\bar{d}_p$ ). The minimum number of studies in the meta-analysis necessary to get that the unbiased index performs better than the uncorrected one is presented as a function of both the sample size in the studies and the magnitude of the population effect size.


**Key-words**: Dichotomization of continuous outcomes, effect size, standardised mean difference, meta-analysis, 2x2 tables.

## 1. INTRODUCCIÓN

La dicotomización de variables de naturaleza continua es una estrategia metodológica muy frecuente en la investigación en Ciencias Sociales y del Comportamiento. Veamos algunos ejemplos. Así, el hábito de fumar puede medirse como una variable cuantitativa, registrando el número de cigarrillos fumados por un sujeto en un determinado período (i.e., al día), y posteriormente dicotomizarse distinguiendo entre fumadores y no fumadores. De forma similar, el grado de rehabilitación de un delincuente ya liberado de la prisión también puede medirse cuantitatitivamente, por ejemplo a través del número de veces que reincide en delinquir, y posteriormente dicotomizarse distinguiendo entre aquéllos que reinciden (no importa cuántas veces) y los que no reinciden. Finalmente, el rendimiento académico frente a un examen puede evaluarse en una escala cuantitativa de 0 a 10 y posteriormente dicotomizarse, registrando únicamente si el sujeto está aprobado o suspenso.

La dicotomización de variables continuas tiene un coste debido a que con ella perdemos buena parte de la información cuantitativa de la variable original, lo que provoca un sesgo negativo (infraestimación) en las estimaciones de cualquier tamaño del efecto obtenidas a partir de la variable dicotomizada (Hunter y Schmidt, 1990; Lipsey y Wilson, 2001; Sánchez-Meca, Marín-Martínez y Chacón-Mocoso, 2003). Un ejemplo elaborado mediante simulación Monte Carlo clarificará esta idea. Supongamos que tenemos dos poblaciones normales, una experimental $N(42,4)$ en la que se aplica un determinado programa de intervención y otra de control en la que no se aplica ningún programa $N(40,4)$. Establecemos un punto de corte para la dicotomización en el valor 41 de la variable dependiente. Obsérvese que ambas poblaciones tienen la misma desviación típica (4 puntos) y diferentes medias (42 la población experimental y 40 la de control) que ilustran la eficacia del programa de intervención. Ante esta situación, un objetivo fundamental sería el de evaluar el grado de eficacia del programa de intervención a través de un índice del tamaño del efecto como la diferencia media tipificada, calculada en la población como la diferencia entre las dos medias experimental y control dividida por la desviación típica común en ambas poblaciones. En el ejemplo, el tamaño del efecto $\delta = (\mu_E - \mu_C) / \sigma = (42\text{-}40)/4 = 0{,}5$, es de 0,5 desviaciones típicas, que indica una eficacia moderada del programa de intervención conforme la clasificación de Cohen (1988) de tamaños del efecto típicamente bajos, medios y altos en las Ciencias Sociales.

Para estimar este tamaño del efecto extraemos mediante simulación dos muestras aleatorias, una de cada población, con el mismo tamaño de 50 casos. La Tabla 1 muestra los resultados a partir de las puntuaciones cuantitiativas, que se resumen en dos medias y en dos desviaciones típicas; y tras la dicotomización de las puntuaciones a partir del punto de corte 41, que da lugar a una tabla de frecuencias 2x2 con el número de casos en cada uno de los dos valores de la variable dicotomizada y en cada grupo. Cuando disponemos de toda la información cuantitativa, la estimación de la diferencia media tipificada se realiza a través del índice $g$ (Hedges y Olkin, 1985), calculado según:

$$g = \frac{\bar{y}_E - \bar{y}_C}{S}, \qquad (1)$$

donde $\bar{y}_E$ e $\bar{y}_C$ son las medias de las muestras experimental y control, y $S$ es es una estimación de la desviación típica intragrupo, obtenida según:

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

siendo $S_E^2$ y $S_C^2$ las varianzas, y $n_E$ y $n_C$ los tamaños muestrales de los grupos experimental y control, respectivamente. Aplicada esta ecuación a nuestros datos muestrales, el resultado, $g = 0{,}528$ (véase Tabla 1), sobreestima ligeramente el valor poblacional δ = 0,5, debido fundamentalmente al error aleatorio de muestreo.

INSERTAR LA TABLA 1 APROXIMADAMENTE AQUÍ, POR FAVOR

Cuando, por otro lado, dicotomizamos y perdemos la información cuantitativa, las medias de ambos grupos se transforman en proporciones de éxito en cada grupo, con lo que la diferencia media tipificada se estima a partir de una diferencia de proporciones tipificada, $d_p$, que es un estimador negativamente sesgado de la diferencia media tipificada en la población, δ. En la Tabla 1 presentamos el cálculo de la diferencia de proporciones tipificada $d_p$ (Johnson, 1989), que se obtiene según la ecuación:

$$d_p = \frac{p_E - p_C}{S'}, \qquad (3)$$

donde $p_E$ y $p_C$ son las proporciones de éxito en las muestras experimental y control, y $S'$ viene dada por:

$$S' = \sqrt{\frac{(n_E - 1)p_E(1 - p_E) + (n_C - 1)p_C(1 - p_C)}{n_E + n_C - 2}}. \qquad (4)$$

El resultado a partir de nuestros datos dicotomizados en la Tabla 1, $d_p$ = 0,417, infraestima el valor $\delta$ = 0,5, y no sólo debido a error de muestreo, sino además al sesgo negativo del estimador. Este sesgo negativo es consecuencia directa de la pérdida de información al dicotomizar la variable cuantitativa, o un efecto del coste de la dicotomización.

**2. ESTIMADORES DE LA DIFERENCIA MEDIA TIPIFICADA A PARTIR DE VARIABLES DEPENDIENTES DICOTOMIZADAS**

En la actualidad disponemos de varios índices para estimar una diferencia media tipificada a partir de una variable dependiente dicotomizada. Podemos dividirlos en dos grupos: (1) los que presentan un sesgo negativo derivado del coste de la dicotomización, y (2) los que tratan de corregir dicho sesgo partiendo en muchos casos de los anteriores.

Dentro del primer grupo tenemos el índice más utilizado hasta la fecha, la diferencia de proporciones tipificada (ecuación 3) que acabamos de aplicar. También se ubica en este grupo el índice $d_{asen}$, basado en la transformación arco seno de las proporciones según: $d_{asen} = 2\text{arcsen}\sqrt{p_E} - 2\text{arcsen}\sqrt{p_C}$ (Cohen, 1988). Este índice presenta los mismos problemas de sesgo negativo que $d_p$, provocados por el coste de la dicotomización (Sánchez-Meca, Marín-Martínez y Chacón-Moscoso, 2003).

Frente a estos dos índices segados se han propuesto otros cuatro que persiguen corregir tal sesgo: $d_{Probit}$, basado en la transformación probit de las proporciones (Glass, McGaw y Smith, 1981), $d_{HH} = L_{OR}/1{,}81$ (Hasselblad y Hedges, 1995) y $d_{Cox} = L_{OR}/1{,}65$ (Cox, 1970; Haddock, Rindskopf y Shadish, 1998), basados en diferentes transformaciones de un odds ratio (*OR*), y $d_{bis}$, basado en la transformación a una diferencia media tipificada del coeficiente *biserial phi* ideado por Thorndike (1949, 1982; véase también Becker y Thorndike, 1988). Siendo el comportamiento de todos estos índices relativamente similar, y a los efectos del propósito de este trabajo centrado en los efectos del coste de la dicotomización, nos centramos en uno de los más conocidos, $d_{Probit}$, que se obtiene según:

$$d_{Probit} = \Phi^{-1}(p_E) - \Phi^{-1}(p_C), \qquad (5)$$

donde $p_E$ y $p_C$ son las proporciones de éxito de los grupos experimental y control, y $\Phi^{-1}$ es la inversa de la función de distribución normal.

En un reciente estudio de simulación (Sánchez-Meca, Marín-Martínez y Chacón-Moscoso, 2003), encontramos que los índices sesgados $d_p$ y $d_{asen}$ infraestimaban sistemáticamente al parámetro en todas las condiciones estudiadas. Por su parte, los índices que pretenden corregir tal sesgo negativo, $d_{Probit}$, $d_{Cox}$, $d_{HH}$ y $d_{bis}$, se comportaron como prácticamente insesgados o con un sesgo mínimo y en cualquier caso inferior al de los índices $d_p$ y $d_{asen}$. Sin embargo, un problema de los índices corregidos es que se comportaban como más variables o menos eficientes en torno al verdadero valor del parámetro δ (diferencia media tipificada en la población), en comparación con los índices sesgados. De tal forma que los índices en principio más adecuados por su carencia de sesgo resultaban ser menos eficientes en su estimación que los índices sesgados. Y la pregunta entonces es obligada: ¿merece la pena emplear los índices corregidos si con ello perdemos eficiencia?

Con el doble propósito de ilustrar este problema y aventurar una primera respuesta, simulamos dos poblaciones normales $N(0,5;1)$ y $N(0;1)$ con una diferencia media tipificada de $\delta = (0,5\text{-}0)/1 = 0,5$, de las que extrajimos dos muestras de 50 puntuaciones cuantitativas, una de cada población, dicotomizadas posteriormente a partir del punto de corte $Y_C = 0,25$, que equidista de las dos medias en ambas poblaciones. Hicimos 100 réplicas de este muestreo y en cada una estimamos δ a partir del índice sesgado $d_p$ (ecuación 3) y del índice corregido $d_{Probit}$ (ecuación 5), ambos obtenidos con las puntuaciones dicotomizadas. Como punto de referencia también calculamos la diferencia media tipificada con las puntuaciones cuantitativas (sin dicotomizar) en cada una de las 100 réplicas, a través del índice $d$ de Hedges (Hedges y Olkin, 1985), que es el resultado de corregir un mínimo sesgo positivo que presenta el índice $g$ (ecuación 1), multiplicándolo por el factor de corrección $c(m)$ según:

$$d = c(m)g \qquad (6)$$

donde $c(m) = 1 - 3/(4N-1)$, siendo $N = n_E + n_C$ el tamaño muestral total del estudio.

INSERTAR LA FIGURA 1 APROXIMADAMENTE AQUÍ, POR FAVOR

En la Figura 1(a) se presentan las 100 estimaciones con cada uno de los tres índices en torno al verdadero valor del parámetro δ = 0,5, así como los valores del sesgo y la varianza estimados para cada índice. Calculamos el sesgo de cada uno de los tres estimadores como la diferencia entre la media aritmética de las 100 estimaciones y el valor verdadero δ = 0,5. La variabilidad de cada índice se evaluó a través de la media aritmética de las diferencias cuadráticas entre las 100 estimaciones correspondientes y el valor δ = 0,5 (promedio de errores cuadráticos en las estimaciones). Obsérvese que el índice $d$ basado en toda la información cuantitativa se muestra casi insesgado (sesgo = 0,002) y es el más eficiente de los tres (varianza = 0,043). Por su parte y como era de esperar, el índice $d_p$ se muestra negativamente sesgado (sesgo = – 0,090), como consecuencia del coste de la dicotomización. El índice $d_{Probit}$ , con un sesgo de 0,007, no plantea los problemas de un sesgo negativo debidos a la dicotomización, pero sin embargo muestra la mayor variabilidad en torno a δ (varianza = 0,069), lo que en la práctica supone que con este índice se cometen los errores de mayor magnitud en las estimaciones. En resumen, con el índice $d_{Probit}$ conseguimos eliminar el sesgo negativo pero aumentamos la magnitud del error global en las estimaciones, de ahí que no parezca ser la solución más adecuada a los problemas de sesgo del índice $d_p$, al menos en las condiciones probadas.

En este punto nos planteamos que una forma de conseguir que el índice $d_{Probit}$ fuese también más eficiente que $d_p$ sería la de aumentar el tamaño muestral de los estudios simulados. Volvimos a repetir la misma simulación modificando el tamaño muestral y nos encontramos que conforme éste aumentaba, disminuía la variabilidad de ambos índices hasta llegar a un punto en que la variabilidad de $d_{Probit}$ pasaba a ser inferior a la de $d_p$, invirtiéndose el orden en eficiencia. Esto sucedió a partir de un tamaño muestral de 200 puntuaciones por grupo. La Figura 1(b) muestra los valores del sesgo y eficiencia en esta nueva condición ($n_E = n_C = 200$). Observe cómo en comparación con los resultados de la Figura 1, el aumento en tamaño muestral ha reducido la variabilidad de los tres índices, hasta el punto de que el índice $d_{Probit}$ pasa a tener menos variabilidad que el índice $d_p$ ($d_{Probit}$ con varianza 0,016 inferior a la de $d_p$: 0,019).

Podemos concluir que el empleo de los índices que corrigen el sesgo negativo provocado por la dicotomización tan sólo es adecuado en condiciones de alta eficiencia o baja variabilidad en los estimadores. Una primera estrategia para aumentar la eficiencia en las estimaciones es el incremento del tamaño muestral de los estudios, con el inconveniente de que éste no es siempre factible debido a las limitaciones de presupuesto, tiempo y de disponibilidad de sujetos, propias de la investigación en Ciencias Sociales y del Comportamiento. Existe, no obstante, una segunda estrategia mucho más factible que la anterior, consistente en hacer un meta-análisis integrando los resultados de diferentes investigaciones o estudios. De hecho, el tamaño del efecto promedio de un meta-análisis es necesariamente menos variable o más eficiente que los tamaños del efecto individuales de cada estudio, debido a que aquél se basa en una mayor evidencia o tamaño muestral, al tomar en consideración todos los estudios revisados.

El objetivo de este trabajo es comparar mediante un estudio de simulación Monte Carlo el comportamiento de los dos índices $d_p$ y $d_{Probit}$ para variables dependientes dicotomizadas, y del índice $d$ de Hedges para variables dependientes cuantitativas, cuando se aplican en un meta-análisis con el propósito de estimar el tamaño del efecto en la población. Esperamos que mediante la metodología meta-analítica se consigan las condiciones de alta eficiencia necesarias para que el índice $d_{Probit}$ se comporte como un índice insesgado y más eficiente que el índice $d_p$, minimizándose de esta forma el coste derivado de la dicotomización.

## 3. MÉTODO

El estudio de simulación se programó en Gauss (Aptech Systems, 1992). Se definieron dos poblaciones normalmente distribuidas con varianzas homogéneas, $N(\mu_E\,;\,\sigma)$ y $N(\mu_C;\,\sigma)$, donde $\mu_E$ y $\mu_C$ son las medias de la poblaciones experimental y control, respectivamente, y $\sigma$ es la desviación típica común. La diferencia media tipificada en la población, $\delta$, se definió como: $\delta = (\mu_E - \mu_C) / \sigma$. Fijamos los valores de la desviación típica para ambas poblaciones y de la media en la población control, siendo $\sigma = 1$ y $\mu_C = 0$ en todos los casos. De tal forma que el valor de la diferencia media tipificada siempre coincidió con el de la media en la población experimental: $\mu_E = \delta$. Asimismo se definió un punto de corte equidistante entre las dos medias de ambas poblaciones que posibilitó la dicotomización. A partir de ambas poblaciones se generaron $k$ pares de muestras aleatorias independientes de tamaños $n_E = n_C = 20$ en todos los casos, que simulaban un meta-análisis de $k$ estudios. Los datos se expresaron en dos métricas: puntuaciones cuantitativas y puntuaciones dicotomizadas a partir del punto de corte. Se calculó el índice $d$ de Hedges (ecuación 6) sobre las puntuaciones cuantitativas y los índices $d_{Probit}$ (ecuación 5) y $d_p$ (ecuación 3) sobre las puntuaciones dicotomizadas (tablas 2x2). Finalmente se obtuvo la media aritmética de los $k$ valores $d$, $d_{Probit}$ y $d_p$, obtenidos en los $k$ estudios del meta-análisis.

Se manipularon la magnitud de la diferencia media tipificada en la población, con los valores $\delta = 0{,}2$, 0,5 y 0,8, que representan una magnitud típicamente baja, media y alta, respectivamente, en las Ciencias Sociales (Cohen, 1988); y el número de estudios, $k$, con los valores $k = 1, 2, 3, 4, 5, 6, 7, 8, 9$

y 10. Cruzando ambos factores se obtuvieron un total de 3 (valores δ) x 10 (valores $k$) = 30 condiciones, de las que se generaron 10,000 réplicas. En consecuencia, se simularon un total de 30 x 10,000 = 30,000 meta-análisis.

Puesto que el valor de δ fue constante para todos los estudios de un meta-análisis, se asumía un modelo de efectos fijos (Hedges y Vevea, 1998; Hunter y Schmidt, 2000). Siendo además el tamaño muestral constante en todos los estudios de cada meta-análisis ($n_E = n_C = 20$), el promedio de los tamaños del efecto a través de los $k$ estudios pudo definirse como una simple media aritmética de los $k$ valores $d$, $d_{Probit}$ ó $d_p$, representados respectivamente como: $\bar{d}, \bar{d}_{Probit}$ y $\bar{d}_p$ . Estos promedios actuaron como estimadores de δ, calculándose el sesgo como la diferencia entre la media de las 10000 estimaciones y el valor de la diferencia media tipificada en la población, δ. También se calculó la varianza de cada estimador, como la media de los errores cuadráticos respecto de δ obtenidos a través de las 10000 réplicas.

## 4. RESULTADOS Y DISCUSIÓN

La Figura 2(a) muestra los valores del sesgo y la varianza cuando el tamaño del efecto en la población fue de una magnitud elevada (δ = 0,8), presentados en función del número de estudios del meta-análisis. Se observa que el aumento del número de estudios en el meta-análisis no afecta al sesgo de los estimadores, pero sí afecta notablemente a su variabilidad, que disminuyó claramente en los tres estimadores $\bar{d}, \bar{d}_{Probit}$ y $\bar{d}_p$. Esto es debido a que cuando acumulamos la evidencia de una serie de estudios a través de un meta-análisis, no alteramos la naturaleza más o menos sesgada de los estimadores. De hecho, si promediamos los valores de un índice negativamente sesgado a través de los $k$ estudios de un meta-análisis, es de esperar que el nuevo estimador promedio conserve el sesgo negativo de los valores individuales. No sucede lo mismo con la varianza, puesto que los estimadores promedio tienen siempre menos variabilidad que los tamaños del efecto individuales de cada estudio sobre los que se obtienen. Así, conforme aumenta el número de estudios y, en consecuencia, la cantidad de tamaños del efecto que entran a formar parte del estimador promedio, disminuye necesariamente su variabilidad.

<u>INSERTAR LA FIGURA 2 APROXIMADAMENTE AQUÍ, POR FAVOR</u>

Los resultados del sesgo muestran al índice $\bar{d}$ de Hedges como prácticamente insesgado, al índice $\bar{d}_p$ como negativamente sesgado y al índice $\bar{d}_{Probit}$ con un sesgo ligeramente positivo, aunque de menor magnitud que en el caso de $\bar{d}_p$ . La ausencia de sesgo del índice *d* de Hedges es un hecho ampliamente conocido y corroborado a través de la literatura (Hedges y Olkin, 1985; Sánchez-Meca y Marín-Martínez, 1998). Incluimos este índice como un punto de referencia que nos permitiese evaluar el coste de la dicotomización en meta-análisis, ya que a diferencia de los otros dos índices, que se obtienen a partir de las puntuaciones dicotomizadas, la *d* de Hedges se calcula sobre las puntuaciones cuantitativas originales. Por su parte, el índice $\bar{d}_p$ (véase Figura 2(a)) muestra un sesgo negativo de una magnitud en torno a 0,1, que supone un 12,5% de infraestimación en relación a la magnitud del parámetro $\delta = 0{,}8$, y que es consecuencia directa del coste de la dicotomización.

No nos esperábamos que el índice $\bar{d}_{Probit}$ , específicamente diseñado para corregir el sesgo provocado por la dicotomización, mostrase un ligero sesgo positivo. Tal sesgo fue debido al hecho de que en nuestra simulación los tamaños muestrales de todos los estudios fuesen de una magnitud relativamente baja, 20 unidades por grupo, y constante. Este tamaño muestral resulta todavía escaso para que $d_{Probit}$ estime de forma insesgada la diferencia media tipificada a partir de los resultados de un estudio. En simulaciones adicionales donde los estudios de un mismo meta-análisis presentaban diferentes tamaños muestrales, de 10, 20, 30 y 40 unidades por grupo, como por otra parte suele ocurrir en la realidad, nos encontramos con que $\bar{d}_{Probit}$ funcionó como un estimador casi insesgado. Esto fue debido a que en meta-análisis, cuando se obtiene un tamaño del efecto medio, se da más peso a los resultados de los estudios con un mayor tamaño muestral, en los que $d_{Probit}$ se muestra prácticamente insesgado.

Los resultados de la varianza (Figura 2(a)) muestran que cuando el meta-análisis tiene pocos estudios (4 ó menos), el índice más eficiente o menos variable es la $\bar{d}$ de Hedges, seguida del índice sesgado $\bar{d}_p$ y del índice corregido $\bar{d}_{Probit}$ . Sin embargo, a partir de 5 estudios el índice $\bar{d}_{Probit}$ pasa a ser menos variable que $\bar{d}_p$ . Esto significa que cuando el tamaño del efecto en la población es de magnitud alta ($\delta = 0{,}8$), se necesita un mínimo de 5 estudios (de unas 40 unidades cada uno) para conseguir que el promedio $\bar{d}_{Probit}$ sea más eficiente que el promedio $\bar{d}_p$ .

La Figura 2(b) muestra los valores del sesgo y la varianza cuando el tamaño del efecto en la población fue de una magnitud media ($\delta = 0{,}5$), presentados en función del número de estudios del meta-análisis. Los resultados del sesgo son similares a los del caso anterior salvo en el hecho de que presentan una magnitud inferior en términos absolutos. Los resultados de la varianza también muestran la misma tendencia que en la Figura 3, salvo en que ahora se necesitan más estudios para conseguir que el índice $\bar{d}_{Probit}$ sea más eficiente que $\bar{d}_p$ : un mínimo de 10 estudios de unas 40 unidades cada uno.

Finalmente, la Figura 2(c) también muestra los valores del sesgo y la varianza para un tamaño del efecto de magnitud baja ($\delta = 0,2$) en función del número de estudios del meta-análisis. Nuevamente el que disminuya el tamaño del efecto en la población reduce la magnitud el sesgo en los estimadores, y en el caso de la varianza, plantea mayores exigencias en cuanto al número de estudios necesario para conseguir que $\bar{d}_{Probit}$ sea más eficiente que $\bar{d}_p$. Obsérvese que ni siquiera con 10 estudios se consigue tal objetivo, precisándose al menos 25 estudios para lograrlo.

## 5. CONCLUSIONES

La estimación del tamaño del efecto a partir de una variable dependiente dicotomizada se asocia a un sesgo negativo (coste de la dicotomización) que se debe a la pérdida de información que conlleva simplificar toda una serie de valores cuantitativos en tan sólo dos niveles. Así, la estimación de una diferencia media tipificada poblacional, $\delta$, a partir de una diferencia de proporciones tipificada, $d_p$, nos conduce a infraestimar sistemáticamente el verdadero valor de $\delta$.

Recientemente se han ideado nuevos índices que corrigen tal sesgo ($d_{Probit}$, $d_{HH}$, $d_{Cox}$ y $d_{bis}$) aunque a costa de una pérdida en eficiencia que no los hace del todo adecuados. De hecho, tales índices resultan ser menos eficientes que el propio índice sesgado $d_p$, con lo que su aplicación en la práctica se asocia a errores de mayor magnitud en la estimación en comparación con $d_p$.

Tan sólo trabajando en condiciones de alta eficiencia o baja variabilidad se consigue que los índices que corrigen el sesgo de la dicotomización, como $d_{Probit}$, funcionen también como más eficientes que el índice tradicional $d_p$. Esta alta eficiencia puede conseguirse aumentando el tamaño muestral de los estudios, pero el elevado número de unidades necesarias va a ser generalmente prohibitivo para la investigación psicológica: un mínimo de 400 unidades por estudio cuando el tamaño del efecto en la población es de magnitud media ($\delta = 0,5$), o de 1000 unidades cuando el tamaño del efecto es bajo ($\delta = 0,2$).

Otra alternativa para conseguir estimaciones altamente eficientes, es la integración en un meta-análisis de los estudios que investigan un mismo efecto. De hecho, el promedio de los tamaños del efecto a través de todos los estudios, se comporta como un estimador tanto más eficiente cuanto mayor sea el número de estudios integrados. Así, siendo el tamaño del efecto en la población de magnitud intermedia ($\delta = 0,5$), conseguimos con un mínimo de 5 estudios (de unas 40 unidades cada uno) que el promedio de los índices $d_{Probit}$ comience a ser un estimador más eficiente que el promedio de los índices sesgados $d_p$.

En resumen, la integración de los índices que corrigen el sesgo derivado de la dicotomización ($d_{Probit}$, $d_{HH}$, $d_{Cox}$ y $d_{bis}$) en un meta-análisis, posibilita una estimación insesgada y más eficiente que la de los índices tradicionales ($d_p$ y $d_{asen}$), siempre y cuando se disponga de un número mínimo de estudios que deberá aumentar conforme disminuyan la magnitud del tamaño del efecto poblacional y/o el tamaño muestral de los estudios.

## REFERENCIAS

Aptech Systems. (1992). *The Gauss system (Version 3.0)* [Software informático]. Kent, WA: Author.

Becker, G. y Thorndike, R.L. (1988). The biserial-phi correlation coefficient. *Journal of Psychology, 122,* 523-526.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). Hillsdale, NJ: Erlbaum.

Cox, D.R. (1970). *Analysis of binary data.* New York: Chapman & Hall/CRC.

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

Haddock, C.K.; Rindskopf, D. y Shadish, W.R. (1998). Using odss ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3,* 339-353.

Hasselblad, V. y Hedges, L.V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117,* 167-178.

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

Hedges, L.V. y Vevea, J.L. (1998). Fixed- and Random-Effects models in meta-analysis. *Psychological Methods, 3,* 486-504.

Hunter, J.E. y Schmidt, F.L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75,* 334-349.

Hunter, J.E. y Schmidt, F.L. (2000). Fixed Effects vs Random Effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*, 275-292.

Johnson, B.T. (1989). *DSTAT: Software for the meta-analysis review of research* [Computer software and manual]. Hillsdale, NJ: Erlbaum.

Lipsey, M.W. y Wilson, D.B. (2001). *Practical meta-analysis.* Thousands Oaks, CA: Sage.

Sánchez-Meca, J. y Marín-Martínez, F. (1998). Weighting by inverse variance or by sample size in meta-analysis: A simulation study. *Educational and Psychological Measurement, 58,* 211-220.

Sánchez-Meca, J.; Marín-Martínez, F. y Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8,* 448-467.

Thorndike, R.L (1949). *Personnel selection*. New York: Willey.

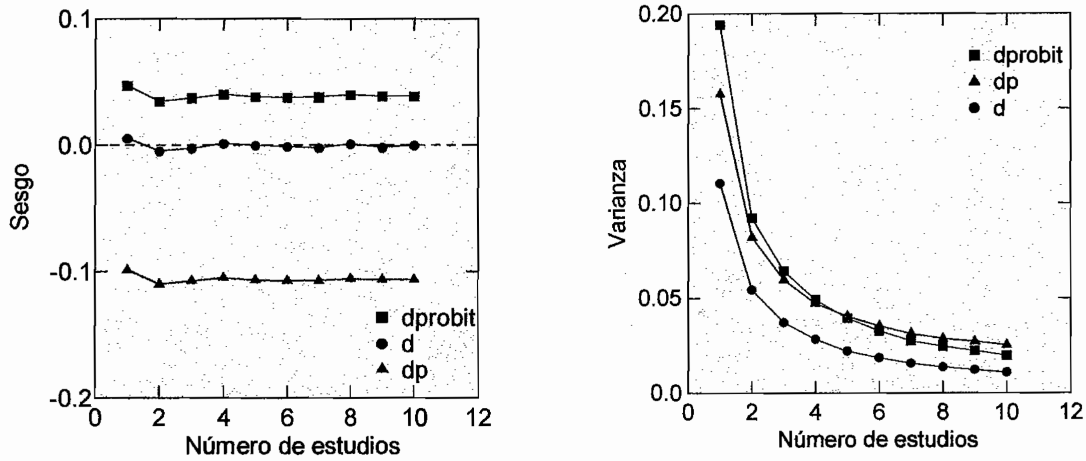
Thorndike, R.L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin.

**TABLAS**

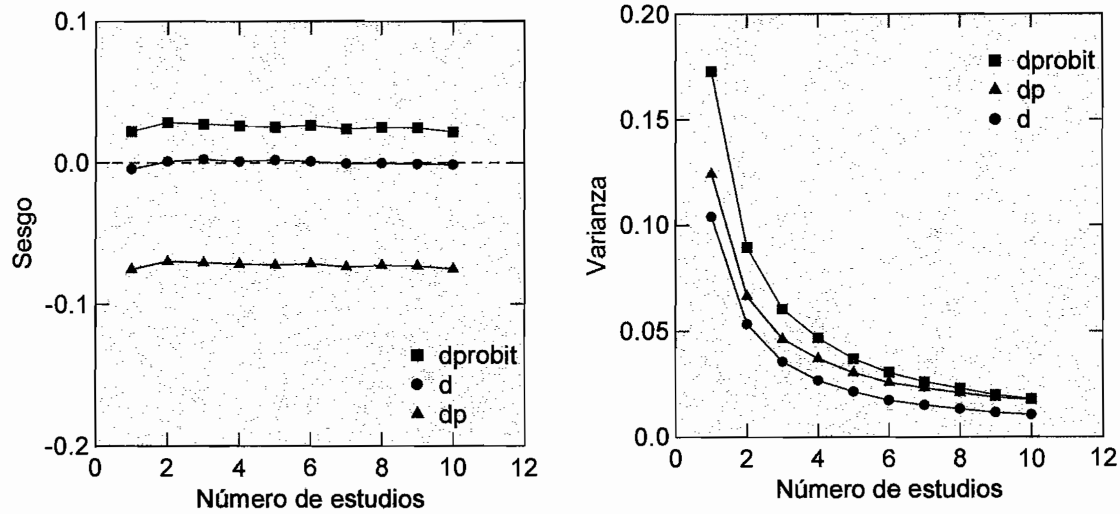
Tabla 1. Resultados de un muestreo aleatorio de dos poblaciones normales, N(42,4) y N(40,4), antes y después de la dicotomización: δ = (42-40)/4 = 0.5.

| *Antes de la dicotomización* | *Grupos* | |
|---|---|---|
| | Experimental | Control |
| Tamaños muestrales | 50 | 50 |
| Medias | 41,796 | 39,587 |
| Desviaciones típicas | 4,468 | 3,877 |
| Estimación de δ: $g = (\bar{Y}_E - \bar{Y}_C)/S = (41{,}796 - 39.587)/4{,}183 = 0{,}528$ | | |
| *Después de la dicotomización* | *Grupos* | |
| Tabla de frecuencias | Experimental | Control |
| Puntuaciones ≥ 41 | 25 | 15 |
| Puntuaciones < 41 | 25 | 35 |
| Totales | 50 | 50 |
| Estimación de δ: $d_p = (p_E - p_C)/S' = (25/50 - 15/50)/0{,}479 = 0{,}417$ | | |

(a)




(b)




(c)

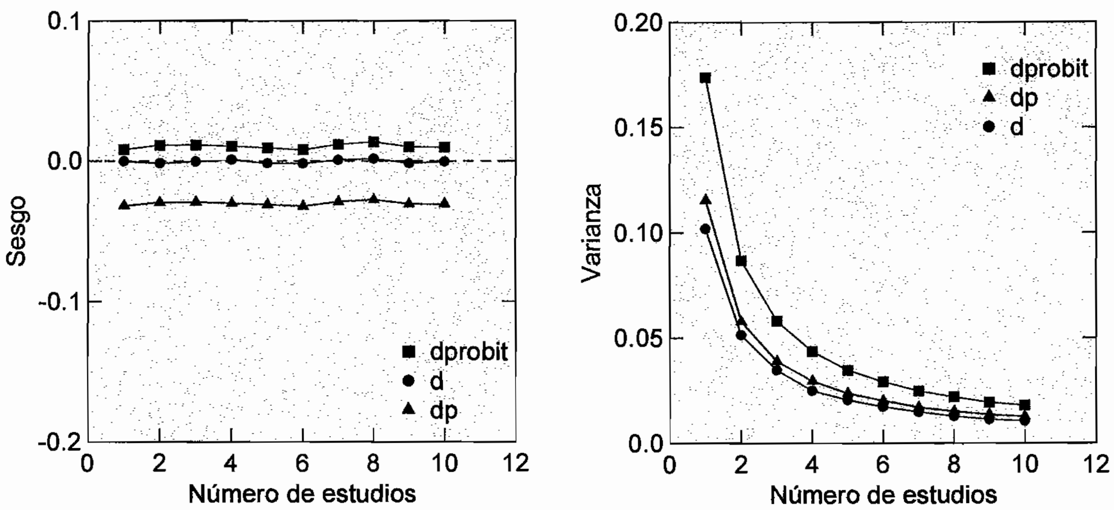



Figura 2. Sesgo y varianza de los tres estimadores, $\bar{d}, \bar{d}_{Probit}$ y $\bar{d}_p$, para un tamaño del efecto: (a) alto ($\delta = 0.8$); (b) medio ($\delta = 0.5$) y (c) bajo ($\delta = 0.2$) y en función del número de estudios.