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

Estableciendo criterios externos para la actualización de un meta-análisis

Santiago Arranz Orlandi

Tutor: Juan Botella Ausina

Modalidad: Investigación

Junio de 2022

Universidad Autónoma de Madrid

Universidad Complutense Madrid

Resumen

El cuerpo de investigaciones que nutren la práctica basada en la evidencia es tan amplio que el meta-análisis, como herramienta de síntesis cuantitativa, se ha constituido como referencia a la hora de tomar decisiones. No obstante, según avanza el tiempo, los meta-análisis correctamente realizados se van quedando desactualizados y cabe preguntarse cuándo merece la pena publicar una actualización. Es importante, de forma similar a las reglas secuenciales en los estudios primarios, establecer criterios a priori o externos que nos ayuden a decidir cuándo es conveniente actualizar un meta-análisis, sobre todo cuando sus resultados no son estadísticamente significativos.

El objetivo del presente estudio es desarrollar una herramienta práctica para el metodólogo a la hora de decidir cuándo actualizar un meta-análisis basándose en dos criterios independientes y complementarios: la potencia y la precisión de la estimación combinada. Se ha desarrollado un *script* para el entorno estadístico R que permite estimar cuántos estudios serían necesarios para alcanzar: (*a*) un valor determinado de potencia, y/o (*b*) una determinada reducción de la amplitud del intervalo de confianza de la estimación combinada del tamaño del efecto. Estos cálculos se podrán realizar tanto para modelos de efecto fijo como de efectos aleatorios.

Abstract

The body of research that feeds evidence-based practice is so broad that meta-analysis, as a quantitative synthesis tool, has become a reference when making decisions. However, as time progresses, well-conducted meta-analyses become out of date, and one may wonder when it is worth publishing an update. It is important, similar to the sequential rules in primary studies, to establish an external criteria that help us decide when it is convenient to update a meta-analysis, especially when its results are not statistically significant.

The objective of the present study is to develop a practical tool for the methodologist when deciding when to update a meta-analysis. That tool is based on two independent and complementary criteria: the power and the precision of the mean effect size estimate. A script has been developed for the statistical environment R that allows estimating how many studies would be necessary to reach: (a) a certain value of power, and/or (b) a specific reduction of the range of the confidence interval of the mean effect estimate. These calculations can be made for both fixed effect and random effects models.

Índice de contenidos

1. Introducción

El meta-análisis (MA) es una metodología para la síntesis y el análisis estadístico de un conjunto de resultados procedentes de estudios primarios con el objetivo de integrar los hallazgos (Botella & Sánchez-Meca, 2015; Cooper et al., 2019; Glass, 1976).

Se trata por tanto de una metodología cuantitativa de síntesis de información, que se sirve de una revisión sistemática para seleccionar los estudios primarios de interés. Una vez realizada la búsqueda de los casos que cumplen los requisitos, se recopilan los resultados de dichos estudios bajo una métrica del tamaño del efecto (TE) común. Esto permite la estimación puntual del tamaño del efecto paramétrico, la cual se encuentra dentro de un rango de valores plausibles, o intervalo de confianza.

El meta-análisis se ha constituido como una técnica de referencia dentro de la corriente de la práctica basada en la evidencia (Botella & Caperos, 2019), ya que es capaz de sintetizar un gran número de estudios primarios, minimizando además los sesgos y limitaciones propios de las revisiones narrativas. El meta-análisis trata de dotar de rigor al proceso de revisión sistemática, al buscar conseguir los objetivos de objetividad, replicabilidad y precisión que ya son exigidos a los estudios primarios incluidos en las revisiones (Botella & Sánchez-Meca, 2015)

No obstante, el MA no debe considerarse únicamente como un punto final en la producción científica (p.ej. decidir si una intervención de tercera generación merece ser utilizada para el tratamiento de la depresión), sino que también puede ser un punto de partida para planificar la dirección de la investigación futura (Rothery et al., 2022). Siguiendo con el ejemplo anterior, un MA podría sugerir sobre qué poblaciones merece la pena seguir investigando los efectos de la intervención.

Como hemos visto, la investigación repercute en la práctica, y la práctica en la investigación. Esto implica un crecimiento continuo de la base de estudios primarios que puede provocar que meta-análisis correctamente realizados se queden desactualizados y, por lo tanto, no sean tan útiles como podrían ser a la hora de tomar decisiones. Es por ello que cabe preguntarse cuándo es conveniente actualizar un MA, ya que es lógico pensar que no cualquier actualización va a ser publicada. Este hecho suscita la idea de tener que establecer unos criterios que permitan al metodólogo tomar la decisión de cuándo actualizar un MA.

Es importante no basar la decisión en cualquier criterio, como puede ser el haber alcanzado la significación estadística o no en el MA original. Basar la decisión en los resultados puede implicar caer en la mala praxis de añadir casos (en este caso estudios) en función de la significación, lo que desemboca en un progresivo aumento de la tasa de error tipo I. Es el mismo efecto que se produce al añadir participantes en un estudio primario, según se haya alcanzado el nivel de significación o no (Botella et al., 2006).

La toma de decisiones relativas a añadir muestras o estudios a un MA para publicar una actualización debe regirse por criterios externos a sus propios resultados. Existen algunas propuestas a modo de tentativa, como la de Valentine et al. (2010), los cuales proponen fijar un aumento determinado de la potencia de los contrastes meta-analíticos, lo que permite estimar el número de estudios necesarios para alcanzarla. Una de las razones que es frecuentemente citada al justificar la realización de un meta-análisis es que la combinación de estudios arrojará una potencia más alta que cada estudio tomado de forma individual (Cohn & Becker, 2003). Esta afirmación resulta evidente si se sigue un modelo de efecto fijo, pero puede no ser cierta siempre si se sigue un modelo de efectos aleatorios (Cohn & Becker, 2003; Jackson & Turner, 2017).

Otro criterio que puede ser de interés, dada su relevancia en los meta-análisis, es fijar una mejora de la precisión de la estimación combinada del TE. Esto es, una determinada reducción de la amplitud del intervalo de confianza (IC). El interés de este criterio es que las decisiones serán más confiables cuanto menor error de estimación haya en los valores en los que se basan dichas decisiones.

El objetivo del presente estudio es desarrollar una herramienta de utilidad para el metodólogo a la hora de decidir cuándo actualizar un meta-análisis basándose en dos criterios independientes y complementarios: la potencia y la precisión de la estimación combinada. También puede ser útil para los consejos editoriales de las revistas científicas, que deben decidir cuándo merece la pena publicar lo que es una mera actualización de resultados de un meta-análisis. Pensemos que por el añadido de un simple estudio ya se gana algo en potencia y precisión. La cuestión es con cuántos estudios (o participantes) nuevos añadidos las mejoras en potencia y/o amplitud justifican la publicación. Hemos desarrollado un *script* para el entorno estadístico R que permite estimar cuántos estudios serían necesarios para: (*a*) alcanzar un valor determinado de potencia, y/o (*b*) una determinada reducción de la amplitud del IC de la estimación combinada del TE. La herramienta permitirá hacer los cálculos tanto para modelos de efecto fijo como de efectos aleatorios.

2. Significación estadística del tamaño del efecto medio

Los meta-análisis sintetizan estudios primarios para alcanzar sus resultados. Dado que estos estudios proceden de muestras de distinto tamaño, la varianza de la estimación de cada estudio será distinta a las demás. Esta peculiaridad aconseja realizar los contrastes meta-analíticos ponderando los estudios según su precisión. El método de ponderación dependerá del modelo estadístico que se decida seguir. Los modelos más utilizados en la aplicación de un MA son el modelo de efecto fijo y el modelo de efectos aleatorios. Pese a que ambos modelos aplican un método de ponderación similar, los factores de ponderación son diferentes.

El modelo de efecto fijo (EF) asume que todos los estudios incluidos en un MA estiman un único TE paramétrico o poblacional. La variabilidad observada entre los estimadores del TE se debe exclusivamente al error de muestreo, o variabilidad intraestudio. De esta forma, la ponderación de los estudios se calcula como el inverso de la varianza:

$$w_i^{EF} = \frac{1}{\hat{\sigma}^2_{m(i)}}, \tag{1}$$

donde w_i^{EF} es el peso atribuido al estudio i-ésimo, y $\hat{\sigma}^2_{m(i)}$ es la varianza de muestreo de dicho estudio.

El modelo de EF se basa en un supuesto poco realista, como es que todos los estudios sean réplicas que estiman un único TE común, siendo su única diferencia la muestra utilizada. Ante esta limitación surge el modelo de efectos aleatorios (EA), el cual asume que los TE estimados constituyen una muestra aleatoria de una población de efectos paramétricos. Por tanto, los estudios no solo difieren en el error de muestreo, sino que también son estimadores de distintos TE paramétricos. Esto implica que la variabilidad proviene de dos fuentes: intraestudio e interestudios. Por lo tanto, el método de ponderación es

$$w_i^{EA} = \frac{1}{\hat{\sigma}^2_{m(i)} + \hat{\tau}^2}, \tag{2}$$

donde w_i^{EA} es el peso atribuido al estudio i-ésimo siguiendo este modelo, y $\hat{\tau}^2$ es la varianza interestudios, la cual es común a todos los estudios y debe ser estimada a través de los propios TE de los estudios integrados en el MA. El método de estimación de τ^2 más adecuado en la mayoría de casos es máxima verosimilitud restringida (Viechtbauer, 2005), por lo que será el método empleado en los cálculos del presente texto. Dado que la única diferencia entre ambos modelos son los factores de ponderación, en los cálculos aquí descritos se utilizará la nomenclatura para efecto fijo o sin especificar el modelo, con el fin de mostrar una exposición más clara. No obstante, todos los cálculos (salvo excepciones determinadas) son aplicables a ambos modelos.

El principal resultado meta-analítico es la estimación combinada del TE común ($\hat{\theta}$, en el modelo de EF) o la media de efectos paramétricos ($\hat{\mu}_\theta$, en el modelo de EA). El estimador óptimo (insesgado y con varianza mínima) de $\theta_\bullet$ se calcula como (Hedges & Olkin, 1985)

$$\hat{\theta}_\bullet = \frac{\sum w_i \cdot \hat{\theta}_i}{\sum w_i}, \tag{3}$$

la cual es la media de los TE ponderada por los pesos de los propios estudios. Si el índice del TE tiene una distribución muestral asintóticamente normal, se puede establecer un intervalo de confianza (IC) del efecto paramétrico. El IC se calcula como:

$$\hat{\theta} \pm \left|z_{1-\alpha/2}\right| \cdot \hat{\sigma}_\theta = \begin{cases} TE_{\text{sup}} \\ TE_{\text{inf}} \end{cases}, \tag{4}$$

donde $z_{1-\alpha/2}$ es el valor de la distribución normal tipificada correspondiente al percentil $(1-\alpha/2)$, asumiendo un nivel de confianza $(1-\alpha)\cdot 100\%$; T_{sup} y T_{inf} son los límites confidenciales superior e inferior, respectivamente, del IC; y $\hat{\sigma}_\theta$ es la estimación del error típico de $\hat{\theta}_\bullet$, la cual se calcula como

$$\hat{\sigma}_\theta = \frac{1}{\sqrt{\sum_i w_i}}. \tag{5}$$

El IC permite contrastar la hipótesis nula de que el TE paramétrico es nulo, evaluando si el valor 0 se encuentra dentro de dicho intervalo de confianza. La hipótesis nula también puede ser evaluada mediante el estadístico de contraste

$$z = \frac{\hat{\theta}}{\hat{\sigma}_\theta}, \tag{6}$$

que bajo hipótesis nula se distribuye según una distribución normal tipificada. Los paquetes estadísticos arrojan un nivel crítico de probabilidad (p) asociado al valor de z, de tal forma que la hipótesis nula será rechazada si se cumple que $p < \alpha$.

3. Actualización de un meta-análisis: Reglas basadas en la potencia

La potencia de un contraste es la probabilidad de rechazar la hipótesis nula (H_0) cuando esta es falsa. Como H_0 y H_1 son exhaustivas, ese rechazo implica la afirmación indirecta de H_1. Esta probabilidad viene determinada por el nivel de significación (α), determinado a priori; el tamaño del efecto paramétrico alternativo para el que se desea conocer la potencia (θ_1) y el error típico del estadístico de contraste ($\hat{\sigma}$). La potencia depende, en primer lugar, de la distribución de la hipótesis nula (H_0); en la mayoría de los contrastes meta-analíticos se asume que, bajo H_0 verdadera, el estadístico de contraste se distribuye aproximadamente según una distribución normal con media igual a 0 y varianza igual a 1. Para que el contraste sea significativo, se debe alcanzar un tamaño del efecto combinado mínimo, o valor crítico ($TE_{\bullet(c)}$), el cual se corresponde con el valor z del nivel de confianza ($1 - \alpha$). En un contraste unilateral derecho $TE_{\bullet(c)}$ se expresa como

$$TE_{\bullet(c)} = z_{(1-\alpha)} \cdot \hat{\sigma}_{\theta_\bullet} . \tag{7}$$

La potencia (*pot*) respecto a un valor alternativo θ_1 es, por tanto, el área de la curva bajo H_1 que el valor de $TE_{\bullet(c)}$ asociado a $z_{(1-\alpha)}$ deja a su derecha (Hedges & Pigott, 2001). Se calcula mediante:

$$pot = 1 - \Phi\left(\frac{TE_{\bullet(c)} - \theta_1}{\hat{\sigma}_{\theta_\bullet}} \right). \tag{8}$$

La magnitud de error típico se relaciona de forma inversa con el número de estudios, ya que cuanto mayor sea el tamaño muestral (número de estudios), mayor será la precisión de la estimación, lo que implica una reducción de la varianza.

3.1. Número de estudios adicionales para alcanzar un determinado valor de potencia

Dado un contraste meta-analítico con $TE_\bullet$ y una potencia determinados, cabe preguntarse cuántos estudios se deben añadir para alcanzar una potencia más alta, fijada a priori, para un valor alternativo concreto del parámetro. Representamos por k al número de estudios del meta-análisis y por j al número de estudios que hay que añadir para conseguir la potencia deseada. De esa forma, el nuevo meta-análisis tendrá $(k + j)$ estudios. Una potencia más alta precisará de un error típico más bajo, el cual se debe extraer de la fórmula (8) como sigue a continuación.

La potencia no es más que la probabilidad complementaria a cometer un error tipo II (β), por lo que se puede expresar como $1-\beta$. Es la probabilidad de que el contraste de nulidad $\left(H_0 : \theta = \theta_0\right)$ resulte significativo cuando el valor paramétrico es uno concreto, diferente de cero, que representaremos por θ_1. También representaremos por $\hat{\sigma}_{\theta_\bullet}$ al error típico con el que se tiene la potencia deseada y por $TE_{\bullet(c)}$ al menor tamaño del efecto combinado con el que el contraste de nulidad resultaría significativo, en un contraste unilateral derecho. En estos términos, la potencia se calcula como

$$1-\beta = 1-\Phi\left(\frac{TE_{\bullet(c)}-\theta_1}{\hat{\sigma}_{\theta_\bullet}}\right) \tag{9}$$

El valor del error típico no puede ser extraído directamente de la fórmula (9), ya que el $TE_{\bullet(c)}$ también es desconocido. No obstante, este valor puede ser extraído del contraste de nulidad en el cual se calcula el valor crítico tipificado para un nivel de confianza $1-\alpha$:

$$\Phi^{-1}\left(1-\alpha\right) = z_{1-\alpha} = \frac{TE_{\bullet(c)}-\theta_0}{\hat{\sigma}_{\theta_\bullet}} \tag{10}$$

Despejando el $TE_{\bullet(c)}$ de la ecuación anterior podemos establecer que para $H_0 : \theta = 0$

$$TE_{\bullet(c)} = z_{1-\alpha} \cdot \hat{\sigma}_{\theta_\bullet} \tag{11}$$

Por lo tanto, si sustituimos el valor de $TE_{\bullet(c)}$ en la ecuación (9) podemos expresar la potencia como

$$1-\beta = 1-\Phi\left(\frac{z_{1-\alpha} \cdot \hat{\sigma}_{\theta_\bullet} - \delta_1}{\hat{\sigma}_{\theta_\bullet}}\right) \tag{12}$$

de donde podemos extraer el error típico y expresarlo como:

$$\hat{\sigma}_{\theta_\bullet} = \frac{\theta_1}{z_{1-\alpha} - z_\beta} \tag{13}$$

donde θ_1 sigue siendo el tamaño del efecto paramétrico alternativo a 0, $z_{1-\alpha}$ corresponde con el valor de la curva normal estandarizada con una probabilidad acumulada igual $1-\alpha$, y z_β corresponde al valor de la misma curva con una probabilidad acumulada igual a la probabilidad de cometer un error tipo II ($\beta = 1 - pot$).

El cuadrado del error típico (varianza) es igual al inverso del sumatorio de los pesos (tanto los del meta-análisis original como los de los estudios que se deben añadir). El sumatorio de los pesos (W_{k+j}) es, por tanto,

$$W_{(k+j)} = \frac{1}{\hat{\sigma}^2_{(k+j)}} . \tag{14}$$

El peso total se descompone en el sumatorio de los pesos originales (W_k) más el peso de los estudios a añadir (W_j). Por lo tanto, el peso que hay que añadir para alcanzar la potencia determinada a priori se obtiene como

$$W_j = W - W_k . \tag{15}$$

Este peso se puede expresar en términos de estudios de varias formas. Nosotros vamos a hacerlo de dos: (a) como un único estudio con la totalidad del peso (que permite estimar un número de participantes, pero solo bajo un modelo de efecto fijos) y (b) como varios estudios de tamaños similares a los ya incluidos en el meta-análisis (válido para ambos modelos).

La primera forma de expresarlo es como un único estudio que en muchas ocasiones tendrá un tamaño muy grande y probablemente poco realista en muchos campos de estudios, pero que permite expresar el resultado en términos del número de participantes a añadir. En concreto, para alcanzar una determinada potencia, debemos añadir un único estudio con una varianza igual a

$$\hat{\sigma}^2_j = \frac{1}{W_j} . \tag{16}$$

Bajo un modelo de efectos aleatorios, esta varianza seguiría descomponiéndose en la varianza de muestreo y en la varianza interestudios, la cual es constante para todos los estudios del meta-análisis $(k+j)$. Dado que el valor de la varianza de este único estudio es presumiblemente muy pequeño, puede ocurrir que sea menor que la varianza interestudios, incurriendo en una imposibilidad teórica. Por lo tanto, en la gran mayoría de ocasiones, esta forma de expresar el peso de los *j* estudios añadidos solo es posible si se sigue un modelo de efecto fijo, ya que en este caso toda la variabilidad es debida a la varianza de muestreo.

Despejando de la fórmula de la varianza de muestreo el índice de tamaño del efecto del que se trate podemos calcular el tamaño de la muestra correspondiente. En la Tabla 1 se recogen las fórmulas para los dos índices del tamaño del efecto más utilizados en psicología: la diferencia de medias tipificada y la transformada de Fisher para correlaciones de Pearson.

Tabla 1

Fórmulas para estimar el tamaño muestral a través de la varianza de muestreo

Índice del TE	Fórmula despejada		Nota
d	$N = \frac{4}{\hat{\sigma}_m^2} + \frac{2d^2}{4\hat{\sigma}_m^2}$	(17)	Se asume tamaños de grupo iguales ($n_1 = n_2$)
Z_r	$N = \frac{1}{\hat{\sigma}_m^2} + 3$	(18)	

La segunda forma de expresarlo es como el número de estudios que hay que añadir para alcanzar la potencia deseada. En realidad es poco realista pensar que se publicaría un único estudio con una varianza tan pequeña como para reducir tan sensiblemente el error típico de la estimación combinada del TE, siendo, además, solo válido para un modelo de EF. Para calcular cuál sería un número razonable de estudios para alcanzar W_j se puede asumir que los tamaños de los estudios ya realizados representan bien a los estudios de ese campo y, por tanto, cabe esperar que los estudios futuros tengan un tamaño parecido. En resumen, se asume que cada estudio a añadir tiene un peso equivalente al peso medio ($\bar{w}$) de los estudios originales:

$$\bar{w} = \frac{W_k}{k}, \tag{19}$$

Por tanto, el número de estudios a añadir es, aproximadamente,

$$j \approx \frac{W_j}{\bar{w}}. \tag{20}$$

También se podría estimar el tamaño muestral de cada estudio añadido extrayendo la varianza muestral de la varianza total de cada estudio:

$$\hat{\sigma}_{m(w)}^2 = \hat{\sigma}_w^2 - \tau^2. \tag{21}$$

Recuérdese que bajo un modelo de efecto fijo la varianza interestudios es nula $\left(\tau^2 = 0\right)$. Una vez obtenido el valor de esta varianza, se calcula el tamaño muestral de cada estudio a través de las fórmulas recogidas en la Tabla 1.

Ejemplo numérico. Para ilustrar lo expuesto hasta aquí con un ejemplo, hemos utilizado los datos del meta-análisis de Velthorst et al. (2015). Este meta-análisis incluye 30 estudios que evalúan el efecto de la terapia cognitivo-conductual sobre los síntomas negativos de la esquizofrenia. Se han seleccionado los 28 estudios donde el tratamiento de dichos síntomas era un objetivo secundario de la terapia. La métrica del tamaño del efecto que aportan es la diferencia de medias estandarizada (*d*), arrojando un $TE_{\bullet} = 0.093$ y un $\hat{\sigma}\left(TE_{\bullet}\right) = 0.0758$. Como se puede apreciar, el contraste arroja una conclusión de no significación, pues sustituyendo en la fórmula (6) obtenemos 0.093 / 0.0758 = 1.227, que es menor que el valor crítico 1.645 ($\alpha = 0.05$).

Calcularemos la potencia respecto a un tamaño del efecto "pequeño" según la taxonomía clásica de Cohen (1992) ($\delta = 0.2$), fijando un nivel de significación de $\alpha = 0.05$. Seguiremos un modelo de efectos aleatorios, siendo la estimación de la varianza interestudios $\tau^2 = 0.0962$.

El tamaño del efecto mínimo necesario para alcanzar la significación estadística y rechazar la H_0 en un contraste unilateral derecho es

$$TE_{\bullet(c)} = 1.645 \cdot 0.0758 = 0.12469. \tag{22}$$

Por lo tanto, el tamaño del efecto mínimo para alcanzar la significación estadística en este caso es 0.12469. La potencia del contraste respecto al valor alternativo del $TE\left(\delta_1\right) = 0.2$ es, por tanto,

$$p = 1 - \Phi\left(\frac{0.12469 - 0.2}{0.0758}\right) = 1 - \Phi\left(-0.99\right) = 0.84. \tag{23}$$

Se plantea ahora la pregunta de qué valor de $\hat{\sigma}$ se necesitará para alcanzar una potencia mayor, por ejemplo, igual a 0.90 respecto al mismo *TE* paramétrico alternativo. El error típico necesario para alcanzar la nueva potencia ($\hat{\sigma}'$) es igual a (recuérdese que 1.645 y −1.28 son los valores de la distribución normal estandarizada con probabilidades acumuladas iguales a 0.95 y 0.10, respectivamente)

$$\hat{\sigma}' = \frac{0.2}{1.645 - (-1.28)} = 0.06838. \tag{24}$$

El sumatorio de los pesos se extrae del inverso del cuadrado del error típico, que es igual a

$$W_{k+j} = \frac{1}{0.06838^2} = 213.866. \tag{25}$$

Calculamos ahora la suma de los pesos con los estudios iniciales: $W_k = 1/0.0758^2 = 174.045$. Por lo tanto, se podría concluir que el peso que hay que añadir es igual a $213.866 - 174.045 = 39.821$. Veamos ahora cómo expresar este resultado en términos de participantes y en términos de estudios.

En términos de participantes se trataría de añadir un único estudio con una varianza igual a

$$\hat{\sigma}_j^2 = \frac{1}{39.821} = 0.0251\,. \tag{26}$$

Dado que la varianza de este estudio es menor que la varianza interestudios del meta-análisis, es teóricamente imposible realizar un estudio primario con esta varianza. Por motivos didácticos, asumiremos un modelo de efecto fijo únicamente para realizar este cálculo. Por lo tanto, para averiguar a qué tamaño muestral se corresponde esa varianza, aquel se debe despejar de la fórmula de la varianza de muestreo para diferencia de medias tipificada (Tabla 1). Representaremos por n_1 al tamaño de una muestra y n_2 al de la otra muestra, siendo el tamaño muestral total igual a $N = n_1 + n_2$. Asumiendo tamaños de cada grupo iguales $(n_1 = n_2)$, la varianza para $\delta = 0.2$ precisaría de un tamaño muestral aproximadamente igual a

$$N = \frac{4}{\text{var}} + \frac{2\delta^2}{4\,\text{var}} = \frac{4}{0.0251} + \frac{2 \cdot 0.2^2}{4 \cdot 0.0251} = 160.159 \approx 161\,. \tag{27}$$

Retomando el modelo original de efectos aleatorios, vamos ahora a expresar el peso en términos de estudios. En concreto, del número de estudios con un peso igual a la media de los pesos originales necesarios para alcanzar W_j. El peso medio de los estudios originales del meta-análisis de Velthorst se calcula como el inverso del error típico original al cuadrado (sumatorio de pesos original) entre el número de estudios originales

$$\bar{w} = \frac{\frac{1}{0.0758^2}}{28} = \frac{174.045}{28} = 6.216\,. \tag{28}$$

Por lo tanto, la estimación del número mínimo de estudios a añadir para alcanzar una potencia de 0.9 es

$$j = \frac{213.866 - 174.045}{6.216} = \frac{39.821}{6.216} = 6.406 \approx 7\,. \tag{29}$$

Si quisiéramos conocer el tamaño muestral de los j estudios hay que calcular, en primer lugar, cuánta varianza es debida al muestreo:

$$\hat{\sigma}^2_{m(w)} = \frac{1}{6.216} - 0.0962 = 0.0647\,. \qquad (30)$$

Para conocer el tamaño muestral debe sustituirse ese valor en la ecuación (17), por lo que el tamaño muestral de los estudios añadidos es aproximadamente igual a

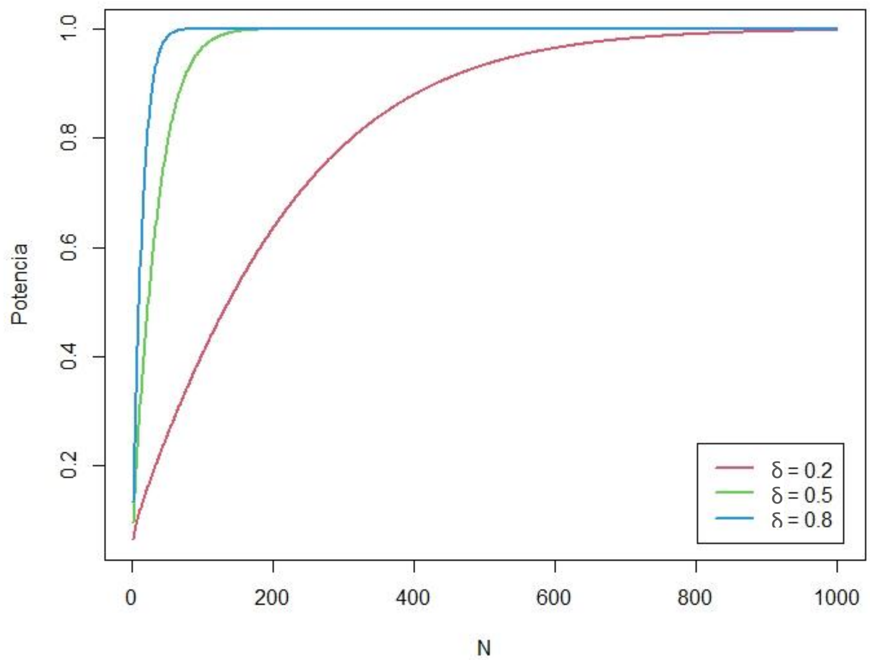
$$N = \frac{4}{0.0647} + \frac{2\cdot 0.2^2}{4\cdot 0.0647} = 62.133 \approx 64\,. \qquad (31)$$

En los términos expresados en nuestro objetivo, podemos concluir que para alcanzar la potencia que hemos establecido que justificaría la publicación del meta-análisis (*pot* = 0.90) harían falta 7 estudios de tamaño medio igual al tamaño medio de los estudios ya publicados (*N* = 64).

Modelos de efecto fijo y de efectos aleatorios. Bajo un modelo de efecto fijo, la agregación de estudios a un meta-análisis siempre supondrá una reducción del error típico, ya que el peso del estudio tan solo depende de la varianza de muestreo (fórmula (1)). Añadir estudios bajo este modelo, aunque tengan un tamaño muestral muy pequeño, supondrá necesariamente un aumento de la magnitud del sumatorio de pesos. En la Figura 1 se muestra, bajo un modelo de efecto fijo, cómo aumenta la potencia de un contraste de diferencias medias tipificadas (DMT) según se añaden casos.

Figura 1

Potencia de un contraste de diferencia de medias tipificadas según el tamaño muestral para varios TE alternativos



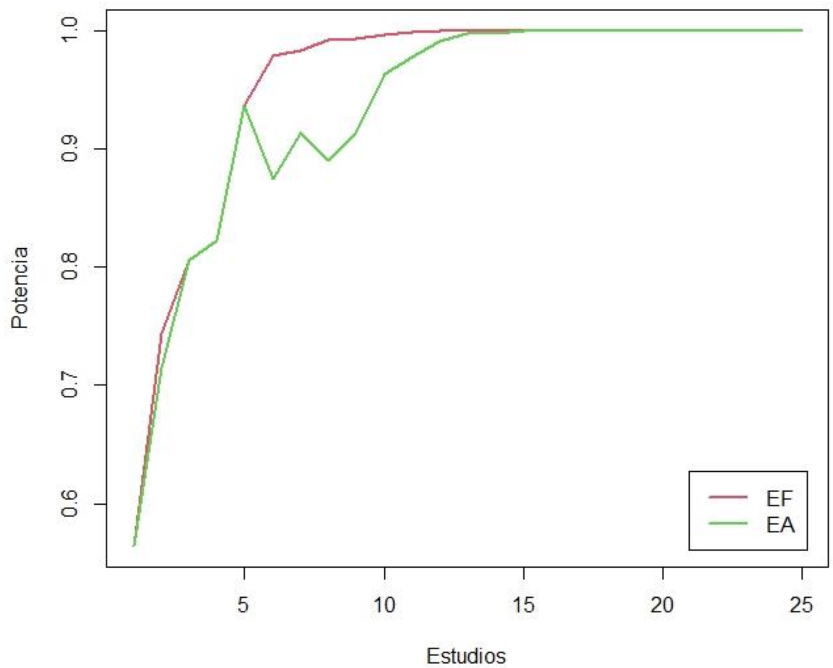
Nota. Se ha utilizado la fórmula para la varianza de muestreo, asumiendo tamaños de grupo iguales.

No obstante, si se sigue un modelo de efectos aleatorios, el peso de los estudios depende tanto de la varianza de muestreo como de la varianza interestudios (τ^2). Esta varianza se estima a partir de la variabilidad de los estudios introducidos en el meta-análisis, por lo que su estimación podría variar si se introducen estudios con un TE y/o una varianza muy distinta a la del resto de estudios. La re-estimación de la varianza interestudios puede implicar retrocesos en el valor de la potencia del meta-análisis según se agregan estudios.

Se han simulado los datos de 25 estudios primarios cuyos TE (DMT) varían aleatoriamente entre −0.2 y 0.5 y cuyos tamaños muestrales varían entre 10 y 200 (asumiendo tamaños de grupo iguales). En la Figura 2 se puede comparar cómo evoluciona el valor de la potencia (para $\delta = 0.2$) según se van agregando estudios. La línea granate sigue un modelo de efecto fijo, mientras que la línea verde sigue un modelo de efectos aleatorios, cuya τ^2 se ha re-estimado cada vez que se añadía un estudio (mediante el método de máxima verosimilitud restringida). Se puede observar cómo, si se sigue un modelo de efecto fijo, la potencia siempre aumenta (aunque la magnitud del aumento varíe), pero si se sigue un modelo de efectos aleatorios aparecen retrocesos en la potencia del contraste.

Figura 2

Comparación del aumento de potencia entre EF y EA en un caso simulado



4. Actualización de un meta-análisis: Reglas basadas en la precisión de la estimación

Como mencionábamos en la introducción, la técnica del meta-análisis ofrece tanto una estimación combinada puntual del tamaño del efecto, como un rango de valores plausibles donde se podría encontrar el TE paramétrico (con un modelo de efecto fijo) o la media de TEs paramétricos (con un modelo de efectos aleatorios). Este rango se conoce como intervalo de confianza (IC) y la amplitud del mismo es una medida de la precisión de la estimación del meta-análisis. Tomando como referencia la fórmula (4), la amplitud (a) se calcula como la diferencia entre el límite superior del TE ($TE_{\sup}$) y el límite inferior ($TE_{\inf}$):

$$a = TE_{\sup} - TE_{\inf} \tag{32}$$

Dado que los límites del IC son equidistantes de la estimación puntual del TE (al menos cuando se asume una distribución asintótica normal), la amplitud puede expresarse como dos veces el error típico por el valor de la distribución normal tipificada correspondiente al percentil $(1-\alpha/2)$:

$$a = 2 \cdot \left(\left| z_{1-\alpha/2} \right| \cdot \hat{\sigma}_{\theta} \right). \tag{33}$$

De esta forma, el cálculo de la amplitud es independiente del tamaño del efecto.

Igual que ocurría con la potencia, la magnitud de error típico se relaciona de forma inversa con el número de estudios, ya que cuanto mayor sea el tamaño muestral (número de estudios), mayor será la precisión de la estimación y, por ende, menor será la amplitud del IC.

Dado que la amplitud se relaciona de forma inversa con la precisión de la estimación, resulta contra-intuitivo pensar que se consigue una mejora reduciendo la amplitud. Ante este caso, se puede emplear el inverso de la amplitud como una medida en sentido positivo de la precisión.

4.1. Número de estudios adicionales para alcanzar una determinada amplitud

Dado un contraste meta-analítico con un error típico estimado y una amplitud determinada, cabe preguntarse cuántos estudios se deben añadir para reducir el intervalo de confianza y, por tanto, aumentar la precisión de la estimación hasta un valor determinado. Representamos de nuevo por *k* al número de estudios del meta-análisis inicial y por *j* al número de estudios que hay que añadir para conseguir la amplitud deseada. De esa forma, el nuevo meta-análisis tendrá *(k + j)* estudios. Una amplitud más pequeña requerirá de un error típico más pequeño, el cual se extrae de la fórmula (33) como

$$\hat{\sigma}_{(k+j)} = \frac{a'}{2 \cdot \left| z_{1-\alpha/2} \right|}, \tag{34}$$

donde a' corresponde a la amplitud que se desea alcanzar, y $z_{1-\alpha/2}$ es el valor de la distribución normal tipificada correspondiente al percentil $(1-\alpha/2)$.

Una vez obtenido el nuevo error típico que contiene los $k + j$ estudios, el procedimiento para calcular el número de estudios necesarios para alcanzarlo es el mismo que el descrito para la potencia a partir de la fórmula (14).

Ejemplo numérico. Retomamos el ejemplo del meta-análisis de Velthorst et al. (2015) para ilustrar el criterio basado en la amplitud como medida de la precisión. Fijando como valor $\alpha = 0.05$, la amplitud de la estimación combinada del TE es

$$a = 2 \cdot (1.96 \cdot 0.0758) = 0.297. \quad (35)$$

Supongamos ahora que los investigadores desean actualizar el meta-análisis cuando el número de estudios nuevos permita alcanzar, por ejemplo, una amplitud igual a aproximadamente la mitad ($a' = 0.15$). La nueva amplitud precisará de un error típico más bajo ($\hat{\sigma}_{(k+j)}$), el cual se calcula como

$$\hat{\sigma}_{(k+j)} = \frac{0.15}{2 \cdot 1.96} = 0.0383. \quad (36)$$

Recuérdese que el valor de la distribución normal tipificada correspondiente al percentil $(1 - 0.05/2)$ es igual a 1.96.

Al igual que con el criterio basado en la potencia, el sumatorio de los pesos se extrae del inverso del cuadrado del error típico, que es igual a

$$W_{k+j} = \frac{1}{0.0383^2} = 681.714. \quad (37)$$

El peso debido a los estudios que hay que añadir es igual a 681.714 – 174.045 = 507.669. Al igual que hacíamos con el criterio de la potencia, vamos a expresar este peso en términos de participantes (asumiendo excepcionalmente un modelo de EF) y en términos de estudios.

En términos de participantes habría que añadir un único estudio con una varianza igual a

$$\mathrm{var}_j = \frac{1}{507.669} = 0.00197. \quad (38)$$

Al igual que con el ejemplo de la potencia, la varianza del hipotético estudio es menor que la varianza interestudios del meta-análisis, por lo que es imposible realizar un estudio con esa varianza. Si asumimos, por motivos didácticos, un modelo de efecto fijo, el total de la varianza es debido a la varianza de muestreo. Por lo tanto, para calcular el tamaño muestral en este meta-análisis este se debe extraer de la fórmula de la varianza de muestreo para el índice *d*. Para ello debemos establecer un tamaño del efecto que, para dar continuidad al ejemplo, será $\delta = 0.2$. Representaremos por n_1 al tamaño de una muestra y n_2 al de la otra muestra, siendo el tamaño muestral total igual a $N = n_1 + n_2$. Asumiendo tamaños de cada grupo iguales $(n_1 = n_2)$, el tamaño muestral es aproximadamente igual a

$$N = \frac{4}{0.00197} + \frac{2 \cdot 0.2^2}{4 \cdot 0.00197} = 2040.61 \approx 2042. \quad (39)$$

Por lo tanto, haría falta un estudio con tamaño total igual a 2042 sujetos; es decir, dos grupos de tamaño 2042 / 2 = 1021 participantes.

Retomamos el modelo original de EA para expresar el peso en términos de estudios. Concretamente, del número de estudios con un peso igual a la media de los pesos originales necesarios para alcanzar W_j. El peso medio de los estudios originales de meta-análisis de Velthorst es $\bar{w} = 6.216$, tal y como se calculó en la ecuación (28). Por lo tanto, la estimación del número mínimo de estudios a añadir para alcanzar un amplitud de 0.15 es

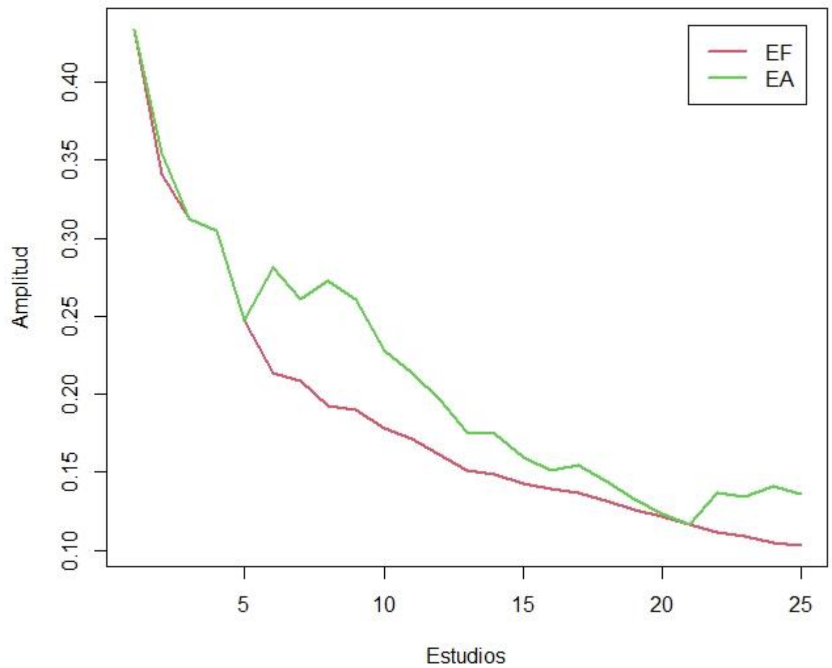
$$j = \frac{507.669}{6.216} = 81.671 \approx 82. \quad (40)$$

Se puede concluir que para que la amplitud del intervalo de confianza alcance el valor que hemos establecido que justificaría la actualización del meta-análisis (0.15), hay que añadir 82 estudios del mismo tipo que los incluidos en el meta-análisis original y con un tamaño muestral medio también igual al de los estudios ($N = 64$ para una $d = 0.2$).

Modelos de efecto fijo y efectos aleatorios. De igual modo que con la potencia, la agregación de estudios bajo un modelo de efecto fijo, independientemente de su variabilidad, siempre implica una reducción de la amplitud del intervalo de confianza de la estimación combinada del tamaño del efecto. No obstante, bajo un modelo de efectos aleatorios la estimación de la varianza interestudios puede aumentar si se re-estima tras agregar estudios con mucha variabilidad, lo que puede implicar un aumento de la amplitud. En la Figura 3 se puede apreciar cómo se reduce la amplitud si se sigue un modelo de EF y un modelo de EA re-estimando τ^2 después de añadir cada estudio. Los datos empleados son los mismos que se simularon para la realización de la Figura 2.

Figura 3

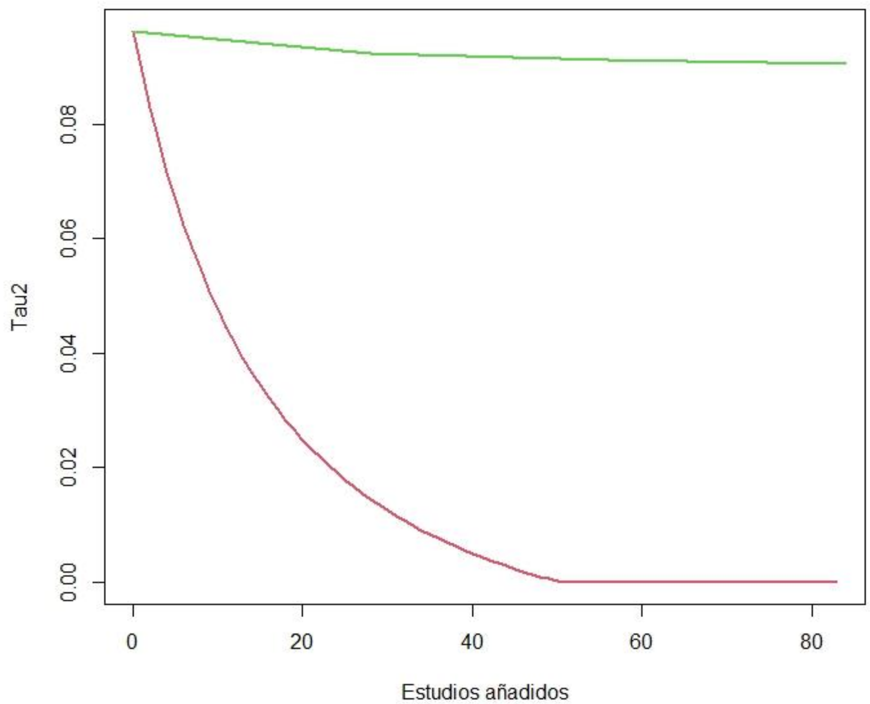
Comparación de la reducción de la amplitud del IC entre EF y EA en un caso simulado



Para los cálculos expuestos en el presente texto, se asume que los datos de los meta-análisis originales son una muestra representativa de la población de estudios, por lo que la estimación de τ^2 debería ser fiable. Además, con nuestro procedimiento de estimación se asume que los estudios a añadir son similares a los ya incluidos en el meta-análisis original, por lo que, a la larga, la varianza interestudios debería ser constante. Si re-estimásemos la τ^2 con cada estudio "similar" añadido, esta se reduciría drásticamente y de forma artificial, sesgando la estimación. Para ilustrarlo con un ejemplo, en la Figura 4 se puede observar cómo cambia la varianza interestudios del meta-análisis de Velthorst et al. (2015) si se re-estima tras añadir estudios similares mediante nuestro procedimiento (con un TE y un peso iguales a la media de los estudios incluidos originalmente en el meta-análisis) y si se re-estima al duplicar, triplicar y cuadruplicar los estudios originales. Ambos procedimientos consisten en la agregación a largo plazo de estudios muy similares, pero la τ^2 se reduce drásticamente si se re-estima mediante nuestro procedimiento. Por lo tanto, para no sesgar artificialmente la varianza interestudios mantenemos este valor constante tras la agregación de los *j* estudios adicionales necesarios para alcanzar una potencia determinada.

Figura 4

Cambio de τ^2 si se re-estima al agregar estudios



Nota. El cambio de τ^2 al re-estimarse tras cada "estudio similar" está representado por la línea de color granate, mientras que el cambio al duplicar, triplicar y cuadriplicar los estudios originales está representado por la línea de color verde.

Si un meta-análisis presenta una gran variabilidad, y existen razones suficientes como para pensar que la estimación de la varianza interestudios puede no ser adecuada, no recomendamos utilizar el criterio aquí expuesto para decidir cuándo publicar una actualización de dicho meta-análisis.

4.2. El método de Hartung y Knapp

El método estándar para realizar los contrastes meta-analíticos asume una distribución normal, cuyo estadístico de contraste es z (fórmula (6)). No obstante, si se sigue un modelo de efectos aleatorios no se tiene en cuenta la incertidumbre respecto a la variabilidad provocada por la estimación de la varianza interestudios (τ^2). Este hecho tiene como consecuencia la posibilidad de que los estadísticos de contraste tengan mayores tasas de error tipo I que el nivel de significación nominal (α).

Para solventar este problema, Hartung y Knapp (2001a, 2001b; véase también Hartung, 1999; Sidik & Jonkman, 2002, 2006) propusieron un estimador mejorado del error típico de la estimación combinada asumiendo una distribución *t* de Student con $k-1$ grados de libertad. Dado que los grados de libertad dependen de los estudios incluidos, sería necesario un proceso iterativo de cálculo para estimar cuántos estudios adicionales serían necesarios para alcanzar tanto un valor de potencia, como de amplitud, fijados a priori.

5. Un *script* en R para analizar y predecir la potencia y la amplitud

Hemos desarrollado un código (*script*) para el entorno estadístico R que permite al metodólogo realizar los cálculos aquí expuestos de forma sencilla (véase el anexo 1). R es un *software* de código libre muy versátil que permite al usuario programar todos los cálculos y procesos que desee realizar, así como utilizar el código desarrollado por otros usuarios, el cual se suele compilar en forma de "paquetes".

Los cálculos pre-programados se suelen ejecutar en forma de funciones, las cuales operan a través de una serie de *inputs* o "argumentos" (en lenguaje de R) y emiten uno o más *outputs* o salidas. Las salidas pueden ser almacenadas en forma de "objetos" para después poder operar con ellos de una forma más cómoda.

Para replicar los cálculos aquí expuestos hemos programado cuatro funciones, todas referidas a la diferencia de medias tipificadas:

- *"power.ma"*. Calcula la potencia de un contraste unilateral derecho de la estimación combinada de un meta-análisis para varios valores alternativos del tamaño del efecto (además de los introducidos manualmente, siempre calcula la potencia para los puntos de corte clásicos del TE propuestos por Cohen: 0.2, 0.5, 0.8).
- *"range.CI"*. Calcula la amplitud del intervalo de confianza de la estimación combinada de un meta-análisis.
- *"studies.power"*. Calcula cuántos estudios similares a los originales serían necesarios para alcanzar una potencia, fijada a priori, para varios valores alternativos del tamaño del efecto (además de los introducidos manualmente,

siempre calcula para los puntos de corte clásicos propuestos por Cohen: 0.2, 0.5, 0.8). También realiza estos cálculos en términos de tamaño muestral para diferencias de medias estandarizadas (asumiendo tamaños de grupo iguales; solo para EF). Las salidas son estos valores, además de un gráfico que estima la potencia que alcanzaría el meta-análisis añadiendo estudios uno a uno (marcando con una línea vertical el número de estudios necesarios para la potencia introducida como *input*).

- *"studies.range*". Calcula cuántos estudios similares a los originales serían necesarios para alcanzar una amplitud del intervalo de confianza, fijada a priori, de la estimación combinada. También realiza estos cálculos en términos de tamaño muestral para diferencias de medias estandarizadas (asumiendo tamaños de grupo iguales e introduciendo como *input* un valor de TE; solo para EF). Las salidas son estos dos valores, además de un gráfico que estima la amplitud que alcanzaría el intervalo de confianza añadiendo estudios uno a uno (marcando con una línea vertical el número de estudios necesarios para la amplitud introducida como *input*).

El código incluye numerosas anotaciones (todo texto escrito tras un símbolo "#") que permiten al usuario utilizar de una forma efectiva todas las funciones desarrolladas. El esquema general del código es el siguiente:

- Nombre y descripción de la función
- Argumentos necesarios para ejecutar la función y, si es necesario, los posibles valores que pueden adoptar.
- Código de la función

Algunos cálculos del código son realizados a través del paquete "metafor" (Viechtbauer, 2010), por lo que se necesitará cargarlo en el espacio de trabajo para que las funciones desarrolladas puedan funcionar correctamente. En el caso de no haber instalado nunca el paquete, será necesario hacerlo mediante el comando: `install.packages("metafor")`.

Ejemplo de uso. Para que el lector se familiarice con las funciones, vamos a replicar los cálculos realizados previamente para el meta-análisis de Velthorst et al. (2015). En primer lugar, cargamos los datos del meta-análisis (incluidos en el anexo 2) como un *data frame*:

```
data.velthorst <- read.table("velthorst.csv", header = T, sep = ";")
```

El *data frame* está compuesto por 28 observaciones (el número de estudios) y tres variables: un número del 1 al 28 como localizador del estudio (n), el tamaño del efecto de cada estudio (ES) y su varianza de muestreo (var). En segundo lugar, cargamos el paquete “metafor” en el sistema mediante el comando:

```
library(metafor)
```

Después ejecutamos el resto del código para cargar las cuatro funciones en el entorno global. Una vez cargadas, ejecutamos la primera función para conocer la potencia del contraste. Como argumentos marcamos (1) la columna del *data frame* con los tamaños del efecto (en nuestro caso ES); (2) la columna del *data frame* con las varianzas de los estudios (var); (3) el nombre del *data frame* que contiene las variables anteriores (data.velthorst); (4) el método de estimación de τ^2 o si se sigue un modelo de efecto fijo (máxima verosimilitud restringida, “REML”); (5) el tamaño del efecto paramétrico alternativo respecto al que se calculará la potencia (0.2); y (6) el nivel de significación (0.05).

```
power.ma(ES = ES, var = var, data = data.velthorst, method = "REML", theta =
    0.2, alpha = 0.05)
```

La función arroja el siguiente resultado:

```
   ES     Power
1 0.2 0.8400466
2 0.5 0.9999996
3 0.8 1.0000000
```

Para calcular cuántos estudios similares a los ya incluidos son necesarios para alcanzar una potencia igual a 0.90 ejecutamos la función “studies.power” con los siguientes argumentos: (1) la columna del *data frame* con los tamaños del efecto (ES); (2) la columna del *data frame* con las varianzas de cada estudio (var); (3) el nombre del *data frame* que contiene las columnas anteriores (data.velthorst); (4) el método de estimación de τ^2 o si se sigue un modelo de efecto fijo (máxima verosimilitud restringida, “REML”); (5) el tamaño del efecto paramétrico alternativo respecto al que se calculará la potencia (0.2); (6) el nivel de significación (0.05); y (7) la potencia que se desea alcanzar (0.9). También se puede incluir el número de estudios que se calculan para la realización del gráfico, pero en este caso lo dejaremos en el valor por defecto (50).

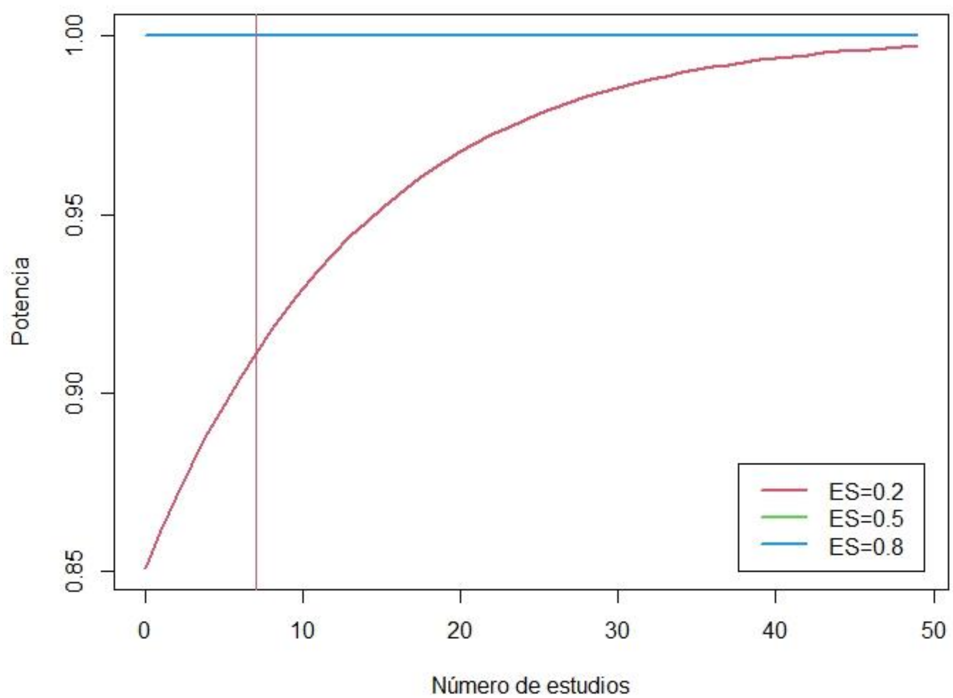
```
studies.power(ES = ES, var = var, data = data.velthorst, method = "REML", theta
    = 0.2, alpha = 0.05, power = 0.9)
```

Los resultados son la Figura 5, además del número de estudios similares a los originales estimados para alcanzar la potencia introducida ($Studies) y los datos empleados para el gráfico ($Data.plot), que se muestran en el anexo 3. Si la potencia que se desea alcanzar ya ha sido alcanzada con los estudios originales, la función devuelve el valor 0 (ya que no se necesita ningún estudio adicional más).

```
$Studies
   ES Studies
1 0.2       7
2 0.5       0
3 0.8       0
```

Figura 5

Aumento de la potencia del meta-análisis de Velthorst et al. (2015) según el número de estudios añadidos



Para calcular la amplitud intervalo de confianza ejecutamos la función “range.CI” con los siguientes argumentos: (1) la columna del *data frame* con los tamaños del efecto (ES); (2) la columna del *data frame* con las varianzas de cada estudio (var); (3) el nombre del *data frame* que contiene las columnas anteriores (data.velthorst); (4) el método de estimación de τ^2 o si se sigue un modelo de efecto fijo (máxima verosimilitud restringida, "REML"); y (5) el nivel de significación (0.05).

```
range.CI(ES = ES, var = var, data = data.velthorst, method = "REML", alpha =
0.05)
```

La función arroja el siguiente resultado:

```
[1] 0.2970202
```

Para calcular cuántos estudios similares a los ya incluidos son necesarios para reducir la amplitud a 0.15 ejecutamos la función "studies.range" con los siguientes argumentos: (1) la columna del *data frame* con los tamaños del efecto (ES); (2) la columna del *data frame* con las varianzas de cada estudio (var); (3) el nombre del *data frame* que contiene las columnas anteriores (data.velthorst); (4) el método de estimación de τ^2 o si se sigue un modelo de efecto fijo (máxima verosimilitud restringida, "REML"); (5) el nivel de significación (0.05); (6) la amplitud que se desea alcanzar (0.15); y (7) el número de estudios para realizar el gráfico (como 50 estudios no serán suficientes para alcanzar el valor deseado, lo ampliamos a 100).

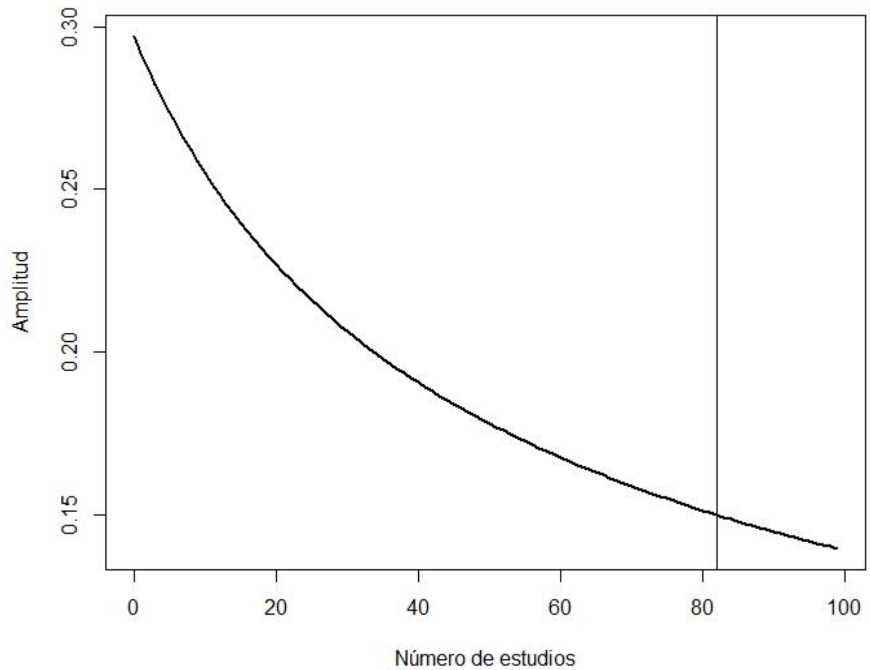
```
studies.range(ES = ES, var = var, data = data.velthorst, method = "REML", alpha
= 0.05, range = 0.15, plot.limit = 100)
```

Los resultados son la Figura 6, además del número de estudios similares a los originales estimados ($Studies) y los datos empleados para el gráfico ($Data.plot), que se muestran en el anexo 4.

```
$Studies
[1] 82
```

Figura 6

Reducción de la amplitud del IC del meta-análisis de Velthorst et al. (2015) según el número de estudios añadidos



Como se puede observar, todos los valores calculados a través del *script* se han replicado (con pequeñas diferencias debidas al redondeo), lo que permite llegar a la misma conclusión que si se realizan los cálculos manualmente.

6. Discusión y líneas futuras de investigación

Existen diversas razones para publicar un meta-análisis sobre una temática sobre la que ya hay otra revisión meta-analítica. Puede suceder que no estemos de acuerdo con las decisiones tomadas en el primer meta-análisis, lo que justificaría una re-elaboración del mismo. No obstante, en este estudio nos hemos centrado en una situación que es cada vez más usual: la aparición de nuevos estudios primarios que hagan dudar de la actualidad de un meta-análisis correctamente desarrollado.

Este desarrollo sirve como una primera aproximación al establecimiento de criterios externos que orienten la actualización de un meta-análisis. Los criterios desarrollados en el presente texto se ofrecen como una posible guía para actualizar un meta-análisis cuando se desea mejorar bien la potencia o bien la amplitud del IC del contraste relativo a la estimación combinada del TE. No obstante, existen otros contrastes meta-analíticos que pueden ser la referencia a la hora de decidir cuándo actualizar un meta-análisis. Uno de esos contrastes es el análisis de moderadoras, ya sean categóricas o continuas. Además, la potencia de estos contrastes suele ser mucho menor que la de la estimación combinada (Valentine et al., 2010). Una futura línea de trabajo debe ir encaminada en este sentido, desarrollando los cálculos necesarios para estimar cuántos estudios son necesarios para mejorar la potencia del contraste de moderadoras.

Otro aspecto a desarrollar es la sensibilidad o impacto que tiene el uso de estos criterios sobre el sesgo de publicación propio de los meta-análisis. Este sesgo se refiere a la mayor probabilidad de incluir en los meta-análisis estudios estadísticamente significativos, en detrimento de posibles investigaciones no publicadas que no arrojan resultados significativos o, en su defecto, arrojan tamaños del efecto muy pequeños (Botella & Sánchez-Meca, 2015). Existen muchos motivos por los que puede aparecer el sesgo de publicación, como la autocensura o las dificultad de publicar resultados no significativos en las revistas científicas (Francis, 2012).

La consecuencia del sesgo de publicación es una estimación combinada del tamaño del efecto irreal, al estar sesgada positivamente. En relación con las guías aquí expuestas, podríamos suponer, tentativamente, que agravarían el impacto del sesgo de publicación, minimizando el número de estudios necesarios para alcanzar la potencia o la precisión deseadas. Esto se debería a que los "estudios similares" que sirven para las estimaciones también estarían sesgados positivamente por el sesgo de publicación. No obstante, uno de los casos en los que sería de más interés la aplicación de estos criterios (especialmente el de la potencia) es la actualización de meta-análisis no significativos que, presumiblemente, adolecerán menos del sesgo de publicación.

Por último, otra línea de trabajo debe ir encaminada a mejorar y depurar el *script* para R, añadiendo funcionalidades nuevas como el cálculo del error típico siguiendo el método de Hartung y Knapp, el cálculo para otros índices del tamaño del efecto, así como estructuras de control de errores. Otro objetivo es desarrollar el código en forma de paquete, para que así pueda ser utilizado por cualquier usuario de R de una forma cómoda.

7. Referencias bibliográficas

Botella, J., & Caperos, J. M. (2019). *Metodología de Investigación en Psicología General Sanitaria*. Síntesis.

Botella, J., & Sánchez-Meca, J. (2015). *Meta-análisis en ciencias sociales y de la salud*. Síntesis.

Botella, J., Ximénez, C., Revuelta, J., & Suero, M. (2006). Optimization of sample size in controlled experiments: The CLAST rule. *Behavior Research Methods*, *38*(1), 65-76. https://doi.org/10.3758/BF03192751

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, *8*(3), 243-253. https://doi.org/10.1037/1082-989X.8.3.243

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis* (3ª). Russell Sage Foundation.

Francis, G. (2012). The Psychology of Replication and Replication in Psychology. *Perspectives on Psychological Science*, *7*, 585-594.

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

Hartung, J. (1999). An alternative method for meta-analysis. *Biometrical Journal*, *41*, 901-906.

Hartung, J., & Knapp, G. (2001a). On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine*, *20*(12), 1771-1782. https://doi.org/10.1002/sim.791

Hartung, J., & Knapp, G. (2001b). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, *20*(24), 3875-3889. https://doi.org/10.1002/sim.1009

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, *6*(3), 203-217.

Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis: Power analysis for meta-analysis. *Research Synthesis Methods*, *8*(3), 290-302. https://doi.org/10.1002/jrsm.1240

Rothery, C., Griffin, S., Koffijberg, H., & Claxton, K. (2022). Using meta-analysis to plan further research. En C. H. Schmid, T. Stijnen, & I. White (Eds.), *Handbook of meta-analysis* (pp. 523-543). CRC Press.

Sidik, K., & Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Statistics in Medicine*, *21*(21), 3153-3159. https://doi.org/10.1002/sim.1262

Sidik, K., & Jonkman, J. N. (2006). Robust variance estimation for random effects meta-analysis. *Computational Statistics & Data Analysis*, *50*(12), 3681-3701. https://doi.org/10.1016/j.csda.2005.07.019

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need?: A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215-247. https://doi.org/10.3102/1076998609346961

Velthorst, E., Koeter, M., van der Gaag, M., Nieman, D. H., Fett, A.-K. J., Smit, F., Staring, A. B. P., Meijer, C., & de Haan, L. (2015). Adapted cognitive–behavioural therapy required for targeting negative symptoms in schizophrenia: Meta-analysis and meta-regression. *Psychological Medicine*, *45*(3), 453-465. https://doi.org/10.1017/S0033291714001147

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261-293.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

8. Anexos

8.1. Anexo 1. Script de R

```
## Paquetes de apoyo necesarios ##
# install.packages("metafor") # Si no se tuviera el paquete instalado
library(metafor) # Se usa para calcular la varianza de la estimación combinada

## -power.ma-       Función para el cálculo de la potencia para un contraste
unilateral derecho ##

# Argumentos #
# -ES-          Columna de un data frame con los tamaños del efecto de cada
estudio (Para metafor)
# -var-         Columna de un data frame con las varianzas de cada estudio
(Para metafor)
# -data-    Data frame con los datos (TEs y varianzas) (Para metafor)
# -method-      Método de estimación de tau^2 (Para metafor). Los códigos más
comunes son (ejecute ?rma.uni para todos los códigos):
  # -"FE"-        (Efecto fijo. No se estima tau^2)
  # -"REML"-      (Máxima Verosimilitud Restringida). Por defecto
# -theta-   TE paramétrico alternativo. Por defecto se calcula la potencia para
los puntos de corte clásicos propuestos por Cohen (0.2, 0.5, 0.8), y siempre
se ofrecen estos valores a mayores.
# -alpha-   Nivel de significación. Por defecto = 0.05

# Programación de la función #
power.ma <- function(ES, var, data, method="REML", theta=NULL, alpha=0.05) {

  # Vector de tamaños del efecto alternativos ordenados
  ES <- sort(c(theta, 0.2, 0.5, 0.8))
  ES <- ES[!duplicated(ES)] # Se eliminan los duplicados si los hubiere

  # Cálculo del error típico
  ma <- rma.uni(yi=ES, vi=var, data=data, method=method, weighted=T)
  SE <- ma$se # Valor del error típico

  # Tamaño del efecto mínimo para alcanzar significación estadística
  ES.alpha <- qnorm(1 - alpha) * SE

  # Cálculo final de la potencia
  power <- 1 - pnorm((ES.alpha - ES)/SE)
```

```
  # Retorno de la función
  print(data.frame(ES = ES,
                   Power = power))
}


## -range.CI-          Función para el cálculo de la amplitud del IC de la
estimación combinada del TE ##

# Argumentos #
# -ES-           Columna de un data frame con los tamaños del efecto de cada
estudio (Para metafor)
# -var-          Columna de un data frame con las varianzas de cada estudio
(Para metafor)
# -data-    Data frame con los datos (TEs y varianzas) (Para metafor)
# -method-      Método de estimación de tau^2 (Para metafor). Los códigos más
comunes son (ejecute ?rma.uni para todos los códigos):
  # -"FE"-        (Efecto fijo. No se estima tau^2)
  # -"REML"-      (Máxima Verosimilitud Restringida). Por defecto
# -alpha-   Nivel de significación. Por defecto = 0.05

range.CI <- function(ES, var, data, method="REML", alpha=0.05) {

  # Cálculo del error típico
  ma <- rma.uni(yi=ES, vi=var, data=data, method=method, weighted=T)
  SE <- ma$se # Valor del error típico

  # Cálculo de la amplitud
  range <- 2 * SE * abs(qnorm(1 - alpha/2))

  # Retorno de la función
  print(range)
}


## -studies.power-   Función para el cálculo prospectivo del número de
estudios/sujetos necesarios para alcanzar una determinada potencia ##

# Argumentos #
# -ES-           Columna de un data frame con los tamaños del efecto de cada
estudio (Para metafor)
# -var-          Columna de un data frame con las varianzas de cada estudio
(Para metafor)
```

```
# -data-    Data frame con los datos (TEs y varianzas) (Para metafor)
# -method-      Método de estimación de tau^2 (Para metafor). Los códigos más comunes son (ejecute ?rma.uni para todos los códigos):
  # -"FE"-        (Efecto fijo. No se estima tau^2)
  # -"REML"-      (Máxima Verosimilitud Restringida). Por defecto
# -theta-   TE paramétrico alternativo. Por defecto se calcula la potencia para los puntos de corte clásicos propuestos por Cohen (0.2, 0.5, 0.8) y siempre se ofrecen estos valores a mayores.
# -alpha-   Nivel de significación. Por defecto = 0.05
# -power-   Valor de potencia que se desea alcanzar.
# -plot.limit-  Total de estudios que se desea estimar para el gráfico. Por defecto = 50

studies.power <- function(ES, var, data, method="REML", theta=NULL, alpha=0.05, power, plot.limit=50) {

  # Vector de tamaños del efecto alternativos ordenados
  es <- sort(c(theta, 0.2, 0.5, 0.8))
  es <- es[!duplicated(es)] # Se eliminan los duplicados si los hubiere

  # Cálculo de los datos del estudio original
  ma <- rma.uni(yi=ES, vi=var, data=data, method=method, weighted=T)
  SE_k <- ma$se # Valor del error típico original

  W_k <- 1/SE_k^2 # Peso total de los estudios originales

  w.mean_k <- W_k/length(data$ES) # Peso medio de los estudios originales

  # Cálculo del error típico de los k+j estudios
  SE_kj <- es/(qnorm(1 - alpha) - qnorm(1 - power))

  # Cálculo del peso de los k+j estudios
  W_kj <- 1/SE_kj^2

  # Cálculo del peso de los j estudios
  W_j <- W_kj - W_k

  # Cálculo de la varianza de los j estudios
  var_j <- 1/W_j

  # Cálculo del tamaño muestral estimado (sólo si method = EF)
  if (method == "FE") {
    Subjets <- 1:length(es)
```

```
  for (i in 1:length(Subjets)) {
    Subjets[i] <- ceiling((4/var_j[i]) + ((2 * es[i]^2)/(4 * var_j[i])))
  }
}

# Cálculo de los estudios estimados
Studies <- ceiling(W_j/w.mean_k)

# Cálculo de la potencia estudio a estudio (para el gráfico)
# Data frame para contener los datos
data.plot <- data.frame(j = 0:(plot.limit-1))

for (i in 1:length(es)) {
  # Vector con las potencias para un TE determinado
  c <- 1 - pnorm(((qnorm(1 - alpha) * SE_k) - es[i])/SE_k)

  # Mecanismo de control del bucle
  meter <- 0

  # Bucle
  repeat {
    # Actualización del contador
    meter <- meter + 1

    # Re-cálculo del error típico con los k+j estudios
    Wkj <- (1/SE_k^2) + (meter * w.mean_k)
    SEkj <- sqrt(1/Wkj)

    # Re-cálculo de la potencia
    powerkj <- 1 - pnorm(((qnorm(1 - alpha) * SEkj) - es[i])/SEkj)

    # Asignación de la potencia al vector
    c[meter] <- powerkj

    # Cierre del bucle
    if (meter == plot.limit)
      break
  }

  # Unión del nuevo vector al data frame
  data.plot <- cbind(data.plot, c)
}
```

```
  # Renombrar columnas del data frame
  cnames <- c("j", paste("ES=", es, sep=""))
  colnames(data.plot) <- cnames

  # Creación del gráfico (Primera línea)
  plot(x=data.plot[,1], y=data.plot[,2],
       type="l", lwd = 2,
       xlab = "Número de estudios", ylab = "Potencia",
       ylim=c(min(data.plot[,2]),max(data.plot[,ncol(data.plot)])), col=2)
  # Siguiente líneas
  for (i in 3:ncol(data.plot)) {
    lines(x=data.plot[,1], y=data.plot[,i], col=i, lwd=2, type="l")
  }
  # Líneas verticales para el número de estudios que alcanzan la potencia
introducida
  for (i in 1:length(cnames)) {
    abline(v=Studies[i], col=(i+1))
  }

  # Leyenda
  legend("bottomright",
         inset=0.03,
         legend=cnames[-1],
         col=c(2:ncol(data.plot)),
         lwd=2)

  # Retorno de la función (para method = FE)
  if (method == "FE"){
    results <- list(Subjets = data.frame(ES = es,
                                         Subjets = Subjets),
                    Studies = data.frame(ES = es,
                                         Studies = Studies),
                    Data.plot = data.plot)
  } else {
    results <- list(Studies = data.frame(ES = es,
                                         Studies = Studies),
                    Data.plot = data.plot)
  }

  # Se coercionan los valores negativos y NAs a 0
  for (i in 1:length(names(results))){
    results[[i]][results[[i]] < 0] <- 0
    results[[i]][is.na(results[[i]])] <- 0
```

```
  }

  print(results)
}


## -studies.range- Función para el cálculo prospectivo del número de estudios/sujetos necesarios para alcanzar una determinada amplitud del IC ##

# Argumentos #
# -ES-          Columna de un data frame con los tamaños del efecto de cada estudio (Para metafor)
# -var-         Columna de un data frame con las varianzas de cada estudio (Para metafor)
# -data-    Data frame con los datos (TEs y varianzas) (Para metafor)
# -method-      Método de estimación de tau^2 (Para metafor). Los códigos más comunes son (ejecute ?rma.uni para todos los códigos):
  # -"FE"-        (Efecto fijo. No se estima tau^2)
  # -"REML"-      (Máxima Verosimilitud Restringida). Por defecto
# -alpha-   Nivel de significación. Por defecto = 0.05
# -range-   Valor de amplitud del IC que se desea alcanzar.
# -theta-     TE paramétrico alternativo (Solo necesario si method="FE" y metric="SMD". Por defecto se calcula la potencia para los puntos de corte clásicos propuestos por Cohen (0.2, 0.5, 0.8) y siempre se ofrecen estos valores a mayores.
# -plot.limit-  Total de estudios que se desea estimar para el gráfico. Por defecto = 50

studies.range <- function(ES, var, data, method="REML", alpha=0.05, range, theta=NULL, plot.limit=50) {

  # Vector de tamaños del efecto alternativos ordenados (Solo se usa si method="FE" y metric="SMD")
  es <- sort(c(theta, 0.2, 0.5, 0.8))
  es <- es[!duplicated(es)] # Se eliminan los duplicados si los hubiere

  # Cálculo de los datos del estudio original
  ma <- rma.uni(yi=ES, vi=var, data=data, method=method, weighted=T)
  SE_k <- ma$se # Valor del error típico original

  W_k <- 1/SE_k^2 # Peso total de los estudios originales

  w.mean_k <- W_k/length(data$ES) # Peso medio de los estudios originales
```

```
  # Cálculo del error típico de los k+j estudios
  SE_kj <- range/(2 * abs(qnorm(1 - alpha/2)))

  # Cálculo del peso de los k+j estudios
  W_kj <- 1/SE_kj^2

  # Cálculo del peso de los j estudios
  W_j <- W_kj - W_k

  # Cálculo de la varianza de los j estudios
  var_j <- 1/W_j

  # Cálculo del tamaño muestral estimado (solo para method="FE")
  if (method == "FE") {
    Subjets <- 1:length(es)
    for (i in 1:length(Subjets)) {
      Subjets[i] <- ceiling((4/var_j) + ((2 * es[i]^2)/(4 * var_j)))
    }
  }

  # Cálculo de los estudios estimados
  Studies <- ceiling(W_j/w.mean_k)

  # Cálculo de la amplitud estudio a estudio (para el gráfico)
  # Data frame que recoge la información necesaria para el gráfico (Nº de
estudios añadidos y la amplitud correspondiente)
  data.plot <- data.frame(j = 0,
                          range = range <- 2 * SE_k * abs(qnorm(1 - alpha/2))
  )

  # Mecanismo de control del bucle
  meter <- 0

  # Bucle
  repeat {
    # Actualización del contador
    meter <- meter + 1

    # Re-cálculo del error típico con los k+j estudios
    Wkj <- (1/SE_k^2) + (meter * w.mean_k)
    SEkj <- sqrt(1/Wkj)
```

```
    # Re-cálculo de la amplitud
    rangekj <- 2 * SEkj * abs(qnorm(1 - alpha/2))

    # Asignación de los nuevos datos al data frame
    data.plot <- rbind(data.plot, c(meter, rangekj))

    # Cierre del bucle
    if (meter == (plot.limit - 1))
      break
  }

  # Gráfico
  plot(data.plot, type="l", xlab = "Número de estudios", ylab = "Amplitud", lwd
= 2)
  abline(v = Studies)


  # Retorno de la función (para method = FE)
  if (method == "FE"){
    if (metric == "SMD"){
      results <- list(Subjets = data.frame(ES = es,
                                            Subjets = Subjets),
                      Studies = Studies,
                      Data.plot = data.plot)
    } else if (metric == "PC"){
      results <- list(Subjets = Subjets,
                      Studies = Studies,
                      Data.plot = data.plot)
    }

  } else {
    results <- list(Studies = Studies,
                    Data.plot = data.plot)
  }

  # Se coercionan los valores negativos y NAs a 0
  for (i in 1:length(names(results))){
    results[[i]][results[[i]] < 0] <- 0
    results[[i]][is.na(results[[i]])] <- 0
  }

  print(results)
}
```

8.2. Anexo 2. Datos del meta-análisis de Velthorst et al. (2015)

Tabla 2

Datos del meta-análisis de Velthorst et al. (2015) necesarios para el ejemplo

n	ES	var
1	0.389	0.027889
2	1.761	0.229441
3	0.364	0.044521
4	1.235	0.123904
5	0.7	0.112896
6	0.372	0.095481
7	0.445	0.081225
8	-0.425	0.046225
9	0.563	0.044944
10	0.013	0.135424
11	-0.495	0.072361
12	-0.313	0.035344
13	-0.044	0.056169
14	0.375	0.097969
15	-0.121	0.033124
16	-0.112	0.018225
17	0.214	0.057121
18	0.188	0.063001
19	-0.096	0.060025
20	-0.123	0.042849
21	0.12	0.0529
22	0.208	0.049284
23	-0.008	0.126736
24	0.066	0.085849
25	0.414	0.097969
26	-0.386	0.062001
27	-0.422	0.039204
28	-0.126	0.051984

8.3. Anexo 3. Output de la función "studies.power" ($Data.plot)

```
$Data.plot
        j      ES=0.2      ES=0.5    ES=0.8
1     0     0.8511480   0.9999998         1
2     1     0.8615460   0.9999999         1
3     2     0.8712782   0.9999999         1
4     3     0.8803810   1.0000000         1
5     4     0.8888896   1.0000000         1
6     5     0.8968378   1.0000000         1
7     6     0.9042578   1.0000000         1
8     7     0.9111807   1.0000000         1
9     8     0.9176361   1.0000000         1
10    9     0.9236521   1.0000000         1
11    10    0.9292558   1.0000000         1
12    11    0.9344727   1.0000000         1
13    12    0.9393269   1.0000000         1
14    13    0.9438415   1.0000000         1
15    14    0.9480382   1.0000000         1
16    15    0.9519376   1.0000000         1
17    16    0.9555592   1.0000000         1
18    17    0.9589212   1.0000000         1
19    18    0.9620410   1.0000000         1
20    19    0.9649348   1.0000000         1
21    20    0.9676178   1.0000000         1
22    21    0.9701044   1.0000000         1
23    22    0.9724082   1.0000000         1
24    23    0.9745417   1.0000000         1
25    24    0.9765169   1.0000000         1
26    25    0.9783448   1.0000000         1
27    26    0.9800358   1.0000000         1
28    27    0.9815996   1.0000000         1
29    28    0.9830453   1.0000000         1
30    29    0.9843815   1.0000000         1
31    30    0.9856159   1.0000000         1
32    31    0.9867561   1.0000000         1
33    32    0.9878088   1.0000000         1
34    33    0.9887805   1.0000000         1
35    34    0.9896771   1.0000000         1
36    35    0.9905043   1.0000000         1
37    36    0.9912672   1.0000000         1
38    37    0.9919705   1.0000000         1
39    38    0.9926188   1.0000000         1
40    39    0.9932162   1.0000000         1
41    40    0.9937666   1.0000000         1
42    41    0.9942735   1.0000000         1
43    42    0.9947403   1.0000000         1
44    43    0.9951700   1.0000000         1
45    44    0.9955655   1.0000000         1
46    45    0.9959293   1.0000000         1
47    46    0.9962641   1.0000000         1
48    47    0.9965719   1.0000000         1
49    48    0.9968550   1.0000000         1
50    49    0.9971152   1.0000000         1
```

8.4. Anexo 4. Output de la función "studies.range" ($Data.plot)

```
$Data.plot
        j        power
1       0      0.2970202
2       1      0.2918542
3       2      0.2869487
4       3      0.2822826
5       4      0.2778369
6       5      0.2735949
7       6      0.2695414
8       7      0.2656629
9       8      0.2619472
10      9      0.2583831
11      10     0.2549606
12      11     0.2516707
13      12     0.2485049
14      13     0.2454556
15      14     0.2425159
16      15     0.2396794
17      16     0.2369401
18      17     0.2342927
19      18     0.2317320
20      19     0.2292535
21      20     0.2268529
22      21     0.2245261
23      22     0.2222695
24      23     0.2200796
25      24     0.2179532
26      25     0.2158873
27      26     0.2138790
28      27     0.2119257
29      28     0.2100250
30      29     0.2081745
31      30     0.2063721
32      31     0.2046157
33      32     0.2029034
34      33     0.2012334
35      34     0.1996039
36      35     0.1980134
37      36     0.1964604
38      37     0.1949433
39      38     0.1934608
40      39     0.1920116
41      40     0.1905946
42      41     0.1892084
43      42     0.1878520
44      43     0.1865245
45      44     0.1852246
46      45     0.1839516
47      46     0.1827044
48      47     0.1814823
49      48     0.1802844
50      49     0.1791099
51      50     0.1779581
52      51     0.1768281
53      52     0.1757195
54      53     0.1746314
55      54     0.1735633
56      55     0.1725146
57      56     0.1714847
58      57     0.1704730
59      58     0.1694789
60      59     0.1685021
```

61	60	0.1675420
62	61	0.1665981
63	62	0.1656699
64	63	0.1647571
65	64	0.1638593
66	65	0.1629759
67	66	0.1621067
68	67	0.1612513
69	68	0.1604092
70	69	0.1595802
71	70	0.1587640
72	71	0.1579601
73	72	0.1571683
74	73	0.1563883
75	74	0.1556198
76	75	0.1548625
77	76	0.1541162
78	77	0.1533806
79	78	0.1526553
80	79	0.1519403
81	80	0.1512353
82	81	0.1505399
83	82	0.1498541
84	83	0.1491775
85	84	0.1485101
86	85	0.1478515
87	86	0.1472016
88	87	0.1465602
89	88	0.1459271
90	89	0.1453021
91	90	0.1446851
92	91	0.1440759
93	92	0.1434744
94	93	0.1428803
95	94	0.1422935
96	95	0.1417139
97	96	0.1411413
98	97	0.1405756
99	98	0.1400166
100	99	0.1394643