

# **Trabajo Final**

David Fernández y Franco Vicario

## **1. Presentación del Problema**

El fenómeno conocido como “Efecto Pigmalión” o tambien conocido como profecía autocomplida en el ámbito educativo propone que las expectativas que un profesor tiene sobre el rendimiento de un estudiante pueden influir causalmente en el desempeño real de este. El estudio pionero de Rosenthal y Jacobson (1968), *Pygmalion in the Classroom*, comentó que si a los maestros se les hacía creer que ciertos alumnos eran “promesas intelectuales” (aunque fueran seleccionados al azar), esos alumnos mostraban un incremento real en su Cociente Intelectual (CI) al final del año escolar.

Este descubrimiento generó una controversia grande en la psicología educativa y la metodología de la investigación. Mientras algunos estudios posteriores lograron replicar los resultados, otros encontraron efectos nulos o incluso negativos. Entonces Raudenbush (1984) recopiló estos estudios para intentar explicar dicha variabilidad.

El problema que afrontar este trabajo final es la necesidad de sintetizar esta evidencia dispersa. Dado que los estudios individuales presentan resultados contradictorios, el metaanálisis es la herramienta correcta para esto.

Trabajamos con el dataset dat.raudenbush1985 de la libreria metadat. El dataset esta conformado por 19 estudios y 10 columnas.

## **2. Metodología Usada**

Siguiendo las pautas de Harrer et al. (2021) en Doing Meta-Analysis with R y la linea del curso, se procedió con la siguiente estrategia metodológica.

## 2.1. Métricas del Tamaño del Efecto

Debido a que los estudios originales miden el desarrollo intelectual(CI) utilizando diferentes medidas, los puntajes directos de cada estudio no son directamente comparables entre ellos. Por eso, se utilizó la Diferencia de Medias Estandarizada (SMD).

Específicamente, se usará la métrica de Hedges' g. A diferencia de la d de Cohen, la g de Hedges incluye un factor de corrección J que reduce el sesgo positivo que ocurre en muestras pequeñas (comunes en este dataset). La estimación para cada estudio k se define como:

$$g_k = \frac{\bar{x}_t - \bar{x}_c}{s_{pooled}} \times J$$

Donde  $s_{pooled}$  es la desviación estándar agrupada. En el código de R, esto se implementa cuando se pone el argumento sm = "SMD".

## 2.2. Selección del Modelo: Efectos Aleatorios

Se seleccionó un Modelo de Efectos Aleatorios debido a que el modelo de efectos fijos asume que todos los estudios estiman el mismo tamaño del efecto verdadero  $\theta$  y que la única fuente de variabilidad es el error de muestreo ( $\epsilon_k$ ). Pero, los estudios de Raudenbush difieren en diseño(semanas de contacto), población y procedimientos. Entonces asumir que hay un único efecto verdadero idéntico para todos es irreal. El modelo de efectos aleatorios asume que los efectos verdaderos  $\theta_k$  siguen una distribución normal:

$$\theta_k \sim N(\mu, \tau^2)$$

Donde  $\tau^2$  representa la varianza entre estudios.

### Estimación de la Heterogeneidad ( $\tau^2$ )

Para estimar  $\tau^2$ , se utilizó el método de Máxima Verosimilitud Restringida. Este método es preferible al estimador de DerSimonian-Laird debido a que evita el sesgo negativo en la estimación de la varianza, dando resultados más robustos cuando el número de estudios es moderado, como este caso( $k = 19$ ).

## 2.3. Evaluación de Datos Faltantes

Se inspeccionó el conjunto de datos `raudenbush1985`.

Se verificó si había valores perdidos (`NA`) en los tamaños del efecto, varianzas y moderadores (`weeks, setting, tester`). El dataset se encuentra completo para las variables de interés. No fue necesario aplicar técnicas de imputación múltiple o análisis de sensibilidad por datos faltantes, ya que todos los estudios reportaron los estadísticos necesarios para el cálculo del tamaño del efecto.

## 2.4. Pertinencia del Análisis Multivariado

Se evaluó si era necesario aplicar un modelo de Metaanálisis Multivariado. Este tipo de análisis se usa cuando los estudios traen múltiples resultados correlacionados para los mismos sujetos o múltiples puntos temporales. En este caso, cada estudio aporta un único tamaño del efecto resumen (CI Global). Dado que los efectos son independientes entre estudios, el modelo univariado de efectos aleatorios es el enfoque estadísticamente correcto.

## 3. Resultados I: Modelo General y Heterogeneidad

A continuación, se presentan los resultados de la estimación del modelo de efectos aleatorios. Se usó el método de estimación **REML** (Máxima Verosimilitud Restringida) para el parámetro de heterogeneidad  $\tau^2$ , siendo considerado el enfoque de menor sesgo y más robusto para la estimación de la varianza inter-estudios en metaanálisis con un número chico de estudios ( $k=19$ ).

### 3.1. Estimación del Efecto Global

Number of studies:  $k = 19$

	SMD	95%-CI	z	p-value
Common effect model	0.0604	[-0.0111; 0.1318]	1.66	0.0979
Random effects model	0.0837	[-0.0175; 0.1849]	1.62	0.1051
Prediction interval		[-0.2243; 0.3917]		

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0188$  [0.0059; 0.1977];  $\tau = 0.1372$  [0.0765; 0.4447]  
 $I^2 = 49.8\%$  [14.7%; 70.4%];  $H = 1.41$  [1.08; 1.84]

Test of heterogeneity:

Q d.f. p-value

35.83 18 0.0074

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q
- Prediction interval based on t-distribution (df = 18)

## Interpretación Estadística

La tabla resumen (**summary**) da la base estadística central del metaanálisis. Se detallan los hallazgos principales divididos por componentes estadísticos:

### 1. Estimación del Efecto Global (Modelo de Efectos Aleatorios)

La fila “Random effects model” presenta el resultado combinado ponderado bajo el supuesto de que hay heterogeneidad.

- **SMD (Tamaño del Efecto):** El valor estimado es 0.0837. Esto indica que, en promedio, el grupo experimental (con expectativas inducidas) superó al grupo control por solo 0.0837 desviaciones estándar. Esto se considera un efecto insignificante.
- **Intervalo de Confianza:** El intervalo [-0.0175; 0.1849] cruza el numero cero. Implicando que con un 95% de confianza el verdadero tamaño del efecto podría ser tanto ligeramente negativo como positivo.
- **Estadístico z:** Este valor de prueba indica que la estimación del efecto se aleja 1.62 errores estándar del cero.
- **p-valor:** Al ser el valor  $p > 0.05$ , no se rechaza la hipótesis nula. No existe evidencia estadística suficiente para afirmar que las expectativas del profesor tengan un efecto sobre el Coeficiente Intelectual de los alumnos.

### 2. Intervalo de Predicción

El intervalo de predicción estimado es [-0.2243; 0.3917]. Que este intervalo sea amplio y que incluya valores negativos demarca la incertidumbre y la dependencia del contexto del efecto Pigmalión.

### 3.2 Análisis de Heterogeneidad

La heterogeneidad cuantifica la variabilidad en los resultados que no puede explicarse solo por el error de muestreo. Se evalúó la variabilidad entre los estudios.

- $\tau^2$  ( $\text{tau}^2$ ): Representa la varianza total entre los efectos verdaderos de los estudios. Un valor distinto de cero confirma la presencia de variabilidad entre estudios.
- $I^2$  : Este estadístico dice qué porcentaje de la variación total en los estudios se debe a la heterogeneidad y no al azar muestral. El valor 49.8% se clasifica como heterogeneidad moderada, indicando que casi la mitad de la dispersión observada se debe a diferencias reales en los diseños o contextos de los estudios, pudiendo justificando el uso del modelo de efectos aleatorios.
- **Interpretación:** Dado que  $p < 0.01$ , se rechaza la hipótesis nula de homogeneidad. Esto confirma que los estudios no comparten un único efecto verdadero entre todos, volviendo a confirmar la decisión de usar un modelo de efectos aleatorios (`method.tau = "REML"`) en lugar de uno de efectos fijos.

## 4. Análisis de Valores Atípicos e Influencia

Debido la heterogeneidad moderada-alta ( $I^2 \approx 50\%$ ), es imprecindible identificar si existen estudios específicos que estén distorsionando el promedio global, Outliers.

### 4.1 Detección de los Outliers

Se utilizó la función `find.outliers`, que identifica estudios donde sus intervalos de confianza no se superponen con el intervalo de confianza del efecto combinado.

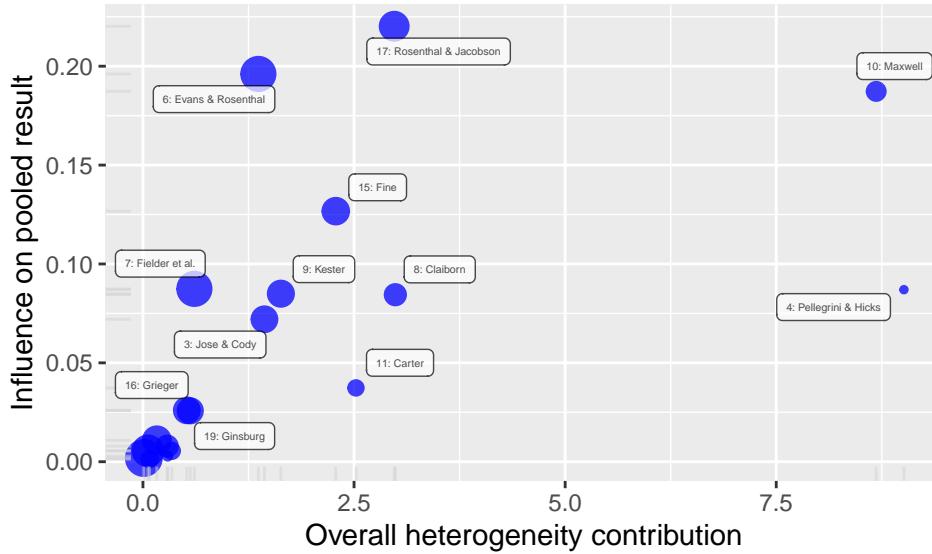
**Hallazgos:** La función identificó dos estudios atípicos:

1. Pellegrini & Hicks (Estudio 4)
2. Maxwell (Estudio 10)

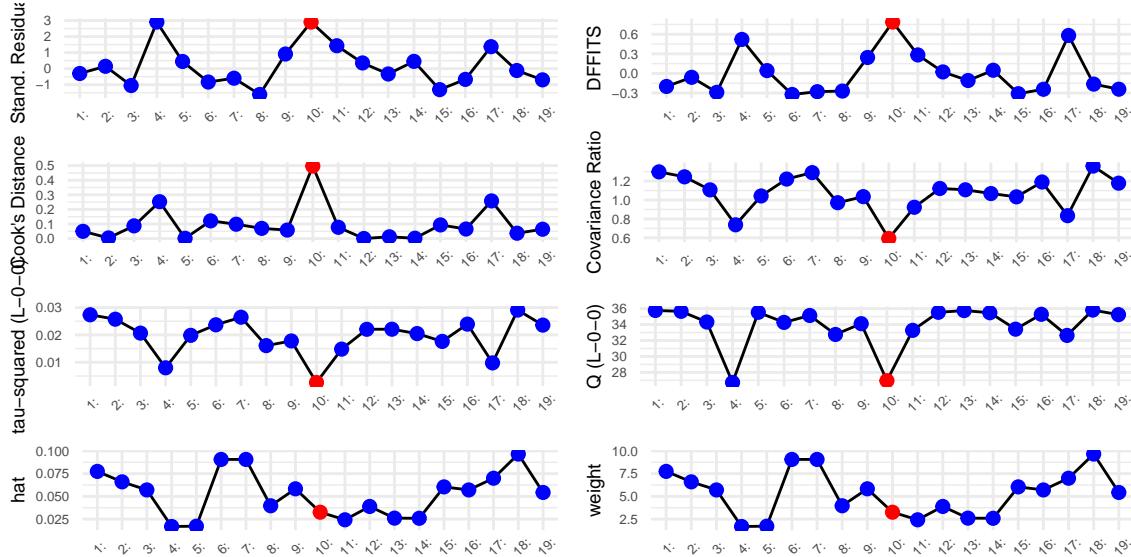
**Análisis de Sensibilidad:** La salida muestra que quitar estos dos estudios, la heterogeneidad  $I^2$  baja del 49.8% al 9.1% y la prueba Q deja de ser significativa ( $p = 0.34$ ). Demostrando que la “inconsistencia” del metaanálisis se debe en gran proporción por estos dos casos extremos. Sin embargo, se puede notar que el efecto global recalculado sin esos outliers ( $SMD=0.03$ ) sigue siendo no significativo.

### 4.2. Diagnósticos de Influencia (Influence Analysis)

Para confirmar visualmente la influencia de estos estudios, utilizamos un conjunto de diagnósticos gráficos.



Este gráfico de baujat muestra simultáneamente cuánto contribuye cada estudio a la heterogeneidad total y cuánto influye sobre el efecto combinado. La mayor parte de los estudios se ubica en la zona inferior izquierda, lo que indica baja influencia y baja contribución a la heterogeneidad; esto sugiere que el conjunto de evidencia es relativamente estable. Sin embargo, algunos estudios se destacan: *Maxwell* presenta la contribución más alta a la heterogeneidad, separándose claramente del resto, mientras que *Pellegrini & Hicks* también aparece como un estudio con contribución moderada. Por otro lado, *Rosenthal & Jacobson* muestra la mayor influencia sobre el efecto global, aunque su contribución a la heterogeneidad es baja.

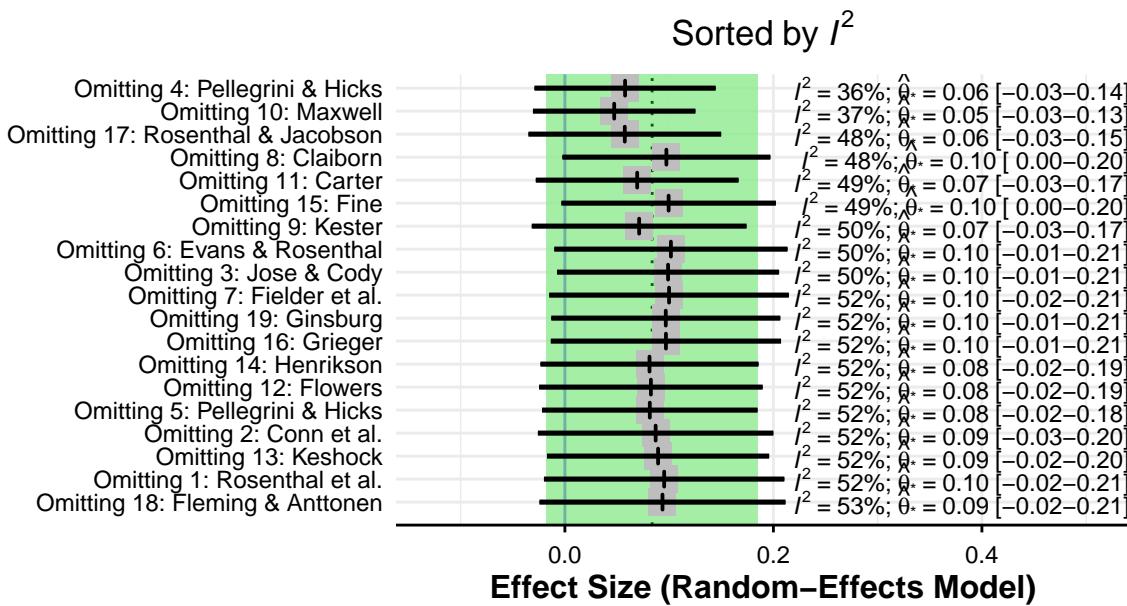


El panel de diagnósticos individuales permite evaluar diferentes dimensiones de influencia,

residuos estandarizados, valores de Cook, DFFITS, covariance ratio, cambios en tau<sup>2</sup>, cambios en Q, hat values y pesos. Podemos ver como algunos estudios muestran picos ocasionales, como *Maxwell* o *Pellegrini & Hicks*.



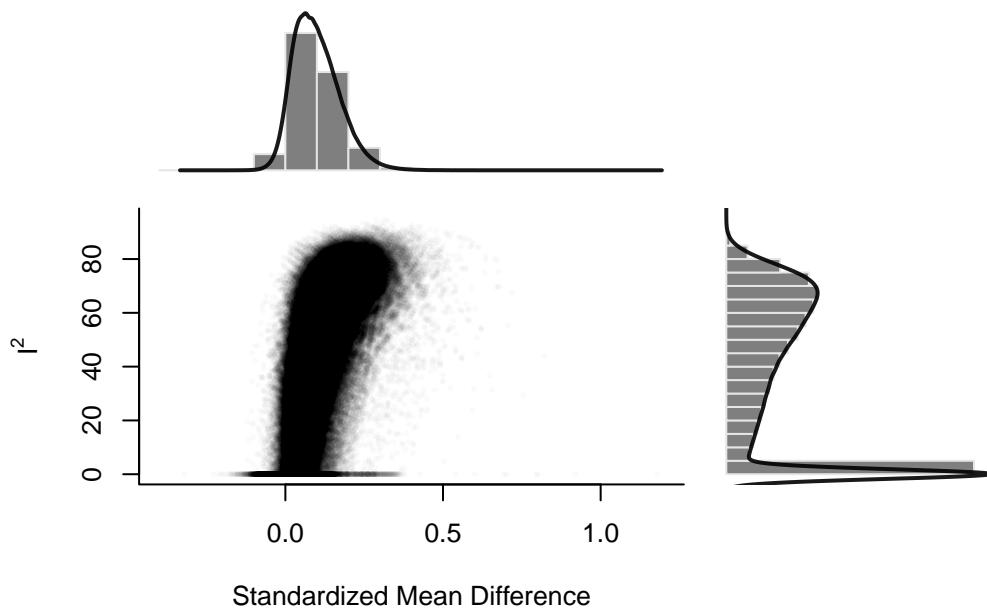
Este gráfico evalúa cómo cambia la estimación del efecto global cuando se omite cada estudio, ordenando los resultados según la magnitud del efecto estimado. El metaanálisis resulta notablemente robusto: la omisión de cualquier estudio produce cambios muy pequeños en la estimación global, que oscila únicamente entre 0.05 y 0.10. Incluso la exclusión de los estudios aparentemente más influyentes (como *Maxwell* o *Rosenthal & Jacobson*) no altera sustancialmente la magnitud ni la dirección del efecto.



Este gráfico muestra cómo cambia la heterogeneidad ( $I^2$ ) del metaanálisis al omitir estudios individualmente y ordena los resultados según el valor de  $I^2$  resultante. Aquí se observa como las variaciones de  $I^2$  se mantienen en un rango aproximadamente entre 36% y 53%.

### 4.3 Análisis GOSH (Graphic Display of Study Heterogeneity)

El análisis GOSH ajusta el modelo de metaanálisis a todos los subconjuntos posibles de estudios (combinatoria de  $2^k - 1$  modelos). Dejando ver si los datos provienen de una única población homogénea o si existen clústeres o patrones distintos.



### Interpretación del GOSH Plot:

El GOSH plot muestra una distribución altamente dispersa y asimétrica de los tamaños de efecto y los niveles de heterogeneidad, lo que indica que distintas combinaciones de estudios producen estimaciones muy diferentes entre sí. La mayor concentración de puntos se ubica cerca de tamaños de efecto pequeños, próximos a cero, pero existe una franja más estrecha que se extiende hacia valores mucho mayores, señalando la presencia de subconjuntos capaces de inflar de forma considerable la magnitud del efecto. En el eje vertical, los valores de  $I^2$  varían desde prácticamente nulos hasta niveles superiores al 80–90%, lo cual sugiere que el grado de heterogeneidad depende fuertemente de qué estudios se incluyan en el modelo. La forma elongada y abierta del gráfico refleja una estructura inestable, donde no existe una única región compacta de soluciones, sino múltiples configuraciones posibles que ponen en duda la robustez del efecto global.

Las salidas numéricas y los resultados de clustering complementan y confirman visualmente lo observado en el GOSH plot. Los estadísticos descriptivos muestran que el tamaño del efecto estimado puede variar desde -0.32 hasta 1.18, mientras que la heterogeneidad presenta un rango extremadamente amplio ( $I^2$  entre 0% y 95%,  $\tau^2$  desde 0 hasta más de 1). Esto implica que el metaanálisis es sensible a la composición del conjunto de estudios, produciendo soluciones muy divergentes según el subconjunto considerado. A su vez, los métodos de agrupamiento —K-means, DBSCAN y GMM— identificaron múltiples clústeres y coincidieron en señalar consistentemente a los estudios **4, 10, 8 y 17** como posibles outliers o puntos influentes. Estos estudios parecen ser los responsables de generar las soluciones extremas observadas en el GOSH plot y de fragmentar el espacio de modelos en varias regiones.

## 5. Resultados II: Análisis de Moderadores

Dada la heterogeneidad significativa que se obtuvo ( $I^2 \approx 50\%$ ) y la presencia de *outliers*, no es suficiente quedarse con el efecto global. Es necesario averiguar si ciertas características del diseño de los estudios actúan como variables moderadoras.

### 5.1 Análisis de Subgrupos

Se analizaron dos variables categóricas: el entorno de instrucción (**setting**) y el conocimiento del evaluador (**tester**). Se compararon modelos asumiendo una varianza entre estudios ( $\tau^2$ ) separada para cada grupo versus una varianza común.

#### A) Subgrupo por Entorno (Setting)

Se evaluó si el efecto del Pigmalión difiere entre intervenciones grupales (Group) e individuales (Indiv). Dado que el subgrupo individual incluye pocos estudios ( $k = 3$ ), se interpretan los modelos con varianza común, que ofrecen estimaciones más estables.

**Resultados principales (  $\tau^2$  común):**

- **Grupal ( $k = 16$ ):** efecto irrelevante (SMD = 0.05; IC95%: -0.04 a 0.14).
- **Individual ( $k = 3$ ):** efecto moderado y significativo (SMD = 0.38; IC95%: 0.05 a 0.71).

El test entre grupos arroja  $Q = 3.58$ ,  $p = 0.0586$ . Aunque no alcanza significación convencional, la diferencia entre efectos ( 0.05 vs. 0.38) muestra una tendencia clara.

El efecto Pigmalión prácticamente desaparece en entornos grupales, pero se mantiene y aumenta en interacciones individuales profesor–alumno. El entorno parece ser un moderador relevante.

#### B) Subgrupo por Evaluador (Tester)

Se analizó si el conocimiento de las expectativas por parte del evaluador (“aware” vs. “blind”) influía en los resultados, bajo la hipótesis de un posible sesgo del observador. Los subgrupos están equilibrados ( $k = 10$  vs.  $k = 9$ ).

**Resultados principales:**

- **Evaluadores conscientes:** SMD = 0.046 (IC95%: -0.09 a 0.19).
- **Evaluadores ciegos:** SMD = 0.149 (IC95%: -0.02 a 0.32).

El p-valor = 0.3606 indica ausencia de diferencias significativas.

En conclusión, no hay evidencia de sesgo del observador: conocer las expectativas no aumenta los puntajes. Además, esta variable no explica la heterogeneidad del metaanálisis ( $I^2$  residual = 51.6%). El conocimiento del evaluador no es un moderador relevante.

## 5.2. Metaregresión

A parte de las categorías presentadas anteriormente, la duración del experimento es una variable continua clave. Se ajustó un modelo de metaregresión utilizando la variable **weeks** (semanas de contacto).

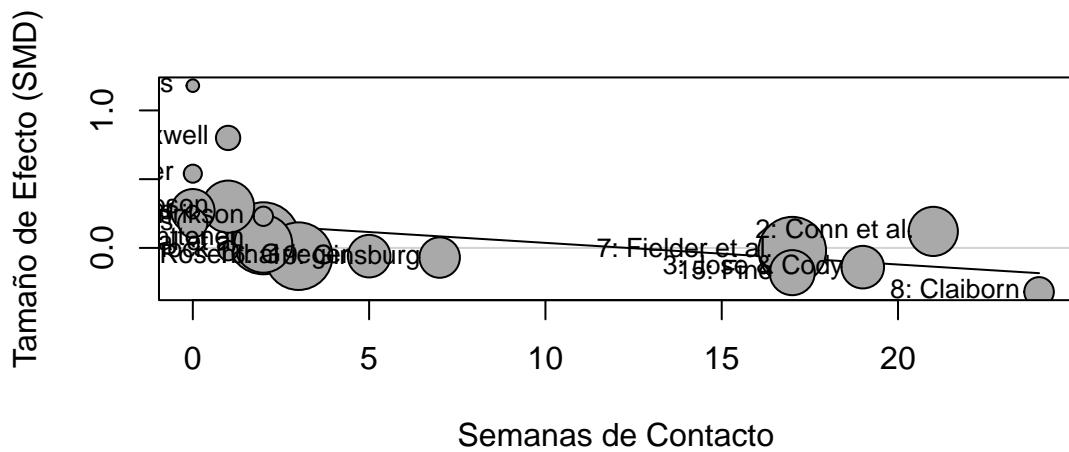
$$\theta_k = \beta_0 + \beta_1 \times \text{Weeks}_k + u_k + \epsilon_k$$

### Análisis de la Regresión:

- **Coeficiente (weeks):** El test de moderadores ( $Q_M = 7.51$ ) confirma una relación significativa. La dirección negativa del coeficiente indica que por cada semana adicional de contacto entre profesor y alumno, el tamaño del “Efecto Pigmalión” cae. Esto sugiere que el efecto es fuerte al inicio pero se apaga con el tiempo.
- **Significancia:** Siendo la hipótesis nula: la duración del contacto (weeks) no tiene ningún efecto sobre el tamaño del efecto Pigmalión, el p-valor obtenido es 0.0061. Al ser este valor muy inferior a 0.05, podemos afirmar que la duración del experimento explica significativamente la heterogeneidad detectada. La variable weeks es un predictor determinante.
- $R^2$ : El modelo muestra un  $R^2 = 40.63\%$ . Esto indica que la variable “duración” es capaz de explicar por si sola más del 40% de la varianza entre estudios que al principio atribuímos a heterogeneidad desconocida. Como consecuencia, la heterogeneidad residual ( $I^2$ ) desciende del 49.8% al 29.40%.

### Gráficos de Burbujas (Bubble Plots)

Estos gráficos permiten visualizar la relación lineal entre la duración y el efecto. El tamaño de cada burbuja es proporcional al peso del estudio (inverso de la varianza).



**Conclusión:** Los estudios cortos tienden a reportar efectos grandes, mientras que los estudios largos reportan efectos nulos. Esto indica que el efecto de las expectativas inducidas es pasajero.

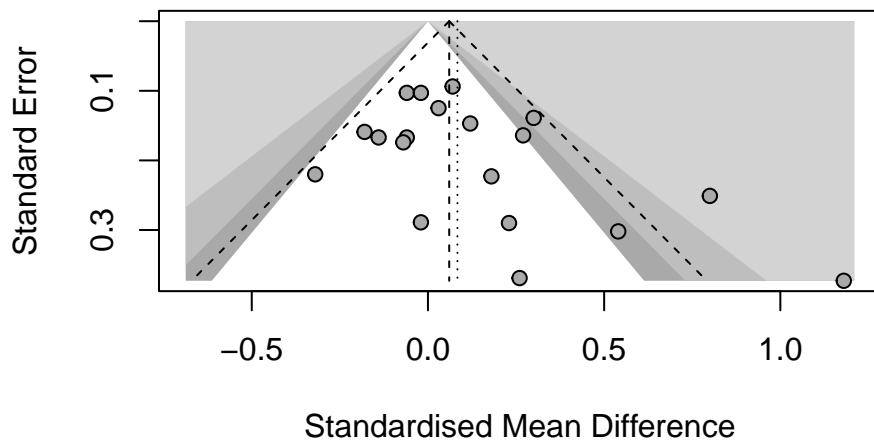
## 6 Evaluación de Sesgos de Publicación

El sesgo de publicación es un tipo de sesgo que aparece cuando los estudios que muestran resultados positivos, significativos o “interesantes” tienen más probabilidad de ser publicados que los estudios que encuentran resultados nulos, pequeños o no significativos.

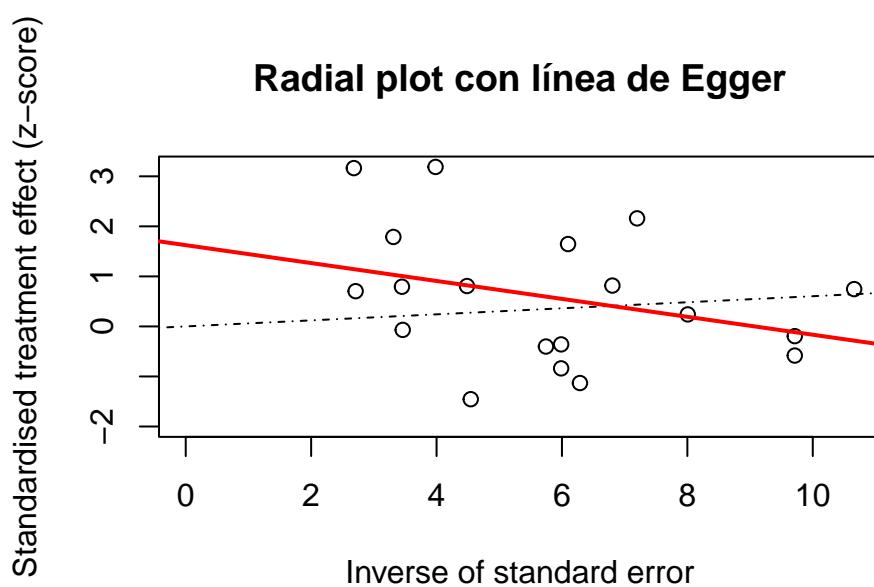
### 6.1 Revisión Gráfica

Para poder evaluar el sesgo de publicación realizamos tanto un funnel plot como un radial plot, este ultimo fue realizado junto con una línea de Egger permite visualizar de manera explícita la relación entre el tamaño del efecto estandarizado y la precisión de los estudios.

### Funnel plot – Raudenbush 1985



### Radial plot con línea de Egger



El funnel plot muestra una asimetría clara: los estudios pequeños tienden a reportar efectos más positivos, mientras que faltan estudios pequeños con efectos negativos, lo que sugiere posible sesgo de publicación.

El radial plot refuerza esta evidencia: la pendiente positiva indica que los estudios menos precisos exhiben efectos más grandes, mientras que los más precisos muestran efectos modestos. La dispersión irregular también apunta a heterogeneidad no explicada.

La línea del test de Egger presenta una pendiente negativa marcada, lo que implica que los estudios con mayor error estándar reportan efectos inflados. Este patrón es típico de small-study effects y respalda la conclusión de que el metaanálisis presenta asimetría y que el efecto global podría estar sobreestimado por la influencia de estudios pequeños.

## 6.2. Pruebas Formales de Asimetría (Metabias)

Se aplicaron las variantes robustas de las pruebas de asimetría disponibles en la librería `meta`. Es decir que fueron realizadas la prueba de correlación de rangos, la prueba de regresión lineal y la prueba con estimador MM.

Los tres métodos de detección de asimetría del funnel plot muestran un patrón consistente: hay señales de posible sesgo de publicación, aunque el grado de evidencia varía según el test.

### Test de Begg (rank correlation):

El test de Begg arroja un resultado marginal ( $z = 1.79$ ,  $p = 0.074$ ).

Esto sugiere una indicación débil de asimetría, pero no alcanza significación convencional ( $p < .05$ ).

Dado que el test de Begg tiene baja potencia con  $k$  reducido (como aquí, con 19 estudios), un  $p$  cercano a 0.05 puede interpretarse como una señal pero no como evidencia concluyente.

### Egger clásico (método = “linreg”):

El test de Egger estándar produce un resultado casi significativo ( $t = 2.04$ ,  $p = 0.057$ ), de nuevo en el límite.

Esto indica una tendencia hacia la asimetría, coherente con lo observado visualmente en el funnel plot, pero que aún no supera el umbral clásico de significación.

Con muestras pequeñas, Egger también tiende a ser inestable y sensible a heterogeneidad alta.

### Egger modificado (método = “mm”, Thompson & Sharp, REML):

Este test, más robusto cuando existe heterogeneidad, sí resulta estadísticamente significativo ( $t = 2.53$ ,  $p = 0.0218$ ).

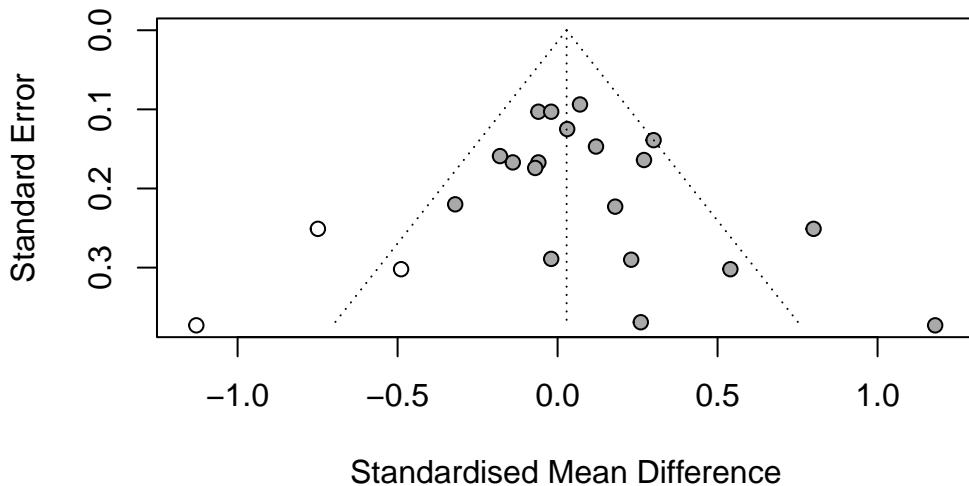
Dado que este método ajusta la heterogeneidad de manera más apropiada mediante un modelo mixto con  $\tau^2$  aditiva, se considera un método más confiable bajo condiciones como las de nuestro metaanálisis.

La significación aquí sugiere asimetría real del funnel plot y, por ende, evidencia más sólida de sesgo de publicación.

## 6.4. Método Trim-and-Fill (Recortar y Rellenar)

Para corregir el impacto del sesgo de publicación, se utilizó el método *Trim-and-Fill*. Este algoritmo imputa los estudios teóricamente “faltantes” para restaurar la simetría del *Funnel Plot* y recalcula el efecto global ajustado.

### Trim and fill – Raudenbush 1985



El método trim-and-fill detecta asimetría en el funnel plot y estima que faltan tres estudios con efectos negativos. Al imputarlos, el efecto combinado cae a 0.028 (IC95%: -0.117 a 0.173;  $p = 0.70$ ), totalmente compatible con cero. Esto indica que el efecto positivo observado inicialmente dependía en parte de estudios pequeños con efectos grandes, en línea con la evidencia previa de Begg, Egger y el radial plot.

Pese al ajuste, la heterogeneidad sigue siendo moderada-alta ( $I^2 = 64\%$ ) y el intervalo de predicción amplio (-0.57 a 0.63) muestra que los efectos futuros podrían ser tanto negativos como positivos. En conjunto, tras corregir la asimetría, no queda evidencia de un efecto promedio consistente.

## 7. Estudio de Tamaño de Muestra y Potencia

En este análisis de potencia se evaluó la capacidad del metaanálisis para detectar efectos del tamaño esperado bajo distintos supuestos. Se trabajó con un total previsto de 19 estudios, cada uno con un tamaño muestral esperado de 102 participantes, y se consideraron dos valores esperados del efecto,  $d = 0.1$  y  $d = 0.2$ . Además, se examinó cómo varía la potencia bajo un modelo de efectos fijos y un modelo de efectos aleatorios con  $I^2 = 50\%$ .

El análisis revela que, para un  $d = 0.1$ , la potencia ~60% y en efectos aleatorios ~36%. Esto implica que un efecto tan débil podría pasar inadvertido, especialmente en presencia de heterogeneidad entre estudios. En contraste, para  $d = 0.2$ , la potencia se vuelve muy alta en el modelo de efectos fijos ~99% y sólida en el modelo de efectos aleatorios ~87%. En conjunto,

estos resultados indican que, con la evidencia disponible, el metaanálisis tiene buena capacidad para detectar efectos moderados, pero no está bien equipado para identificar efectos muy pequeños.

## **8. Discusión y Alcance de los Resultados**

El metaanálisis de 19 estudios muestra que el efecto Pigmalión, medido sobre desempeño cognitivo, es pequeño, inconsistente y altamente dependiente del contexto. El efecto promedio no es significativo y presenta heterogeneidad moderada, evidenciada tanto en los tests de heterogeneidad como en los análisis gráficos e influenciales. Aunque algunos estudios individuales influyen fuertemente en el tamaño del efecto, incluso al removerlos el resultado global se mantiene cercano a cero.

Los análisis de moderadores indican que el entorno individual muestra efectos mayores que el grupal, aunque con pocos estudios. El moderador más importante es la duración del experimento, que explica una proporción sustancial de la variabilidad: los estudios breves reportan efectos más altos, mientras que los de mayor duración no muestran impacto. Además, existe evidencia consistente de sesgo de publicación, ya que los estudios pequeños tienden a inflar el efecto y los métodos de corrección reducen el tamaño del efecto a valores prácticamente nulos.

En conjunto, los resultados sugieren que el efecto Pigmalión no es un fenómeno robusto, sino uno que aparece en condiciones específicas, tiende a ser efímero y está afectado por sesgos de la literatura.

## **9. Conclusiones y Futuros Pasos**

En base al análisis estadístico realizado, se concluye que el efecto global es pequeño, no significativo y sensible al contexto, la duración del experimento es el moderador más claro: efectos iniciales desaparecen con el tiempo., las intervenciones individuales muestran mayor efecto que las grupales, aunque con evidencia limitada, hay sesgo de publicación que amplifica artificialmente el efecto en estudios pequeños y que la heterogeneidad es moderada y solo parcialmente explicada por los moderadores analizados.

Se recomienda que futuras investigaciones abandonen el uso del CI global como variable dependiente, ya que es una medida estable y difícil de modificar, y se enfoquen en variables más sensibles como el rendimiento académico específico, la autoeficacia o la motivación del estudiante, controlando rigurosamente el tiempo de exposición. Realizar estudios contemporáneos y preregistrados para reducir sesgo de publicación y explorar directamente cómo las expectativas modifican el comportamiento docente.

## 10. Referencias Bibliográficas

- Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Chapman & Hall/CRC Press.
- Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, 76(1), 85–97.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. *Holt, Rinehart & Winston*.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.

Citas de paquetes de R utilizados:

To cite package 'meta' in publications use:

Baldazzi S, Rücker G, Schwarzer G (2019), How to perform a meta-analysis with R: a practical tutorial, *Evidence-Based Mental Health*; 22: 153–160.

DOI: 10.1136/ebmental-2019-300117

To cite 'dmetar' in publications use:

Harrer, M., Cuijpers, P., Furukawa, T. & Ebert, D. D. (2019). dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'. R package version 0.1.0. URL <http://dmetar.protectlab.org/>.

To cite package 'metadat' in publications use:

Viechtbauer W, White T, Noble D, Senior A, Hamilton W (2025). `_metadat: Meta-Analysis Datasets_`. doi:10.32614/CRAN.package.metadat <<https://doi.org/10.32614/CRAN.package.metadat>>, R package version 1.4-0, <<https://CRAN.R-project.org/package=metadat>>.

To cite metapower in publications use:

Griffin, J.W. (2020). metapoweR: an R package for computing meta-analytic statistical power. R package version 0.2.1

## Apéndice: Sentencias de R Utilizadas

Para garantizar la reproducibilidad de este estudio, se adjunta la totalidad del código utilizado, el cual se ha presentado y ejecutado a lo largo de las secciones previas de este documento dinámico.

```
library(meta)
library(metafor)
library(dmetar)
```

```
dat <- dat.raudenbush1985
dat
```

	study	author	year	weeks	setting	tester	n1i	n2i	yi	vi
1	1	Rosenthal et al.	1974	2	group	aware	77	339	0.0300	0.0156
2	2	Conn et al.	1968	21	group	aware	60	198	0.1200	0.0216
3	3	Jose & Cody	1971	19	group	aware	72	72	-0.1400	0.0279
4	4	Pellegrini & Hicks	1972	0	group	aware	11	22	1.1800	0.1391
5	5	Pellegrini & Hicks	1972	0	group	blind	11	22	0.2600	0.1362
6	6	Evans & Rosenthal	1969	3	group	aware	129	348	-0.0600	0.0106
7	7	Fielder et al.	1971	17	group	blind	110	636	-0.0200	0.0106
8	8	Claiborn	1969	24	group	aware	26	99	-0.3200	0.0484
9	9	Kester	1969	0	group	aware	75	74	0.2700	0.0269
10	10	Maxwell	1970	1	indiv	blind	32	32	0.8000	0.0630
11	11	Carter	1970	0	group	blind	22	22	0.5400	0.0912
12	12	Flowers	1966	0	group	blind	43	38	0.1800	0.0497
13	13	Keshock	1970	1	indiv	blind	24	24	-0.0200	0.0835
14	14	Henrikson	1970	2	indiv	blind	19	32	0.2300	0.0841
15	15	Fine	1972	17	group	aware	80	79	-0.1800	0.0253
16	16	Grieger	1970	5	group	blind	72	72	-0.0600	0.0279
17	17	Rosenthal & Jacobson	1968	1	group	aware	65	255	0.3000	0.0193
18	18	Fleming & Anttonen	1971	2	group	blind	233	224	0.0700	0.0088
19	19	Ginsburg	1970	7	group	aware	65	67	-0.0700	0.0303

```
dat$seTE <- sqrt(dat$vi)
m.gen <- metagen(TE = yi,
                  seTE = seTE,
                  data = dat,
                  studlab = paste(study, author, sep=": "),
                  sm = "SMD",
```

```

method.tau = "REML",
prediction = TRUE)

summary(m.gen)

```

	SMD	95%-CI	%W(common)	%W(random)
1: Rosenthal et al.	0.0300	[ -0.2148; 0.2748]	8.5	7.7
2: Conn et al.	0.1200	[ -0.1681; 0.4081]	6.2	6.6
3: Jose & Cody	-0.1400	[ -0.4674; 0.1874]	4.8	5.7
4: Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	1.0	1.7
5: Pellegrini & Hicks	0.2600	[ -0.4633; 0.9833]	1.0	1.7
6: Evans & Rosenthal	-0.0600	[ -0.2618; 0.1418]	12.5	9.1
7: Fielder et al.	-0.0200	[ -0.2218; 0.1818]	12.5	9.1
8: Claiborn	-0.3200	[ -0.7512; 0.1112]	2.7	4.0
9: Kester	0.2700	[ -0.0515; 0.5915]	4.9	5.8
10: Maxwell	0.8000	[ 0.3081; 1.2919]	2.1	3.3
11: Carter	0.5400	[ -0.0519; 1.1319]	1.5	2.4
12: Flowers	0.1800	[ -0.2569; 0.6169]	2.7	3.9
13: Keshock	-0.0200	[ -0.5864; 0.5464]	1.6	2.6
14: Henrikson	0.2300	[ -0.3384; 0.7984]	1.6	2.6
15: Fine	-0.1800	[ -0.4918; 0.1318]	5.3	6.0
16: Grieger	-0.0600	[ -0.3874; 0.2674]	4.8	5.7
17: Rosenthal & Jacobson	0.3000	[ 0.0277; 0.5723]	6.9	7.0
18: Fleming & Anttonen	0.0700	[ -0.1139; 0.2539]	15.1	9.7
19: Ginsburg	-0.0700	[ -0.4112; 0.2712]	4.4	5.4

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	0.0604	[ -0.0111; 0.1318]	1.66	0.0979
Random effects model	0.0837	[ -0.0175; 0.1849]	1.62	0.1051
Prediction interval		[ -0.2243; 0.3917]		

Quantifying heterogeneity (with 95%-CIs):

tau^2 = 0.0188 [0.0059; 0.1977]; tau = 0.1372 [0.0765; 0.4447]  
I^2 = 49.8% [14.7%; 70.4%]; H = 1.41 [1.08; 1.84]

Test of heterogeneity:

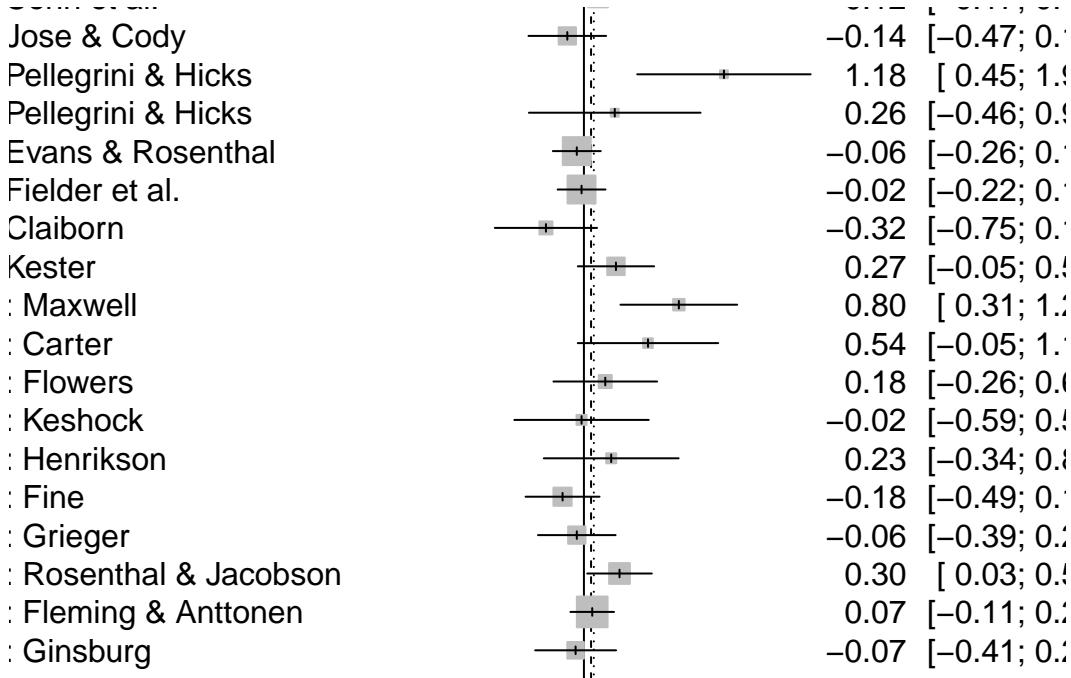
Q	d.f.	p-value
35.83	18	0.0074

Details of meta-analysis methods:

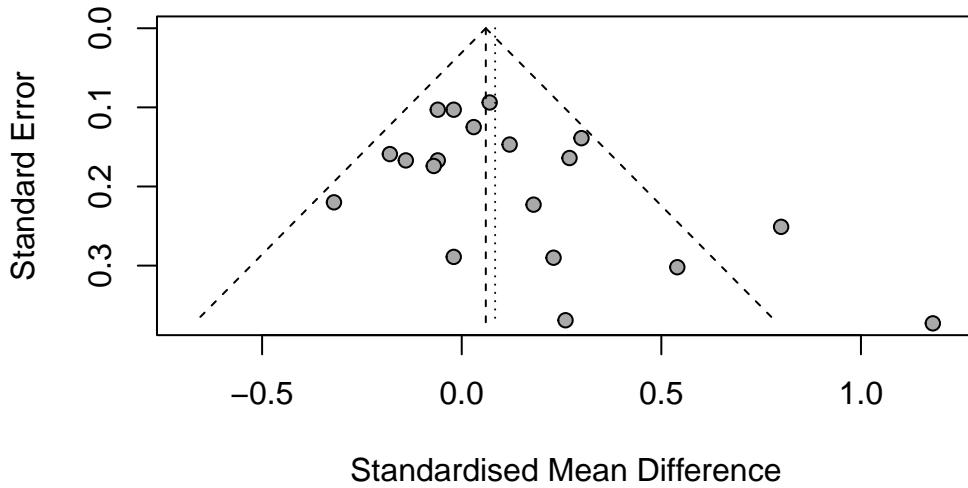
- Inverse variance method

- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q
- Prediction interval based on t-distribution (df = 18)

```
forest(m.gen, leftcols = c("studlab"), rightcols = c("effect", "ci"))
```



```
funnel(m.gen)
```



```
rma_obj <- rma(yi = dat$yi, vi = dat$vi, method = "REML")
regtest(rma_obj, model = "rma", predictor = "sei")
```

Regression Test for Funnel Plot Asymmetry

Model: mixed-effects meta-regression model  
 Predictor: standard error

Test for Funnel Plot Asymmetry:  $z = 2.5261$ ,  $p = 0.0115$   
 Limit Estimate (as sei  $\rightarrow 0$ ):  $b = -0.2169$  (CI:  $-0.4642, 0.0304$ )

```
outliers <- find.outliers(m.gen)
outliers
```

Identified outliers (fixed-effect model)

-----  
 "4: Pellegrini & Hicks", "10: Maxwell"

Results with outliers removed

-----  
 Number of studies:  $k = 17$

	SMD	95%-CI	z	p-value
Common effect model	0.0332	[-0.0394; 0.1058]	0.90	0.3699
Random effects model	0.0332	[-0.0394; 0.1058]	0.90	0.3699
Prediction interval		[-0.0454; 0.1118]		

Quantifying heterogeneity (with 95%-CIs):  
 $\tau^2 < 0.0001$  [0.0000; 0.0535];  $\tau = 0.0017$  [0.0000; 0.2312]  
 $I^2 = 9.1\%$  [0.0%; 45.5%];  $H = 1.05$  [1.00; 1.35]

Test of heterogeneity:

Q	d.f.	p-value
17.60	16	0.3480

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q
- Prediction interval based on t-distribution (df = 16)

Identified outliers (random-effects model)

-----  
"4: Pellegrini & Hicks", "10: Maxwell"

Results with outliers removed

-----  
Number of studies: k = 17

	SMD	95%-CI	z	p-value
Common effect model	0.0332	[-0.0394; 0.1058]	0.90	0.3699
Random effects model	0.0332	[-0.0394; 0.1058]	0.90	0.3699
Prediction interval		[-0.0454; 0.1118]		

Quantifying heterogeneity (with 95%-CIs):  
 $\tau^2 < 0.0001$  [0.0000; 0.0535];  $\tau = 0.0017$  [0.0000; 0.2312]  
 $I^2 = 9.1\%$  [0.0%; 45.5%];  $H = 1.05$  [1.00; 1.35]

Test of heterogeneity:

Q	d.f.	p-value
17.60	16	0.3480

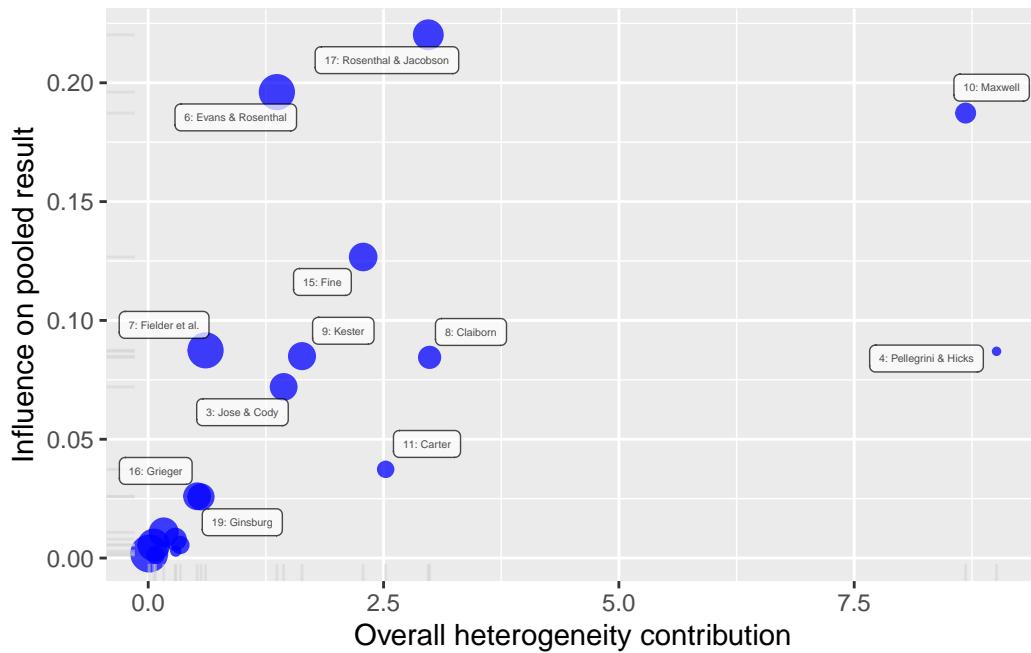
Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q
- Prediction interval based on t-distribution (df = 16)

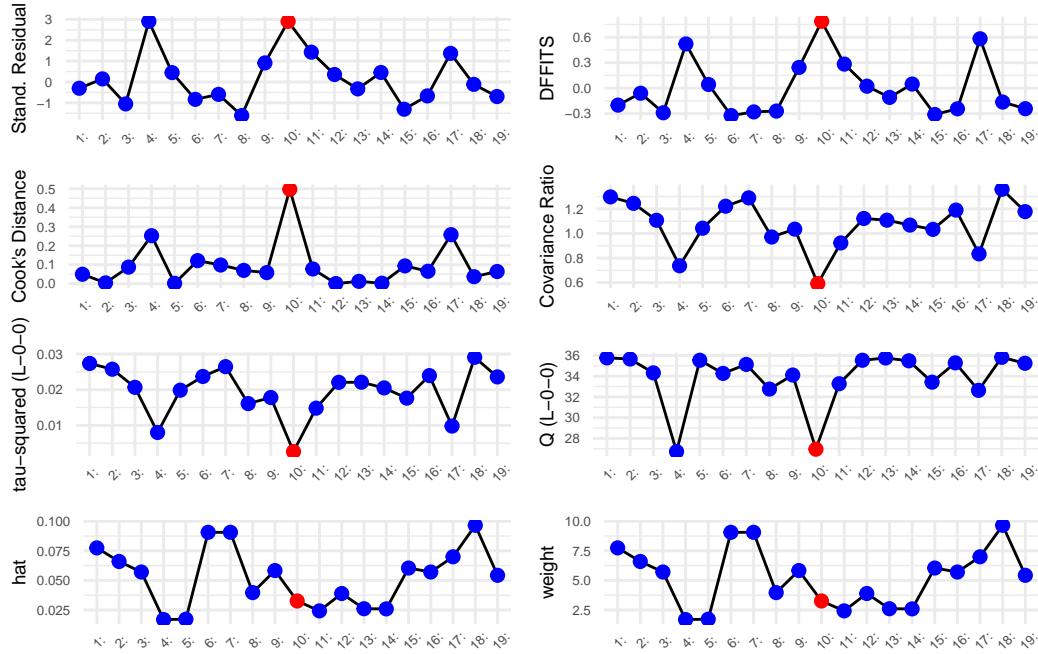
```
m.gen.inf <- InfluenceAnalysis(m.gen, random = TRUE)
```

[=====] DONE

```
plot(m.gen.inf, "baujat")
```



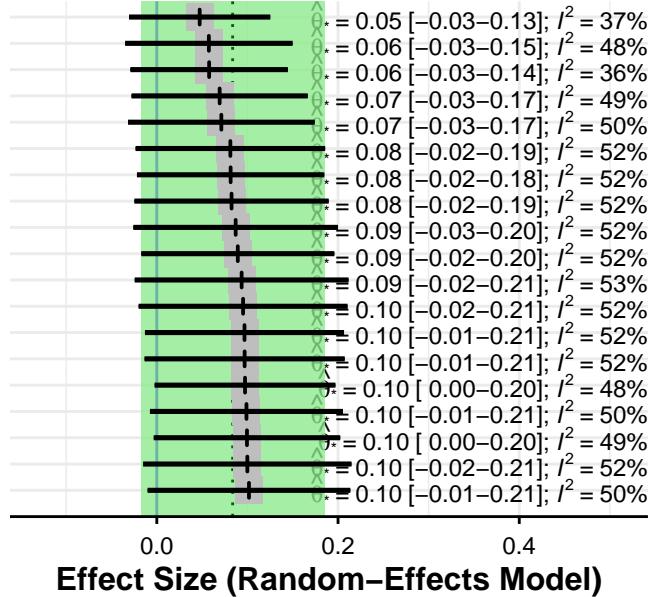
```
plot(m.gen.inf, "influence")
```



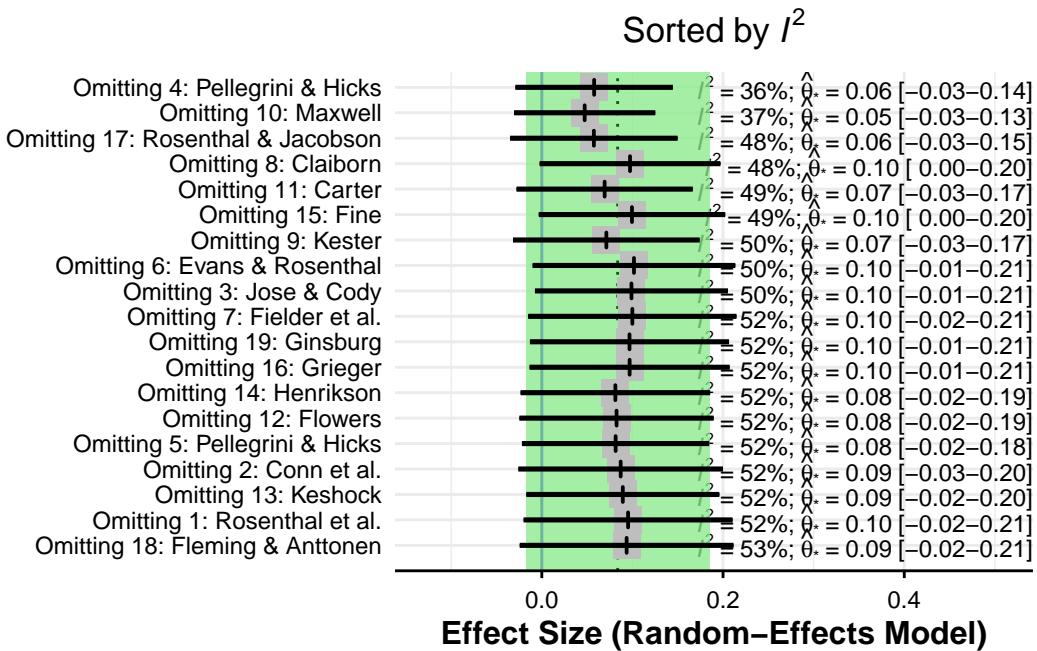
```
plot(m.gen.inf, "ES")
```

Omitting 10: Maxwell  
 Omitting 17: Rosenthal & Jacobson  
 Omitting 4: Pellegrini & Hicks  
 Omitting 11: Carter  
 Omitting 9: Kester  
 Omitting 14: Henrikson  
 Omitting 5: Pellegrini & Hicks  
 Omitting 12: Flowers  
 Omitting 2: Conn et al.  
 Omitting 13: Keshock  
 Omitting 18: Fleming & Anttonen  
 Omitting 1: Rosenthal et al.  
 Omitting 19: Ginsburg  
 Omitting 16: Grieger  
 Omitting 8: Claiborn  
 Omitting 3: Jose & Cody  
 Omitting 15: Fine  
 Omitting 7: Fielder et al.  
 Omitting 6: Evans & Rosenthal

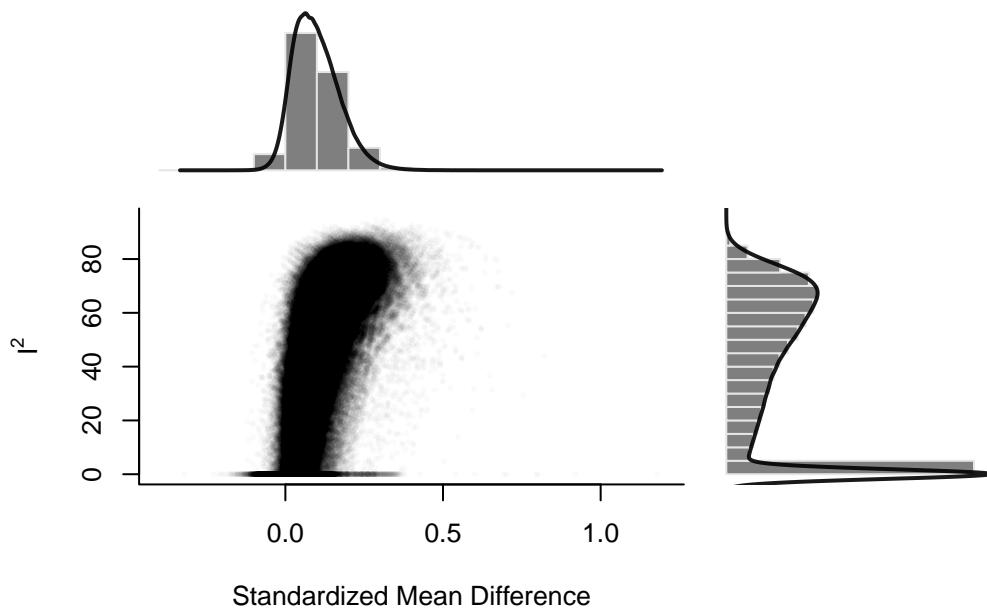
### Sorted by Effect Size



```
plot(m.gen.inf, "I2")
```



```
m.rma <- rma(yi = dat$yi, sei = dat$seTE, method = "REML", test = "knha")
set.seed(123)
res.gosh <- gosh(m.rma)
plot(res.gosh, alpha = 0.02)
```



```
res.gosh.diagnos <- gosh.diagnostics(res.gosh, km.params = list(centers = 2), db.params = li
```

Perform Clustering...

```
|=====
```

```
res.gosh.diagnos
```

GOSH Diagnostics

```
=====
```

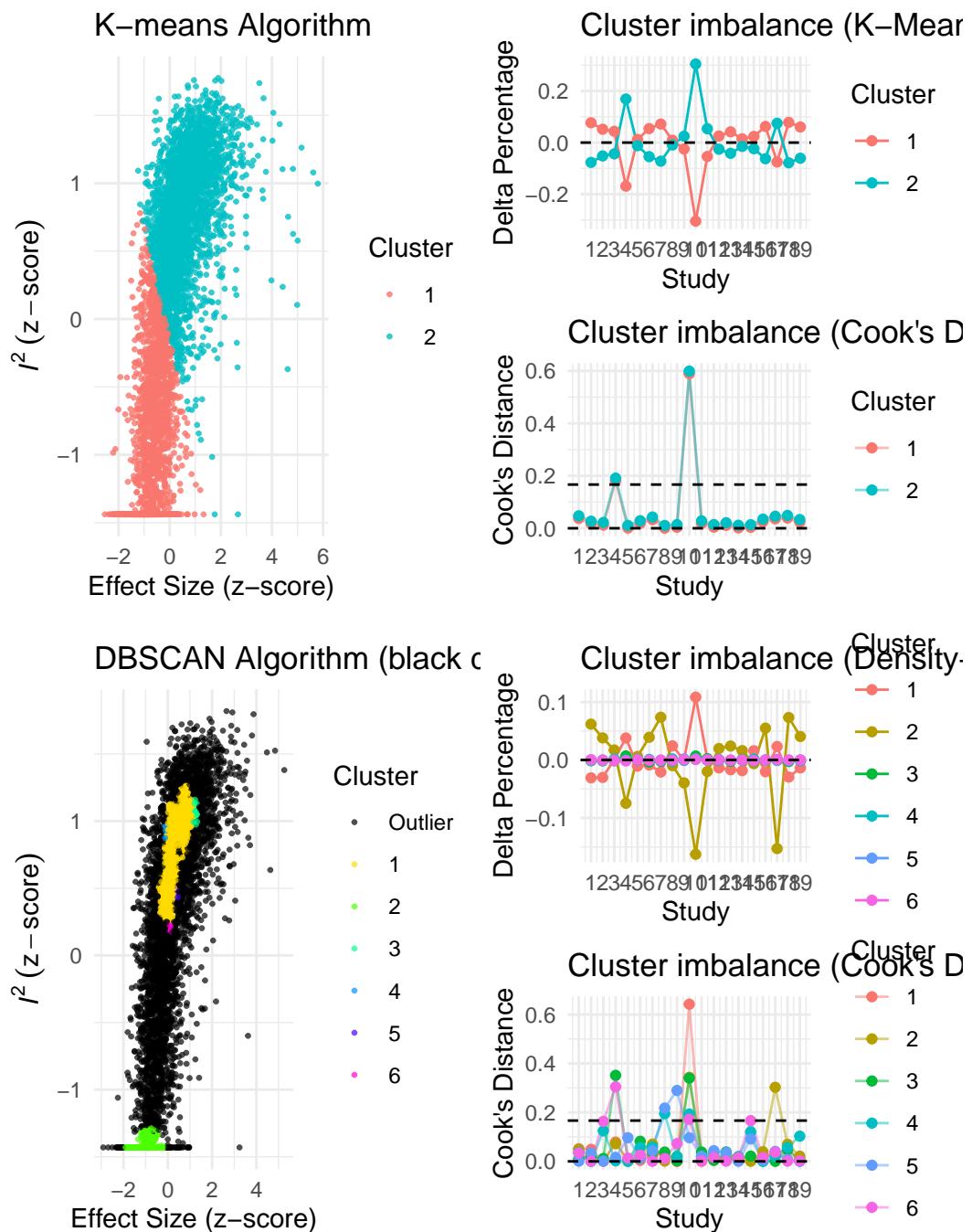
- Number of K-means clusters detected: 2
- Number of DBSCAN clusters detected: 6
- Number of GMM clusters detected: 7

Identification of potential outliers

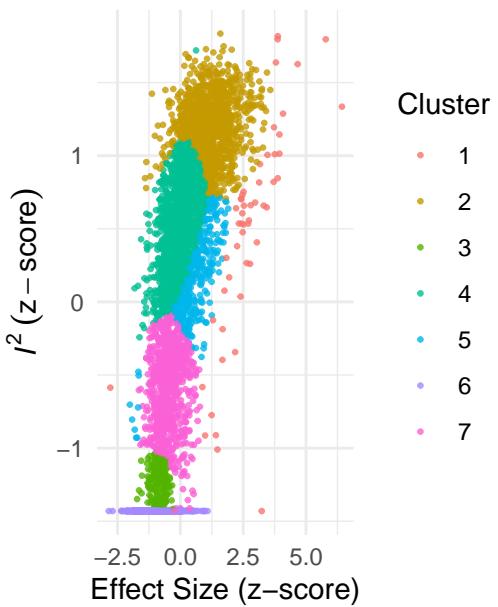
```
=====
```

- K-means: Study 4, Study 10
- DBSCAN: Study 10, Study 17, Study 4, Study 8, Study 9, Study 15
- Gaussian Mixture Model: Study 4, Study 10, Study 8, Study 17

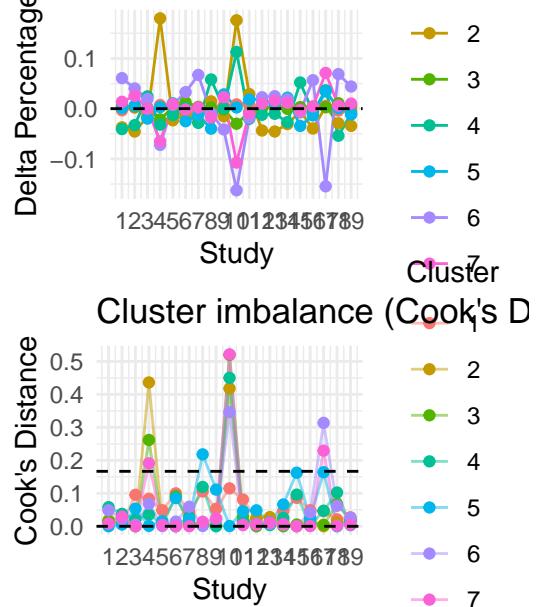
```
plot(res.gosh.diagnos)
```



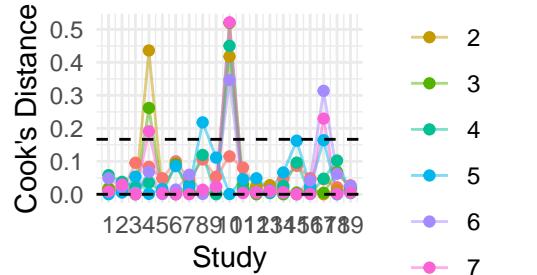
Gaussian Mixture Model



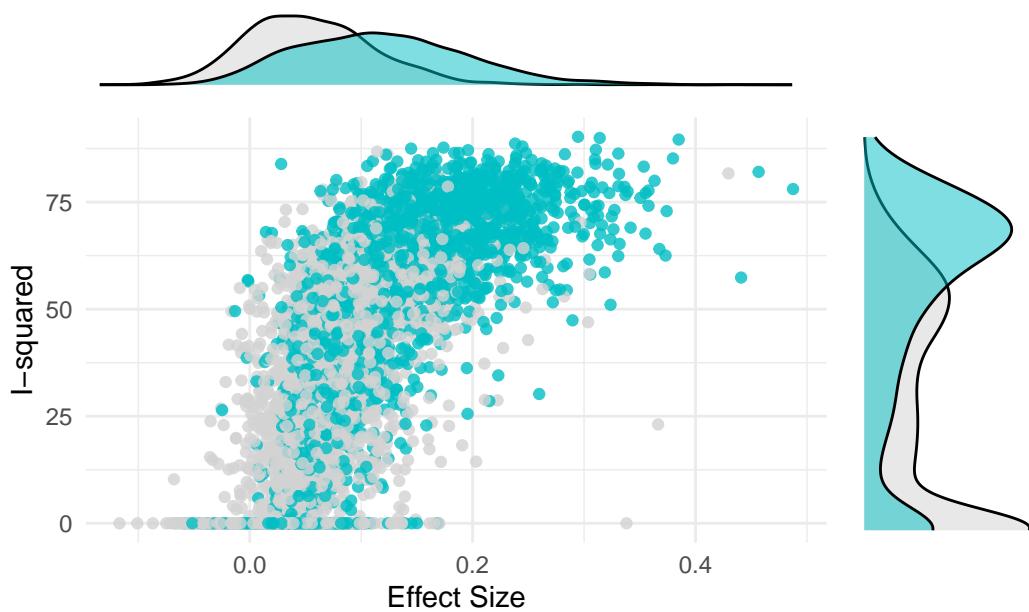
Cluster imbalance (GMM)



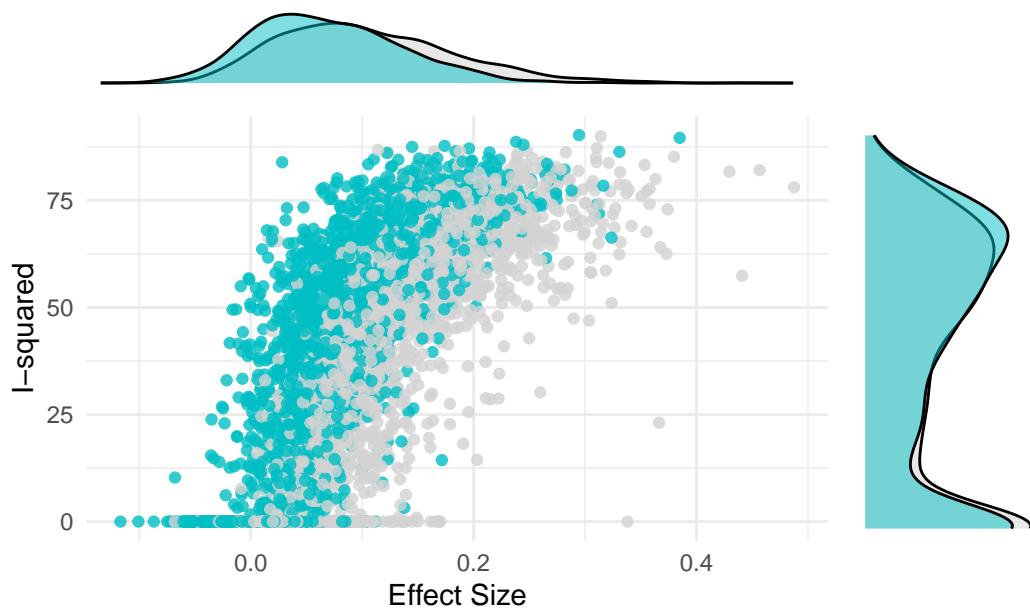
Cluster imbalance (Cook's D)



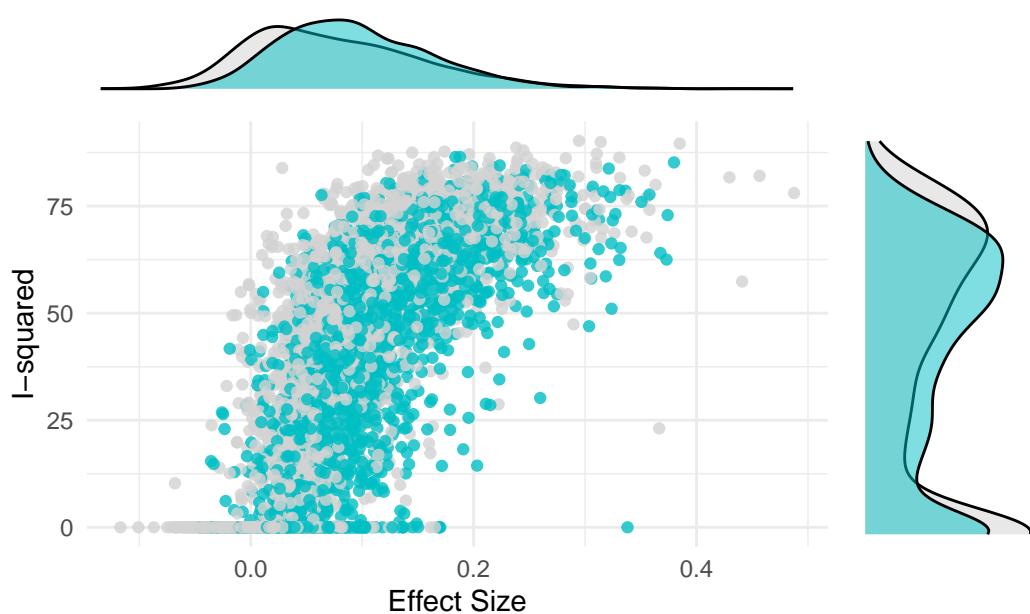
Study 4



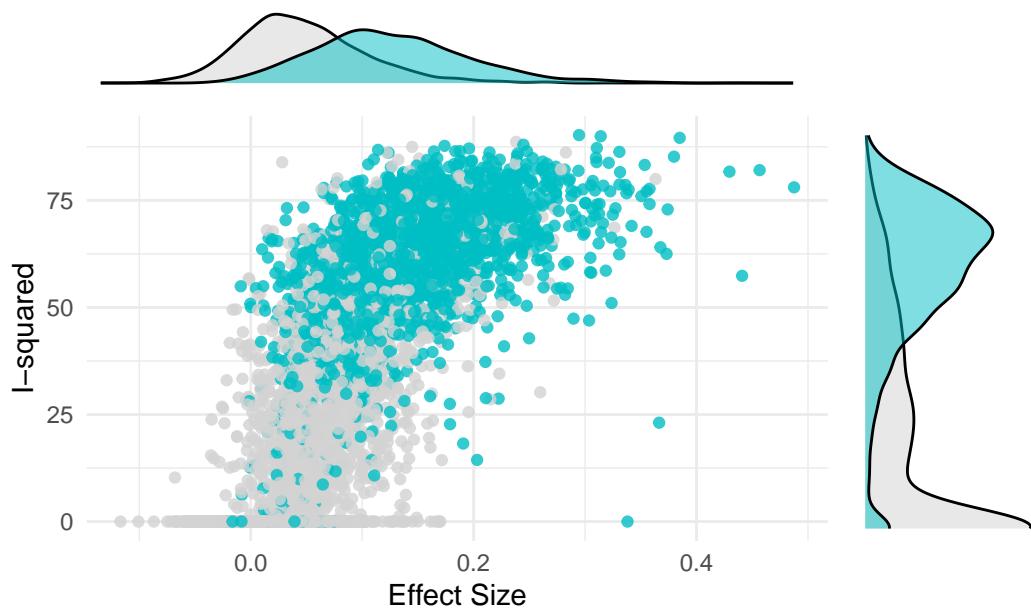
Study 8



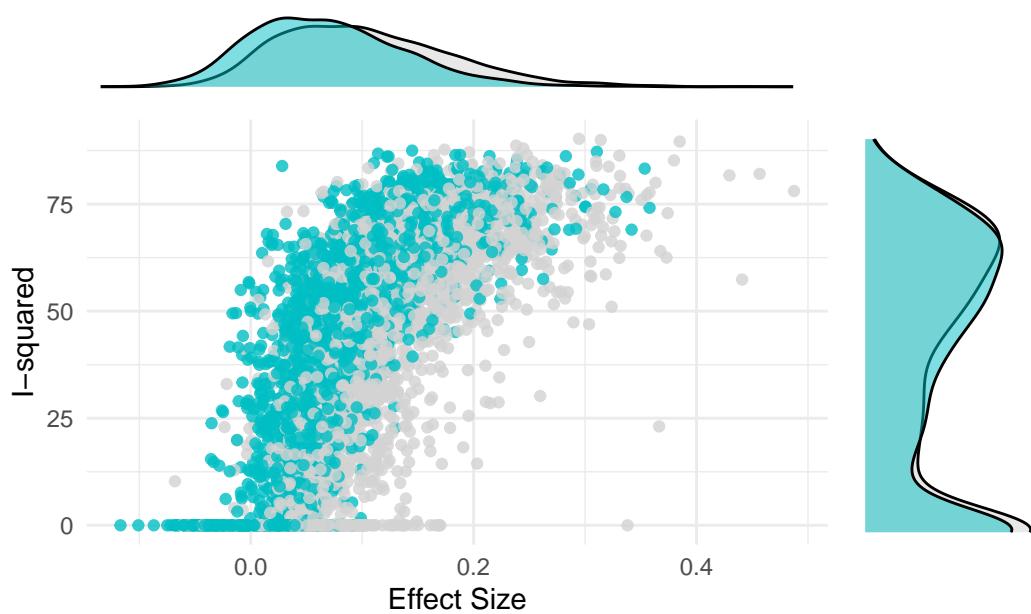
Study 9



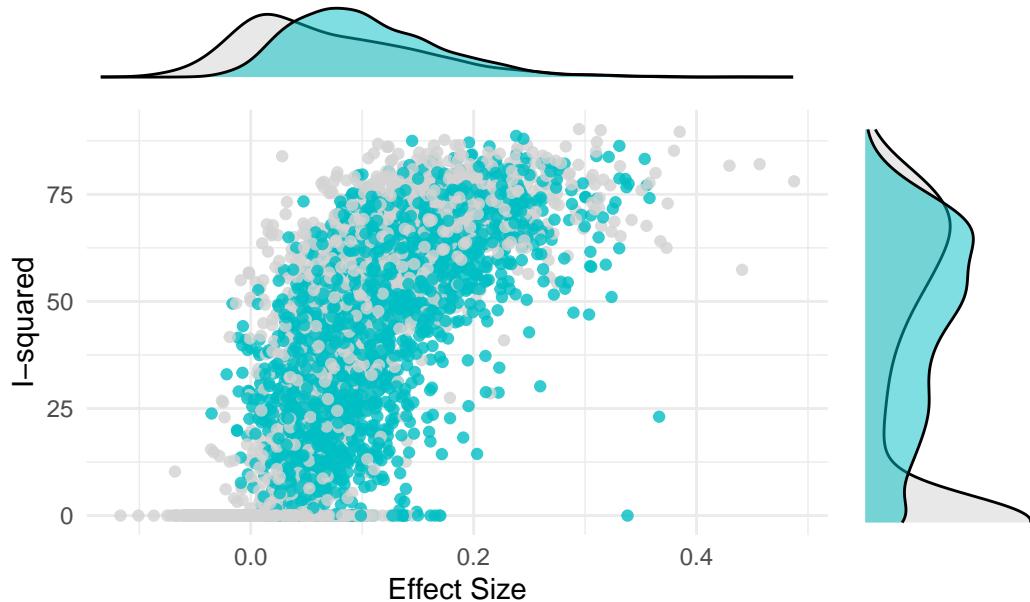
Study 10



Study 15



## Study 17



```
# A) Subgrupo por 'setting'
#Primero con tau.common = FALSE (estimaciones separadas de tau^2 por subgrupo)
m.setting_separateTau <- update(m.gen, subgroup = setting, tau.common = FALSE)
summary(m.setting_separateTau)
```

	SMD	95%-CI	%W(common)	%W(random)
1: Rosenthal et al.	0.0300	[-0.2148; 0.2748]	8.5	7.7
2: Conn et al.	0.1200	[-0.1681; 0.4081]	6.2	6.6
3: Jose & Cody	-0.1400	[-0.4674; 0.1874]	4.8	5.7
4: Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	1.0	1.7
5: Pellegrini & Hicks	0.2600	[-0.4633; 0.9833]	1.0	1.7
6: Evans & Rosenthal	-0.0600	[-0.2618; 0.1418]	12.5	9.1
7: Fielder et al.	-0.0200	[-0.2218; 0.1818]	12.5	9.1
8: Claiborn	-0.3200	[-0.7512; 0.1112]	2.7	4.0
9: Kester	0.2700	[-0.0515; 0.5915]	4.9	5.8
10: Maxwell	0.8000	[ 0.3081; 1.2919]	2.1	3.3
11: Carter	0.5400	[-0.0519; 1.1319]	1.5	2.4
12: Flowers	0.1800	[-0.2569; 0.6169]	2.7	3.9
13: Keshock	-0.0200	[-0.5864; 0.5464]	1.6	2.6
14: Henrikson	0.2300	[-0.3384; 0.7984]	1.6	2.6
15: Fine	-0.1800	[-0.4918; 0.1318]	5.3	6.0
16: Grieger	-0.0600	[-0.3874; 0.2674]	4.8	5.7

17: Rosenthal & Jacobson	0.3000 [ 0.0277; 0.5723]	6.9	7.0
18: Fleming & Anttonen	0.0700 [-0.1139; 0.2539]	15.1	9.7
19: Ginsburg	-0.0700 [-0.4112; 0.2712]	4.4	5.4
	setting		
1: Rosenthal et al.	group		
2: Conn et al.	group		
3: Jose & Cody	group		
4: Pellegrini & Hicks	group		
5: Pellegrini & Hicks	group		
6: Evans & Rosenthal	group		
7: Fielder et al.	group		
8: Claiborn	group		
9: Kester	group		
10: Maxwell	indiv		
11: Carter	group		
12: Flowers	group		
13: Keshock	indiv		
14: Henrikson	indiv		
15: Fine	group		
16: Grieger	group		
17: Rosenthal & Jacobson	group		
18: Fleming & Anttonen	group		
19: Ginsburg	group		

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	0.0604 [-0.0111; 0.1318]	1.66	0.0979	
Random effects model	0.0837 [-0.0175; 0.1849]	1.62	0.1051	
Prediction interval		[-0.2243; 0.3917]		

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0188$  [0.0059; 0.1977];  $\tau = 0.1372$  [0.0765; 0.4447]  
 $I^2 = 49.8\%$  [14.7%; 70.4%];  $H = 1.41$  [1.08; 1.84]

Test of heterogeneity:

Q	d.f.	p-value
35.83	18	0.0074

Results for subgroups (common effect model):

	k	SMD	95%-CI	Q	$I^2$
setting = group	16	0.0424	[-0.0310; 0.1158]	26.50	43.4%
setting = indiv	3	0.3823	[ 0.0714; 0.6932]	4.98	59.9%

Test for subgroup differences (common effect model):

	Q	d.f.	p-value
Between groups	4.35	1	0.0370
Within groups	31.48	17	0.0174

Results for subgroups (random effects model):

	k	SMD	95%-CI	tau <sup>2</sup>	tau
setting = group	16	0.0464	[-0.0358; 0.1286]	0.0040	0.0633
setting = indiv	3	0.3543	[-0.1357; 0.8443]	0.1112	0.3334

Test for subgroup differences (random effects model):

	Q	d.f.	p-value
Between groups	1.48	1	0.2245

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for tau<sup>2</sup>
- Q-Profile method for confidence interval of tau<sup>2</sup> and tau
- Calculation of I<sup>2</sup> based on Q
- Prediction interval based on t-distribution (df = 18)

```
#Luego con tau.common = TRUE (misma estimación tau^2 para todos los subgrupos)
m.setting_commonTau <- update(m.gen, subgroup = setting, tau.common = TRUE)
summary(m.setting_commonTau)
```

	SMD	95%-CI	%W(common)	%W(random)
1: Rosenthal et al.	0.0300	[-0.2148; 0.2748]	8.5	7.7
2: Conn et al.	0.1200	[-0.1681; 0.4081]	6.2	6.6
3: Jose & Cody	-0.1400	[-0.4674; 0.1874]	4.8	5.7
4: Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	1.0	1.7
5: Pellegrini & Hicks	0.2600	[-0.4633; 0.9833]	1.0	1.7
6: Evans & Rosenthal	-0.0600	[-0.2618; 0.1418]	12.5	9.1
7: Fielder et al.	-0.0200	[-0.2218; 0.1818]	12.5	9.1
8: Claiborn	-0.3200	[-0.7512; 0.1112]	2.7	4.0
9: Kester	0.2700	[-0.0515; 0.5915]	4.9	5.8
10: Maxwell	0.8000	[ 0.3081; 1.2919]	2.1	3.3
11: Carter	0.5400	[-0.0519; 1.1319]	1.5	2.4
12: Flowers	0.1800	[-0.2569; 0.6169]	2.7	3.9
13: Keshock	-0.0200	[-0.5864; 0.5464]	1.6	2.6
14: Henrikson	0.2300	[-0.3384; 0.7984]	1.6	2.6
15: Fine	-0.1800	[-0.4918; 0.1318]	5.3	6.0

16: Grieger	-0.0600	[-0.3874; 0.2674]	4.8	5.7
17: Rosenthal & Jacobson	0.3000	[ 0.0277; 0.5723]	6.9	7.0
18: Fleming & Anttonen	0.0700	[-0.1139; 0.2539]	15.1	9.7
19: Ginsburg	-0.0700	[-0.4112; 0.2712]	4.4	5.4
		setting		
1: Rosenthal et al.	group			
2: Conn et al.	group			
3: Jose & Cody	group			
4: Pellegrini & Hicks	group			
5: Pellegrini & Hicks	group			
6: Evans & Rosenthal	group			
7: Fielder et al.	group			
8: Claiborn	group			
9: Kester	group			
10: Maxwell	indiv			
11: Carter	group			
12: Flowers	group			
13: Keshock	indiv			
14: Henrikson	indiv			
15: Fine	group			
16: Grieger	group			
17: Rosenthal & Jacobson	group			
18: Fleming & Anttonen	group			
19: Ginsburg	group			

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	0.0604	[-0.0111; 0.1318]	1.66	0.0979
Random effects model	0.0837	[-0.0175; 0.1849]	1.62	0.1051
Prediction interval		[-0.2243; 0.3917]		

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0188$  [0.0059; 0.1977];  $\tau = 0.1372$  [0.0765; 0.4447]  
 $I^2 = 49.8\%$  [14.7%; 70.4%];  $H = 1.41$  [1.08; 1.84]

Quantifying residual heterogeneity (with 95%-CIs):

$\tau^2 = 0.0082$  [0.0017; 0.2009];  $\tau = 0.0904$  [0.0417; 0.4482]  
 $I^2 = 46.0\%$  [6.3%; 68.9%];  $H = 1.36$  [1.03; 1.79]

Test of heterogeneity:

Q	d.f.	p-value
35.83	18	0.0074

```
Results for subgroups (common effect model):
      k      SMD      95%-CI      Q     I^2
setting = group  16  0.0424 [-0.0310; 0.1158] 26.50 43.4%
setting = indiv   3  0.3823 [ 0.0714; 0.6932]  4.98 59.9%
```

```
Test for subgroup differences (common effect model):
      Q d.f. p-value
Between groups 4.35    1  0.0370
Within groups 31.48   17  0.0174
```

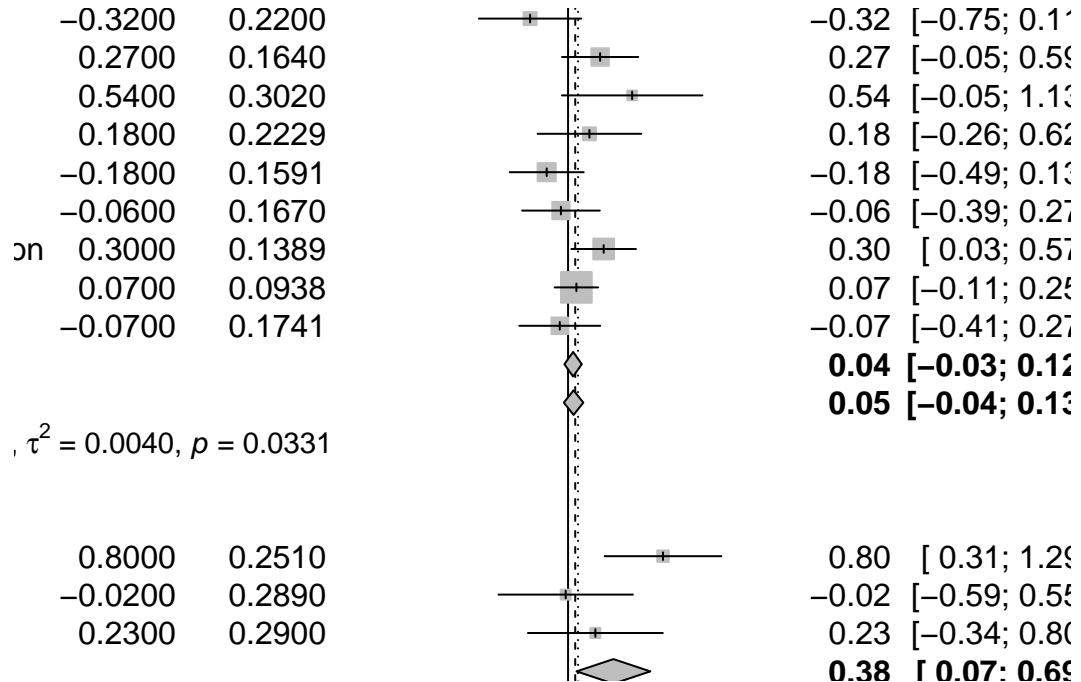
```
Results for subgroups (random effects model):
      k      SMD      95%-CI tau^2      tau
setting = group 16  0.0498 [-0.0400; 0.1397] 0.0082 0.0904
setting = indiv   3  0.3776 [ 0.0500; 0.7051] 0.0082 0.0904
```

```
Test for subgroup differences (random effects model):
      Q d.f. p-value
Between groups 3.58    1  0.0586
Within groups 31.48   17  0.0174
```

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$   
(assuming common  $\tau^2$  in subgroups)
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q
- Prediction interval based on t-distribution (df = 18)

```
# Forest plot por subgrupos (ajustá xlim/cex si es necesario)
forest(m.setting_separateTau, xlab = "Diferencia de Media Estandarizada")
```



```
print.subgroup.name = TRUE
```

```
# B) Subgrupo por 'tester' (aware vs blind)
m.tester_separateTau <- update(m.gen, subgroup = tester, tau.common = FALSE)
summary(m.tester_separateTau)
```

	SMD	95%-CI	%W(common)	%W(random)	tester
1: Rosenthal et al.	0.0300	[-0.2148; 0.2748]	8.5	7.7	aware
2: Conn et al.	0.1200	[-0.1681; 0.4081]	6.2	6.6	aware
3: Jose & Cody	-0.1400	[-0.4674; 0.1874]	4.8	5.7	aware
4: Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	1.0	1.7	aware
5: Pellegrini & Hicks	0.2600	[-0.4633; 0.9833]	1.0	1.7	blind
6: Evans & Rosenthal	-0.0600	[-0.2618; 0.1418]	12.5	9.1	aware
7: Fielder et al.	-0.0200	[-0.2218; 0.1818]	12.5	9.1	blind
8: Claiborn	-0.3200	[-0.7512; 0.1112]	2.7	4.0	aware
9: Kester	0.2700	[-0.0515; 0.5915]	4.9	5.8	aware
10: Maxwell	0.8000	[ 0.3081; 1.2919]	2.1	3.3	blind
11: Carter	0.5400	[-0.0519; 1.1319]	1.5	2.4	blind
12: Flowers	0.1800	[-0.2569; 0.6169]	2.7	3.9	blind
13: Keshock	-0.0200	[-0.5864; 0.5464]	1.6	2.6	blind
14: Henrikson	0.2300	[-0.3384; 0.7984]	1.6	2.6	blind
15: Fine	-0.1800	[-0.4918; 0.1318]	5.3	6.0	aware
on	0.3000				
	0.0700				
	-0.0700				

16: Grieger	-0.0600	[-0.3874; 0.2674]	4.8	5.7	blind
17: Rosenthal & Jacobson	0.3000	[ 0.0277; 0.5723]	6.9	7.0	aware
18: Fleming & Anttonen	0.0700	[-0.1139; 0.2539]	15.1	9.7	blind
19: Ginsburg	-0.0700	[-0.4112; 0.2712]	4.4	5.4	aware

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	0.0604	[-0.0111; 0.1318]	1.66	0.0979
Random effects model	0.0837	[-0.0175; 0.1849]	1.62	0.1051
Prediction interval		[-0.2243; 0.3917]		

Quantifying heterogeneity (with 95%-CIs):

$$\begin{aligned}\tau^2 &= 0.0188 [0.0059; 0.1977]; \tau = 0.1372 [0.0765; 0.4447] \\ I^2 &= 49.8\% [14.7\%; 70.4\%]; H = 1.41 [1.08; 1.84]\end{aligned}$$

Test of heterogeneity:

Q	d.f.	p-value
35.83	18	0.0074

Results for subgroups (common effect model):

	k	SMD	95%-CI	Q	I <sup>2</sup>
tester = aware	10	0.0345	[-0.0600; 0.1290]	22.20	59.5%
tester = blind	9	0.0949	[-0.0143; 0.2041]	12.96	38.3%

Test for subgroup differences (common effect model):

	Q	d.f.	p-value
Between groups	0.67	1	0.4122
Within groups	35.16	17	0.0059

Results for subgroups (random effects model):

	k	SMD	95%-CI	$\tau^2$	$\tau$
tester = aware	10	0.0466	[-0.0954; 0.1887]	0.0256	0.1599
tester = blind	9	0.1474	[-0.0193; 0.3140]	0.0236	0.1535

Test for subgroup differences (random effects model):

	Q	d.f.	p-value
Between groups	0.81	1	0.3672

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$

- Calculation of  $I^2$  based on Q
- Prediction interval based on t-distribution (df = 18)

```
m.tester_commonTau <- update(m.gen, subgroup = tester, tau.common = TRUE)
summary(m.tester_commonTau)
```

	SMD	95%-CI	%W(common)	%W(random)	tester
1: Rosenthal et al.	0.0300	[-0.2148; 0.2748]	8.5	7.7	aware
2: Conn et al.	0.1200	[-0.1681; 0.4081]	6.2	6.6	aware
3: Jose & Cody	-0.1400	[-0.4674; 0.1874]	4.8	5.7	aware
4: Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	1.0	1.7	aware
5: Pellegrini & Hicks	0.2600	[-0.4633; 0.9833]	1.0	1.7	blind
6: Evans & Rosenthal	-0.0600	[-0.2618; 0.1418]	12.5	9.1	aware
7: Fielder et al.	-0.0200	[-0.2218; 0.1818]	12.5	9.1	blind
8: Claiborn	-0.3200	[-0.7512; 0.1112]	2.7	4.0	aware
9: Kester	0.2700	[-0.0515; 0.5915]	4.9	5.8	aware
10: Maxwell	0.8000	[ 0.3081; 1.2919]	2.1	3.3	blind
11: Carter	0.5400	[-0.0519; 1.1319]	1.5	2.4	blind
12: Flowers	0.1800	[-0.2569; 0.6169]	2.7	3.9	blind
13: Keshock	-0.0200	[-0.5864; 0.5464]	1.6	2.6	blind
14: Henrikson	0.2300	[-0.3384; 0.7984]	1.6	2.6	blind
15: Fine	-0.1800	[-0.4918; 0.1318]	5.3	6.0	aware
16: Grieger	-0.0600	[-0.3874; 0.2674]	4.8	5.7	blind
17: Rosenthal & Jacobson	0.3000	[ 0.0277; 0.5723]	6.9	7.0	aware
18: Fleming & Anttonen	0.0700	[-0.1139; 0.2539]	15.1	9.7	blind
19: Ginsburg	-0.0700	[-0.4112; 0.2712]	4.4	5.4	aware

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	0.0604	[-0.0111; 0.1318]	1.66	0.0979
Random effects model	0.0837	[-0.0175; 0.1849]	1.62	0.1051
Prediction interval		[-0.2243; 0.3917]		

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0188$  [0.0059; 0.1977];  $\tau = 0.1372$  [0.0765; 0.4447]  
 $I^2 = 49.8\%$  [14.7%; 70.4%];  $H = 1.41$  [1.08; 1.84]

Quantifying residual heterogeneity (with 95%-CIs):

$\tau^2 = 0.0248$  [0.0068; 0.2115];  $\tau = 0.1574$  [0.0825; 0.4599]  
 $I^2 = 51.6\%$  [17.1%; 71.8%];  $H = 1.44$  [1.10; 1.88]

Test of heterogeneity:

Q	d.f.	p-value
35.83	18	0.0074

Results for subgroups (common effect model):

	k	SMD	95%-CI	Q	I <sup>2</sup>
tester = aware	10	0.0345	[-0.0600; 0.1290]	22.20	59.5%
tester = blind	9	0.0949	[-0.0143; 0.2041]	12.96	38.3%

Test for subgroup differences (common effect model):

	Q	d.f.	p-value
Between groups	0.67	1	0.4122
Within groups	35.16	17	0.0059

Results for subgroups (random effects model):

	k	SMD	95%-CI	tau <sup>2</sup>	tau
tester = aware	10	0.0463	[-0.0945; 0.1872]	0.0248	0.1574
tester = blind	9	0.1488	[-0.0198; 0.3174]	0.0248	0.1574

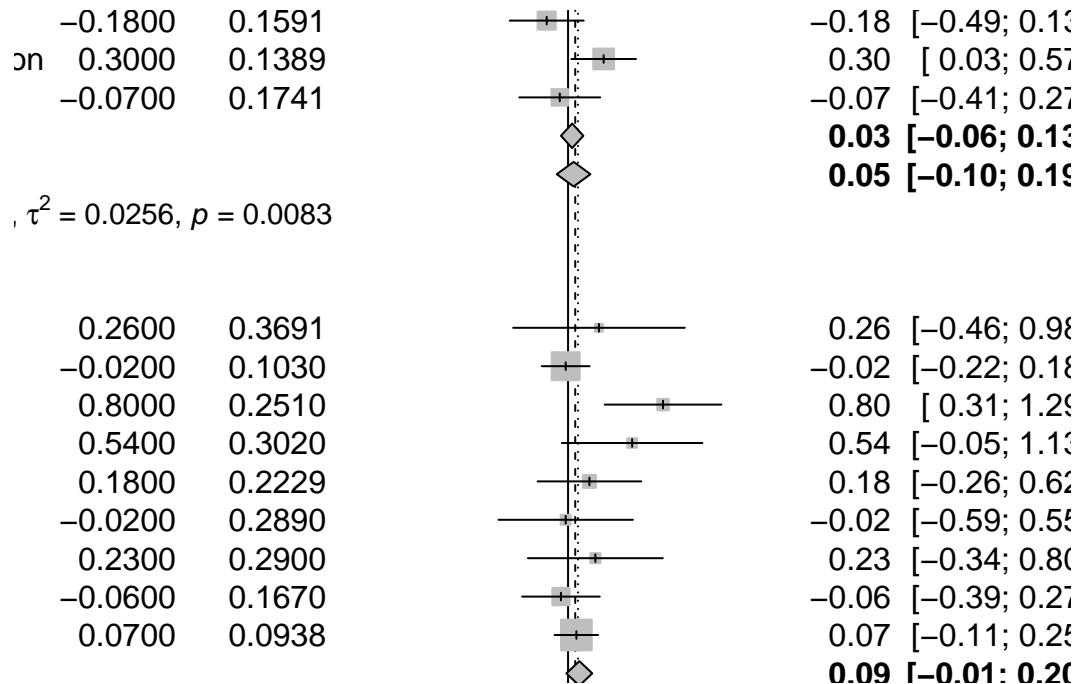
Test for subgroup differences (random effects model):

	Q	d.f.	p-value
Between groups	0.84	1	0.3606
Within groups	35.16	17	0.0059

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for tau<sup>2</sup>  
(assuming common tau<sup>2</sup> in subgroups)
- Q-Profile method for confidence interval of tau<sup>2</sup> and tau
- Calculation of I<sup>2</sup> based on Q
- Prediction interval based on t-distribution (df = 18)

```
forest(m.tester_separateTau, xlab ="Diferencia de Media Estandarizada")
```



```
mr.weeks <- metareg(m.gen, ~ weeks)
summary(mr.weeks)
```

Mixed-Effects Model (k = 19; tau^2 estimator: REML)

logLik	deviance	AIC	BIC	AICc
-0.7144	1.4287	7.4287	9.9284	9.2749

tau^2 (estimated amount of residual heterogeneity): 0.0112 (SE = 0.0128)  
 tau (square root of estimated tau^2 value): 0.1057  
 I^2 (residual heterogeneity / unaccounted variability): 29.40%  
 H^2 (unaccounted variability / sampling variability): 1.42  
 R^2 (amount of heterogeneity accounted for): 40.63%

Test for Residual Heterogeneity:  
 QE(df = 17) = 27.6730, p-val = 0.0489

Test of Moderators (coefficient 2):  
 QM(df = 1) = 7.5094, p-val = 0.0061

Model Results:

```

            estimate      se     zval    pval    ci.lb    ci.ub
intrcpt    0.1942  0.0633  3.0649  0.0022   0.0700   0.3183 ** 
weeks     -0.0157  0.0057 -2.7403  0.0061  -0.0270  -0.0045 ** 

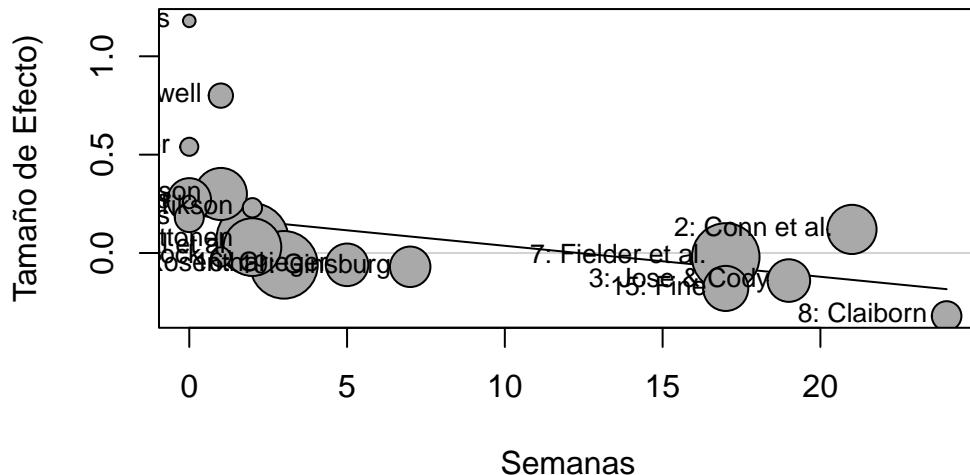
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

bubble(mr.weeks,
       xlab = "Semanas",
       ylab = "Tamaño de Efecto",
       studlab = TRUE)

```

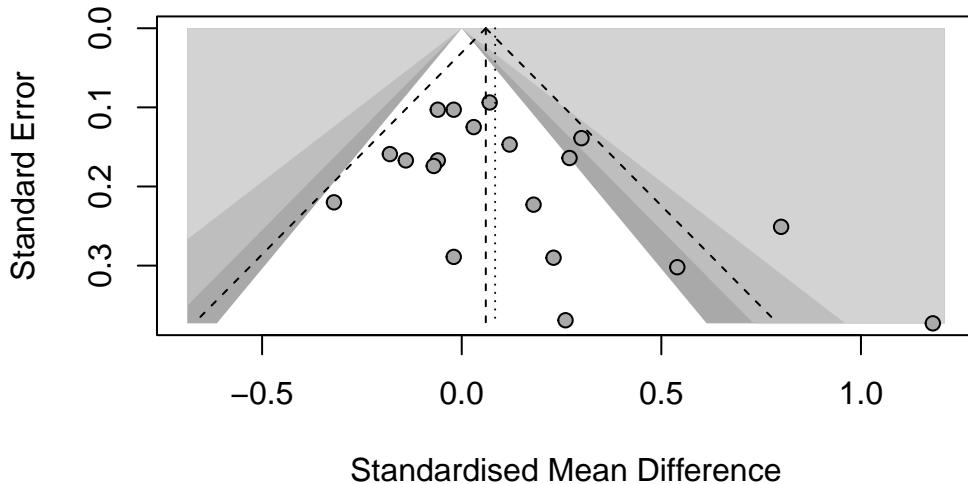


```

funnel(m.gen, main = "Funnel plot - Raudenbush 1985", contour = c(0.90, 0.95, 0.99), col.cont

```

## Funnel plot – Raudenbush 1985



```
# Radial con Egger-
radial(m.gen, main = "Radial plot con línea de Egger")

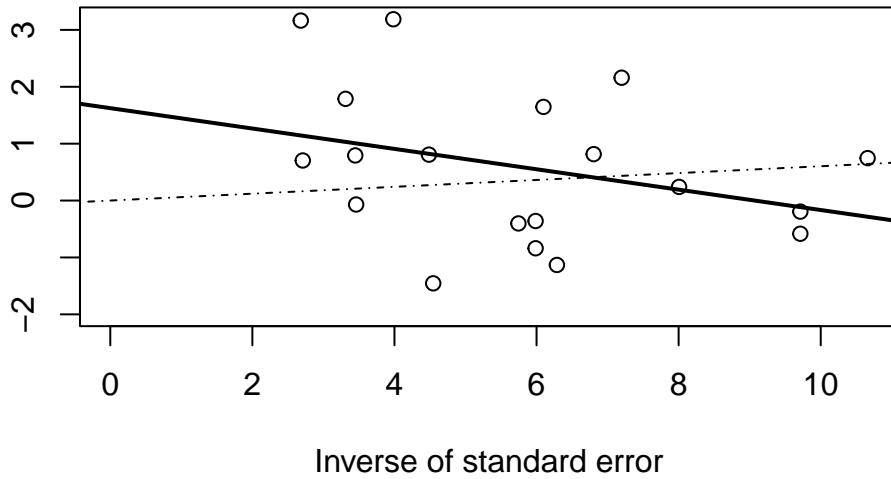
# Crear variables de Egger
yi_over_se <- m.gen$TE / m.gen$seTE      # yi / SEi
inv_se     <- 1 / m.gen$seTE                # 1 / SEi

# Ajustar modelo de Egger (regresión lineal)
egger_model <- lm(yi_over_se ~ inv_se)

# Agregar la línea de Egger al radial plot
abline(egger_model, lwd = 2)
```

Standardised treatment effect (z-score)

Radial plot con línea de Egger



```
metabias(m.gen, method = "rank")
```

Rank correlation test of funnel plot asymmetry

Test result:  $z = 1.79$ , p-value = 0.0740  
Bias estimate: 51.0000 (SE = 28.5482)

Reference: Begg & Mazumdar (1993), Biometrics

```
metabias(m.gen, method = "linreg")
```

Linear regression test of funnel plot asymmetry

Test result:  $t = 2.04$ , df = 17, p-value = 0.0574  
Bias estimate: 1.6243 (SE = 0.7970)

Details:

- multiplicative residual heterogeneity variance ( $\tau^2 = 1.6938$ )
- predictor: standard error
- weight: inverse variance
- reference: Egger et al. (1997), BMJ

```
metabias(m.gen, method = "mm")
```

Linear regression test of funnel plot asymmetry

Test result: t = 2.53, df = 17, p-value = 0.0218  
Bias estimate: 1.8281 (SE = 0.7237)

Details:

- additive residual heterogeneity variance ( $\tau^2 = 0.0129$ )
- restricted maximum-likelihood estimator for  $\tau^2$
- predictor: standard error
- weight: inverse variance
- reference: Thompson & Sharp (1999), Stat Med

```
tf <- trimfill(m.gen)  
summary(tf)
```

	SMD	95%-CI	%W(random)
1: Rosenthal et al.	0.0300	[ -0.2148; 0.2748]	5.9
2: Conn et al.	0.1200	[ -0.1681; 0.4081]	5.6
3: Jose & Cody	-0.1400	[ -0.4674; 0.1874]	5.2
4: Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	2.5
5: Pellegrini & Hicks	0.2600	[ -0.4633; 0.9833]	2.6
6: Evans & Rosenthal	-0.0600	[ -0.2618; 0.1418]	6.3
7: Fielder et al.	-0.0200	[ -0.2218; 0.1818]	6.3
8: Claiborn	-0.3200	[ -0.7512; 0.1112]	4.4
9: Kester	0.2700	[ -0.0515; 0.5915]	5.3
10: Maxwell	0.8000	[ 0.3081; 1.2919]	3.9
11: Carter	0.5400	[ -0.0519; 1.1319]	3.3
12: Flowers	0.1800	[ -0.2569; 0.6169]	4.3
13: Keshock	-0.0200	[ -0.5864; 0.5464]	3.4
14: Henrikson	0.2300	[ -0.3384; 0.7984]	3.4
15: Fine	-0.1800	[ -0.4918; 0.1318]	5.4
16: Grieger	-0.0600	[ -0.3874; 0.2674]	5.2
17: Rosenthal & Jacobson	0.3000	[ 0.0277; 0.5723]	5.7
18: Fleming & Anttonen	0.0700	[ -0.1139; 0.2539]	6.4
19: Ginsburg	-0.0700	[ -0.4112; 0.2712]	5.1
Filled: 11: Carter	-0.4891	[ -1.0809; 0.1028]	3.3
Filled: 10: Maxwell	-0.7491	[ -1.2410; -0.2571]	3.9
Filled: 4: Pellegrini & Hicks	-1.1291	[ -1.8600; -0.3981]	2.5

Number of studies: k = 22 (with 3 added studies)

SMD	95%-CI	z	p-value
Random effects model 0.0282	[-0.1170; 0.1734]	0.38	0.7035
Prediction interval	[-0.5689; 0.6253]		

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0769$  [0.0440; 0.3714];  $\tau = 0.2774$  [0.2098; 0.6094]  
 $I^2 = 64.3\%$  [43.7%; 77.3%];  $H = 1.67$  [1.33; 2.10]

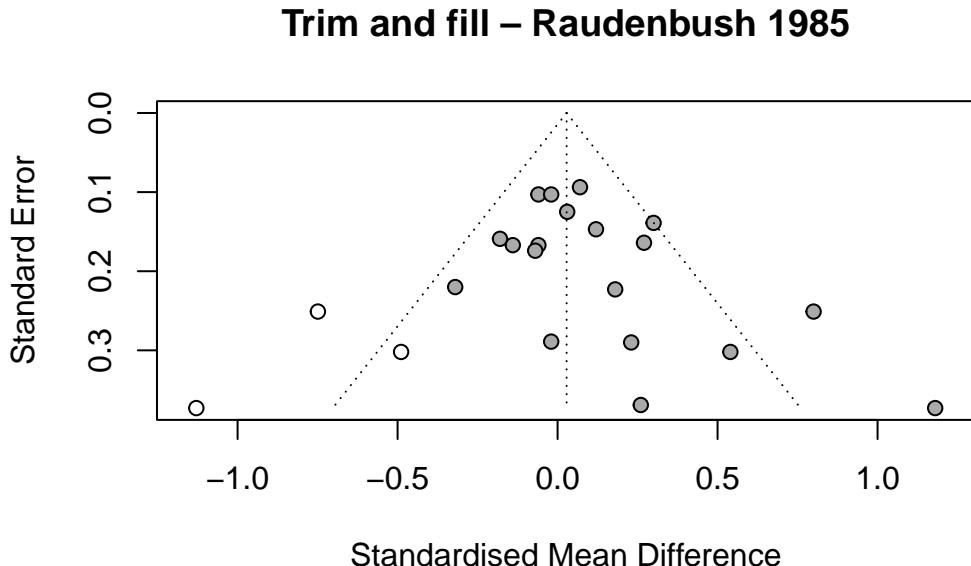
Test of heterogeneity:

Q	d.f.	p-value
58.75	21	< 0.0001

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q
- Prediction interval based on t-distribution (df = 21)
- Trim-and-fill method to adjust for funnel plot asymmetry (L-estimator)

```
funnel(tf, main = "Trim and fill - Raudenbush 1985")
```



```

#al no haber datos faltantes seguimos el procedimiento del libro y modificamos los datos agrupados

set.seed(123)
dat2 <- dat

# columnas básicas
dat2$n1i <- as.integer(dat2$n1i)
dat2$n2i <- as.integer(dat2$n2i)
dat2$TE    <- dat2$yi
dat2$seTE <- sqrt(dat2$vi)

# Elegimos estudios más precisos
ord <- order(dat2$seTE)
sel7 <- ord[1:7]
sel_missing <- sel7[c(3,5)]

# Definir pérdidas
dat2$Nem <- 0
dat2$Ncm <- 0
dat2$Nem[sel_missing] <- c(8,5)
dat2$Ncm[sel_missing] <- dat2$Nem[sel_missing]

# Observados por brazo
dat2$Neo <- dat2$n1i - dat2$Nem
dat2$Nco <- dat2$n2i - dat2$Ncm

# Proporciones faltantes
dat2$Pe <- dat2$Nem / dat2$n1i
dat2$Pc <- dat2$Ncm / dat2$n2i

# Metaanálisis basado en los datos observados (pre-ajuste)
m.gen.miss <- metagen(
  TE = dat2$TE,
  seTE = dat2$seTE,
  data = dat2,
  studlab = paste(dat2$study, dat2$author, sep=": "),
  sm = "SMD",
  method.tau = "REML",
  prediction = TRUE
)

summary(m.gen.miss)

```

	SMD	95%-CI	%W(common)	%W(random)
1: Rosenthal et al.	0.0300	[-0.2148; 0.2748]	8.5	7.7
2: Conn et al.	0.1200	[-0.1681; 0.4081]	6.2	6.6
3: Jose & Cody	-0.1400	[-0.4674; 0.1874]	4.8	5.7
4: Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	1.0	1.7
5: Pellegrini & Hicks	0.2600	[-0.4633; 0.9833]	1.0	1.7
6: Evans & Rosenthal	-0.0600	[-0.2618; 0.1418]	12.5	9.1
7: Fielder et al.	-0.0200	[-0.2218; 0.1818]	12.5	9.1
8: Claiborn	-0.3200	[-0.7512; 0.1112]	2.7	4.0
9: Kester	0.2700	[-0.0515; 0.5915]	4.9	5.8
10: Maxwell	0.8000	[ 0.3081; 1.2919]	2.1	3.3
11: Carter	0.5400	[-0.0519; 1.1319]	1.5	2.4
12: Flowers	0.1800	[-0.2569; 0.6169]	2.7	3.9
13: Keshock	-0.0200	[-0.5864; 0.5464]	1.6	2.6
14: Henrikson	0.2300	[-0.3384; 0.7984]	1.6	2.6
15: Fine	-0.1800	[-0.4918; 0.1318]	5.3	6.0
16: Griege	-0.0600	[-0.3874; 0.2674]	4.8	5.7
17: Rosenthal & Jacobson	0.3000	[ 0.0277; 0.5723]	6.9	7.0
18: Fleming & Anttonen	0.0700	[-0.1139; 0.2539]	15.1	9.7
19: Ginsburg	-0.0700	[-0.4112; 0.2712]	4.4	5.4

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	0.0604	[-0.0111; 0.1318]	1.66	0.0979
Random effects model	0.0837	[-0.0175; 0.1849]	1.62	0.1051
Prediction interval		[-0.2243; 0.3917]		

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0188$  [0.0059; 0.1977];  $\tau = 0.1372$  [0.0765; 0.4447]  
 $I^2 = 49.8\%$  [14.7%; 70.4%];  $H = 1.41$  [1.08; 1.84]

Test of heterogeneity:

Q	d.f.	p-value
35.83	18	0.0074

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q
- Prediction interval based on t-distribution (df = 18)

```

# semiss: función del libro Schwarzer et al. (2015), cap. 6

semiss <- function(TE, seTE, mu.e, mu.c, nu.e, nu.c, rho){

  # Delta para cada estudio
  delta.e <- rnorm(length(TE), mean = mu.e, sd = nu.e)
  delta.c <- rnorm(length(TE), mean = mu.c, sd = nu.c)

  # Correlación entre brazos
  if (rho != 0) {
    delta.c <- rho * delta.e + sqrt(1 - rho^2) * delta.c
  }

  # Ajustar TE
  TE.star <- TE + delta.e - delta.c

  # Ajustar varianza
  seTE.star <- sqrt(seTE^2 + nu.e^2 + nu.c^2 - 2*rho*nu.e*nu.c)

  return(list(TE = TE.star, seTE = seTE.star))
}

mu <- 8
nu <- sqrt(5)

# Convenience:
TE <- dat2$yi
seT <- dat2$seTE

#fixed equal
S1 <- semiss(TE, seT, mu.e = mu, mu.c = mu, nu.e = 0, nu.c = 0, rho = 0)

m.S1 <- metagen(
  TE = S1$TE,
  seTE = S1$seTE,
  studlab = paste(dat2$study, dat2$author),
  sm = "SMD",
  method.tau = "REML"
)
summary(m.S1)

```

SMD                  95%-CI %W(common) %W(random)

1 Rosenthal et al.	0.0300	[ -0.2148; 0.2748]	8.5	7.7
2 Conn et al.	0.1200	[ -0.1681; 0.4081]	6.2	6.6
3 Jose & Cody	-0.1400	[ -0.4674; 0.1874]	4.8	5.7
4 Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	1.0	1.7
5 Pellegrini & Hicks	0.2600	[ -0.4633; 0.9833]	1.0	1.7
6 Evans & Rosenthal	-0.0600	[ -0.2618; 0.1418]	12.5	9.1
7 Fielder et al.	-0.0200	[ -0.2218; 0.1818]	12.5	9.1
8 Claiborn	-0.3200	[ -0.7512; 0.1112]	2.7	4.0
9 Kester	0.2700	[ -0.0515; 0.5915]	4.9	5.8
10 Maxwell	0.8000	[ 0.3081; 1.2919]	2.1	3.3
11 Carter	0.5400	[ -0.0519; 1.1319]	1.5	2.4
12 Flowers	0.1800	[ -0.2569; 0.6169]	2.7	3.9
13 Keshock	-0.0200	[ -0.5864; 0.5464]	1.6	2.6
14 Henrikson	0.2300	[ -0.3384; 0.7984]	1.6	2.6
15 Fine	-0.1800	[ -0.4918; 0.1318]	5.3	6.0
16 Grieger	-0.0600	[ -0.3874; 0.2674]	4.8	5.7
17 Rosenthal & Jacobson	0.3000	[ 0.0277; 0.5723]	6.9	7.0
18 Fleming & Anttonen	0.0700	[ -0.1139; 0.2539]	15.1	9.7
19 Ginsburg	-0.0700	[ -0.4112; 0.2712]	4.4	5.4

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	0.0604	[ -0.0111; 0.1318]	1.66	0.0979
Random effects model	0.0837	[ -0.0175; 0.1849]	1.62	0.1051

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0188$  [0.0059; 0.1977];  $\tau = 0.1372$  [0.0765; 0.4447]  
 $I^2 = 49.8\%$  [14.7%; 70.4%];  $H = 1.41$  [1.08; 1.84]

Test of heterogeneity:

Q	d.f.	p-value
35.83	18	0.0074

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q

```
#fixed opposite
S2 <- semiss(TE, seT, mu.e = mu, mu.c = -mu, nu.e = 0, nu.c = 0, rho = 0)
```

```

m.S2 <- metagen(
  TE = S2$TE,
  seTE = S2$seTE,
  studlab = paste(dat2$study, dat2$author),
  sm = "SMD",
  method.tau = "REML"
)
summary(m.S2)

```

	SMD	95%-CI	%W(common)	%W(random)
1 Rosenthal et al.	16.0300	[15.7852; 16.2748]	8.5	7.7
2 Conn et al.	16.1200	[15.8319; 16.4081]	6.2	6.6
3 Jose & Cody	15.8600	[15.5326; 16.1874]	4.8	5.7
4 Pellegrini & Hicks	17.1800	[16.4490; 17.9110]	1.0	1.7
5 Pellegrini & Hicks	16.2600	[15.5367; 16.9833]	1.0	1.7
6 Evans & Rosenthal	15.9400	[15.7382; 16.1418]	12.5	9.1
7 Fielder et al.	15.9800	[15.7782; 16.1818]	12.5	9.1
8 Claiborn	15.6800	[15.2488; 16.1112]	2.7	4.0
9 Kester	16.2700	[15.9485; 16.5915]	4.9	5.8
10 Maxwell	16.8000	[16.3081; 17.2919]	2.1	3.3
11 Carter	16.5400	[15.9481; 17.1319]	1.5	2.4
12 Flowers	16.1800	[15.7431; 16.6169]	2.7	3.9
13 Keshock	15.9800	[15.4136; 16.5464]	1.6	2.6
14 Henrikson	16.2300	[15.6616; 16.7984]	1.6	2.6
15 Fine	15.8200	[15.5082; 16.1318]	5.3	6.0
16 Grieger	15.9400	[15.6126; 16.2674]	4.8	5.7
17 Rosenthal & Jacobson	16.3000	[16.0277; 16.5723]	6.9	7.0
18 Fleming & Anttonen	16.0700	[15.8861; 16.2539]	15.1	9.7
19 Ginsburg	15.9300	[15.5888; 16.2712]	4.4	5.4

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	16.0604	[15.9889; 16.1318]	440.40	0
Random effects model	16.0837	[15.9825; 16.1849]	311.42	0

Quantifying heterogeneity (with 95%-CIs):

tau^2 = 0.0188 [0.0059; 0.1977]; tau = 0.1372 [0.0765; 0.4447]  
I^2 = 49.8% [14.7%; 70.4%]; H = 1.41 [1.08; 1.84]

Test of heterogeneity:

Q	d.f.	p-value
35.83	18	0.0074

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q

```
#random equal
S3 <- semiss(TE, seT, mu.e = mu, mu.c = mu, nu.e = nu, nu.c = nu, rho = 1)

m.S3 <- metagen(
  TE = S3$TE,
  seTE = S3$seTE,
  studlab = paste(dat2$study, dat2$author),
  sm = "SMD",
  method.tau = "REML"
)
summary(m.S3)
```

	SMD	95%-CI	%W(common)	%W(random)
1 Rosenthal et al.	0.0300	[ -0.2148; 0.2748]	8.5	7.7
2 Conn et al.	0.1200	[ -0.1681; 0.4081]	6.2	6.6
3 Jose & Cody	-0.1400	[ -0.4674; 0.1874]	4.8	5.7
4 Pellegrini & Hicks	1.1800	[ 0.4490; 1.9110]	1.0	1.7
5 Pellegrini & Hicks	0.2600	[ -0.4633; 0.9833]	1.0	1.7
6 Evans & Rosenthal	-0.0600	[ -0.2618; 0.1418]	12.5	9.1
7 Fielder et al.	-0.0200	[ -0.2218; 0.1818]	12.5	9.1
8 Claiborn	-0.3200	[ -0.7512; 0.1112]	2.7	4.0
9 Kester	0.2700	[ -0.0515; 0.5915]	4.9	5.8
10 Maxwell	0.8000	[ 0.3081; 1.2919]	2.1	3.3
11 Carter	0.5400	[ -0.0519; 1.1319]	1.5	2.4
12 Flowers	0.1800	[ -0.2569; 0.6169]	2.7	3.9
13 Keshock	-0.0200	[ -0.5864; 0.5464]	1.6	2.6
14 Henrikson	0.2300	[ -0.3384; 0.7984]	1.6	2.6
15 Fine	-0.1800	[ -0.4918; 0.1318]	5.3	6.0
16 Grieger	-0.0600	[ -0.3874; 0.2674]	4.8	5.7
17 Rosenthal & Jacobson	0.3000	[ 0.0277; 0.5723]	6.9	7.0
18 Fleming & Anttonen	0.0700	[ -0.1139; 0.2539]	15.1	9.7
19 Ginsburg	-0.0700	[ -0.4112; 0.2712]	4.4	5.4

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	0.0604	[-0.0111; 0.1318]	1.66	0.0979
Random effects model	0.0837	[-0.0175; 0.1849]	1.62	0.1051

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0188$  [0.0059; 0.1977];  $\tau = 0.1372$  [0.0765; 0.4447]  
 $I^2 = 49.8\%$  [14.7%; 70.4%];  $H = 1.41$  [1.08; 1.84]

Test of heterogeneity:

Q	d.f.	p-value
35.83	18	0.0074

Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q

```
#random equal
S4 <- semiss(TE, seT, mu.e = mu, mu.c = -mu, nu.e = nu, nu.c = nu, rho = 0)

m.S4 <- metagen(
  TE = S4$TE,
  seTE = S4$seTE,
  studlab = paste(dat2$study, dat2$author),
  sm = "SMD",
  method.tau = "REML"
)
summary(m.S4)
```

	SMD	95%-CI	%W(common)	%W(random)
1 Rosenthal et al.	14.0386	[ 7.8358; 20.2414]	5.3	5.3
2 Conn et al.	14.9923	[ 8.7877; 21.1969]	5.3	5.3
3 Jose & Cody	13.8237	[ 7.6171; 20.0303]	5.3	5.3
4 Pellegrini & Hicks	15.8662	[ 9.6253; 22.1071]	5.2	5.2
5 Pellegrini & Hicks	14.5537	[ 8.3137; 20.7937]	5.2	5.2
6 Evans & Rosenthal	21.5350	[15.3338; 27.7362]	5.3	5.3
7 Fielder et al.	20.9587	[14.7575; 27.1599]	5.3	5.3
8 Claiborn	15.5653	[ 9.3523; 21.7782]	5.3	5.3
9 Kester	14.6904	[ 8.4841; 20.8967]	5.3	5.3

10 Maxwell	14.7543 [ 8.5369; 20.9717]	5.3	5.3
11 Carter	18.1655 [11.9394; 24.3917]	5.2	5.2
12 Flowers	13.9313 [ 7.7180; 20.1447]	5.3	5.3
13 Keshock	11.9623 [ 5.7385; 18.1861]	5.2	5.2
14 Henrikson	17.2641 [11.0402; 23.4881]	5.2	5.2
15 Fine	20.8876 [14.6818; 27.0934]	5.3	5.3
16 Grieger	16.7514 [10.5448; 22.9580]	5.3	5.3
17 Rosenthal & Jacobson	17.3810 [11.1771; 23.5849]	5.3	5.3
18 Fleming & Anttonen	20.9994 [14.7987; 27.2000]	5.3	5.3
19 Ginsburg	10.1736 [ 3.9663; 16.3810]	5.3	5.3

Number of studies: k = 19

	SMD	95%-CI	z	p-value
Common effect model	16.2295	[14.8041; 17.6548]	22.32	< 0.0001
Random effects model	16.2294	[14.7987; 17.6601]	22.23	< 0.0001

Quantifying heterogeneity (with 95%-CIs):

$\tau^2 = 0.0758$  [0.0000; 12.0318];  $\tau = 0.2753$  [0.0000; 3.4687]  
 $I^2 = 0.6\%$  [0.0%; 49.2%];  $H = 1.00$  [1.00; 1.40]

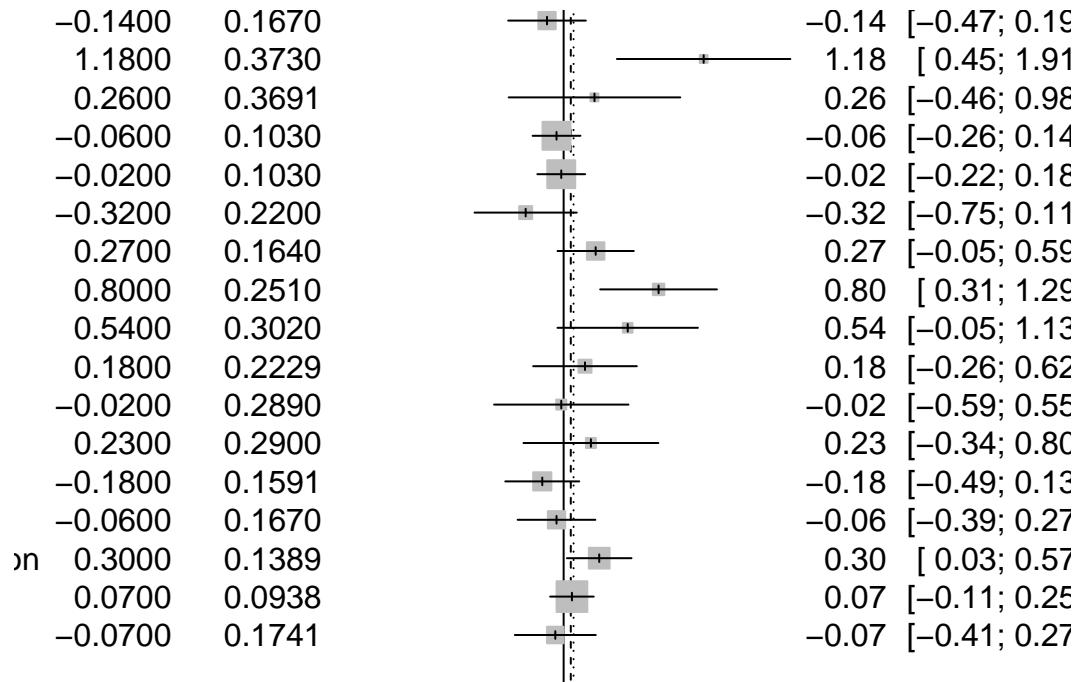
Test of heterogeneity:

Q	d.f.	p-value
18.10	18	0.4488

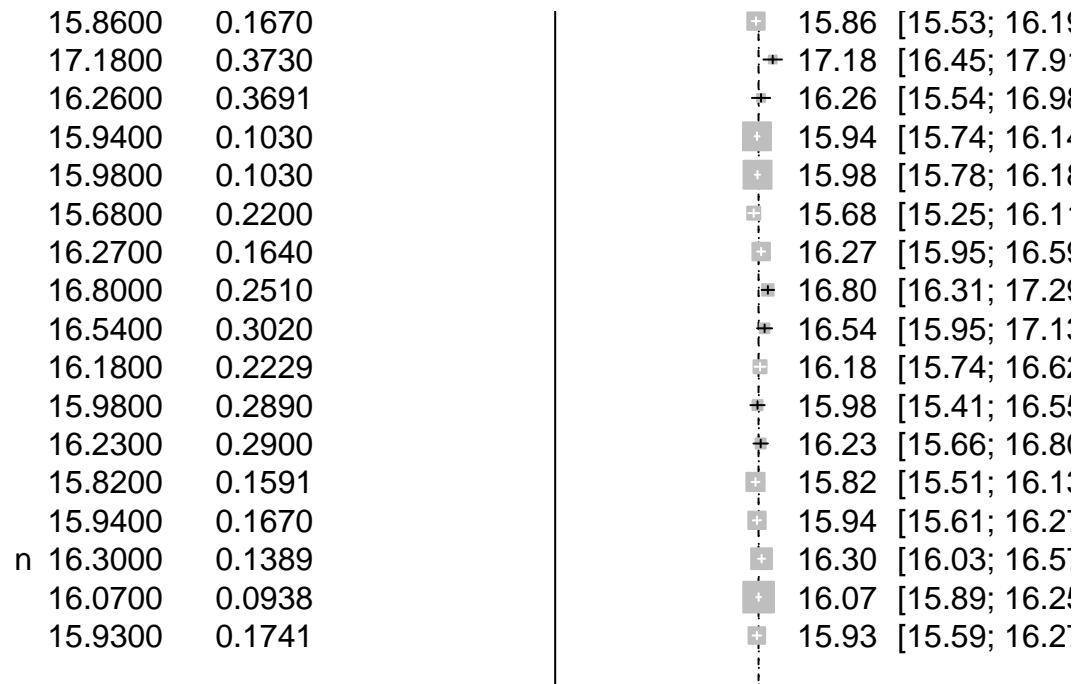
Details of meta-analysis methods:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Q-Profile method for confidence interval of  $\tau^2$  and  $\tau$
- Calculation of  $I^2$  based on Q

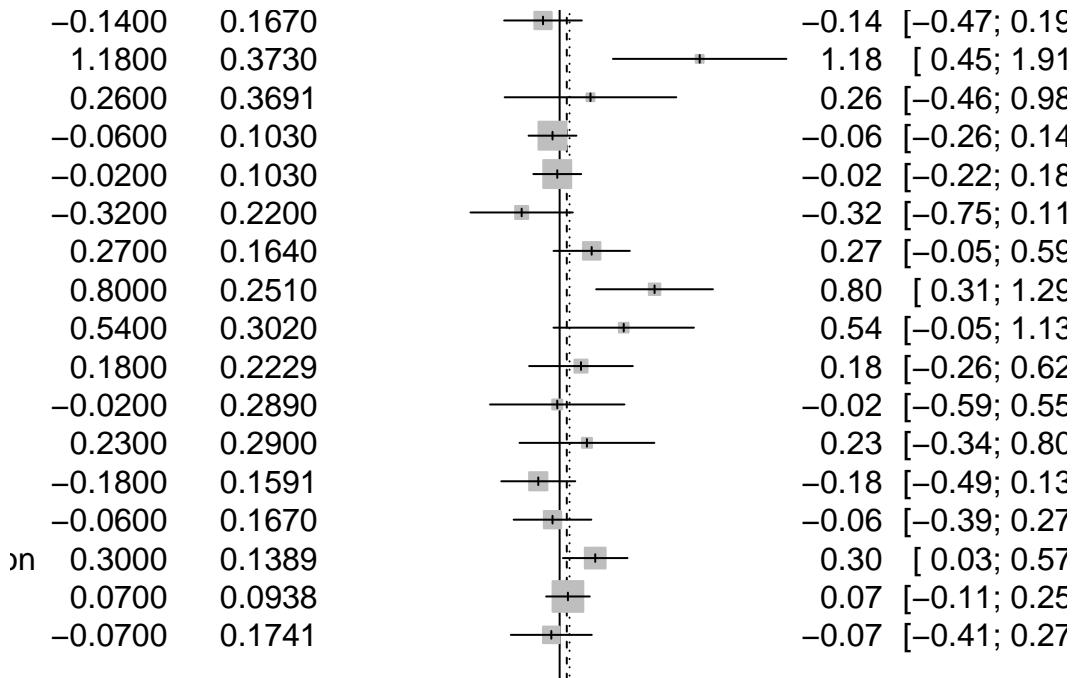
```
#graficos
forest(m.S1, main="Missing Data - Fixed Equal")
```



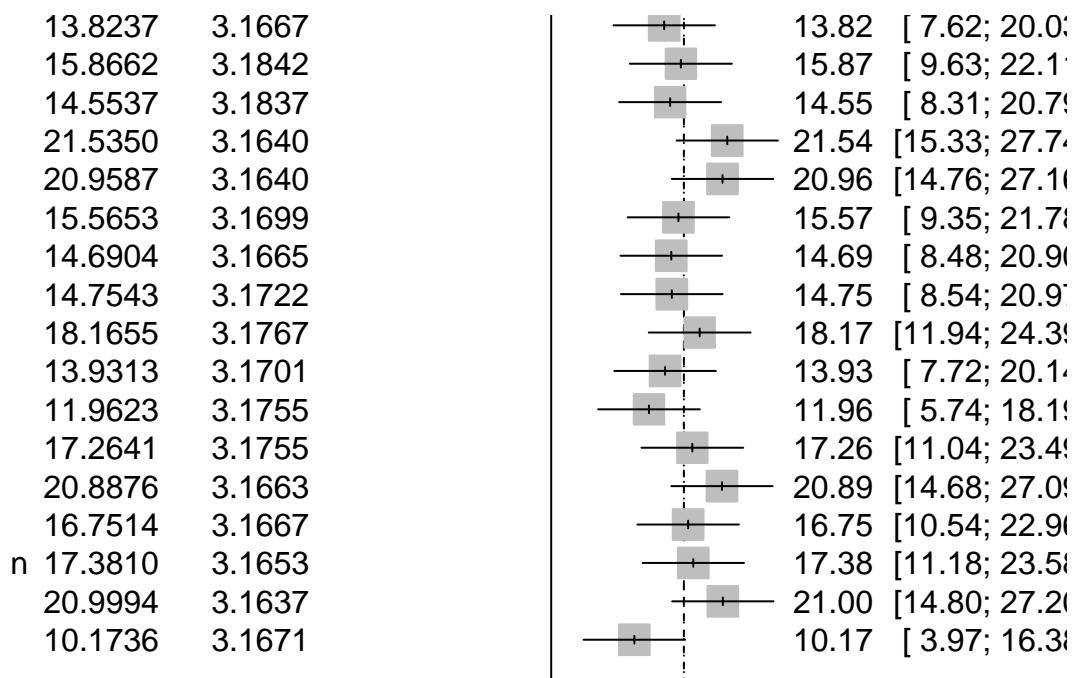
```
forest(m.S2, main="Missing Data - Fixed Opposite")
```



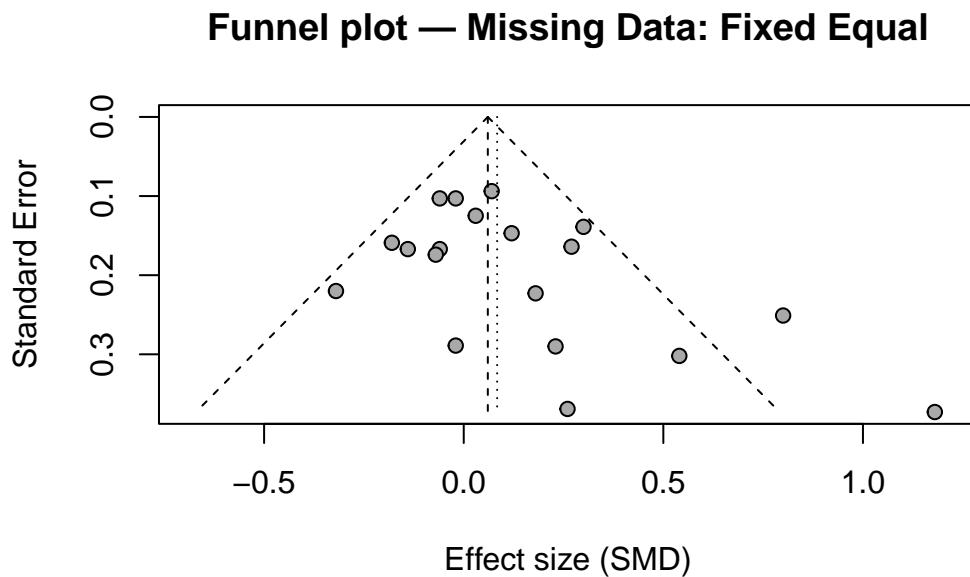
```
forest(m.S3, main="Missing Data - Random Equal")
```



```
forest(m.S4, main="Missing Data - Random Uncorrelated")
```

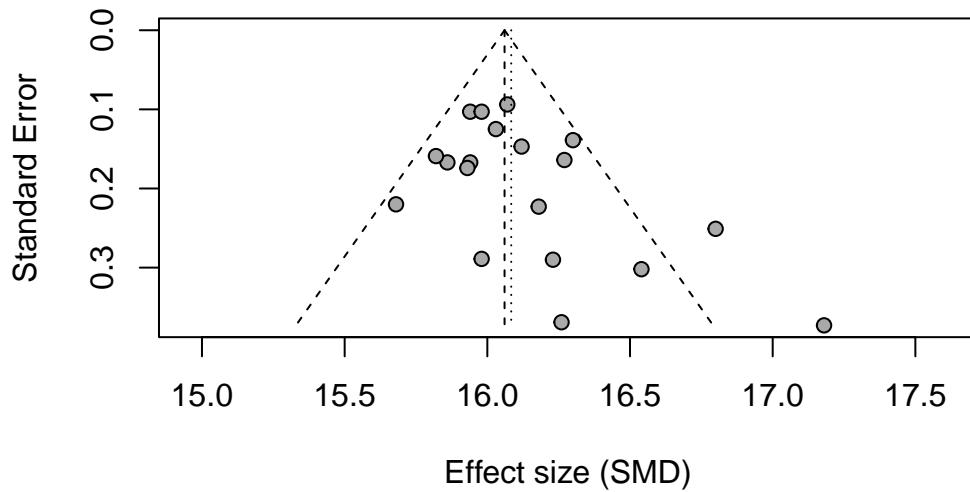


```
funnel(m.S1,
       main = "Funnel plot - Missing Data: Fixed Equal",
       xlab = "Effect size (SMD)")
```



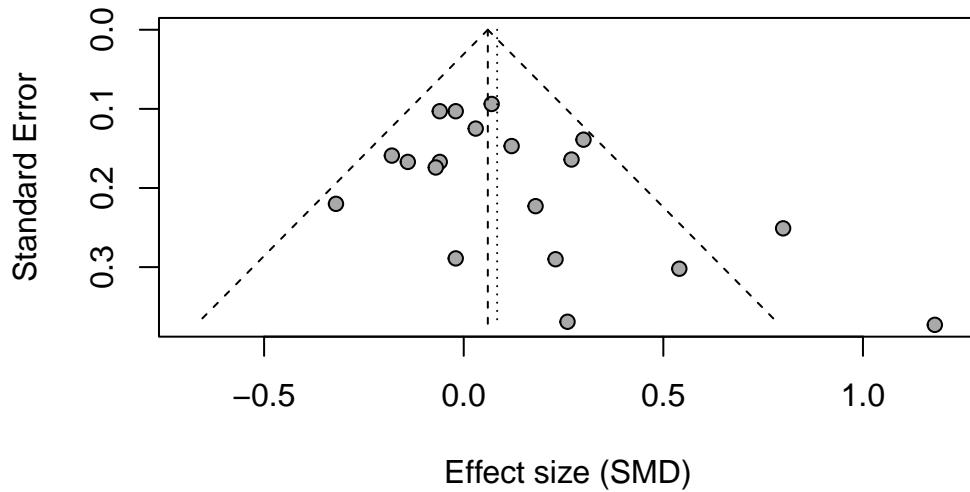
```
funnel(m.S2,
       main = "Funnel plot - Missing Data: Fixed Opposite",
       xlab = "Effect size (SMD)")
```

### Funnel plot — Missing Data: Fixed Opposite



```
funnel(m.S3,
       main = "Funnel plot - Missing Data: Random Equal",
       xlab = "Effect size (SMD)")
```

### Funnel plot — Missing Data: Random Equal



```
funnel(m.S4,
       main = "Funnel plot - Missing Data: Random Uncorrelated",
       xlab = "Effect size (SMD)")
```



```
library(metapower)

n_exp <- mean((dat$n1i + dat$n2i)/2)

my_power1<- mpower(effect_size = .1, study_size =n_exp, k = 19,i2 = .50, es_type = "d")
my_power1
```

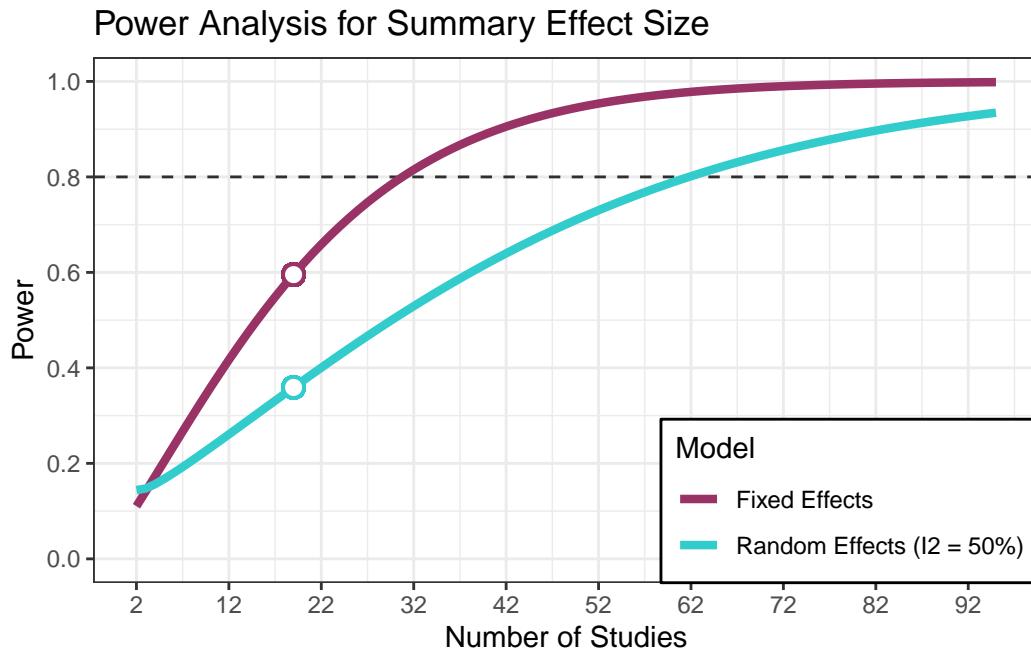
Power Analysis for Meta-analysis

```
Effect Size Metric: d
Expected Effect Size: 0.1
Expected Study Size: 102.1316
Expected Number of Studies: 19
```

Estimated Power: Mean Effect Size

```
Fixed-Effects Model           0.5953668
Random-Effects Model (i2 = 50%): 0.3591031
```

```
plot_mpower(my_power1)
```



```
my_power2<- mpower(effect_size = .2, study_size =n_exp, k = 19, i2 = .50, es_type = "d")
my_power2
```

### Power Analysis for Meta-analysis

```
Effect Size Metric:          d
Expected Effect Size:        0.2
Expected Study Size:         102.1316
Expected Number of Studies:  19
```

### Estimated Power: Mean Effect Size

```
Fixed-Effects Model           0.9925383
Random-Effects Model (i2 = 50%): 0.867599
```

```
plot_mpower(my_power2)
```

