anales de psicología / annals of psychology
2024, vol. 40, nº 2 (may), 344-354
https://doi.org/10.6018/analesps.591831

ISSN online: 1695-2294.
https://revistas.um.es/analesps

Meta-análisis: Intervalos de confianza e Intervalos de predicción

Juan Botella[*,1] y Julio Sánchez-Meca[2]

1 Universidad Autónoma de Madrid, Madrid (España)
2 Universidad de Murcia, Murcia (España)

Resumen: En los informes meta-analíticos se suelen reportar varios tipos de intervalos, hecho que ha generado cierta confusión a la hora de interpretarlos. Los intervalos de confianza reflejan la incertidumbre relacionada con un número, el tamaño del efecto medio paramétrico. Los intervalos de predicción reflejan el tamaño paramétrico probable en cualquier estudio de la misma clase que los incluidos en un meta-análisis. Su interpretación y aplicaciones son diferentes. En este artículo explicamos su diferente naturaleza y cómo se pueden utilizar para responder preguntas específicas. Se incluyen ejemplos numéricos, así como su cálculo con el paquete *metafor* en R.
Palabras clave: Intervalo de confianza. Intervalo de predicción. Meta-análisis.

Title: Meta-analysis: Confidence intervals and Prediction intervals.
Abstract: Several types of intervals are usually employed in meta-analysis, a fact that has generated some confusion when interpreting them. Confidence intervals reflect the uncertainty related to a single number, the parametric mean effect size. Prediction intervals reflect the probable parametric effect size in any study of the same class as those included in a meta-analysis. Its interpretation and applications are different. In this article we explain in detail their different nature and how they can be used to answer specific questions. Numerical examples are included, as well as their computation with the *metafor* R package.
Keywords: Confidence interval. Prediction interval. Meta-analysis.

Introducción

En meta-análisis se utilizan varios tipos de intervalos, una circunstancia que ha generado cierta confusión en su interpretación, ya que su naturaleza es diferente. En este escrito vamos a exponer en qué consisten los dos tipos principales de intervalos, incluyendo ejemplos numéricos. Empezaremos por recordar la diferencia entre los *modelos de efecto fijo* (*EF*) y los *modelos de efectos aleatorios* (*EA*), cuestión clave para comprender bien la cuestión que tratamos aquí. Después expondremos las características de los dos tipos principales de intervalos que se usan en meta-análisis, pasando por un antecedente de amplio uso en los estudios de generalización de la validez. Luego ilustraremos todo ello con un ejemplo numérico. Después de destacar el papel de los intervalos de predicción para reflejar la heterogeneidad de los efectos abordaremos la discusión y una recomendación general.

Para esta exposición representaremos al parámetro que refleja el efecto que estamos estudiando por θ y asumiremos que disponemos de k estimaciones independientes.

Modelos de efecto fijo y de efectos aleatorios

Tal y como ya hemos explicado en otros lugares (e.g., Botella y Sánchez-Meca, 2015), en un modelo de *EF* (también llamado *modelo de efecto común*) se asume que el análisis que se está haciendo se refiere a un único valor paramétrico (θ). Cada estudio primario aporta una estimación del parámetro, $(\hat{\theta}_i)$.

Las estimaciones tienen una varianza, que llamamos *varianza de muestreo*, que interpretamos como imprecisión en la estimación, ya que es debida a que en cada estudio se trabaja con una muestra aleatoria concreta, distinta de las de los demás estudios. Aunque en ocasiones se asume que es conocida (e.g., Higgins, Thompson y Spiegelhalter, 2009), en realidad esta varianza de muestreo es a su vez un valor estimado, que representamos por $\text{var}(\hat{\theta}_i)$. Como cada estudio tiene un tamaño muestral diferente, contamos con una varianza distinta en cada estudio; de ahí el subíndice *i* de las varianzas de muestreo. Por tanto,

a) si los tamaños muestrales de los estudios fueran iguales, las varianzas de muestreo paramétricas de los estudios serían todas iguales (aunque sus estimaciones podrían ser distintas), y
b) en un caso hipotético en el que los tamaños muestrales tuvieran un tamaño indefinidamente grande, tendiendo a infinito, la varianza empírica tendería a 0.

En un modelo de *EA* se acepta que el valor paramétrico de cada estudio es diferente. Esos valores paramétricos tienen una distribución que habitualmente se asume normal, con media μ_θ y varianza σ^2_θ (esta varianza también se suele representar como τ^2 y recibe los nombres de *varianza inter-estudios, varianza específica* o *varianza de heterogeneidad*). Eso significa que la varianza del estimador del tamaño del efecto de cada estudio tiene dos fuentes de variación. Por un lado, la varianza de los efectos paramétricos; por otro lado, la varianza de muestreo del efecto paramétrico de cada estudio, que es similar a la varianza del modelo de *EF* para ese estudio en particular. Por tanto,

a) si los tamaños muestrales de los estudios fueran iguales, las varianzas de muestreo paramétricas de los estudios serían todas iguales[1] (aunque sus estimaciones podrían ser distintas), y

*** Correspondence address [Dirección para correspondencia]:**
Juan Botella. Universidad Autónoma de Madrid, Facultad de Psicología, Campus de Cantoblanco, c/ Ivan Pavlov, 6, 28049 Madrid (España).
E-mail: juan.botella@uam.es
(Artículo recibido: 6-11-2023; revisado: 15-01-2024; aceptado: 20-01-2024)

[1] Siempre y cuando el propio valor de θ_i no esté involucrado en la varianza, como ocurre con la diferencia media tipificada (*d* de Cohen; Suero, Botella, Durán y Blázquez-Rincón, 2023). No ocurre así con otros índices de tamaño del efecto, como por ejemplo la correlación de Pearson transformada mediante la fórmula de Fisher.

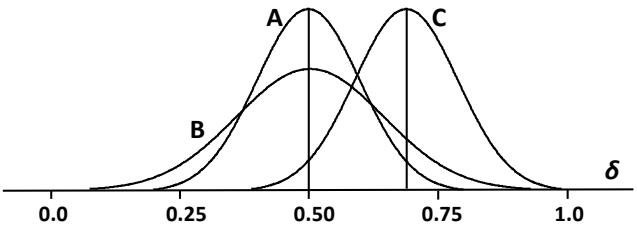
b) en un caso hipotético en el que los tamaños muestrales tuvieran un tamaño indefinidamente grande, tendiendo a infinito, la varianza empírica tendería a τ^2.

Intervalo de confianza (del efecto medio)

El intervalo de confianza más extendido en meta-análisis es el que se refiere a la magnitud del efecto de interés. En un modelo de *EF* la magnitud del efecto es única: θ. Su estimador puntual es la combinación ponderada de las *k* estimaciones, $\hat{\theta}_i$, representando por $\hat{\theta}_\bullet$ a dicha estimación combinada. Sin embargo, en el modelo de *EA* no es así, sino que se asume que existe una distribución de valores paramétricos. Una magnitud de máximo interés en este modelo es la estimación del *valor medio* de los efectos paramétricos, μ_θ, ya que con frecuencia interesa conocer el "efecto medio" de una intervención. Esta magnitud se estima también a través de un promedio ponderado de las *k* estimaciones independientes, $\hat{\mu}_\theta$. Una de las fuentes habituales de confusión está precisamente en que en ambos tipos de modelos el valor de interés se estima a través de una combinación ponderada de las estimaciones independientes proporcionadas por los *k* estudios. Pero en el caso del modelo de *EA* se trata de estimar en qué valor está centrada la distribución de efectos paramétricos; por tanto, también se estima un valor concreto.

Sin embargo, ese valor no nos dice nada acerca de la dispersión de esos efectos paramétricos. Un único valor paramétrico, μ_θ, es insuficiente para describir eficazmente unos efectos que son heterogéneos (Borenstein, 2019b). En la figura 1 aparecen 3 distribuciones de efectos paramétricos de la *diferencia de medias tipificada* (o *d* de Cohen). Las curvas A y B están centradas en el mismo valor ($\mu_{\theta(A)} = \mu_{\theta(B)}$), pero tienen diferentes varianzas ($\tau_A^2 < \tau_B^2$). Por el contrario, las curvas A y C tienen diferentes valores centrales ($\mu_{\theta(C)} > \mu_{\theta(A)}$), pero la misma varianza ($\tau_A^2 = \tau_C^2$).

Figura 1
Representación de tres distribuciones de efectos paramétricos.



En resumen, este primer intervalo es un intervalo de confianza clásico, con el que se estima un valor único y concreto: el único valor paramétrico en el modelo de *EF* o el valor central de la distribución de valores paramétricos en el modelo de *EA*. Las fórmulas son[2]:

$$EF:\ \hat{\theta}_\bullet \pm z_{\alpha/2} \cdot \hat{\sigma}_{\hat{\theta}_\bullet} \quad (1)$$

$$EA:\ \hat{\mu}_\theta \pm z_{\alpha/2} \cdot \hat{\sigma}_{\hat{\mu}_\theta} \quad (2)$$

En el modelo de *EF*, el efecto paramétrico se estima mediante una media ponderada de los efectos individuales:

$$\hat{\theta}_\bullet = \frac{\sum \hat{w}_{EF,i} \cdot \hat{\theta}_i}{\sum \hat{w}_{EF,i}} \quad (3)$$

Siendo $\hat{w}_{FE,i}$ el factor de ponderación de cada estudio, que se obtiene mediante el inverso de la varianza de muestreo, $\text{var}(\hat{\theta}_i)$:

$$\hat{w}_{EF,i} = \frac{1}{\text{var}\left(\hat{\theta}_i\right)} \quad (4)$$

El error típico del efecto estimado se calcula mediante:

$$\hat{\sigma}_{\hat{\theta}_\bullet} = \frac{1}{\sqrt{\sum \hat{w}_{EF,i}}} \quad (5)$$

Y $z_{\alpha/2}$ es la puntuación de la distribución normal tipificada que se corresponde con el percentil $(\alpha/2)100\%$ (en valor absoluto).

En el modelo de *EA*, el efecto medio paramétrico también se estima mediante una media ponderada de los efectos individuales, pero el factor de ponderación es diferente:

$$\hat{\mu}_\theta = \frac{\sum \hat{w}_{EA,i} \cdot \hat{\theta}_i}{\sum \hat{w}_{EA,i}} \quad (6)$$

siendo $\hat{w}_{EA,i}$ el factor de ponderación de cada estudio, que se obtiene mediante el inverso de la suma de la varianza de muestreo, $\text{var}(\hat{\theta}_i)$ y la varianza inter-estudios, $\hat{\tau}^2$:

$$\hat{w}_{EA,i} = \frac{1}{\text{var}\left(\hat{\theta}_i\right) + \hat{\tau}^2} \quad (7)$$

La varianza inter-estudios se estima por alguno de los métodos propuestos en la literatura, por ejemplo, por máxima verosimilitud restringida (cf. e.g., Sánchez-Meca y Marín-Martínez, 2008; véase también Suero, Botella y Durán, 2023). El error típico del efecto medio estimado se calcula mediante:

[2] Asumiremos que la diferencia de medias tipificada se distribuye según el modelo normal, algo que sabemos positivamente que es incorrecto, aunque aproximado. Lo haremos por comodidad y simplicidad, pero también porque en el programa R que usaremos para los ejemplos se hace así. En realidad, la distribución de la *d* de Cohen es la *t* de *Student* no centrada (Suero, Botella, Durán y Blázquez-Rincón, 2023).

$$\hat{\sigma}_{\hat{\mu}_\theta} = \frac{1}{\sqrt{\sum \hat{w}_{EA,i}}} \quad (8)$$

Su interpretación es la tradicional para este tipo de intervalos: el IC95% proporciona una horquilla de valores respecto al que tenemos una confianza del 95% de que incluye el valor de interés (θ bajo el modelo de *EF* y μ_θ bajo el modelo de *EA*). Dicho de otra forma, si repitiésemos las acciones que nos han conducido a ese intervalo un número indefinidamente grande de veces, en las mismas condiciones, entonces aproximadamente el 95% de los intervalos incluirían al valor de interés.

El IC95% presentado en la ecuación (2) para el modelo de *EA* no tiene en cuenta la incertidumbre en la estimación del error estándar del efecto medio, $\hat{\sigma}_{\hat{\mu}_\theta}$, ni de la varianza inter-estudios, τ^2. Como consecuencia, la amplitud confidencial tiende a infraestimar la verdadera amplitud confidencial de dicho intervalo. Para resolver este problema, Hartung y Knapp (2001; cf. también Sidik y Jonkman, 2002) propusieron una fórmula alternativa para calcular el IC95% que tiene en cuenta dicha incertidumbre. Por una parte, el método de Hartung-Knapp utiliza una distribución *t* de *Student* con $k - 1$ grados de libertad en lugar de la distribución normal tipificada. En segundo lugar, aplica un factor de corrección a la varianza del efecto medio. Siendo *q* el factor de corrección, éste se obtiene mediante:

$$q = \frac{1}{k-1} \sum \hat{w}_{EA,i} \left(\hat{\theta}_i - \hat{\mu}_\theta \right)^2 \quad (9)$$

Aunque es poco probable, el factor de corrección, *q*, puede ser menor que 1, en cuyo caso la varianza por este método sería menor que la varianza original, dando lugar a intervalos de confianza más estrechos. Con objeto de evitar esta circunstancia, Hartung y Knapp (2001) recomiendan truncar el valor de *q*, de forma que hacen $q^* = \text{máx}[1, q]$. Así pues, la varianza del efecto medio según el método de Hartung-Knapp viene dada por (Partlett y Riley, 2017):

$$\hat{\sigma}^2_{HK,\hat{\mu}_\theta} = q \cdot \hat{\sigma}^2_{\hat{\mu}_\theta} \quad (10)$$

En resumen, el intervalo de confianza por el método de Hartung-Knapp se obtiene mediante:

$$EA_{HK} : \hat{\mu}_\theta \pm t_{(k-1),\alpha/2} \cdot \sqrt{\hat{\sigma}^2_{HK,\hat{\mu}_\theta}} \quad (11)$$

Conviene tener en cuenta, no obstante, que hay autores que no recomiendan truncar el valor *q* cuando sea menor que 1. Estudios de simulación han demostrado un mejor ajuste al nivel de confianza y mayor potencia cuando se utiliza el método de Hartung-Knapp sin truncar que cuando se sigue la recomendación de los autores de truncar el valor de *q* (Viechtbauer et al., 2015). Así, el método de cálculo del IC95% propuesto por Hartung y Knapp no incluye dicho truncamiento en el programa *metafor* de R. El módulo de meta-análisis que incorpora la versión 28 del programa IBM SPSS incluye las dos opciones, truncada y sin truncar (cf. IntHout, Ioannidis y Borm, 2014; Jackson et al., 2017 para una revisión de las alternativas en el uso del método de Hartung-Knapp).

Intervalo de predicción

Como los intervalos de predicción tienen un antecedente directo en los llamados intervalos de credibilidad o de validez, vamos a exponer primero estos y luego expondremos el desarrollo que lleva desde los intervalos de validez a los intervalos de predicción. En la literatura clásica sobre la generalización de la validez se usan indistintamente los términos *intervalo de credibilidad* e *intervalo de validez* (Hunter y Schmidt, 1990). Pero queremos subrayar que más recientemente se ha extendido un uso del primer término (intervalo de credibilidad) dentro del enfoque bayesiano (Schmid, Carlin y Welton, 2021), aunque aquí no nos detendremos en él.

Los *intervalos de credibilidad* fueron propuestos por Hunter y Schmidt (1990) para un objetivo diferente del de los intervalos de confianza, en el marco de un tipo de meta-análisis conocido como *generalización de la validez*. Sólo es posible calcularlos bajo modelos de *EA*, pero en las ciencias sociales y de la salud en general, y en la psicología en particular, se asume que los modelos de *EA* reflejan mejor el escenario de los fenómenos que nos interesan y son los modelos que se asumen por defecto. Una consecuencia de asumir un escenario de *EA* tiene que ver con la expectativa de los efectos en estudios futuros. Como se aprecia en la figura 1, para caracterizar una distribución de efectos hacen falta como mínimo dos magnitudes: el valor central o efecto medio y la varianza de los efectos. El primero nos dice qué valor promedio tendrán los efectos de los estudios futuros del tipo al que se refiere la población de estudios que tenemos entre manos. El segundo nos dice cómo de heterogéneos son dichos efectos. Supongamos que hablamos del impacto de una terapia según se refleja en una variable cuantitativa, en comparación con pacientes no tratados, en lista de espera. Cada estudio con un diseño de grupos aleatorios que implique estas dos condiciones tendría un efecto paramétrico perteneciente a una distribución del índice *d* de Cohen, con media μ_θ y varianza τ^2. Estimar μ_θ mediante $\hat{\mu}_\theta$ y su intervalo de confianza sólo nos permitirá establecer cuál sería el *efecto paramétrico medio* en un número infinitamente grande de futuros estudios similares. En otras palabras, nos permite conjeturar dónde está centrada la distribución de los efectos.

Sin embargo, nos interesa saber también qué oscilaciones cabe esperar en dichos efectos, ya que lo que nos interesa es cada una de las aplicaciones futuras. En particular, si los efectos paramétricos fueran muy homogéneos en torno a μ_θ,

entonces la toma de decisiones sería muy directa y poco controvertida. Pero, ¿qué ocurre si esos efectos paramétricos fueran muy heterogéneos? Cabría la posibilidad de que en el próximo estudio de ese tipo el efecto fuese muy grande, mucho mayor que el efecto medio, pero también podría ser muy pequeño, o incluso nulo o negativo. Dependiendo del tipo de problema y de sus consecuencias, podría ser inaceptable que la intervención tuviese un efecto muy pequeño o contrario a su eficacia. Conviene tener una idea, por tanto, de cuál es el efecto más pequeño que es razonable esperar.

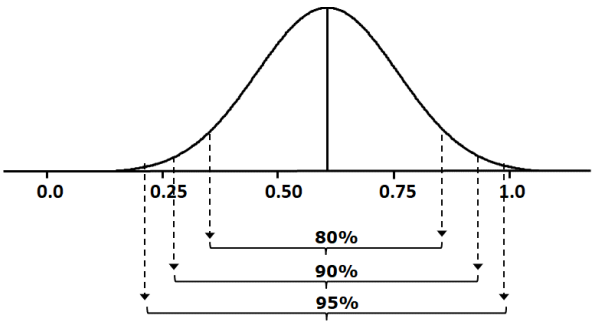
Hunter y Schmidt (1990) propusieron llamar *intervalos de credibilidad* a los que informan de horquillas referidas a valores paramétricos. Se obtienen mediante:

$$\hat{\mu}_\theta \pm z_{\alpha/2} \cdot \hat{\tau} \qquad (12)$$

Por ejemplo, supongamos que estamos trabajando con la diferencia de medias tipificada y que estimamos que el efecto medio es 0.60 y la varianza de dichos efectos paramétricos es 0.04 (la desviación típica es 0.2). La distribución estimada asumiendo la normalidad, mediante (12), es la que aparece en la figura 2 para tres niveles de confianza alternativos. Entre los valores 0.208 y 0.992 se encuentra el 95% central de los valores paramétricos, entre los valores 0.271 y 0.929 el 90% y entre los valores 0.344 y 0.856 el 80%. Con esos resultados se puede concluir con afirmaciones como las siguientes, que como se puede apreciar no se refieren exclusivamente al valor medio de la distribución de efectos:

a) con probabilidad aproximadamente .90, en un nuevo estudio sobre una intervención de este tipo el efecto estará entre 0.271 y 0.929, y con probabilidad .80 estará entre 0.344 y 0.856.
b) con probabilidad aproximadamente .95, en un nuevo estudio sobre una intervención de este tipo el efecto será igual o mayor que 0.271, y con probabilidad .90 será de al menos 0.344 (intervalos de predicción unilaterales).

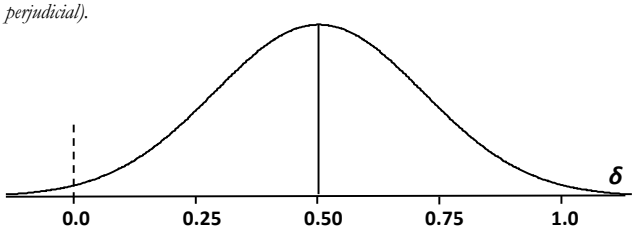
Figura 2
Representación de tres intervalos de credibilidad.



Es importante volver a destacar que estas afirmaciones no se refieren a un valor medio, sino al efecto de un nuevo ejemplar de esta intervención concreta, como por ejemplo la del próximo estudio. Lo que se afirma en el ejemplo b) tiene la importancia de que nos proporciona un *valor suelo* de los efectos. A veces no es aconsejable aplicar una intervención si no va a tener un cierto efecto mínimo, ya que no queremos arriesgarnos a que el efecto sea demasiado pequeño, o incluso nulo o negativo. Las afirmaciones de este tipo solo implican al límite inferior del intervalo, ya que se refieren al efecto que se obtendrá, como mínimo, con la probabilidad indicada.

Por otro lado, una distribución de efectos muy variable es probable que incluya efectos contrarios. Por ejemplo, en la figura 3 se muestra una distribución de efectos con media 0.50 y varianza 0.09 (desviación típica 0.3). Los valores que están a dos desviaciones típicas del valor central son 0.50 ±2·0.30: [1.10; -0.10]. Por tanto, cabe la posibilidad de que el efecto sea negativo, lo que quiere decir que la intervención no sólo podría no ser beneficiosa, sino que podría resultar perjudicial. En el ejemplo esa probabilidad es pequeña [P(z ≤ (0 - 0.5)/ 0.3) = .0475], pero aun así podría resultar inaceptable.

Figura 3
Representación de una distribución de efectos que se extiende desde valores negativos (efecto perjudicial).



En resumen, los intervalos de credibilidad, creados en el contexto del tipo de meta-análisis conocido como generalización de la validez, proporcionan horquillas estimadas de valores paramétricos. Reflejan probabilidades de que un futuro nuevo estudio tenga un efecto comprendido entre dos valores o que dicho efecto sea al menos igual a un cierto valor.

Pasemos ahora a los *intervalos de predicción* (Higgins, Thompson, & Spiegelhalter, 2009; Riley, Higgins, & Deeks, 2011). Son conceptualmente idénticos a los de credibilidad y se pueden considerar un desarrollo y sofisticación de éstos. Sin embargo, mientras los intervalos de credibilidad apenas son utilizados fuera del ámbito de los estudios de generalización de la validez, los de predicción son utilizados en campos muy variados y su presencia es cada vez mayor en los meta-análisis tanto de psicología como de otras disciplinas, como la medicina en general.

El objetivo de los intervalos de predicción es el mismo que el de los intervalos de credibilidad: ofrecer un rango de valores probables para el efecto paramétrico de un futuro nuevo estudio del mismo tipo de los que se han incluido en el meta-análisis. Las diferencias con los intervalos de credibilidad son más bien técnicas: asumen e incluyen en el modelo la incertidumbre asociada a la estimación tanto de μ_θ como de τ^2. En el intervalo de credibilidad, al calcular $\hat{\mu}_\theta \pm z_{\alpha/2} \cdot \hat{\tau}$ se está asumiendo que ambos valores son exactos. Reconociendo que son estimaciones y, por tanto, valores imprecisos, se tiene en

cuenta la incertidumbre derivada de dicho proceso de estimación. En concreto, si asumimos que el valor central de la distribución está en el rango que proporciona su intervalo de confianza, entonces el menor valor del efecto que queremos identificar estará a una cierta distancia de su límite inferior, no del valor central del intervalo. Lo mismo ocurre con el límite superior del intervalo. Por otro lado, al reconocer la incertidumbre asociada a esas dos magnitudes el modelo de distribución ya no es el normal, sino la *t* de *Student* con (*k*-2) grados de libertad. Hay cierta controversia respecto a los grados de libertad apropiados, pero muchos autores siguen la sugerencia de Higgins et al (2009) de utilizar la distribución t_{k-2} como opción razonable y práctica (e.g.., Borenstein et al, 2021; Stijnen, White y Schmid, 2021).

En resumen, el intervalo de predicción se puede obtener a través de (Higgins et al, 2009),

$$\hat{\mu}_\theta \pm {}_{\alpha/2}t_{k-2} \cdot \sqrt{\hat{\tau}^2 + \hat{\sigma}^2_{\hat{\mu}_\theta}} \qquad (13)$$

El intervalo de predicción definido en (13) utiliza la varianza originalmente propuesta para el efecto medio en un modelo de *EA*, $\hat{\sigma}^2_{\hat{\mu}_\theta}$, que se calcula mediante el cuadrado de la ecuación (8). En su lugar, es más recomendable utilizar la varianza propuesta por Hartung y Knapp, $\hat{\sigma}^2_{HK,\hat{\mu}_\theta}$, definida en la ecuación (10), que tiene en cuenta la incertidumbre en la estimación de la varianza inter-estudios y de la varianza muestral del efecto medio (Partlett y Riley, 2017):

$$\hat{\mu}_\theta \pm {}_{\alpha/2}t_{k-2} \cdot \sqrt{\hat{\tau}^2 + \hat{\sigma}^2_{HK,\hat{\mu}_\theta}} \qquad (14)$$

Una diferencia importante entre los intervalos de confianza y los de predicción es que, si se va aumentando el número de estudios progresivamente, el de confianza va reduciendo su amplitud y dicha amplitud tiende a 0 cuando el número de estudios tiende a infinito. Esto es así porque el único componente de la varianza involucrada, $\hat{\sigma}^2_{\hat{\mu}_\theta}$, disminuye al aumentar el número de estudios. Por el contrario, por muchos estudios que se añadan la amplitud del intervalo de predicción tiene un valor suelo que no puede sobrepasar. Esto es así porque uno de los componentes de la varianza involucrada, τ^2 , no cambia con el número de estudios (aunque sí mejora su estimación).

Un ejemplo

En este apartado ilustramos lo expuesto hasta aquí a través de un ejemplo numérico (el código R del apéndice permite reproducir los cálculos). En la tabla siguiente aparecen las estimaciones de la *diferencia media tipificada* (*g*, o *d* de Cohen corregida) de 15 estudios independientes y los tamaños de sus muestras. Con esos tres valores hemos obtenido las varianzas de muestreo que aparecen en la última columna con la fórmula aproximada de Hedges y Olkin (1985; fórmula [2.8] en Botella y Sánchez-Meca, 2015). También aparecen en la tabla los resultados que ofrece *metafor* (Viechtbauer, 2010) al ajustar los modelos de *EF* y de *EA*, estimando la varianza específica por máxima verosimilitud restringida. Disponemos de numerosos métodos alternativos para estimar τ^2 , pero no hay todavía un acuerdo suficiente sobre las mejores opciones (Blázquez-Rincón, Sánchez-Meca, Botella y Suero, 2023; Langan, et al, 2017, 2019; Veroniki et al, 2016). Para este ejemplo hemos elegido un método usado y recomendado con frecuencia, máxima verosimilitud restringida, que no tiene las desventajas del bien conocido método de los momentos (Viechtbauer, 2005). Detallamos en la parte derecha de la tabla los cálculos para los tipos de intervalos que hemos expuesto.

Est	N1	N2	*g*	Var_g
1	11	11	0.2700	0.18350
2	123	135	0.2991	0.01571
3	14	16	0.0584	0.13399
4	36	35	-0.0198	0.05635
5	20	20	0.7645	0.10731
6	24	23	0.9341	0.09443
7	20	12	0.1560	0.13371
8	80	75	0.2388	0.02602
9	18	18	0.0293	0.11112
10	16	16	1.2087	0.14783
11	32	20	0.4728	0.08340
12	16	18	0.7811	0.12703
13	20	20	1.0977	0.11506
14	76	58	0.8551	0.03313
15	24	16	0.2156	0.10475

Fixed-Effects Model (k = 15)
I^2 (total heterogeneity / total variability): 49.74%
H^2 (total variability / sampling variability): 1.99
Test for Heterogeneity:
Q(df = 14) = 27.8570, *p*-val = .0149
Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.4326	0.0640	6.7593	< .0001	0.3071	0.5580

Random-Effects Model (k = 15; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0690 (SE = 0.0551)
tau (square root of estimated tau^2 value): 0.2627
I^2 (total heterogeneity / total variability): 51.24%
H^2 (total variability / sampling variability): 2.05
Test for Heterogeneity:
Q(df = 14) = 27.8570, *p*-val = .0149
Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.4707	0.1007	4.6760	< .0001	0.2734	0.6680	***

Model Results (Hartung-Knapp method):

	estimate se tval df pval ci.lb ci.ub 0.4707 0.1012 4.6493 14 .0004 0.2535 0.6878 *** 95% Prediction Interval (assuming a standard normal distr.): pred se ci.lb ci.ub pi.lb pi.ub 0.4707 0.1007 0.2734 0.6680 -0.0808 1.0221 95% Prediction Interval (Hartung-Knapp method): pred se ci.lb ci.ub pi.lb pi.ub 0.4707 0.1012 0.2535 0.6878 -0.1332 1.0746

El *intervalo de confianza* es el único que se puede calcular bajo un modelo de *EF*. Las estimaciones para ambos modelos son los que proporciona directamente *metafor*, que detallamos a continuación aplicando las fórmulas (1), (2) y (11):

IC95%(EF): 0.4326 ± 1.96·0.0640: **[0.307; 0.558]**
IC95%(EA): 0.4707 ± 1.96·0.1007: **[0.273; 0.668]**
IC95%(EA por HK): 0.4707 ± 2.145·0.1012: **[0.254; 0.688]**

El valor 2.145 se corresponde con el percentil 97.5% de la distribución *t* de *Student* con k - 1 = 15 - 1 = 14 grados de libertad. El valor 0.1012 es la raíz cuadrada del valor obtenido con la ecuación (10).

Como ya hemos explicado, los intervalos de credibilidad y de predicción sólo se pueden calcular para el modelo de *EA*. En concreto, el intervalo de credibilidad o validez sería el que proporciona la fórmula (12):

IV95%: 0.4707 ± 1.96·0.2627: **[-0.044; 0.986]**

Por su parte, los intervalos de predicción son los que proporcionan las fórmulas (13) y (14) (el valor 2.16 es el que ocupa el percentil 97.5 en la distribución *t* con k-2 = 13 grados de libertad):

IP95%: 0.4707 ± 2.16·sqrt(0.2627² + 0.1007²) =
= 0.4707 ± 2.16·0.2813: **[-0.137; 1.078]**
IP95% por HK: 0.4707 ± 2.16·sqrt(0.2627² + 0.01024) =
= 0.4707 ± 2.16·0.2813: **[-0.137; 1.079]**

Adviértase que el IP95% calculado con (13), [-0.137; 1.078], no coincide con el reportado en la salida del programa *metafor*: IP95%[-0.0808; 1.0221]. Ello se debe a que la fórmula que incorpora *metafor* no es la ecuación (13), sino que asume una distribución normal tipificada en lugar de una distribución *t* de *Student*. La fórmula implementada en *metafor* es, pues (Viechtbauer, 2023, p. 32):

$$\hat{\mu}_\theta \pm z_{\alpha/2} \cdot \sqrt{\hat{\tau}^2 + \hat{\sigma}^2_{\hat{\mu}_\theta}} \tag{15}$$

Así mismo, obsérvese que el IP95% por HK calculado con la ecuación (14), [-0.137; 1.079], tampoco coincide con el reportado en la salida de *metafor*: IP95%HK[-0.1332; 1.0746]. Ello se debe a que *metafor* utiliza como grados de libertad de la distribución *t* de *Student* (k – 1) en lugar de (k – 2). Así pues, la fórmula que implementa *metafor* para construir un IP95% por el método de Hartung-Knapp es (Viechtbauer, 2023, p. 32):

$$\hat{\mu}_\theta \pm {}_{\alpha/2}t_{k-1} \cdot \sqrt{\hat{\tau}^2 + \hat{\sigma}^2_{HK,\hat{\mu}_\theta}} \tag{16}$$

Adviértase que el intervalo de predicción es más impreciso (más amplio) que el de validez o credibilidad y, a su vez, el intervalo de predicción por Hartung-Knapp es más amplio que el que no aplica este método. Los resultados se pueden interpretar de la siguiente forma. El IC95% indica, bajo el modelo de *EF*, que podemos concluir, con una confianza del 95%, que el valor paramétrico (único) implicado está en el rango 0.307 – 0.558. Bajo el modelo de *EA* el IC95% indica que podemos concluir, con una confianza del 95%, que el valor medio del efecto en futuros estudios de este tipo tenderá a ser un valor en el rango 0.273 – 0.668.

Respecto al IP95%, se interpreta concluyendo que, con una confianza del 95%, el efecto de un futuro estudio de este tipo estará en el rango -0.137 – 1.078, o bien en el rango -0.137 – 1.079, según hayamos utilizado o no el método de Hartung-Knapp. Adviértase la diferencia entre esta conclusión y la del IC95%. En el IP95% se afirma *sobre el efecto de una futura intervención*, mientras que en el IC95% se afirma *sobre el efecto medio de las futuras intervenciones*.

Cuando es importante no aplicar o recomendar intervenciones con efectos que podrían ser demasiado pequeños (o, por supuesto, intervenciones que pudieran ser perjudiciales) conviene proporcionar interpretaciones en ese sentido. Ya hemos visto en el ejemplo de la figura 2 que podemos hacerlo mediante intervalos unilaterales. Por ejemplo, en este caso podemos concluir que el efecto de la próxima intervención de este tipo será, con una confianza del 97.5%, superior a -0.137. Podemos calcular e interpretar también otros valores, por ejemplo de las siguientes formas:

a) El efecto de una futura intervención de este tipo será, con una confianza del 95%, superior a (-1.771 es el valor correspondiente al percentil 5 de la distribución *t* con 13 grados de libertad):

$$0.4707 - 1.771 \cdot 0.2813 = -0.027$$

b) Se puede estimar la probabilidad de obtener un efecto positivo en la próxima intervención. Para ello calculamos el valor del estadístico como:

$$(0 - 0.4707) / 0.2813 = -1.6733$$

Como en la distribución *t* de *Student* con 13 grados de libertad este valor tiene una probabilidad acumulada de .059, la probabilidad buscada es 1 – .059 = .941.

c) Si consideramos que hay un tamaño mínimo del efecto, irrenunciable para que sea razonable aplicar la intervención, podemos estimar la probabilidad de que se alcance dicho valor y así tomar mejores decisiones. Supongamos que en este ejemplo dicho tamaño mínimo es $\delta = 0.25$. Estimamos la probabilidad de que el efecto sea al menos de esa magnitud de la siguiente forma. Primero calculamos el valor del estadístico para ese valor de efecto:

$$(0.25 - 0.4707) / 0.2813 = -0.7846$$

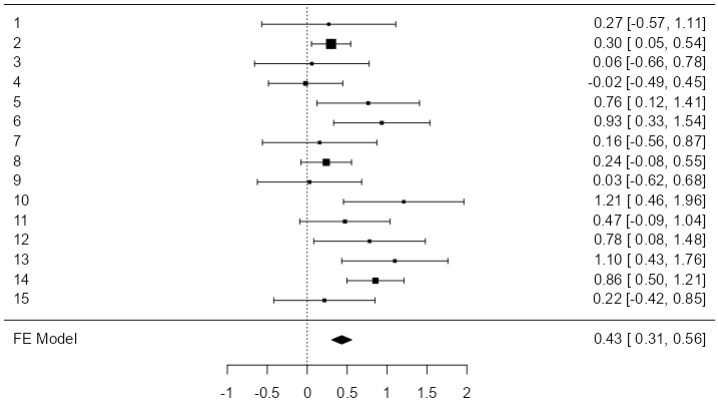
En la distribución *t* de *Student* con 13 grados de libertad este valor tiene una probabilidad acumulada de .2234. Por tanto, la probabilidad buscada es 1 – .2234 = .7766. En resumen, la probabilidad estimada de que el efecto sea de al menos $\delta = 0.25$ es igual a .7766.

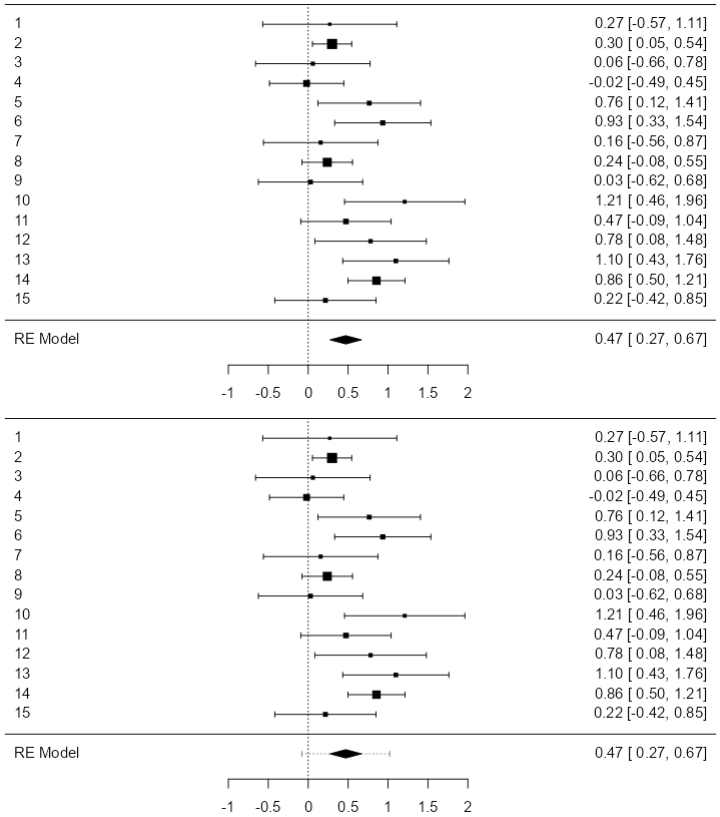
Análisis como los que acabamos de describir pueden ser especialmente apropiados cuando se estudia el papel de moderadoras, especialmente las categoriales. Es posible que bajo una de las categorías se alcance muy probablemente el mínimo establecido, mientras que en otra la probabilidad sea muy baja. Los intervalos de predicción pueden ayudar a tomar mejores decisiones en el contexto de una práctica basada en la evidencia.

En la figura 4 se presentan tres gráficos *forest plot*. Los tres muestran en su parte central las estimaciones que proporcionan los estudios individuales del ejemplo anterior, con intervalos de confianza idénticos a nivel de estudio. Lo que los distingue son las estimaciones combinadas que proporcionan y sus intervalos. En el primero se presenta la estimación combinada bajo un modelo de *EF* (el rombo que aparece debajo del último estudio; su eje vertical representa el valor combinado y los extremos de su eje horizontal los límites del IC95%). En el segundo y el tercero el rombo también representa al IC95%, pero bajo un modelo de *EA* (sin el ajuste Hartung y Knapp). Naturalmente, el IC95% bajo un modelo de *EA* es más amplio que bajo un modelo de *EF*. En el tercero se ha añadido al rombo un segmento horizontal intermitente, que representa al IP95%. En este tercer *forest plot* se aprecia más fácilmente la diferente interpretación de los componentes. Mientras el rombo muestra el rango en el que estaría el valor central de los efectos en el 95% de las ocasiones, el segmento muestra el rango central en el que estará el efecto paramétrico del 95% de los futuros estudios individuales pertenecientes a la población de estudios de referencia en este meta-análisis. Con frecuencia se interpreta incorrectamente la amplitud del rombo como un reflejo del rango de los efectos paramétricos. Este error ha sido señalado a veces como uno de los errores comunes o típicos en meta-análisis (Borenstein, 2019a). Adviértase que mientras que el intervalo de confianza se construye mediante el *error típico* del estimador del valor central, el intervalo de predicción se construye mediante la *desviación típica* de los efectos paramétricos.

Figura 4

Forest plots *con los estudios del ejemplo. El primero con los resultados de un modelo de EF. El segundo y el tercero con un modelo de EA. El tercero incluye la representación del intervalo de predicción tal y como lo calcula metafor con la ecuación (15).*





El intervalo de predicción como índice de la heterogeneidad

Ya sabemos de la importancia de reportar alguna estimación de la dispersión de los efectos en meta-análisis. Comunicar la estimación del tamaño del efecto medio es insuficiente, incluso aunque dicha estimación sea muy precisa (intervalo de confianza estrecho). Es imprescindible acompañar dicha estimación de algún indicador de cómo de heterogéneos son los efectos. Especialmente en ámbitos aplicados, sólo se pueden adoptar buenas decisiones si se anticipan las variaciones esperables del efecto (Borenstein, 2019b). Se han propuesto varias formas de reflejar esta dispersión, pero no todas son correctas ni las que son correctas son igual de eficaces. IntHout et al (2016) ponen el foco justamente en esta cuestión, revisando los índices más utilizados y evaluando sus ventajas e inconvenientes. Finalmente proponen utilizar precisamente el intervalo de predicción para esta función. Remitimos al lector interesado a dicha fuente, mientras que aquí resumimos las conclusiones principales.

La varianza específica, τ^2, no es un índice adecuado porque no está en la métrica natural del efecto estudiado, mientras que su raíz cuadrada puede estarlo, pero sólo si los valores no se han transformado para los cálculos (e.g., transformación logarítmica de *OR* o *Z* de Fisher de la correlación de Pearson). El índice I^2 tampoco refleja bien el constructo buscado, ya que es un índice de la proporción relativa de variación debida a los efectos paramétricos (Higgins y Thompson, 2002). Como consecuencia, un conjunto de estudios con tamaños muestrales muy elevados puede dar lugar a un índice I^2 muy alto, aunque no exista una gran heterogeneidad entre los tamaños del efecto; y viceversa, un conjunto de estudios con tamaños muestrales pequeños con la misma varianza inter-estudios a la del caso anterior, puede dar lugar a un índice I^2 bajo (Borenstein et al, 2017). Por supuesto, el estadístico Q tampoco es útil en este sentido, pues sólo es un estadístico para contrastar si es sostenible la hipótesis nula H_0: $\tau^2 = 0$ (Huedo-Medina et al, 2006). Ya hemos visto que no se debe usar el *intervalo de confianza* para hacer este tipo de interpretaciones, pues sólo refleja la incertidumbre en la estimación del efecto medio, no las variaciones en dichos efectos. Por el contrario, IntHout el al (2016) proponen el uso del *intervalo de predicción* para evaluar y comunicar el grado de dispersión de los efectos verdaderos. El intervalo de predicción refleja exactamente lo que se pretende y está en la métrica del índice de tamaño del efecto empleado. Recomiendan, como hacemos nosotros también, reportar en los meta-análisis de forma rutinaria los intervalos de predicción. Se debe tener en cuenta, no obstante, que si la síntesis meta-analítica se

ha llevado a cabo previa transformación de los tamaños del efecto de los estudios (e.g., *Z* de Fisher para coeficientes de correlación, el logaritmo natural del *odds ratio* o de la razón de riesgos, o la función logit aplicada sobre prevalencias), entonces los límites inferior y superior del intervalo de predicción se deben retro-transformar para devolverlos a la escala métrica del índice del tamaño del efecto de interés (e.g., transformarlos a coeficientes de correlación, *odds ratios*, razones de riesgos o prevalencias). También se debe ser consciente de que los intervalos de predicción se obtienen mediante una estimación de τ^2 y que estas estimaciones son muy imprecisas con un número bajo de estudios. En consecuencia, también la interpretación de los intervalos de predicción se debe tomar con cautela cuando se han obtenido con pocos estudios (o incluso evitarse con un número muy bajo).

Discusión y conclusiones

Uno de los objetivos principales del meta-análisis es el análisis de resultados orientado a la síntesis de la evidencia. Esta síntesis sirve para comprender mejor los fenómenos y las inter-relaciones entre las variables bajo estudio. Pero también debe servir para tomar mejores decisiones en nuevas investigaciones y nuevas intervenciones. La predicción basada en el conocimiento adquirido (o práctica basada en la evidencia) refleja mejor que cualquier otra cosa el papel del meta-análisis en el mundo aplicado.

En el modelo de *EA* se ha asumido un modelo de distribución determinado de los efectos paramétricos, el modelo normal. Este supuesto puede no ser adecuado, pero se asume por comodidad, porque otros modelos son también arbitrarios y porque cuando el número de estudios es razonablemente alto la aproximación es bastante buena. Higgins et al (2009) discuten otras distribuciones dentro del enfoque frecuentista, así como algunas alternativas dentro del enfoque bayesiano.

En este artículo hemos descrito los intervalos de predicción, distinguiéndolos tanto de los intervalos de confianza (que se refieren a un concepto diferente) como de su antecedente, los intervalos de credibilidad. Hemos destacado las diferencias en su cálculo y, sobre todo, en su interpretación. Creemos, siguiendo a IntHout et al (2016), que los intervalos de predicción deberían ser informados con mucha más frecuencia, y de forma rutinaria cuando se trata de estudios sobre intervenciones, una recomendación que también se hace en otras fuentes (e.g., Borenstein, 2019a, 2019b; Borenstein et al, 2017, 2021; Higgins et al, 2019; Schmid et al, 2021).

Información complementaria

Apoyo financiero.- Esta investigación ha sido financiada por el Ministerio de Ciencia e Innovación de España (referencia del proyecto: PID2021-122404NB-I00).

Conflicto de intereses.- Los autores declaran que no hay conflicto de intereses.

Referencias

Blázquez-Rincón, D., Sánchez-Meca, J., Botella, J., & Suero, M. (2023). Heterogeneity estimation in meta-analysis of standardized mean differences when the distribution of random effects departs from normal: A Monte Carlo simulation study. *BMC Medical Research Methodology*, 23(1), 19. https://doi.org/1 .1186/s12874-022-01809-0

Borenstein, M. (2019a). *Common mistakes in meta-analysis*. Englewood, NJ: Biostat inc.

Borenstein, M. (2019b). Heterogeneity in meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (eds.), *The Handbook of Research Synthesis and Meta-analysis*, 3rd ed. (pp. 453-468). New York: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8*(1), 5-18. https://doi.org/1 .1002/jrsm.1230

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2ª ed.). Chichester, UK: John Wiley and sons. [Chapter 17]

Botella, J., & Sánchez-Meca, J. (2015). *Meta-análisis en Ciencias Sociales y de la Salud* [Meta-analysis in Social and Health Sciences]. Madrid: Editorial Síntesis.

Hartung, J., & Knapp, G. (2001). On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine, 20*(12), 1771-1782. https://doi.org/1 .1002/sim.791

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-analysis*. San Diego, CA: Academic Press.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds) (2019). *Cochrane handbook for systematic reviews of interventions*. (2nd edition). Wiley.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11), 1539-1558. https://doi.org/1 .1002/sim.1186

Higgins, J. P., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1), 137-159. https://doi.org/1 .1111/j.1467-985X.2008.00552.x

Huedo-Medina, T., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistics or I^2 index? *Psychological Methods*, 11, 193–206. https://doi.org/1 .1037/1082-989X.11.2.193

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology, 14*(25). http://www.biomedcentral.com/1471-2288/14/25

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ open*, *6*(7), e010247. http://doi:1 .1136/bmjopen-2015-010247

Jackson, D., Law, M., Rücker, G., & Schwarzer, G. (2017). The Hartung-Knapp modification for random-effects meta-analysis: A useful refinement but are there any residual concerns? *Statistics in Medicine, 36*, 3923–3934. https://doi.org/1 .1002/sim.7411

Langan, D., Higgins, J. P., & Simmonds, M. (2017). Comparative performance of heterogeneity variance estimators in meta-analysis: a review of simulation studies. *Research Synthesis Methods*, 8(2), 181-198. https://doi.org/1 .1002/jrsm.1198

Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., ... & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, 10(1), 83-98. https://doi.org/1 .1002/jrsm.1316

Partlett, C., & Riley, R.D. (2017). Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. *Statistics in Medicine, 36*, 301-317. https://doi.org/1 .1002/sim.7140

Riley, R. D., Higgins, J. P., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *BMJ*, *342*. https://doi.org/1 .1136/bmj.d549

Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods, 13*, 31-48. https://doi.org/1 .1037/1082-989X.13.1.31

Schmid, C. H., Stijnen, T., & White, I. (Eds.). (2021). *Handbook of Meta-analysis*. CRC Press.

Schmid, C. H, Carlin, B. P., & Welton, N. J. (2021). Bayesian Methods for Meta-analysis. En *Handbook of Meta-Analysis* (pp. 41-64). Chapman and Hall/CRC.

Sidik, K., & Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Statistics in Medicine, 21*(21), 3153–3159. https://doi.org/1 .1002/sim.1262

Stijnen, T., White, I. R., & Schmid, C. H. (2021). Analysis of univariate study-level summary data using normal models. In *Handbook of Meta-Analysis* (pp. 41-64). Chapman and Hall/CRC. [Section 4.4.4.2]

Suero, M., Botella, J., & Durán, J. I. (2023). Methods for estimating the sampling variance of the standardized mean difference. *Psychological Methods*, 28(4), 895-904. https://doi.org/1 .1037/met0000446

Suero, M., Botella, J., Durán, J. I., & Blázquez-Rincón, D. (2023, September 12). Reformulating the meta-analytical random effects model as a mixture model. https://doi.org/1 .17605/OSF.IO/V2FDE

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., ... & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, 7(1), 55-79. https://doi.org/1 .1002/jrsm.1164

Viechtbauer W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects Model. *Journal of Educational and Behavioral Statistics*, 30, 261–293. https://doi.org/1 .3102/10769986030003261

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48. https://doi.org/1 .18637/jss.v036.i03

Viechtbauer, W. (2023). *Package 'metafor'*. Unpublished document. University de Maastricht.

Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., & Marín-Martínez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods, 20*, 360-374. https://doi.org/1 .1037/met0000023

Apéndice

Código R empleado para el ejemplo del texto. Antes de ejecutarlo se debe crear un archivo de SPSS (.sav) con las columnas que aparecen en el panel izquierdo de la tabla del ejemplo. Las columnas de los tamaños del efecto y sus varianzas están en las columnas llamadas "g" y "Var_g", respectivamente.

```
#EJEMPLO DEL TEXTO, CON 15 ESTUDIOS
#Lectura de datos, en el archivo SPSS "Ejemplo_15_estudios.sav"
library(foreign)  #El paquete foreign debe estar ya instalado
Datos <- read.spss("Ejemplo_15_estudios.sav")

#Se carga el paquete "metafor"
library("metafor")

#Modelo de Efecto Fijo
resEF <- rma.uni(yi=g, vi=Var_g, data=Datos, method="FE")
resEF

#Modelo de Efectos Aleatorios, sin ajuste
resEA <- rma.uni(yi=g, vi=Var_g, data=Datos, method="REML")
resEA

#Modelo de Efectos Aleatorios, ajuste de Knapp y Hartung
resEA_KH <- rma.uni(yi=g, vi=Var_g, data=Datos, method="REML", test="knha")
resEA_KH

#Intervalo de predicción, Modelo de EA sin ajuste
predict(resEA)

#Intervalo de prediccion, Modelo de EA, ajuste Knapp y Hartung
predict(resEA_KH)

#Los tres forest plot de la figura 4
forest(resEF)
forest(resEA)
forest(resEA, addpred=TRUE)
```