



# MÁSTERES de la UAM

Facultad de Psicología  
/ 15-16

Metodología  
de las Ciencias  
del Comportamiento  
y de la Salud



excelencia Campus Internacional  
UAM  
CSIC+



**Alcance  
del meta-análisis  
restringido  
a estudios  
significativos: un  
ejemplo práctico**  
*Desirée Blázquez  
Rincón*



**FACULTAD DE PSICOLOGÍA**  
**Departamento de Psicología Social y Metodología**

# Alcance del meta-análisis restringido a estudios significativos: un ejemplo práctico

Trabajo Fin de Máster  
Máster en Metodología de las Ciencias del Comportamiento y de la Salud

Autora: Desirée Blázquez Rincón  
Tutores: Juan Botella y Manuel Suero Suñe  
Madrid, junio 2016



## ÍNDICE

<b>Resumen</b>	4
<b>1 Introducción: El Sesgo de Publicación</b>	5
<i>1.1 La prevalencia de los resultados estadísticamente significativos</i>	5
<i>1.2 El sesgo de publicación en meta-análisis</i>	6
<b>2 Una estrategia alternativa: meta-análisis con los resultados significativos</b>	8
<i>2.1 Distribución uniforme de los valores <math>p</math></i>	8
<i>2.2 Contraste de hipótesis en meta-análisis con la distribución de <math>p</math></i>	10
<i>2.3 Estimación del efecto poblacional con la distribución de Irwin-Hall</i>	12
<b>3 La presente investigación: Simulaciones</b>	14
<b>4 Estudio 1: Desempeño en condiciones óptimas</b>	15
<i>4.1 Método</i>	15
<i>4.2 Resultados y Discusión</i>	17
<b>5 Estudio 2: Desempeño en condiciones con sesgo de publicación</b>	17
<i>5.1 Método</i>	18
<i>5.2 Resultados y Discusión</i>	20
<b>6 Estudio 3: Desempeño del Procedimiento basado en los valores <math>p</math> significativos</b>	20
<i>6.1 Método</i>	23
<i>6.2 Resultados y Discusión</i>	24
<b>7 Una aplicación práctica: Meta-análisis sobre el efecto del <i>Ego-Depletion</i></b>	27
<i>7.1 El Ego Depletion en meta-análisis: La amenaza del Sesgo de Publicación</i>	28
<i>7.2 Nueva evidencia basada en los estudios estadísticamente significativos.</i>	29
<b>Referencias bibliográficas</b>	34
<b>Apéndice</b>	38

## RESUMEN

Con el propósito de hacer frente a las conclusiones sobrestimadas, fruto del sesgo de publicación a favor de los estudios con resultados estadísticamente significativos, se han desarrollado nuevos procedimientos estadísticos que permiten al meta-analista trabajar sólo con la muestra de resultados significativos a partir de los valores  $p$  asociados a los estadísticos de contraste. La presente investigación tiene como objetivo comparar el rendimiento de estas nuevas técnicas, empleando la distribución de Irwin-Hall, frente a los modelos convencionales de Efecto Fijo y Efectos Aleatorios aplicados a todos los estudios localizados. Tres estudios de simulación llevados a cabo por el método Monte Carlo, sirvieron para constatar la ausencia de diferencias, en términos de sesgo y eficiencia, entre las estimaciones de estas tres técnicas bajo condiciones óptimas (estudio 1), y bajo distintos niveles de sesgo de publicación (estudio 2). Sin embargo, al emplear exclusivamente los resultados significativos (estudio 3), esta nueva técnica aporta estimaciones menos sesgadas, aunque menos eficientes, que los métodos tradicionales. Finalmente, se ha realizado una aplicación práctica sobre dos meta-análisis previos que bajo los métodos tradicionales alcanzaron resultados muy dispares, mientras que con esta nueva técnica se han obtenido estimaciones casi idénticas.

## **1 Introducción: El Sesgo de Publicación**

Todos los campos de investigación que emplean la inferencia estadística confían en la replicación como árbitro para decidir qué efectos seguir manteniendo y cuales poner en tela de juicio. No obstante, la capacidad de la replicación se ve mermada a causa de la tendencia conocida como sesgo de publicación, por la que no todos los estudios tienen la misma probabilidad de ser publicados. La principal consecuencia del sesgo de publicación es que la investigación publicada no es representativa del efecto que se persigue conocer.

La distancia que separa a un artículo de su publicación en una revista científica se puede ver influida por diversos factores, entre ellos el diseño del estudio, si es experimental u observacional, si la fuente de financiación es una empresa privada o gubernamental, quiénes sean los autores, el género de los mismos, su país de origen, el tamaño de la muestra, el número de estudios que incluye o sus conclusiones. Scherer y Langenberg (2007) comprobaron que en su mayoría se publican más fácilmente aquellos manuscritos referentes a la investigación básica, que cuentan con un diseño aleatorio y favorecen al tratamiento experimental, habiendo sido presentados con anterioridad en congresos y conferencias.

Uno de los principales motivos por los que un manuscrito alcanza más rápidamente la meta de la publicación es el sesgo a favor de los resultados estadísticamente significativos (Francis, 2012). Como fruto de esta predisposición, la literatura científica disponible concluye con una idea sobrestimada de los efectos poblacionales reales (Smith, 1980), llegando incluso a encontrar falsos positivos de efectos verdaderamente nulos. El objetivo del presente trabajo es abordar el problema del sesgo de publicación supeditado a la comunicación de los resultados estadísticamente significativos desde una perspectiva meta-analítica, analizando nuevas formas de atajar los impedimentos que representa.

### *1.1 La prevalencia de los resultados estadísticamente significativos*

Abundante investigación apoya la convicción de que tanto la fuerza como la dirección de los resultados del estudio influyen en la elección de la revista por parte del autor y en la probabilidad de aceptación. Cooper, DeNeve y Charlton (1997) encontraron que de 117 estudios aprobados por el comité ético del Departamento de Psicología de la Universidad de Missouri, el 74% de los que obtuvieron resultados estadísticamente significativos fueron enviados para su publicación en libros o revistas científicas, comparado con el 4% de los estudios que no obtuvieron resultados estadísticamente significativos. Al mismo tiempo, los manuscritos con resultados significativos se publican en revistas con un índice de impacto

mayor (Smart, 1964; Tierney, Clarke y Stewart, 2000), como ocurre con las revistas no especializadas (Ioannidis, Cappelleri, Sacks y Lau, 1997). Finalmente, cabe mencionar que los estudios financiados con fondos privados tienen una mayor probabilidad de obtener resultados estadísticamente significativos que la de los estudios financiados gracias a otro tipo de fondos (Yaphe, Edman, Knishkowy y Herman, 2001).

Ante esta situación, los investigadores tienen menos incentivos para trabajar en los manuscritos con resultados nulos. Y, ya sea porque tardan más en ser enviados o porque directamente no se envían, el resultado final es que los estudios estadísticamente significativos aparecen en la literatura antes que los estudios con resultados no significativos, de forma que la literatura científica tenderá a sobrestimar los verdaderos efectos poblacionales. Esta comprensión errónea de los fenómenos estudiados trae consigo repercusiones que, como han puesto de manifiesto Rothstein, Sutton y Borenstein (2005), abarcan desde la inversión en programas de investigación no fructíferos y la promoción de cambios sociopolíticos erróneamente fundamentados, hasta el impulso de terapias poco o nada efectivas que pueden llegar a poner en peligro la calidad de vidas humanas.

### *1.2 El sesgo de publicación en meta-análisis*

Para que los resultados de un meta-análisis sean válidos, se requiere que la literatura relevante para el fenómeno contenga todos los estudios o una muestra representativa de los mismos, independientemente de la significación estadística de los resultados y su dirección. De esta manera, la muestra de estimaciones del tamaño del efecto con que contará el meta-analista será aleatoria, y por tanto, representativa del parámetro correspondiente, donde los resultados no significativos son una consecuencia esperada del muestreo aleatorio.

No obstante, el sesgo a favor de la publicación de los resultados significativos se ha convertido en la principal amenaza para las conclusiones de la investigación en ciencias sociales. La inclusión de la literatura no controlada por las editoriales comerciales, conocida como literatura gris, en los estudios meta-analíticos conduce a una reducción de las estimaciones combinadas de los tamaños del efecto (McAuley, Pham y San Martín, 2000). Ante tal amenaza, se ha invertido mucho esfuerzo tanto en desarrollar métodos eficientes para identificar la literatura no publicada y la literatura gris, como en procedimientos estadísticos para valorar la amenaza del sesgo de publicación e inferir su eventual impacto en los resultados del meta-análisis. En lo referente a este último empeño, existen distintos análisis estadísticos que se recomienda llevar a cabo de manera conjunta.

Con el objetivo de ordenar visualmente los datos, las figuras conocidas como *forest plot* y *funnel plot* permiten hacerse una primera idea acerca de la variabilidad de los tamaños del efecto, resaltando los valores atípicos que requieren atención. Si existe mayor concentración de efectos a un lado del valor del efecto medio que al otro, el *funnel plot* nos está indicando que los estudios con muestras más pequeñas se han publicado en mayor medida en tanto que cumplían el criterio de significación estadística.

Por su parte, Begg y Mazumdar (1994) propusieron una prueba de correlación por rangos para informar de la presencia de sesgo en el caso de hallar una relación no nula entre los tamaños del efecto y sus varianzas. Una prueba más potente es el método de regresión de Egger, que ajusta una ecuación de regresión para los efectos estandarizados en función del inverso de su error típico mediante el método de mínimos cuadrados, indicando la existencia de sesgo de publicación si la ordenada en el origen difiere significativamente de cero.

Los números de seguridad (Orwin, 1983; Rosenberg, 2005; Rosenthal, 1979) nos informan del número de estudios no localizados o no publicados con efecto medio nulo que necesitaríamos añadir a nuestro meta-análisis para que la estimación combinada del tamaño del efecto dejara de ser estadísticamente significativa.

Duval y Tweedie (2000) desarrollaron el procedimiento conocido como *trim and fill*, por el que se recalcula de manera iterativa la estimación del efecto medio hasta que el *funnel plot* se vuelve simétrico. Después, se añaden los estudios originales, anteriormente excluidos por ser atípicos, y su imagen especular, para que la nueva estimación combinada del efecto tenga un mayor intervalo de confianza.

Finalmente, otra opción es realizar un meta-análisis acumulativo, calculando el efecto medio a partir de ir añadiendo estudio por estudio a los análisis, desde el de mayor muestra al de menor. Si el valor estimado se estabiliza y no varía con la adición de los estudios pequeños, entonces no hay razón para pensar que estos últimos han introducido un sesgo en el resultado.

Todas las técnicas citadas anteriormente nos informan de si las conclusiones de un estudio meta-analítico podrían estar influidas o no por el sesgo de publicación. Pero ante una respuesta afirmativa, la única alternativa que nos presentan es imputar valores no reales para generar un efecto medio igualmente irreal. En este contexto y con el objetivo de acercarnos a una estimación lo más próxima posible al verdadero efecto poblacional a partir de las restricciones que nos supone el sesgo de publicación, nacen técnicas alternativas basadas en la distribución de los valores  $p$  asociados a los estadísticos de contraste de los distintos estudios incluidos en el meta-análisis.

## 2 Una estrategia alternativa: meta-análisis con los resultados significativos

Cuando realizamos un meta-análisis partimos de la idea de que tenemos una muestra representativa de estimaciones del verdadero tamaño del efecto. A partir de ahí, aplicamos principalmente dos modelos estadísticos que funcionan bien bajo este supuesto: el modelo de Efecto Fijo y el modelo de Efectos Aleatorios. No obstante, como hemos visto, la muestra de tamaños del efecto puede estar sesgada, en tanto que los estudios con resultados significativos tienen más probabilidad de ser publicados que aquellos con resultados no significativos. Ante esta perspectiva, en la que la actual distribución de los tamaños del efecto queda parcialmente representada, el efecto medio que resulte de los procedimientos convencionales estará igualmente sesgado. Nos vemos entonces en la necesidad de trabajar con nuevos modelos estadísticos que nos permitan obtener una estimación fiable del efecto poblacional medio a partir de la muestra de efectos de la que realmente disponemos.

Bajo el supuesto de que todos los estudios con resultados significativos tienen la misma probabilidad de ser publicados, las pérdidas de éstos son mínimas y se pueden asumir aleatorias. De esta forma la muestra de valores significativos publicados es más representativa de ese segmento de la distribución muestral de estudios hipotéticos. En los últimos años se han retomado modelos estadísticos que permiten al meta-analista trabajar exclusivamente con la muestra de resultados significativos. La idea básica que comparten estas nuevas aproximaciones es que la distribución de los valores  $p$  asociados a los tamaños del efecto es uniforme si la hipótesis contrastada es verdadera, y por ende, también será uniforme la distribución de los valores  $p$  que en los contrastes de nulidad serían estadísticamente significativos.

### 2.1 Distribución uniforme de los valores $p$

Como veremos a continuación, los valores  $p$  siguen una distribución uniforme  $U(0; 1)$  al tratarse de la probabilidad complementaria a la acumulada asociada a los valores de una v.a.c.. Si reunimos distintas estimaciones del tamaño del efecto representado en un parámetro, y obtenemos de ellos sus respectivos estadísticos de contraste respecto al valor de ese mismo parámetro, los valores críticos o valores  $p$  asociados a dichos estadísticos se distribuirán de manera uniforme.

Aunque se han desarrollado diferentes demostraciones para estudiar la distribución de los valores  $p$  (Bhattacharya y Habtzghi, 2002; Donahue, 1999; Hung, O'Neill, Bauer y Kohne, 1997), en el presente trabajo vamos a seguir la demostración propuesta por Suero y

Botella (2015). Como demuestran estos autores, la distribución de densidad de los valores críticos es uniforme siempre y cuando la hipótesis confrontada en los estadísticos de contraste sea verdadera. Esto quiere decir que la distribución de los valores  $p$  será uniforme siempre que éstos sean resultado de la función primitiva de la densidad de sus correspondientes estadísticos de contraste.

La demostración se asienta en entender los valores  $p$  como la transformación de una v.a.c.. Para entender la demostración, partamos de la idea de que un estadístico de contraste es una v.a.c., a la que llamaremos  $X$ , que se distribuye según la función de densidad de probabilidad  $f$ . Sirviéndonos del Teorema de la Transformación (Gut, 2009), los valores  $p$  se obtienen aplicando una regla de transformación, a la que llamaremos  $G$ , a los valores  $x$  de los estadísticos de contraste, y por lo tanto, son a su vez una variable aleatoria continua a la que llamaremos  $P$ .

Para conseguir la función de densidad de probabilidad de los valores  $p$ ,  $h(p)$ , emplearemos la fórmula del Teorema de la Transformación para variables aleatorias, asumiendo que dicha transformación es derivable e inyectiva. Así pues, la distribución de probabilidad de  $P$  viene dada por el producto de dos factores:

$$h(p) = f(G^{-1}(p)) \cdot \frac{d}{dy} G^{-1}(p) \quad [1]$$

El primer factor hace referencia a la función de densidad de probabilidad de la variable  $X$  con respecto a  $P$ , si obtenemos los valores de  $X$  en función de los valores de  $P$ ;  $G^{-1}(p) = X$ .

El segundo factor hace referencia a la derivada de la inversa de la función con respecto a  $P$ . Ahora bien, se sabe que la derivada de la inversa de una función es igual al recíproco de la derivada de la función (Spivak, 1988):

$$(G^{-1}(p))' = \frac{1}{G'(p)}$$

De esta manera, al interpretar el segundo factor como el recíproco de la derivada de  $f(x)$  con respecto a  $X$ , la función [1] se ve modificada del siguiente modo:

$$h(p) = f(G^{-1}(p)) \cdot \frac{1}{G'(p)}$$

Esta ecuación, donde  $G'(p)$  se puede entender también como  $g(G^{-1}(p))$ , serviría para cualquier transformación inyectiva y derivable, independientemente de la regla que se desee seguir.

Finalmente, si asumimos que la regla  $G$  hace referencia a la función de distribución de la variable  $X$  (estadístico de contraste), las reglas  $G$  y  $F$  se vuelven homólogas, y por consiguiente:

$$h(p) = \frac{f(G^{-1}(p))}{g(G^{-1}(p))} = \frac{f(F^{-1}(p))}{f(F^{-1}(p))}$$

Donde  $p$  está incluido en el intervalo  $(0; 1)$ .

Así queda probado que los valores  $p$  ligados a un estadístico de contraste siguen una distribución uniforme si las estimaciones del tamaño del efecto a partir de los cuales se obtienen estos estadísticos siguen la distribución de la hipótesis contrastada.

## 2.2 Contraste de hipótesis en meta-análisis con la distribución de $p$

Centrémonos por ahora en el contraste de hipótesis sobre la diferencia de medias para muestras independientes y en la diferencia de medias tipificada como índice del tamaño del efecto ( $\delta$ ; Cohen, 1988). El estadístico de contraste en condiciones de normalidad y homocedasticidad (distribuido según la  $t$  de Student) se corresponde con la siguiente fórmula (Pardo, Ruiz y San Martín, 2009).

$$EC = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2} \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

La hipótesis de nulidad o de no efecto establece que  $H_0: \mu_1 - \mu_2 = 0$ . Si el efecto poblacional hipotetizado, en este caso cero, fuera verdadero, los valores de  $p$  asociados a los estadísticos de contraste se distribuirían de manera uniforme  $U(0; 1)$ , según hemos visto en el apartado anterior. Así se puede apreciar en el primer histograma de la figura 1. En él se recogen las frecuencias de diez mil valores de  $p$  generados por el método Monte Carlo bajo hipótesis nula verdadera ( $\delta = \frac{\mu_1 - \mu_2}{\sigma} = 0$ )<sup>1</sup>.

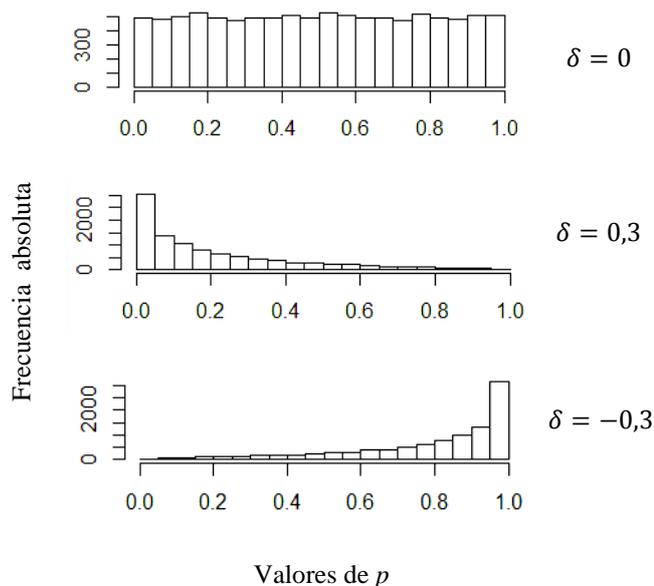


Figura 1. Histogramas de 10.000 valores de  $p$  obtenidos bajo un delta de 0, 0,3 y -0,3.

Cuando el verdadero efecto poblacional es superior al contrastado bajo hipótesis nula, como ocurre en el segundo histograma ( $\delta = 0,3$ ), se observa mayor cantidad o frecuencia de valores  $p$  cercanos a cero (existe mayor frecuencia de contrastes estadísticamente significativos). En cambio, en el tercer histograma ( $\delta = -0,3$ ) podemos observar que cuando el verdadero efecto poblacional es inferior al efecto contrastado, existe mayor frecuencia de valores  $p$  cercanos a uno (recordemos que se trata de contrastes unilaterales derechos).

Por tanto, esta es una forma alternativa de contrastar hipótesis cuando se dispone de una muestra de estudios independientes, como ocurre en el meta-análisis. Para contrastar un valor concreto como el verdadero efecto poblacional, existen pruebas estadísticas que nos informan del grado de desviación de una distribución empírica de valores de  $p$  respecto de la distribución teórica uniforme (Anderson y Darling, 1954; Hedges, 1984; Massey, 1951).

Fisher (1932) propuso una transformación logarítmica de los valores  $p$  que seguían una distribución  $\chi^2$  con 2 grados de libertad, a partir de la cual se pueden obtener dos estadísticos de contraste (van Assen, van Aert y Wicherts, 2014). De manera similar, Stouffer, Suchman, Devinney, Star y Williams (1949) propusieron la transformación inversa normal de los valores  $p$  en puntuaciones típicas, y un nuevo estadístico de contraste, distribuido  $N(0, 1)$ , fruto de su división entre el error típico de una nueva variable, suma de dichas puntuaciones típicas. George y Mudholkar (1977), por su parte, utilizó una transformación *logit* de los valores  $p$ , de tal manera que pudiera comprobarse su desviación respecto de la distribución teórica uniforme gracias a un estadístico que se distribuyera aproximadamente  $t_{5k+4}$ .

1. Si disponemos de una muestra de valores independientes de  $t$ , los valores de  $p$  asociados a ellos deben tender a la distribución uniforme en una sencilla simulación.

Si los estadísticos mencionados ocupan un percentil inferior a  $P_{95}$  en la distribución de referencia, consideraremos que no existe evidencia de que la distribución de los valores  $p$  difiera significativamente de una distribución uniforme  $U(0; 1)$ , y por lo tanto, que el valor de  $\delta$  empleado en los cálculos no es incorrecto (no se rechaza el valor hipotetizado). Ahora bien, además de informar del grado de disparidad existente entre la distribución empírica de valores  $p$  y la distribución uniforme teórica, estos métodos podrían permitirnos estimar un valor del parámetro ( $\hat{\delta}$ ) para el que la distribución de la muestra aleatoria de valores  $p$  con que cuenta el meta-analista fuera uniforme  $U(0; 1)$  o compatible con ella.

Persiguiendo este último objetivo, van Assen et al. (2014) y Kocak (2015) han propuesto recientemente estimar por el método de los momentos el valor del efecto poblacional utilizando la distribución de Irwin-Hall (1927). Este método no requiere ninguna transformación de los valores  $p$ , como ocurre con los procedimientos mencionados más arriba, y por ello será el que utilicemos en el presente trabajo, dado que es el más cómodo de implementar.

### *2.3 Estimación del efecto poblacional con la distribución de Irwin-Hall*

El objetivo del meta-analista es concluir con una muestra de  $k$  estudios de una población de valores  $p$  en relación a un efecto poblacional cualquiera. Antes de que los investigadores de cada estudio primario obtengan el estadístico de contraste de la hipótesis de nulidad pertinente, no se conoce cuál será el valor concreto que tomará cada valor  $p$ . Así, la primera observación de su muestra o primer valor de  $p$ , puede ser considerada una variable aleatoria continua  $P_1$ , la segunda observación, otra v.a.c.  $P_2$ , etc. Cada una de estas v.a.c. será un valor concreto de  $p$ , por lo que, si  $H_0$  es verdadera, todas ellas seguirán una distribución de probabilidad uniforme  $U(0; 1)$ , como hemos comprobado en el apartado 2.1 de este mismo trabajo. Cada intervalo de la misma longitud comprendido entre los extremos 0 y 1 posee la misma probabilidad, lo que implica que la función de densidad toma el mismo valor para todos los puntos dentro de este intervalo.

Una vez obtenida la muestra de  $k$  valores de  $p$ , los valores observados los denotaremos por  $p_1, p_2, p_3 \dots p_k$ . Definimos ahora una nueva variable aleatoria, a la que llamaremos  $X$ , resultado de la suma de  $k$  variables aleatorias independientes, todas ellas con distribución  $U(0; 1)$ .

$$P_i \sim U(0, 1) \quad ; \quad X = \sum_{i=1}^K P_i$$

La distribución de Irwin-Hall, denominada así por Joseph Oscar Irwin y Philip Hall (1927), nos informa de cómo se distribuyen los valores de esta nueva variable. Dado que la esperanza de las variables  $P_i$  toma el valor  $\frac{1}{2}$ , pues  $E(P_i) = \left(\frac{b+a}{2}\right)$  y su varianza toma el valor  $\frac{1}{12}$ , pues  $\sigma^2(P_i) = \left(\frac{(b-a)^2}{12}\right)$ , el valor esperado y la varianza de esta nueva variable  $X$  se obtienen mediante la suma de las  $k$  esperanzas y varianzas, respectivamente, de dichas variables aleatorias  $U(0; 1)$ , pues asumimos que las covarianzas son nulas ya que son estimaciones independientes.

$$E(X) = E(\sum_{i=1}^K P_i) = \sum_{i=1}^K E(P_i) = k \cdot \frac{1}{2} = \frac{k}{2} \quad [2]$$

$$V(X) = V(\sum_{i=1}^K P_i) = \sum_{i=1}^K V(P_i) = k \cdot \frac{1}{12} = \frac{k}{12} \quad [3]$$

Gracias a la estimación puntual podemos utilizar una muestra aleatoria de estimaciones independientes para obtener el valor que, en algún sentido, mejor represente al parámetro de su distribución poblacional. Mediante el procedimiento conocido como *Método de los Momentos*, un método de estimación de parámetros, igualamos ciertas características muestrales con las correspondientes características poblacionales. Este método consiste en relacionar los valores de la muestra empírica con el parámetro que pretendemos estimar, mediante una función determinada por el momento concreto de la distribución que se desea utilizar (Revuelta y Ponsoda, 2005; Wickens, 1982). Con ello conseguimos un estimador puntual, un valor representativo del parámetro de interés. En este caso, ponemos en relación el valor esperado de la distribución de Irwin-Hall con el estadístico obtenido a partir de las  $k$  variables  $P_i$  mediante la función del momento de primer orden tal y como hemos visto en [2]. Por simplicidad restringiremos la siguiente formulación a un contraste unilateral derecho, pero es fácilmente extensible a un contraste bilateral o uno unilateral izquierdo. Para este caso el valor de  $p$  se obtiene como  $1 - F(EC)$ . Por tanto, sumando esta expresión para los  $k$  estudios e igualando esa suma al valor esperado de la distribución,

$$k - \sum_{i=1}^K P \left( Z < \frac{d_i - \delta}{\sigma_{d_i}} \right) = \frac{k}{2} \quad [4]$$

Como se puede observar, en esta ecuación no tenemos ninguna incógnita cuando los valores  $p$ , los valores concretos que toma cada variable  $P_i$ , vienen asociados a los estadísticos

de contraste que prueban la hipótesis nula (el efecto poblacional contrastado,  $\delta$ , toma el valor cero). Sin embargo, el objetivo con el Método de los Momentos es conseguir un estimador puntual del parámetro,  $\hat{\delta}$ , para el que la distribución de Irwin-Hall se equipare a su valor esperado de la siguiente manera (despejando en [4]).

$$\sum_{i=1}^K P\left(Z \geq \frac{d_i - \hat{\delta}}{\sigma_{d_i}}\right) = \frac{k}{2} \quad [5]$$

Esta ecuación nos servirá cuando la muestra de estudios primarios de la que parte el meta-analista sea una muestra realmente aleatoria. Especialmente importante es que no esté afectada por el sesgo de publicación a favor de los resultados significativos. Mostraremos su comportamiento en los estudios 1 y 2 de simulación. No obstante, como sabemos, el sesgo de publicación afecta a la probabilidad con que se publican los estudios con resultados no significativos, haciendo que la población del efecto quede sólo parcialmente representada. Ante la posible amenaza de sesgo de publicación, veremos cómo el método basado en la distribución de los valores  $p$  permite la opción de trabajar exclusivamente con la muestra de valores  $p$  significativos, ya que todos los tamaños del efecto asociados a resultados significativos tienen la misma probabilidad de ser publicados. Volveremos sobre este punto más adelante, en el estudio 3.

### 3 La presente investigación: Simulaciones

En el presente trabajo se han llevado a cabo tres estudios de simulación por el método Monte Carlo con el fin de comparar en distintos escenarios el desempeño de la técnica basada en la distribución de los valores  $p$ . Con ellos examinaremos su rendimiento frente al de las técnicas meta-analíticas convencionales de Efecto Fijo y Efectos Aleatorios, especialmente cuando se trata de un campo con un nivel importante de sesgo de publicación.

Hemos empleado el lenguaje de programación R (Ihaka y Gentleman, 1996; versión 3.2.2, 2015) y los paquetes *Truncnorm* (Trautmann, Steuer, Mersmann y Bornkamp, 2014), con la función *rtruncnorm*( ) para obtener datos aleatorios en el intervalo (a, b) de una distribución normal<sup>2</sup>; y *Metafor* (Viechtbauer, 2010; 2016), para calcular la estimación combinada mediante la función *rma.uni*( ), utilizando el argumento *method*="FE" para Efecto Fijo, y *method*="DL" en el caso de Efectos Aleatorios (método de DerSimonian-Laird, 1986) para estimar la varianza específica). Los  $k$  tamaños del efecto para cada meta-análisis

2. Los programas han sido desarrollados por la autora de este trabajo.

simulado se generaron a partir de una distribución  $N(\delta, \sigma_d)$ , con valores de delta concretos para cada estudio y valores de sigma o bien constantes, cuando se mantiene el mismo tamaño muestral para todos los estudios primarios, o bien aleatorios, a partir de una distribución ji-cuadrado con valor esperado 60 (López-López, Rubio-Aparicio, Sánchez-Meca y Martín-Martínez, 2013). En último lugar, para el cálculo de  $\hat{\delta}$  basado en la distribución de los valores de  $p$ , se programaron primero las funciones pertinentes para cada una de las simulaciones, basadas en la distribución de Irwin-Hall, que igualan la diferencia entre la esperanza de esta distribución  $\left(\frac{k}{2}\right)$  y la suma de los valores  $p$ , a cero. Y a continuación, mediante la función *uniroot*( ) se resuelve el conjunto de  $k$  ecuaciones para los tamaños del efecto aleatorios obtenidos de la distribución  $N(\delta, \sigma_d)$ , obteniendo así un valor para el estimador puntual del parámetro,  $\hat{\delta}$ .

#### 4 Estudio 1: Desempeño en condiciones óptimas

Los modelos de Efecto Fijo y Efectos Aleatorios aportan resultados bajo el supuesto de que se posee una muestra representativa de estimaciones del verdadero tamaño del efecto, esto es, una muestra aleatoria de la población del efecto. En este primer análisis queremos comprobar si existen diferencias en el rendimiento de la técnica basada en la distribución de los valores  $p$  con respecto a las técnicas tradicionales en un contexto sin sesgo de publicación para varios valores de  $\delta$ , donde únicamente se altera el número  $k$  de estudios que componen el meta-análisis y los tamaños muestrales de cada uno de ellos.

##### 4.1 Método

Tomando los valores de delta 0,25, 0,5 y 0,75, generamos diez mil meta-análisis con un número variable,  $k$ , de estudios (de 5 a 50 estudios). De cada una de esas diez mil réplicas obtenemos la estimación combinada bajo los modelos de Efecto Fijo y de Efectos Aleatorios (Botella y Sánchez-Meca, 2015) y el cálculo mediante *uniroot*( ) del efecto poblacional en base a la función que sigue (procedimiento de Irwin-Hall bajo el modelo basado en la distribución uniforme de los valores  $p$ ), que procede de despejar en [5]:

$$k/2 - \sum_{i=1}^K P\left(Z \geq \frac{d_i - \hat{\delta}}{\sigma_{d_i}}\right) = 0 \quad [6]$$

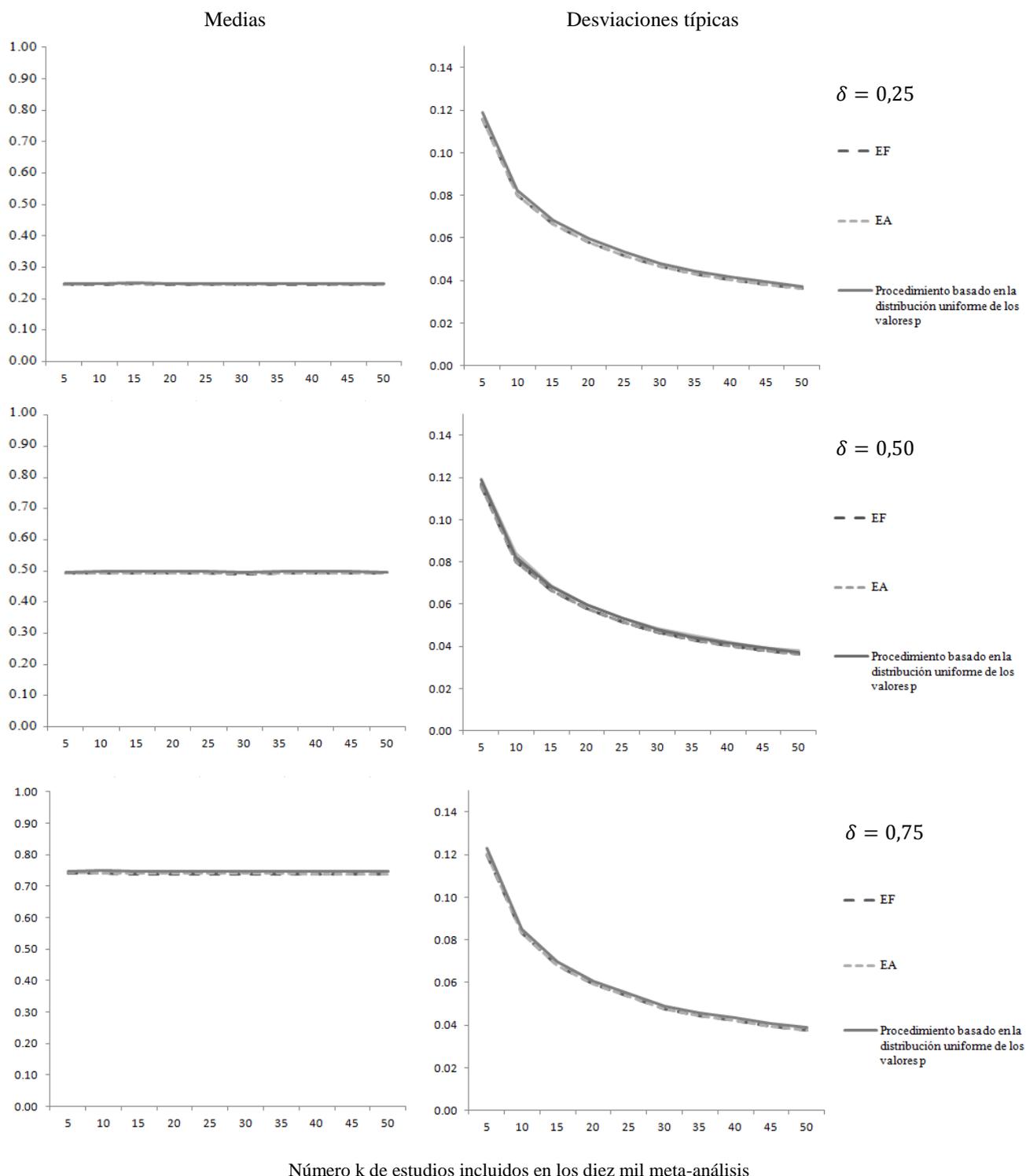


Figura 2. Medias y desviaciones típicas de las estimaciones obtenidas por los modelos de Efecto Fijo, Efectos Aleatorios y el procedimiento basado en los valores  $p$ , en 10.000 meta-análisis de  $k$  estudios obtenidos bajo un delta de 0,25 , 0,50 y 0,75. “EF” hace referencia a Efecto Fijo y “EA” a Efectos Aleatorios (método de Dersimonian y Laird). El tercer procedimiento hace referencia a la estimación por el método de los momentos utilizando la distribución de Irwin-Hall como suma de los valores  $p$ . Los datos de estas gráficas pueden consultarse en las tablas 1, 2 y 3 incluidas en el apéndice del presente trabajo.

En este procedimiento se estima el valor de  $\delta$  con el criterio ya adelantado. Se obtiene el valor del estimador  $\hat{\delta}$  bajo el cual la suma de los valores  $p$  es igual al valor esperado de la distribución Irwin-Hall cuando dicho valor es correcto,  $\frac{k}{2}$ . Una vez obtenemos las tres estimaciones del efecto poblacional para cada uno de los meta-análisis, nos disponemos a calcular la media y desviación típica de las estimaciones obtenidas en las diez mil réplicas.

#### 4.2 Resultados y Discusión

En la figura 2 se advierte que las diferencias entre las estimaciones por los métodos convencionales y el cálculo a partir de la distribución de los valores  $p$  son mínimas y se mantienen a través de distintos valores de delta. En condiciones óptimas, donde no existe sesgo de publicación, la media del procedimiento basada en la distribución de  $p_i$  toma valores, en la práctica, despreciablemente más cercanos al parámetro real que los métodos de Efecto Fijo y Efectos Aleatorios. Esta diferencia alcanza un valor máximo de 0,009 en unidades del tamaño del efecto entre las estimaciones de la técnica basada en los valores  $p$  y el método de Efecto Fijo cuando delta toma un valor de 0,75.

Por otro lado, la variabilidad de las estimaciones obtenidas con estas tres técnicas disminuye conforme aumenta el número de estudios que componen el meta-análisis. El procedimiento basado en los valores  $p_i$ , presenta una variabilidad mínimamente mayor que los métodos de Efecto Fijo y Efectos Aleatorios. La diferencia entre la eficiencia de este nuevo procedimiento frente a las otras dos técnicas alcanza un valor máximo de 0,002 en desviaciones típicas.

Cuando la muestra de estudios primarios es realmente aleatoria, podemos concluir que utilizar la técnica basada en la distribución uniforme de los valores  $p$  proporciona estimadores igualmente buenos, en términos de sesgo y eficiencia, que los métodos habituales bajo los modelos de Efecto fijo y Efectos Aleatorios, tradicionalmente utilizados en los estudios meta-analíticos (medias de los tamaños del efecto ponderadas por el inverso de sus varianzas).

### 5 Estudio 2: Desempeño en condiciones con sesgo de publicación

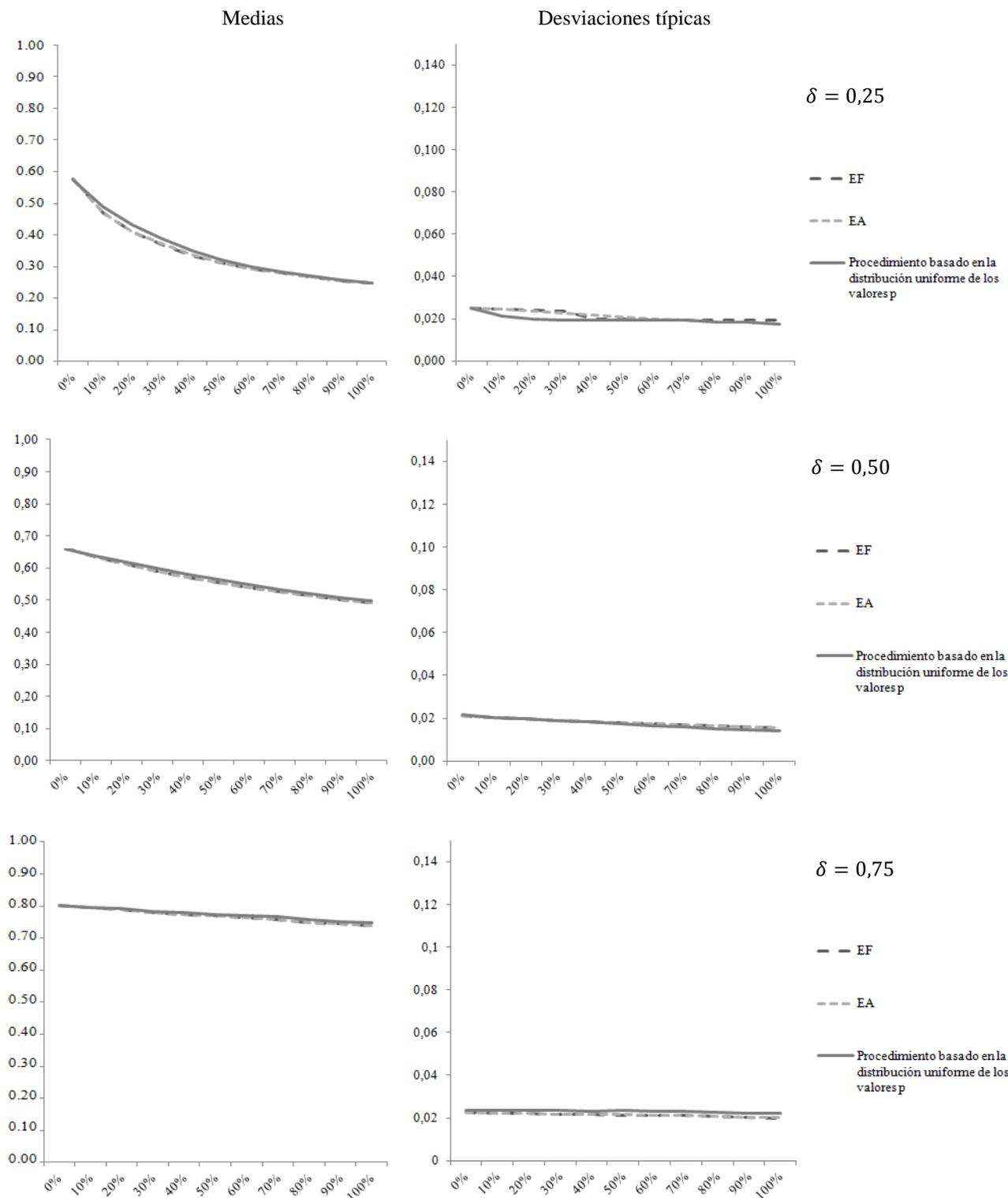
En el primer estudio hemos analizado el comportamiento de los procedimientos convencionales de estimación en meta-análisis y de la nueva técnica para el cálculo del efecto poblacional, basada en el método de los momentos utilizando la distribución de Irwin-Hall, en

condiciones óptimas, donde la muestra de estudios primarios no está afectada por el sesgo de publicación. Lo que desconocemos es cómo actúan estos tres modelos estadísticos bajo distintos niveles de sesgo, es decir, cuando el número de estudios no significativos que aparecen en la literatura científica varía. Muchas veces se ha denunciado que el sesgo de publicación contra los estudios con resultados no significativos provoca una sobreestimación del tamaño del efecto. Vamos a constatar este hecho, valorando a su vez en qué medida ese sesgo afecta a las estimaciones obtenidas mediante esta nueva técnica basada en la distribución uniforme de los valores  $p$ . Adviértase que todavía vamos a trabajar con todos los estudios “publicados”. En el estudio 3 abordaremos la estimación empleando sólo los estudios significativos.

### 5.1 Método

Para tres efectos poblacionales distintos, con unos valores de  $\delta$  de 0,25, 0,5 y 0,75, generamos meta-análisis que incluyan distintos porcentajes de estudios con resultados no significativos. Para ello, tomamos como número de estudios máximo un  $k = 100$ , y, dependiendo de la potencia que deja a la derecha de su distribución cada valor de delta en función de una  $N$  total de 60, mantendremos el número de estudios significativos constante mientras varía el porcentaje de estudios no significativos que permanece incluido en el meta-análisis. Así por ejemplo, si partimos de un  $\delta$  de 0,5 y un  $N$  total de 60, la potencia del contraste es de 0,60 mientras que  $\beta$  toma un valor de 0,40. En esta situación, mantenemos una cantidad constante de sesenta estudios con resultados significativos, mientras que, de los cuarenta estudios que no lo son, incluimos en cada condición distintos porcentajes, que comprenden desde un 10% (4 estudios no significativos) hasta el 100% de los mismos (40 estudios).

Generando por cada valor de  $\delta$  diez mil meta-análisis para cada condición de sesgo bajo esta lógica, obtenemos el cálculo mediante *uniroot()* del efecto poblacional en base a la función presentada en [6] y las estimaciones del valor del parámetro bajo Efecto Fijo y Efectos Aleatorios. De este modo, los tres procedimientos puestos a prueba actúan como si no existiera sesgo en la muestra de estudios, y podremos valorar el impacto que tiene en nuestros resultados el hecho de desconocer o ignorar el sesgo de publicación. Las gráficas de la figura 3 recogen los cálculos del efecto medio y las desviaciones típicas de estas diez mil estimaciones para cada condición de sesgo en los valores de delta anteriormente convenidos.



Nivel de Sesgo de Publicación: Porcentaje de estudios no significativos incluidos en el meta-análisis

Figura 3. Medias y desviaciones típicas de las estimaciones obtenidas por los modelos de Efecto Fijo, Efectos Aleatorios y el procedimiento basado en la distribución de los valores  $p$  en 10.000 meta-análisis en función del porcentaje de estudios no significativos incluidos. “EF” hace referencia a Efecto Fijo, “EA” a Efectos Aleatorios (método de Dersimonian y Laird) y el tercer procedimiento a la técnica basada en la distribución de valores  $p$  por el método de los momentos con la distribución de Irwin-Hall. Los datos de estas gráficas pueden consultarse en la tabla 4, 5 y 6 incluidas en el apéndice del presente trabajo.

## 5.2 Resultados y Discusión

Las gráficas que presentan las medias de estas diez mil réplicas por condición nos muestran que los tres procedimientos dan como resultado estimadores que se aproximan en el mismo grado al verdadero valor del parámetro a medida que aumenta el número de estudios con resultados no significativos incorporados en el meta-análisis. La diferencia entre el cálculo del tamaño del efecto poblacional basado en la distribución de los valores  $p$  y el basado en los procedimientos de Efecto Fijo y Efectos Aleatorios es mínima, comprendiendo valores entre  $-0,004$  y  $0,01$  y alcanzando su valor máximo en la condición donde el sesgo es del 50% cuando  $\delta$  toma los valores de  $0,25$  y  $0,5$ , y en la condición en que es del 100% cuando el parámetro toma un valor de  $0,75$ . De los tres procedimientos puestos a prueba, la técnica basada en la distribución uniforme de los valores  $p$  es la que más se acerca ligeramente al verdadero valor del parámetro a medida que el sesgo de publicación es menor y el efecto poblacional se aleja de cero.

Es importante recalcar que el hecho de que el sesgo en la estimación se perciba menor en las gráficas a medida que aumenta el valor de  $\delta$ , se debe a que, por la lógica que hemos empleado en la simulación, aumenta el número de estudios significativos constantes a medida que el valor del efecto poblacional es mayor.

Por otro lado, las desviaciones típicas de los valores obtenidos por los tres métodos disminuyen de manera ínfima a medida que aumenta el número de estudios totales que componen el meta-análisis, y varían en el rango  $-0,004$  y  $0,002$ .

Podemos concluir que, aun existiendo sesgo de publicación a favor de los estudios con resultados significativos, utilizar el procedimiento basado en la distribución de los valores  $p$  es tan buena o mala opción como decantarse por las técnicas estadísticas de Efecto Fijo y Efectos Aleatorios.

## 6 Estudio 3: Desempeño del Procedimiento basado en los valores $p$ significativos

En los apartados anteriores hemos comprobado que los métodos tradicionalmente utilizados en meta-análisis no ofrecen mejores resultados que esta nueva técnica basada en la distribución de los valores  $p$ , ni cuando las condiciones en que se recogen los estudios primarios son óptimas, ni cuando se publican en mayor medida los estudios con resultados significativos.

La principal razón para que los métodos basados en los valores  $p$  estén siendo recuperados es que proporcionan una oportunidad para afrontar la amenaza del sesgo de publi-

cación que no ofrecen los métodos habituales. Como ya hemos adelantado en la introducción, el procedimiento se basa en el supuesto de que los estudios con resultados estadísticamente significativos están poco afectados por el sesgo de publicación. Es decir, que a diferencia de los no significativos publicados, los estudios publicados que sí son significativos representan bien a los estudios realizados y que son significativos.

Como vienen mostrando diversos autores (Hedges, 1984; Kocak, 2015; Simonsohn, Nelson y Simmons, 2013; van Assen et al., 2014), ante condiciones donde existe sesgo de publicación, este nuevo procedimiento basado en la distribución uniforme de los valores  $p$ , permite conseguir estimadores más cercanos al verdadero valor del parámetro utilizando únicamente los estudios significativos, mientras que, por otro lado, los métodos de Efecto Fijo y Efectos Aleatorios, haciendo uso de toda la literatura publicada, proporcionan una estimación sesgada como hemos constatado en el estudio 2.

Sin embargo, para trabajar únicamente con los valores  $p$  iguales o inferiores a 0,05, la ecuación empleada en los estudios anteriores, [6], tiene que sufrir ciertas modificaciones. Ahora las variables aleatorias  $P_i$  se distribuyen bajo  $H_0$  verdadera de manera  $U(0; 0,05)$ , y la variable  $X$  ya no se puede contemplar como la suma de  $k$  v.a.c.  $U(0; 1)$ . Por tanto, la variable  $X$  ya no sigue la distribución de Irwin-Hall y no podemos relacionar su esperanza con las  $k$  variables  $P_i$ , como se ha planteado para cuando se emplea todo el rango de valores de  $p$ .

Para que las variables  $P_i$  vuelvan a distribuirse  $U(0; 1)$ , es necesario re-escalar los valores  $p$  en el rango  $(0, 1)$ . Si la hipótesis nula fuera verdadera, bastaría con dividir los valores  $p$  entre el valor máximo que puede tomar cualquier  $P_i$ , es decir, 0,05. Ahora bien, como el objetivo es estimar el efecto poblacional y este no tiene por qué ser nulo, deberemos dividir los valores  $p$  entre el posible valor máximo que pueda tomar cualquier  $P_i$  en función del valor del efecto poblacional. Este valor máximo es el tamaño del efecto crítico,  $d_{critica}$ , a partir del cual los resultados se asumen significativos, y haciendo uso de la aproximación a la normal, se calcula como  $d_{critica} = 1,645 \cdot \hat{\sigma}_{d_i}$

El re-escalamiento se produce entonces cuando los valores  $p_i$  se dividen entre la probabilidad que deja por encima de sí el valor  $Z = \frac{d_{critica} - \hat{\delta}}{\sigma_{d_{critica}}}$ , obteniendo la siguiente ecuación que permite conseguir un estimador puntual del parámetro,  $\hat{\delta}$ , para el que la distribución de Irwin-Hall se equipara a su valor esperado.

$$\frac{k}{2} - \sum_{i=1}^K \frac{P\left(Z \geq \frac{d_i - \delta}{\sigma_{d_i}}\right)}{P\left(Z \geq \frac{d_{critica} - \delta}{\sigma_{d_{critica}}}\right)} = 0 \quad [7]$$

Partiendo de esta modificación de la ecuación inicial, podremos obtener un estimador del parámetro a partir únicamente de los estudios significativos de la muestra total. En el presente estudio abordaremos también la cuestión de con qué número mínimo de estudios significativos resulta más rentable utilizar este nuevo método en lugar de las técnicas convencionales. Las estimaciones proporcionadas por los métodos tradicionales pueden resultar sesgadas y desviarse del verdadero valor del parámetro cuando los estudios significativos se publican en mayor medida que aquellos con resultados no significativos, pero al mismo tiempo pueden resultar más homogéneas entre sí, lo que aportaría menos variabilidad a la estimación.

Con el objetivo de responder a esta pregunta se ha llevado a cabo el tercer estudio, con el que pretendemos valorar con qué cantidad de estudios exclusivamente significativos resulta mejor recurso este nuevo procedimiento, y qué aspectos de nuestra estimación mejorarán o empeorarán al utilizarlo.

Pero antes veamos qué ocurre con los datos generados en el ejemplo del apartado 2.2 (Figura 1) si aplicamos una selección estricta de los estudios estadísticamente significativos, es decir, de aquellos para los que el valor de  $p$  al contrastar la hipótesis de nulidad es menor de 0,05. Los resultados (una vez re-escalados) aparecen en la figura 4 para los tres valores de  $\delta$  utilizados (0, 0,3 y -0,3).

Si el efecto poblacional contrastado tomara un valor real de cero, los valores de  $p$  asociados a los estadísticos de contraste significativos se distribuirían de manera  $U(0; 1)$  una vez re-escalados, como se puede apreciar en el primer histograma. Del mismo modo que en la figura 1, en él se recogen las frecuencias de diez mil valores de  $p$ , esta vez entre los valores 0 y 0,05, obtenidos bajo hipótesis nula verdadera ( $\delta = 0$ ), y después re-escalados en el intervalo (0; 1). En el segundo histograma, comprobamos que al quedarnos con los valores  $p$  significativos cuando el verdadero efecto poblacional es superior al contrastado bajo hipótesis nula ( $\delta = 0,3$ ), sigue existiendo mayor cantidad o frecuencia de valores  $p$  cercanos a cero. En cambio, cuando el verdadero efecto poblacional es inferior al efecto contrastado ( $\delta = -0,3$ ), como se puede apreciar en el tercer histograma, continuará existiendo mayor frecuencia de valores  $p$  cercanos a uno.

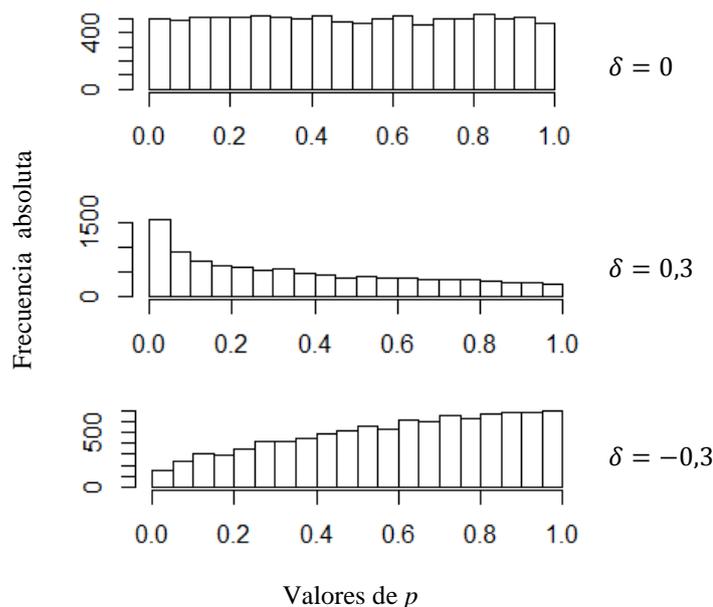


Figura 4. Histogramas de 10.000 valores de  $p$  entre 0 y 0,05 obtenidos bajo un delta de 0, 0,3 y -0,3, posteriormente reescalados en el intervalo (0;1).

Comparando las figuras 1 y 4 observamos que la desviación respecto de la uniforme (pendiente y curvatura) es sensiblemente menor en la figura 4. Es lo esperado, ya que en realidad estamos aumentando de tamaño el pequeño efecto que la figura 1 se daba dentro del intervalo de valores de  $p$  entre 0 y 0,05.

### 6.1 Método

Hemos llevado a cabo una nueva simulación en la cual, para tres valores distintos de  $\delta$  (0,25, 0,5 y 0,75), hemos generado diez mil meta-análisis para diez condiciones distintas en las que varía el número  $k$  de estudios significativos que componen el meta-análisis (de 5 a 50 estudios). Hemos utilizado la función `rtruncnorm()` para obtener los  $k$  tamaños del efecto significativos que componen cada meta-análisis fijando como límite inferior el valor a partir del cual, para un delta concreto y una  $N$  total variable, resultan en estadísticos de contraste significativos con sus valores  $p$  asociados. De cada uno de esos diez mil meta-análisis obtenemos el cálculo mediante `uniroot()` del efecto poblacional en base a la función expresada en [7].

Una vez obtenidas las diez mil estimaciones del parámetro en función de  $k$  y para cada valor de  $\delta$ , nos disponemos a calcular la media y desviación típica de las diez mil réplicas de cada condición.

## 6.2 Resultados y Discusión

Atendiendo a la figura 5 observamos que a medida que aumenta el número de estudios significativos que componen el meta-análisis, las estimaciones obtenidas por el procedimiento basado en los valores  $p$  re-escalados se desvían menos del verdadero efecto poblacional. Al mismo tiempo comprobamos que la media de las diez mil estimaciones se aproxima al verdadero valor del parámetro cuanto más diferencia existe entre éste y el efecto nulo. Esto es, cuanto más se aleja de cero el verdadero valor del efecto poblacional, menos estudios significativos se requieren para obtener una estimación menos sesgada. Este hecho se puede observar fácilmente bajo la condición más desfavorable en que  $k$  toma el valor de 5. En este caso, cuando  $\delta$  es 0,25, la media de los diez mil meta-análisis toma un valor de 0,183, por lo que se aleja 0,067 unidades del verdadero efecto poblacional. Tomando  $\delta$  un valor de 0,50, la media de las estimaciones es de 0,463, alejándose 0,037 unidades del verdadero efecto. Finalmente, cuando el parámetro es de 0,75, la media de los diez mil meta-análisis se aleja 0,013 unidades del verdadero valor del parámetro, siendo ésta de 0,737.

En cuanto a la variabilidad de las estimaciones, comprobamos que, aún partiendo de la misma cantidad de estudios significativos, cuanto mayor es el valor de  $\delta$ , más homogéneas resultan las diez mil estimaciones. Esta diferencia entre las desviaciones típicas va disminuyendo de un valor de delta a otro, a medida que aumenta el número de  $k$  estudios significativos. Volviendo a la condición en que  $k$  es 5, la variabilidad de las diez mil estimaciones obtenidas bajo un delta de 0,25 superan en 0,096 desviaciones típicas a las obtenidas bajo un delta de 0,50, y en 0,165 a las obtenidas bajo un valor de 0,75.

Recordemos que las estimaciones obtenidas por el procedimiento basado en los valores  $p$  en éste y en el primer estudio, difieren únicamente en que aquí estamos aprovechando sólo los estudios con resultados significativos. Cuando en el primer estudio seleccionamos aleatoriamente  $k$  tamaños del efecto tanto significativos como no significativos, la desviación típica máxima de las diez mil estimaciones obtenidas para cualquiera de los tres valores de delta se da en la condición en que  $k$  es 5. No obstante, este valor no supera las 0,12 desviaciones típicas, mientras que al escoger exclusivamente estudios significativos, este valor se triplica cuando delta vale 0,25, y se duplica cuando delta es 0,50. Ni siquiera cuando delta toma un valor de 0,75, este valor máximo en desviaciones típicas se iguala (0,154).

Así pues, podemos concluir que, el rendimiento del procedimiento basado en la distribución de los valores  $p$  utilizando exclusivamente los resultados significativos y re-escalándolos en el intervalo (0; 1), proporciona estimaciones menos sesgadas que cuando se

utilizan los resultados de todos los estudios, tanto significativos como no significativos, en presencia de sesgo de publicación, es decir, cuando la muestra de estudios no es representativa de la distribución poblacional del efecto. Sin embargo, su menor sesgo viene acompañado de una menor precisión (mayor varianza). Para valorar de una manera gráfica este hecho, la figura 6 muestra los resultados del estudio 2 (figura 3: medias y desviaciones típicas de los valores obtenidos en función del nivel de sesgo de publicación por los modelos de Efecto Fijo, Efectos Aleatorios y el procedimiento basado en la distribución de los valores  $p$ , utilizando la muestra total de estudios) junto con las medias y desviaciones típicas de los valores obtenidos por el procedimiento basado en la distribución de los valores  $p$ , utilizando esta vez únicamente los datos de los estudios con resultados estadísticamente significativos.

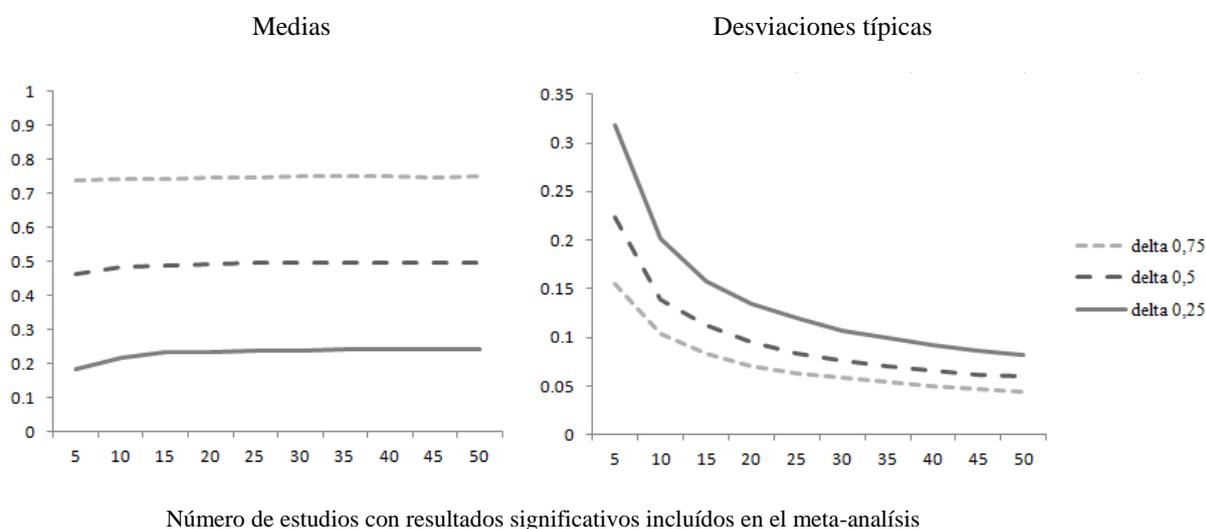
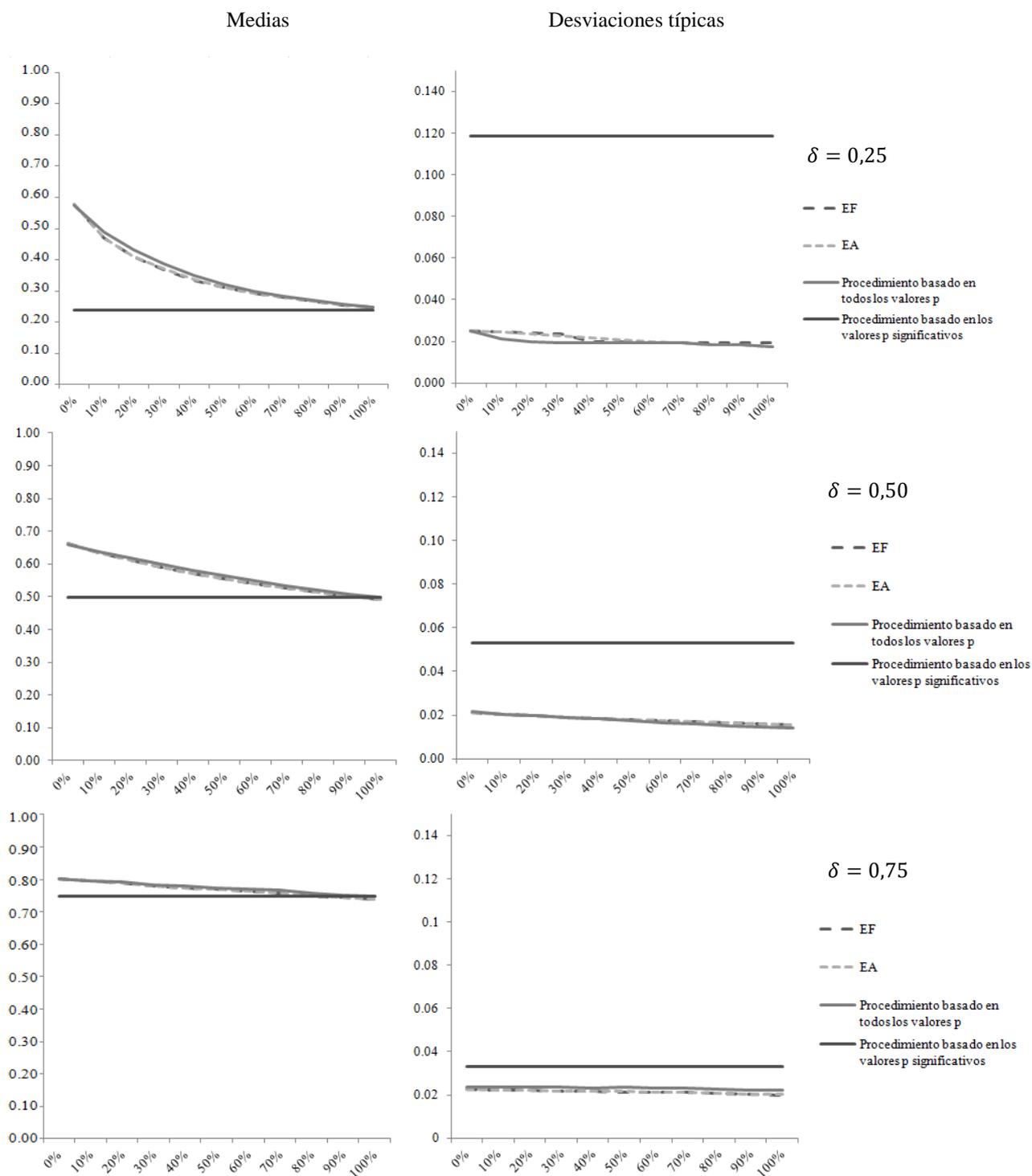


Figura 5. Medias y desviaciones típicas de las 10.000 estimaciones obtenidas por el procedimiento basado en los valores  $p$  re-escalados en función del número de estudios con resultados significativos incluidos en el meta-análisis. Los datos de estas dos gráficas pueden consultarse en la table 7 incluida en el apéndice del presente trabajo.

Recordemos que en el estudio 2, para los valores de delta empleados, 0,25, 0,5 y 0,75, tomamos como número de estudios máximo un  $k = 100$ , y, dependiendo de la potencia que deja a la derecha de su distribución cada valor de delta en función de una  $N$  total de 60, mantendremos un número constante de estudios significativos, que será el utilizado por la técnica basada en los valores  $p$  significativos, mientras varía el porcentaje de estudios no significativos incluido en el meta-análisis (conformando una muestra total de estudios bajo distintos niveles de sesgo de publicación, utilizada en los métodos de Efecto Fijo, Efectos Aleatorios y la técnica basada en todos los valores  $p$ ). Partiendo de la misma explicación empleada en el estudio 2, bajo un  $\delta$  de 0,5 y un  $N$  total de 60, la potencia del contraste es de 0,60 mientras que  $\beta$  toma un valor de 0,40, por lo que mantenemos una cantidad constante de



Nivel de Sesgo de Publicación: Porcentaje de estudios no significativos incluidos en el meta-análisis

Figura 6. Medias y desviaciones típicas de las estimaciones obtenidas por los modelos de Efecto Fijo, Efectos Aleatorios y el procedimiento basado en la distribución de los valores  $p$ , bien con la muestra total de estudios total, bien utilizando un número constante de estudios con resultados significativos, en 10.000 meta-análisis en función del porcentaje de estudios no significativos incluidos.

sesenta estudios significativos, mientras que, de los cuarenta estudios que no lo son, incluimos en cada condición distintos porcentajes, que comprenden desde un 10% (4 estudios no significativos) hasta el 100% de los mismos (40 estudios).

Atendiendo a las gráficas de la figura 6 referidas a las medias de las 10.000 estimaciones obtenidas por cada uno de los cuatro procedimientos, constatamos que el procedimiento basado en la distribución de los valores  $p$  que utiliza exclusivamente los resultados significativos re-escalados, proporciona estimaciones menos sesgadas que los técnicas tradicionales de estimación del tamaño medio, en condiciones con presencia de sesgo de publicación. No obstante, como conclusión podemos afirmar que la técnica basada en la distribución uniforme de los valores  $p$  significativos no depende únicamente del número de estudios significativos con que cuente el meta-analista, sino que además las estimaciones que ofrezca serán más exactas y precisas (esto es, menos sesgadas y variables), utilizando el mismo número de estudios, cuanto mayor sea el verdadero efecto poblacional.

## **7 Una aplicación práctica: Meta-análisis sobre el efecto del *Ego-Depletion***

En 1998 los psicólogos Roy Baumeister y Dianne Tice, de la Case Western Reserve University, dieron el primer paso en el desarrollo de lo que se convertiría en un concepto fundamental en el campo de la Psicología, mediante un experimento en el que presentaron a la vez un cuenco con galletas de chocolate y otro con rábanos a dos grupos de voluntarios. En la segunda fase, todos tenían que intentar completar un puzle irresoluble. A aquellos sujetos a los que se les había permitido comer lo que quisieran, tardaban de media 11 minutos más en darse por vencidos que aquellos a los que sólo se les permitió comer rábanos. Baumeister y Tice llamaron a este efecto *ego depletion* al considerar que el autocontrol es una capacidad limitada que mengua con su uso y, era debido a este motivo, que aquellos sujetos que mostraron suficiente fuerza de voluntad como para no comer las galletas de chocolate fueron quienes peor sobrellevaron la frustración de no completar el puzle, abandonando antes la tarea.

Esta idea evolucionó hasta convertirse en lo que conocemos como *strength model* (Baumeister, Bratslavsky, Muraven y Tice, 1998), teoría que predice que cualquier acto que requiera auto-control dependerá de un recurso común y limitado, que mengua con su uso y puede ejercitarse con entrenamiento. De acuerdo con este modelo, una vez agotadas las reservas de auto-control, éstas pueden restaurarse mediante el descanso o la ingesta de una bebida azucarada. De estas ideas se sirvieron Baumeister y John Tierney, del New York

Times, para publicar en 2011 *Willpower: Rediscovering the Greatest Human Strength*, un libro de divulgación acerca de cómo la ciencia del *ego depletion* podía ayudar al día a día de las personas. Desde entonces, numerosos investigadores han descubierto que distintas tareas, que abarcan desde el control de los pensamientos y la atención, al control de los impulsos, las emociones y la voluntad, pueden consumir la energía de una persona y dejarla cognitivamente exhausta.

### 7.1 El Ego Depletion en meta-análisis: La amenaza del Sesgo de Publicación

Aunque ciertos estudios realizados en el campo no aportaron resultados significativos, para 2010 el *ego depletion* parecía ser un efecto robusto y estar respaldado por una extensa parte de la comunidad científica. Ese mismo año, dos investigadores, Hagger y Chatzisarantis (2010), realizaron un meta-análisis con los datos de 83 estudios (198 estimaciones independientes), concluyendo que el *ego depletion* parecía ser un fenómeno real y fiable (con un valor de  $d = 0,62$  [IC95%: 0,57; 0,67] bajo Efecto Fijo y de  $d = 0,68$  [IC95%: 0,63; 0,73] bajo Efectos Aleatorios), aunque presentaba una cierta variabilidad ( $Q = 310,068$   $p < ,001$ ,  $I^2 = 36,466$ ), lo que podía estar indicando la existencia de distintos efectos subyacentes. Por esta razón, analizaron a su vez los datos a través de distintas variables moderadoras, como el ámbito del auto-control manejado, el tipo de tareas utilizadas para fatigar y medir el auto-control, si la medida de la tarea de seguimiento era conductual o intencional, etc.

Sin embargo, los resultados obtenidos en el meta-análisis de 2010 no contentaron a todo el mundo. Carter y McCullough (2013, 2014) analizaron los datos recopilados por Hagger y Chatzisarantis con distintos métodos para estimar y corregir la posible influencia del sesgo de publicación a favor de los estudios con resultados estadísticamente significativos, debido a que en su meta-análisis de 2010 sólo habían incluido estudios publicados. Lo que Carter y McCullough encontraron fue que existían fuertes indicios de sesgo de publicación y que, al corregirlos con esos procedimientos, el efecto del *ego depletion* no difería significativamente de cero.

Esta y otras razones, basadas en la idea de que los criterios de inclusión de Hagger y Chatzisarantis no habían sido los más apropiados, llevaron a publicar un segundo meta-análisis sobre el *ego depletion* (Carter, Kofler y McCullough, 2015). Con nuevos criterios de inclusión, que sólo permitían la incorporación de aquellos estudios experimentales con una medida comportamental del auto-control en la tarea de seguimiento, y haciendo una búsqueda exhaustiva de la literatura no publicada, reunieron 116 estudios (188 estimaciones

independientes). En esta nueva muestra, sólo mantuvieron 28 de los estudios incluidos por Hagger y Chatzisarantis en 2010, e incluyeron 48 nuevos estudios que no estaban publicados. Las conclusiones que se extraen de los datos recopilados por estos autores bajo los modelos de Efecto Fijo y Efectos Aleatorios son que el *ego depletion* toma un valor significativo general mucho menor,  $d = 0,37$  [IC95%: 0,32; 0,42] y  $d = 0,43$  [IC95%: 0,35; 0,52], respectivamente (aunque su valor se reduce, sin dejar de ser significativo, al aplicar el método *Trim and Fill*; Duval y Tweedie, 2000). Además, se sigue apreciando una alta variabilidad entre los diferentes estudios ( $Q = 333,1475$   $p < ,001$ ,  $I^2 = 64,88$ ) que indicaría la coexistencia de distintos efectos subyacentes. No obstante, estos autores aplicaron a sus propios datos los mismos análisis que habían aplicado a los de Hagger en su artículo de 2014, con la intención de evaluar si sus resultados seguían viéndose influidos por el sesgo de publicación. Observaron que al corregir la influencia del exceso de significación estadística de sus resultados, había muy poca evidencia de que el *ego depletion* fuera un fenómeno real.

No todo el mundo ha concedido la credibilidad pertinente a las conclusiones del equipo de Carter. Diferentes investigadores de prestigio en el campo opinan que sus resultados son prematuros, ya que los análisis que utilizan para valorar la influencia del sesgo de publicación (*Test for Excess Significance*, o TES; Ioannidis y Trikalinos, 2007 y *Funnel Plot Asymmetry Test*, o FAT; Egger, Davey-Smith, Schneider y Minder, 1997) y para corregir su influencia (*Precision Effect Test*, o PET; Stanley, 2005 y *Precision Effect Estimate with Standard Error*, o PEESE; Stanley y Doucouliagos, 2007) son muy novedosos y no existe suficiente investigación para determinar completamente la calidad de sus conclusiones.

En este marco de incertidumbre que rodea al paradigma del *ego depletion*, buscamos con el presente trabajo arrojar algo de luz sobre la influencia del sesgo de publicación en los resultados de ambos meta-análisis presentados anteriormente. Para ello, vamos a aplicar la técnica basada en la distribución uniforme de los valores  $p$  sobre los datos de la fracción de los estudios significativos de la muestra total de estudios seleccionados tanto por Hagger y Chatzisarantis, como por Carter y su equipo, y a valorar el alcance de estos nuevos resultados.

### 7.2 Nueva evidencia basada en los estudios estadísticamente significativos.

Para valorar el efecto global, nos remitimos a la muestra de estudios total reunida por ambos grupos de investigación<sup>3</sup>. Por un lado, el meta-análisis de Hagger y Chatzisarantis (2010) incluye 198 estimaciones independientes, de las cuales 169 son estadísticamente significativas y 19 no lo son. Por su parte, Carter, Kofler y McCullough, con criterios de

3. Para ello, la autora de este trabajo se ha puesto en contacto con los autores de ambos meta-análisis, obteniendo de ellos mismos las bases de datos de sus estudios.

inclusión distintos a los del primer meta-análisis, seleccionaron 118 diferencias de medias, siendo 57 estadísticamente significativas y 61 no. Recordemos que para evitar la influencia del sesgo de publicación, por el que los resultados no significativos están infrarrepresentados en la muestra de estudios, la técnica basada en la distribución uniforme de los valores  $p$  utiliza exclusivamente los valores de  $p$  re-escalados de los estudios con resultados estadísticamente significativos.

Como hemos visto en el estudio 3, cuanto menor sea el número de estudios significativos del que partamos, mayor variabilidad tendrá el proceso de estimación a través de la técnica basada en la distribución de los valores  $p$ . Aún así y a pesar de la diferencia entre el número de estudios significativos con que cuenta el meta-análisis de Hagger y Chatzisarantis frente al de Carter et al., esperamos que los resultados aportados por esta técnica sean similares en ambas ocasiones. Esto estaría indicando que, a pesar de que los criterios de inclusión adoptados por el grupo de Carter fueron más restrictivos que los utilizados por Hagger y Chatzisarantis, el efecto global que miden los primeros no se vería alterado por la inclusión de estudios llevada a cabo por los segundos.

En la tabla 9 se pueden observar los porcentajes de estudios significativos y no significativos incluidos en la muestra total de cada uno de ambos estudios meta-analíticos, los resultados obtenidos de ellos a partir de la aplicación del método tradicional de Efecto Fijo y el cálculo obtenido al aplicar la técnica basada en la distribución uniforme de los valores  $p$  sobre los dos grupos de estudios significativos. Al analizar los datos del meta-análisis de Hagger y Chatzisarantis (2010), encontramos que el valor que proporciona la ecuación [7] para los valores re-escalados de las 169 diferencias de medias estadísticamente significativas, es 0,644; mientras que para las 57 diferencias de medias significativas del meta-análisis de Carter et al. (2015), este valor es de 0,670. En la figura 8 se pueden observar las distribuciones de los valores  $p$  asociados a los estadísticos de contraste originales de los estudios incluidos en ambos meta-análisis (el valor poblacional contrastado es cero,  $\delta = 0$ ), la distribución de los valores  $p$  significativos de cada meta-análisis re-escalados en el intervalo (0; 1), y finalmente, la distribución de los valores  $p$  asociados a los estadísticos de contraste cuando se obtienen con los valores para el efecto poblacional antes mencionados (0,644 para Hagger y Chatzisarantis, 2010; y 0,670 para Carter et al., 2015).

Los valores obtenidos en ambos casos son muy similares, lo que podría estar indicando que el efecto medido por Hagger y Chatzisarantis es el mismo efecto que más tarde tratará de valorar el equipo de Carter. Sin embargo, los resultados obtenidos en ambos casos por los métodos tradicionales de Efecto Fijo y Efectos Aleatorios difieren: sus intervalos de

confianza, tanto para un método como para el otro, son excluyentes. A su vez, como hemos visto en el estudio 1, al emplear la técnica basada en la distribución de los valores  $p$  con los datos de todos los estudios incluidos en el meta-análisis, las estimaciones no difieren de aquellas obtenidas bajo Efecto Fijo y Efectos Aleatorios. Cuando utilizamos esta técnica con todos los estudios incluidos en ambos meta-análisis, obtenemos que el valor que proporciona la ecuación [6] para los valores re-escalados de las 198 diferencias de medias empleadas por Hagger y Chatzisarantis, es 0,706; mientras que para las 118 diferencias de medias del meta-análisis de Carter et al., este valor es de 0,426. Bajo nuestro punto de vista, y partiendo del supuesto de que ambos equipos de investigación están indagando el mismo efecto, esta diferencia en los resultados se debe a los procesos de inclusión empleados por cada uno y a la lógica en la que se basan los métodos convencionales de estimación del efecto medio (medias de los tamaños del efecto ponderadas por el inverso de sus varianzas).

Tabla 9.

Número de estudios incluidos y estimación del efecto poblacional del *ego depletion* obtenida por el método de Efecto Fijo y la técnica basada en la distribución uniforme de los valores  $p$  en los

	Base de datos	
	Hagger y Chatzisarantis, 2010	Carter et al., 2014
<b>Estudios incluidos</b>		
Total	198	118
Significativos	169 (85,35%)	57 (48,30%)
No significativos	29 (14,65%)	61 (51,70%)
<b>Estimación tradicional</b>		
Modelo de Efectos Fijos	0,62 [0,57; 0,67]	0,37 [0,32; 0,42]
<b>Análisis de valores <math>p</math></b>		
Todos los estudios	0,706	0,426
Solo estudios significativos	0,644	0,670

Aunque el efecto medido sea el mismo, en cuanto que la muestra de estudios deja de ser aleatoria, es decir, se incluyen más estudios significativos que no significativos con respecto a su proporción real en la distribución poblacional del efecto, los métodos convencionales proporcionarán un resultado sobrestimado con respecto al verdadero valor que toma el efecto en la población. Pero a su vez, si por distintas razones se incluye una proporción de estudios no significativos mayor que su correspondiente proporción de estudios significativos, ocurrirá lo que parece ocurrir en el meta-análisis de Carter et al. (2015): el efecto medio obtenido resulta en una infraestimación del verdadero valor del efecto poblacional.

Si empleáramos la técnica basada en la distribución de los valores  $p$  con toda la muestra de estudios recopilada para un efecto, es decir, sin tener presente que la muestra

puede no ser aleatoria (ya sea debido al sesgo de publicación a favor de los estudios significativos, o debido a los criterios de inclusión utilizados por el meta-analista), obtendríamos un valor del efecto poblacional inferior (0,426) o superior (0,706) al verdadero valor del parámetro, como ocurre si cogemos toda la muestra de estudios empleada por Carter et al. (2015) y por Hagger y Chatzisarantis (2010), respectivamente. Pero en cambio, si partimos del supuesto de que todos los estudios con resultados estadísticamente significativos tienen la misma probabilidad de ser publicados, y seleccionamos sólo esta fracción de estudios de la muestra total, el valor del efecto poblacional obtenido por la técnica basada en la distribución uniforme de los valores  $p$ , tenderá a ser similar (0,64 para Hagger y Chatzisarantis, 2010; y 0,67 para Carter, Kofler y McCullough, 2015) si se está valorando el mismo efecto, y conforme el número de estudios significativos sea mayor.

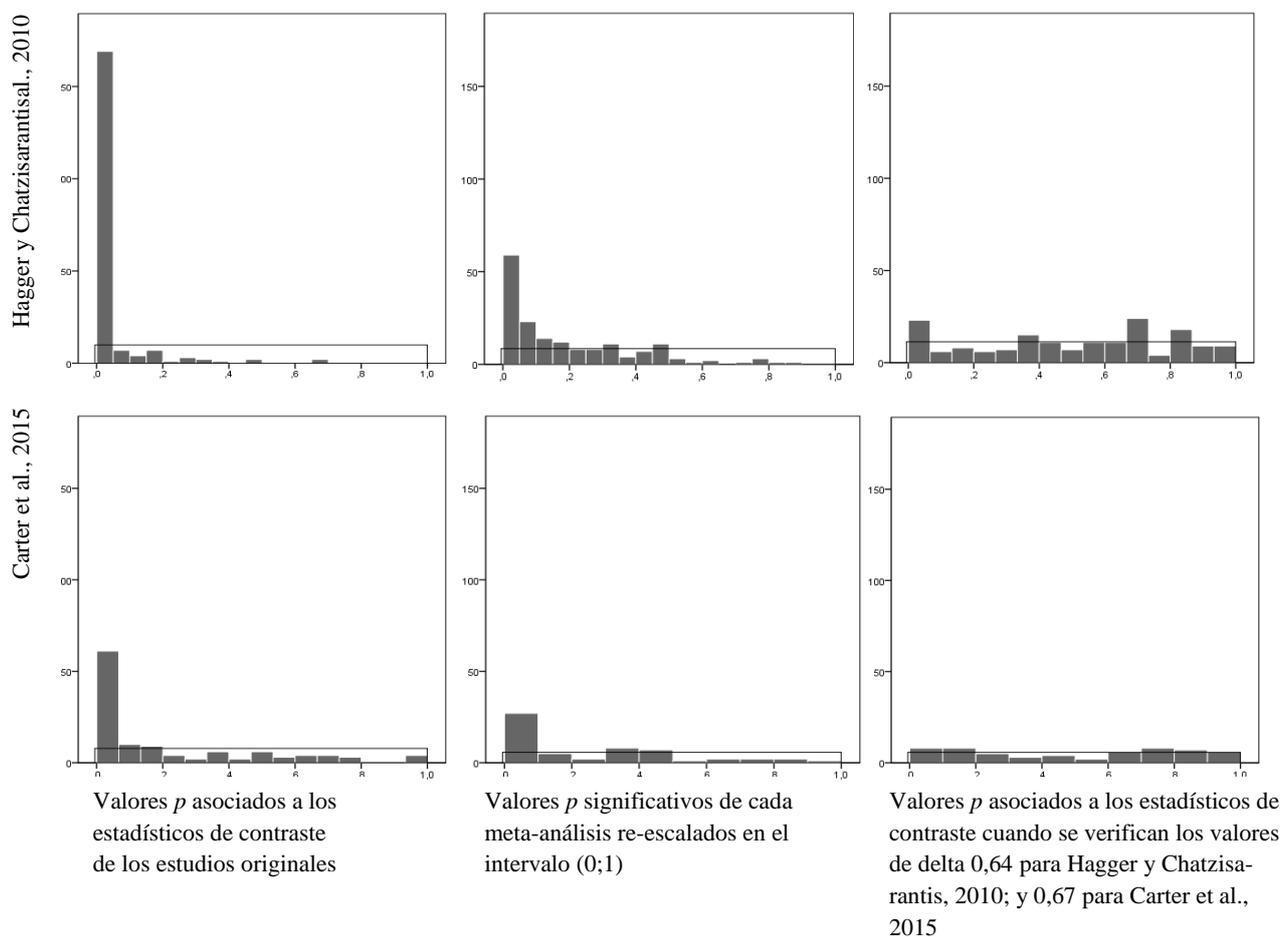


Figura 8. Distribuciones de frecuencias de los valores  $p$  asociados a los estadísticos de contraste de los estudios originales incluidos en ambos meta-análisis, cuando el valor poblacional contrastado es cero,  $\delta = 0$  (primera columna), cuando se re-escalan en el intervalo (0;1) (segunda columna), y cuando se contrastan los valores 0,64 para Hagger et al, 2010, y 0,67 para Carter y McCullough, 2015.

En resumen, aplicar el método basado en la distribución uniforme de los valores  $p$  en los estudios significativos utilizados en distintos meta-análisis de un mismo paradigma, nos permite extraer dos conclusiones. En primer lugar, este método nos proporciona estimaciones muy parecidas desde ambas bases de datos. En segundo lugar, nos revela cómo la aplicación de reglas de inclusión de los estudios en un meta-análisis puede tener un impacto determinante en las estimaciones. Cualquier regla de inclusión que afecte al equilibrio entre las proporciones de estudios significativos y no significativos que conforman la base final de estudios analizada tendrá un impacto en las estimaciones obtenidas por los procedimientos tradicionales.

**REFERENCIAS BIBLIOGRÁFICAS**

- Anderson, T.W., y Darling, D.A. (1954). A Test of Goodness-of-Fit. *Journal of the American Statistical Association*, 49, 765–769.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., y Tice, D. M. (1998). Ego depletion: Is the active self a limited resource?. *Journal of Personality and Social Psychology*, 74(5) 1252–1265.
- Begg, C. B., y Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088–1101.
- Bhattacharya y Habtzghi (2002). Median of the p Value Under the Alternative Hypothesis. *The American Statistician*, 56(3), 202-206.
- Botella, J. y Sánchez-Meca, J. (2015). *Meta-análisis en ciencias sociales y de la salud*. Madrid, España: Editorial Síntesis.
- Carter, E. C., y McCullough, M. E. (2013). Is ego depletion too incredible? Evidence for the overestimation of the depletion effect. *Behavioural and Brain Sciences*, 36(6), 683-684.
- Carter, E. C., y McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: has the evidence for ego depletion been overestimated?. *Frontiers in Psychology*, 5, 1-11.
- Carter, E. C., Kofler, L. M, y McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on limited resource. *Journal of Experimental Psychology: General*, 144(4), 796-815.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Nueva Jersey, Estados unidos de América: Lawrence Erlbaum Associates.
- Cooper, H., DeNeve, K., y Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods*, 2, 447–452.
- DerSimonian, R. y Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177–88.
- Donahue, R. M. J. (1999). A Note on Information Seldom Reported via the P Value. *The American Statistician*, 53(4), 303-306.
- Duval, S. J., y Tweedie, R. L. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463.

- Egger, M., Davey Smith, G., Schneider, M., y Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634.
- Fisher, R. A. (1932). *Statistical methods for research workers*. Londres, Reino Unido: Oliver & Boyd.
- Francis, G. (2012). The psychology of replication and replication in psychology. *Perspectives on Psychological Science*, 7, 585–594.
- George, E. O., y Mudholkar, G. S. (1977). The logit method for combining independent tests. *Institute of Mathematical Statistics Bulletin*, 6, 212.
- Gut, A. (2009) *An intermediate course in probability*. (2<sup>nd</sup> ed.) Estados Unidos, New York: Springer-Verlag.
- Hagger, M. S., y Chatzisarantis, N. L. D. (2010). Ego depletion and strength model of self-control: A meta-analysis. *Psychological Bulletin*, 136(4), 495-525.
- Hall, P. (1927). The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika*, 19(3/4), 240–245.
- Hedges, L. V., (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, 9(1), 61-85.
- Hung, H. M. J., O'Neill, R. T., Bauer, P., y Kohne, K. (1997). The Behavior of the P Value When the Alternative Hypothesis is True. *Biometrics*, 53, 11-22.
- Ihaka, R. y Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314.
- Ioannidis, J. P. A., Cappelleri, J. C., Sacks, H. S. y Lau, J. (1997). The relationship between study design, results, and reporting of randomized clinical trials of HIV infection. *Controlled Clinical Trials*, 18, 431-444.
- Ioannidis, J. P. A., y Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, 4, 245-253.
- Irwin, J. O. (1927). On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's type II. *Biometrika*, 19(3/4), 225–239.
- Kocak, M. (2015). Meta-analysis of univariate p-values. *Communication in Statistics: Simulation and Computation*, 2(4), 179-185.
- López-López, J. A., Rubio-Aparicio, M., Sánchez-Meca, J., y Marín-Martínez, F. (2013, Septiembre). Distribution of effect size and sample size in meta-analysis in the

- psychological field. Paper presented at the XIII Congress of Methodology of the Social and Health Sciences (Tenerife, Spain).
- Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253), 68-78.
- McAuley, L., Pham, B., Tugwell, P., y Moher, D. (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses?. *Lancet*, 356, 1228-1231.
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8, 157-159.
- Pardo, A., Ruiz, M. A., y San Martín, R. (2009). *Análisis de datos en ciencias sociales y de la salud I*. Madrid. España: Síntesis.
- Revuelta, J., y Ponsoda, V., (2005). *Fundamentos de Estadística*. Madrid, España: UNED.
- Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, 59, 464-468.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rothstein, H. R., Sutton, A. J., y Borenstein, M. (2005). *Publication bias in meta-analyses: prevention, assessment and adjustments*. Chichester, Inglaterra: Wiley.
- Scherer, R. W. y Langenberg, P. (2007). Full publication of results initially presented in abstracts. *The Cochrane Database of Methodology Reviews*, 2.
- Simonsohn, U., Nelson, L. D., y Simmons, J. P., (2013). P-Curve: A key to the file drawer. *Journal of Experimental Psychology: General*, Forthcoming.
- Smart, R. G. (1964). The importance of negative results in psychological research. *Canadian Psychologist*, 5, 225-232.
- Smith, M. L. (1980). Sex bias in counseling and psychotherapy. *Psychological Bulletin*, 87, 392-407.
- Spivak, M. (1988). *Calculus: Cálculo Infinitesimal*. Barcelona, España: Reverté.
- Stanley, T. D. (2005). Beyond publication bias, *Journal of Economic Surveys*, 19, 309-345.
- Stanley, T. D., y Doucouliagos, C. (2007). Identifying and correcting publication selection bias in the efficacy-wage literature: heckman meta-regression. *Economic Series*, 11.
- Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., y Williams, R. M. (1949). *The american soldier: Adjustment ring army life*. Princeton University Press, 113(1), 43-53.

- Suero, M., y Botella, J. (2015). *Demostración de la distribución uniforme de los valores p*. Manuscrito no publicado, Facultad Autónoma de Madrid, España.
- Tierney, J. F., Clarke, M., y Stewart, L. A. (2000). Is there bias in the publication of individual patient data meta-analyses?. *International Journal of Technology Assessment in Health Care*, 16, 657-667.
- Trautmann, H., Steuer, D., Mersmann, O., y Bornkamp, B. (2014). Truncnorm: Truncated normal distribution (Versión 1.0-7). Recuperado de <http://CRAN.R-project.org/package=truncnorm>
- Van Assen, M. A. L. M., van Aert, R. C. M., y Wicherts, J. M. (2014). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, advanced online publication.
- Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, 36, 3.
- Viechtbauer, W. (2016). *Meta-analysis with R: The metafor package*. GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany.
- Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco, Estados Unidos de América: WH Freeman.
- Yaphe, J., Edman, R., Knishkowsky, B., y Herman, J. (2001). The association between funding by commercial interests and study outcome in randomized controlled drug trials. *Family Practice*, 18, 565-568.

## APÉNDICE

Tabla 1.

Medias y desviaciones típicas de las estimaciones por Efectos fijos, Efectos aleatorios y el cálculo basado en los valores  $p$  no reescalados, para una delta de 0,25 en función del número de estudios incluidos en el meta-análisis.

		Número de estudios									
		5	10	15	20	25	30	35	40	45	50
Efecto Fijo	Media	0,2448	0,2461	0,2472	0,2466	0,2462	0,2462	0,2466	0,2466	0,2457	0,2460
	Desv. típica	0,1156	0,0799	0,0664	0,0578	0,0516	0,0466	0,0428	0,0404	0,0380	0,0360
Efectos Aleatorios	Media	0,2456	0,2466	0,2477	0,2470	0,2465	0,2466	0,2469	0,2469	0,2459	0,2463
	Desv. típica	0,1160	0,0801	0,0665	0,0579	0,0517	0,0466	0,0429	0,0404	0,0380	0,0360
Proc. Valores $p$	Media	0,2468	0,2488	0,2500	0,2495	0,2491	0,2490	0,2495	0,2496	0,2486	0,2489
	Desv. típica	0,1191	0,0825	0,0685	0,0597	0,0534	0,0482	0,0443	0,0418	0,0392	0,0373

Tabla 2.

Medias y desviaciones típicas de las estimaciones por Efectos fijos, Efectos aleatorios y el cálculo basado en los valores  $p$  no reescalados, para una delta de 0,5 en función del número de estudios incluidos en el meta-análisis.

		Número de estudios									
		5	10	15	20	25	30	35	40	45	50
Efecto Fijo	Media	0,4920	0,4932	0,4933	0,4923	0,4926	0,4914	0,4924	0,4919	0,4928	0,4916
	Desv. típica	0,1169	0,0816	0,0664	0,0578	0,0516	0,0468	0,0437	0,0407	0,0382	0,0367
Efectos Aleatorios	Media	0,4936	0,4944	0,4942	0,4932	0,4933	0,4921	0,4931	0,4925	0,4933	0,4922
	Desv. típica	0,1172	0,0817	0,0665	0,0578	0,0516	0,0468	0,0437	0,0407	0,0382	0,0367
Proc. Valores $p$	Media	0,4963	0,4982	0,4989	0,4979	0,4981	0,4971	0,4982	0,4976	0,4987	0,4975
	Desv. típica	0,1196	0,0841	0,0687	0,0596	0,0534	0,0485	0,0451	0,0421	0,0396	0,0380

Tabla 3.

Medias y desviaciones típicas de las estimaciones por Efectos fijos, Efectos aleatorios y el cálculo basado en los valores  $p$  no reescalados, para una delta de 0,75 en función del número de estudios incluidos en el meta-análisis.

		Número de estudios									
		5	10	15	20	25	30	35	40	45	50
Efecto Fijo	Media	0,7401	0,7404	0,7387	0,7388	0,7393	0,7384	0,7386	0,7383	0,7383	0,7385
	Desv. típica	0,1200	0,0832	0,0678	0,0591	0,0533	0,0474	0,0443	0,0419	0,0393	0,0377
Efectos Aleatorios	Media	0,7425	0,7422	0,7402	0,7401	0,7404	0,7394	0,7396	0,7392	0,7392	0,7393
	Desv. típica	0,1201	0,0832	0,0678	0,0591	0,0532	0,0474	0,0442	0,0419	0,0393	0,0377
Proc. Valores $p$	Media	0,7470	0,7487	0,7469	0,7472	0,7477	0,7468	0,7473	0,7470	0,7471	0,7470
	Desv. típica	0,1230	0,0852	0,0698	0,0608	0,0548	0,0488	0,0457	0,0432	0,0407	0,0389

Tabla 4.

Medias y desviaciones típicas de las estimaciones por Efectos fijos, Efectos aleatorios y el cálculo basado en los valores  $p$  no reescalados, para una delta de 0,25 en función de distinto niveles de sesgo de publicación.

		Número de estudios										
		0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Efecto Fijo	Media	0,578	0,469	0,410	0,370	0,335	0,310	0,293	0,279	0,266	0,255	0,247
	Desv. típica	0,025	0,025	0,024	0,024	0,020	0,020	0,020	0,020	0,019	0,019	0,019
Efectos Aleatorios	Media	0,578	0,470	0,411	0,370	0,336	0,311	0,294	0,279	0,266	0,255	0,247
	Desv. típica	0,025	0,025	0,024	0,023	0,022	0,021	0,020	0,020	0,019	0,018	0,017
Proc. Valores $p$	Media	0,576	0,491	0,432	0,388	0,349	0,320	0,300	0,285	0,270	0,258	0,250
	Desv. típica	0,025	0,021	0,020	0,019	0,020	0,019	0,019	0,019	0,019	0,018	0,018

Tabla 5.

Medias y desviaciones típicas de las estimaciones por Efectos fijos, Efectos aleatorios y el cálculo basado en los valores  $p$  no reescalados, para una delta de 0,50 en función de distinto niveles de sesgo de publicación.

		Número de estudios										
		0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Efecto Fijo	Media	0,663	0,635	0,611	0,590	0,571	0,554	0,539	0,525	0,513	0,501	0,491
	Desv. típica	0,021	0,020	0,020	0,019	0,018	0,018	0,017	0,017	0,016	0,016	0,015
Efectos Aleatorios	Media	0,663	0,635	0,611	0,590	0,571	0,554	0,539	0,526	0,513	0,502	0,491
	Desv. típica	0,021	0,020	0,020	0,019	0,018	0,018	0,017	0,017	0,016	0,016	0,015
Proc. Valores $p$	Media	0,659	0,637	0,617	0,598	0,581	0,564	0,548	0,534	0,521	0,508	0,497
	Desv. típica	0,022	0,020	0,020	0,019	0,018	0,017	0,017	0,016	0,015	0,015	0,014

Tabla 6.

Medias y desviaciones típicas de las estimaciones por Efectos fijos, Efectos aleatorios y el cálculo basado en los valores  $p$  no reescalados, para una delta de 0,75 en función de distinto niveles de sesgo de publicación.

		Número de estudios										
		0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Efecto Fijo	Media	0,803	0,797	0,791	0,780	0,774	0,769	0,764	0,759	0,749	0,744	0,739
	Desv. típica	0,023	0,022	0,022	0,022	0,022	0,021	0,021	0,021	0,021	0,020	0,020
Efectos Aleatorios	Media	0,803	0,797	0,791	0,780	0,774	0,769	0,764	0,759	0,749	0,744	0,740
	Desv. típica	0,023	0,022	0,022	0,022	0,022	0,021	0,021	0,021	0,021	0,020	0,020
Proc. Valores $p$	Media	0,801	0,797	0,792	0,783	0,779	0,775	0,770	0,766	0,757	0,753	0,749
	Desv. típica	0,024	0,024	0,024	0,023	0,023	0,023	0,023	0,023	0,023	0,022	0,022

Tabla 7.

Medias y desviaciones típicas de las estimaciones por el procedimiento basado en los valores  $p$  reescalados para un delta de 0,25, 0,50 y 0,75 en función del número de estudios significativos incluidos en el meta-análisis.

		Número de estudios									
		5	10	15	20	25	30	35	40	45	50
$\delta = 0,25$	Media	0,1833	0,2184	0,2320	0,2356	0,2375	0,2395	0,2407	0,2422	0,2442	0,2443
	Desv. típica	0,3190	0,2020	0,1574	0,1351	0,1200	0,1071	0,0991	0,0922	0,0867	0,0822
$\delta = 0,50$	Media	0,4625	0,4856	0,4898	0,4943	0,4953	0,4946	0,4958	0,4965	0,4959	0,4966
	Desv. típica	0,2229	0,1382	0,1130	0,0950	0,0836	0,0766	0,0708	0,0657	0,0621	0,0593
$\delta = 0,75$	Media	0,7372	0,7440	0,7448	0,7479	0,7487	0,7491	0,7492	0,7493	0,7485	0,7496
	Desv. típica	0,1544	0,1043	0,0837	0,0706	0,0631	0,0583	0,0538	0,0497	0,0475	0,0444

Tabla 8.

Medias y desviaciones típicas de las estimaciones por el procedimiento basado en los valores  $p$  no reescalados, en su versión ponderada y sin ponderar, para una delta de 0,50 en función del número de estudios incluidos en el meta-análisis (10, 20 o 40) y de la heterogeneidad de la muestra de los estudios primarios.

	$N_1$	20	20	20	20	20	20	20	20	20	20
	$N_2$	20	40	60	80	100	120	140	160	180	200
k = 40	Media	0,4932	0,4958	0,4965	0,4973	0,4970	0,4976	0,4976	0,4972	0,4981	0,4984
Sin ponderar	Desv. típica	0,0730	0,0604	0,0532	0,0490	0,0450	0,0427	0,0399	0,0381	0,0369	0,0353
k = 40	Media	0,4932	0,4965	0,4978	0,4983	0,4979	0,4984	0,4986	0,4990	0,4993	0,4995
Ponderado	Desv. típica	0,0730	0,0603	0,0525	0,0472	0,0433	0,0405	0,0373	0,0350	0,0336	0,0320
k = 20	Media	0,4924	0,4963	0,4970	0,4961	0,4965	0,4974	0,4979	0,4979	0,4986	0,4980
Sin ponderar	Desv. típica	0,1029	0,0860	0,0758	0,0692	0,0640	0,0601	0,0566	0,0541	0,0519	0,0493
k = 20	Media	0,4924	0,4965	0,4977	0,4977	0,4982	0,4985	0,4995	0,4992	0,4991	0,4991
Ponderado	Desv. típica	0,1029	0,0862	0,0747	0,0673	0,0617	0,0572	0,0535	0,0496	0,0476	0,0449
k = 10	Media	0,4937	0,4954	0,4968	0,4966	0,4958	0,4965	0,4969	0,4985	0,4985	0,4983
Sin ponderar	Desv. típica	0,1457	0,1206	0,1060	0,0991	0,0900	0,0845	0,0805	0,0766	0,0727	0,0702
k = 10	Media	0,4937	0,4951	0,4973	0,4976	0,4969	0,4969	0,4982	0,4996	0,4996	0,4991
Ponderado	Desv. típica	0,1457	0,1201	0,1046	0,0962	0,0858	0,0794	0,0747	0,0712	0,0667	0,0634