

Aquiles y el bosque de los efectos

Aquiles. ¡Qué tío! Sin duda, uno de los más famosos de todo el follón que armaron en Troya por culpa de Helena la guapa. Ya sabéis su historia. El tío era la leche porque su madre, que era nada menos que la ninfa Tetis, lo bañó en ambrosía y lo sumergió en la laguna Estigia para que fuese invulnerable. Pero cometió un error que una ninfa no debiera haberse permitido: lo agarró por el talón derecho, que no se mojó en la laguna. Así que de ahí le viene a Aquiles su punto débil. Héctor no se dio cuenta a tiempo pero Paris, bastante más espabilado, le metió un flechazo en el talón y lo mandó otra vez a la laguna, pero no al agua, sino al otro lado. Y sin barquero.

Este cuento es el origen de la expresión “talón de Aquiles”, que suele referirse al punto más débil o vulnerable de alguien o algo que, por lo demás, suele ser conocido por su fortaleza.

Por ejemplo, algo tan robusto y formidable como el metanálisis tiene su talón de Aquiles: el [sesgo de publicación](#). Y eso se debe a que en el mundo de la ciencia no hay justicia social.

Todos los trabajos científicos deberían tener las mismas oportunidades de ser publicados y alcanzar la fama, pero la realidad no es en absoluto así y los trabajos pueden verse discriminados por cuatro razones: significación estadística, popularidad del tema que tratan, el tener alguien que los apadrine y el idioma en que están escritos.

Estos son los principales factores que pueden contribuir a este sesgo de publicación. En primer lugar, es más probable que se publiquen los estudios con resultados significativos y, dentro de estos, es más probable que se publiquen cuando el efecto es mayor. Esto hace que los estudios con resultados negativos o con efectos de pequeña magnitud puedan no llegar a ser publicados, con lo que sacaremos una conclusión sesgada del análisis solo de los estudios grandes con resultado positivo. De igual manera, los trabajos sobre temas de interés público tienen más probabilidad de ser publicados con independencia de la importancia de sus resultados. Además, el padrino también influye: una casa comercial que financie un estudio de un producto suyo y le salga mal, encima no va a publicarlo para que todos sepamos que su producto no es útil.

En segundo lugar, como es lógico, los estudios publicados tienen más probabilidad de llegar a nuestras manos que los que no se publican en revistas científicas. Es el caso de tesis doctorales, comunicaciones a

congresos, informes de agencias gubernamentales o, incluso, estudios pendientes de publicar realizados por investigadores del tema que estemos tratando. Por este motivo es tan importante hacer una búsqueda que incluya este tipo de trabajos, que se engloban dentro del término de [literatura gris](#).

Por último, pueden enumerarse una serie de sesgos que influyen en la probabilidad de que un trabajo sea publicado o recuperado por el investigador que realiza la revisión sistemática tales como el [sesgo de lenguaje](#) (limitamos la búsqueda por idioma), el [sesgo de disponibilidad](#) (se incluyen solo los estudios que son fáciles de recuperar por parte del investigador), el [sesgo de coste](#) (se incluyen estudios que son gratis o baratos), el [sesgo de familiaridad](#) (solo se incluyen los de la disciplina del investigador), el [sesgo de duplicación](#) (los que tienen resultados significativos tienen más probabilidad de ser publicados más de una vez) y el [sesgo de citación](#) (los estudios con resultado significativo tienen más probabilidad de ser citados por otros autores).

Uno puede pensar que esto de perder trabajos durante la revisión no puede ser tan grave, ya que podría argumentarse, por ejemplo, que los estudios no publicados en revistas con revisión por pares suelen ser de peor calidad, por lo que no merecen ser incluidos en el metanálisis. Sin embargo, no está claro ni que las revistas científicas aseguren la calidad metodológica del trabajo ni que este sea el único método para hacerlo. Hay investigadores, como los de las agencias gubernamentales, que no están interesados en publicar en revistas científicas, sino en elaborar informes para quienes los encargan. Además, la revisión por pares no es garantía de calidad ya que, con demasiada frecuencia, ni el investigador que realiza el trabajo ni los encargados de revisarlo tienen una formación en metodología que asegure la calidad del producto final.

Todo esto puede verse empeorado por el hecho de que estos mismos factores pueden influir en los criterios de inclusión y exclusión de los estudios primarios del metanálisis, de tal forma que obtenemos una muestra de trabajos que puede no ser representativa del conocimiento global sobre el tema del que trate la revisión sistemática y el metanálisis.

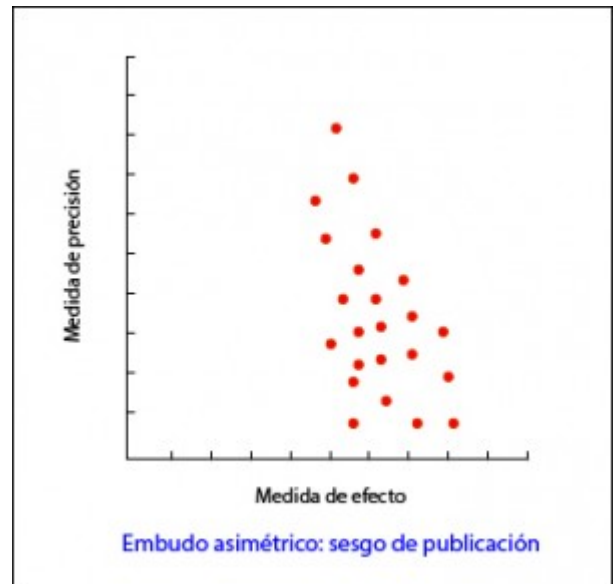
Si tenemos un sesgo de publicación la aplicabilidad de los resultados se verá seriamente comprometida. Por esto decimos que el sesgo de publicación es el verdadero talón de Aquiles del metanálisis.

Si delimitamos correctamente los criterios de inclusión y exclusión de los estudios y hacemos una búsqueda global y sin restricciones de la literatura habremos hecho todo lo posible para minimizar el riesgo de sesgo, pero nunca podremos estar seguros de haberlo evitado. Por eso se han ideado técnicas y herramientas para su detección.



La más usada tiene el simpático nombre de **gráfico en embudo**, aunque es más conocido por su nombre en inglés: *funnel plot*. En él se representa la magnitud del efecto medido (eje X) frente a una medida de precisión (eje Y), que suele ser el tamaño muestral, pero que puede también ser el inverso de la varianza o el error estándar. Representamos cada estudio primario con un punto y observamos la nube de puntos.

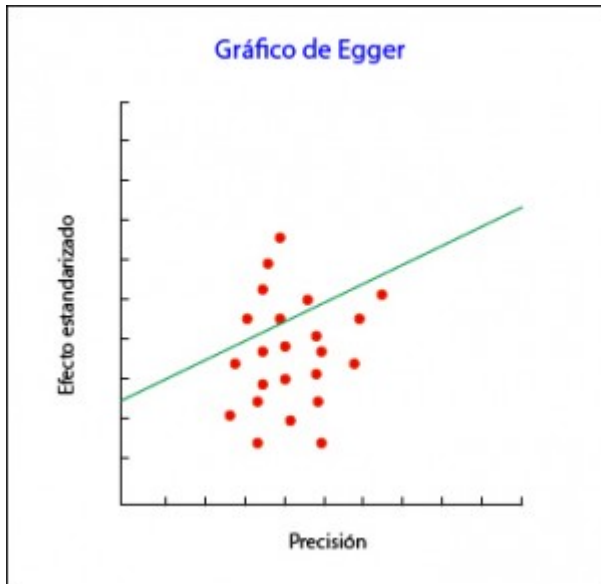
En la forma más habitual, con el tamaño de la muestra en el eje Y, la precisión de los resultados será mayor en los estudios de muestra más grande, con lo que los puntos estarán más juntos en la parte alta del eje y se irán dispersando al acercarnos al origen del eje Y. De esta forma, se observa una nube de puntos en forma de embudo, con la parte ancha hacia abajo. Este gráfico debe ser simétrico y, en caso de que no sea así, debemos sospechar siempre un sesgo de publicación. En el segundo ejemplo que os pongo podéis ver como "faltan" los estudios que están hacia el lado de falta de efecto: esto puede significar que solo se publican los estudios con resultado positivo.



Este método es muy sencillo de utilizar pero, en ocasiones, podemos tener dudas acerca de la asimetría de nuestro embudo, sobre todo si el número de estudios es pequeño. Además, el embudo puede ser asimétrico por defectos de la calidad de los estudios o porque estemos tratando con intervenciones cuyo efecto varíe según el tamaño de la muestra de cada estudio. Para estos casos se han ideado otros métodos más objetivos como la prueba de correlación de rangos de Begg y la prueba de regresión lineal

de Egger.

La [prueba de Begg](#) estudia la presencia de asociación entre las estimaciones de los efectos y sus varianzas. Si existe correlación entre ellos, mal asunto. El problema de esta prueba es que tiene poca potencia estadística, por lo que es poco de fiar cuando el número de estudios primarios es pequeño.

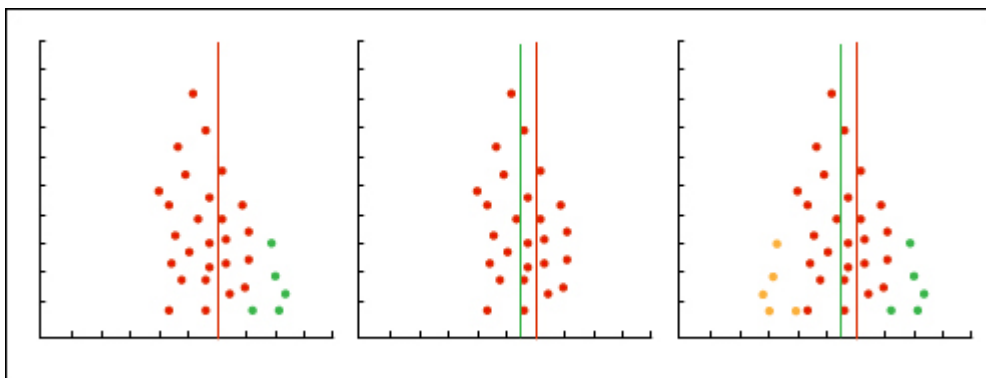


La [prueba de Egger](#), más específica que la de Begg, consiste en representar gráficamente la recta de regresión entre la precisión de los estudios (variable independiente) y el efecto estandarizado (variable dependiente). Esta regresión hay que ponderarla por el inverso de la varianza, así que no os recomiendo que la hagáis por vuestra cuenta, salvo que seáis estadísticos consumados. Cuando no hay sesgo de publicación la recta de regresión se origina en el cero del eje Y. Cuánto más se aleje del cero, mayor evidencia de sesgo de publicación.

Como siempre, existen programas informáticos que hacen estas pruebas con rapidez sin que tengamos que quemarnos el cerebro con sus cálculos.

¿Y si después de hacer el trabajo vemos que hay sesgo de publicación? ¿Podemos hacer algo para ajustarlo? Como siempre, podemos.

La forma más sencilla es utilizar un método gráfico que se llama de [ajuste y relleno](#) (*trim and fill* para los amantes del inglés). Consiste en lo siguiente: a) dibujamos el *funnel plot*; b) quitamos los estudios pequeños para que el embudo sea simétrico; c) se determina el nuevo centro del gráfico; d) se vuelven a poner los estudios quitados y añadimos su reflejo al otro lado de la línea central; e) reestimamos el efecto.



Otra actitud muy conservadora que podemos adoptar es dar por hecho que existe un sesgo de

publicación y preguntarnos cuánto afecta nuestros resultados, dando por hecho que nos hemos dejado estudios sin incluir en el análisis.

La única forma de saber si el sesgo de publicación afecta a nuestras estimaciones sería comparar el efecto en los estudios recuperados y en los no recuperados pero, claro está, entonces no tendríamos que preocuparnos por el sesgo de publicación.

Para saber si el resultado observado es robusto o, por el contrario, es susceptible de estar sesgado por un sesgo de publicación, se han ideado dos métodos de la N de seguridad, los conocidos en inglés como los métodos *fail-safe N*.

El primero es el método de la [N de seguridad de Rosenthal](#). Supongamos que tenemos un metanálisis con un efecto que es estadísticamente significativo, por ejemplo, un riesgo relativo mayor que uno con una $p < 0,05$ (o un intervalo de confianza del 95% que no incluye el valor nulo, el uno). Entonces nos hacemos una pregunta: ¿cuántos estudios con $RR = 1$ (valor nulo) tendremos que incluir hasta que la p no sea significativa? Si necesitamos pocos estudios (menos de 10) para hacer nulo el valor del efecto, podemos preocuparnos porque puede que el efecto sea nulo en realidad y nuestra significación sea producto de un sesgo de publicación. Por el contrario, si hacen falta muchos estudios, probablemente el efecto sea significativo de verdad. Este número de estudios es lo que significa la letra N del nombre del método.

El problema de este método es que se centra en la significación estadística y no en la importancia de los resultados. Lo correcto sería buscar cuántos estudios hacen falta para que el resultado pierda importancia clínica, no significación estadística. Además, asume que los efectos de los estudios faltantes es nulo (uno en caso de riesgos relativos y odds ratios, cero en casos de diferencias de medias), cuando el efecto de los estudios faltantes puede ir en sentido contrario que el efecto que detectamos o en el mismo sentido pero de menor magnitud.

Para evitar estos inconvenientes existe una variación de la fórmula anterior que valora la significación estadística y la importancia clínica. Con este método, que se denomina el de la [N de seguridad de Orwin](#), se calcula cuántos estudios hacen falta para llevar el valor del efecto a un valor específico, que será generalmente el menor efecto que sea clínicamente importante. Este método permite también especificar el efecto medio de los estudios faltantes.

Para terminar con el metanálisis, veamos cuál es la forma correcta de expresar los resultados del análisis de los datos. Para hacerlo bien, podemos seguir las recomendaciones de la declaración [PRISMA](#), que dedica siete de sus 27 ítems a darnos consejos de cómo presentar los resultados de

un metanálisis.

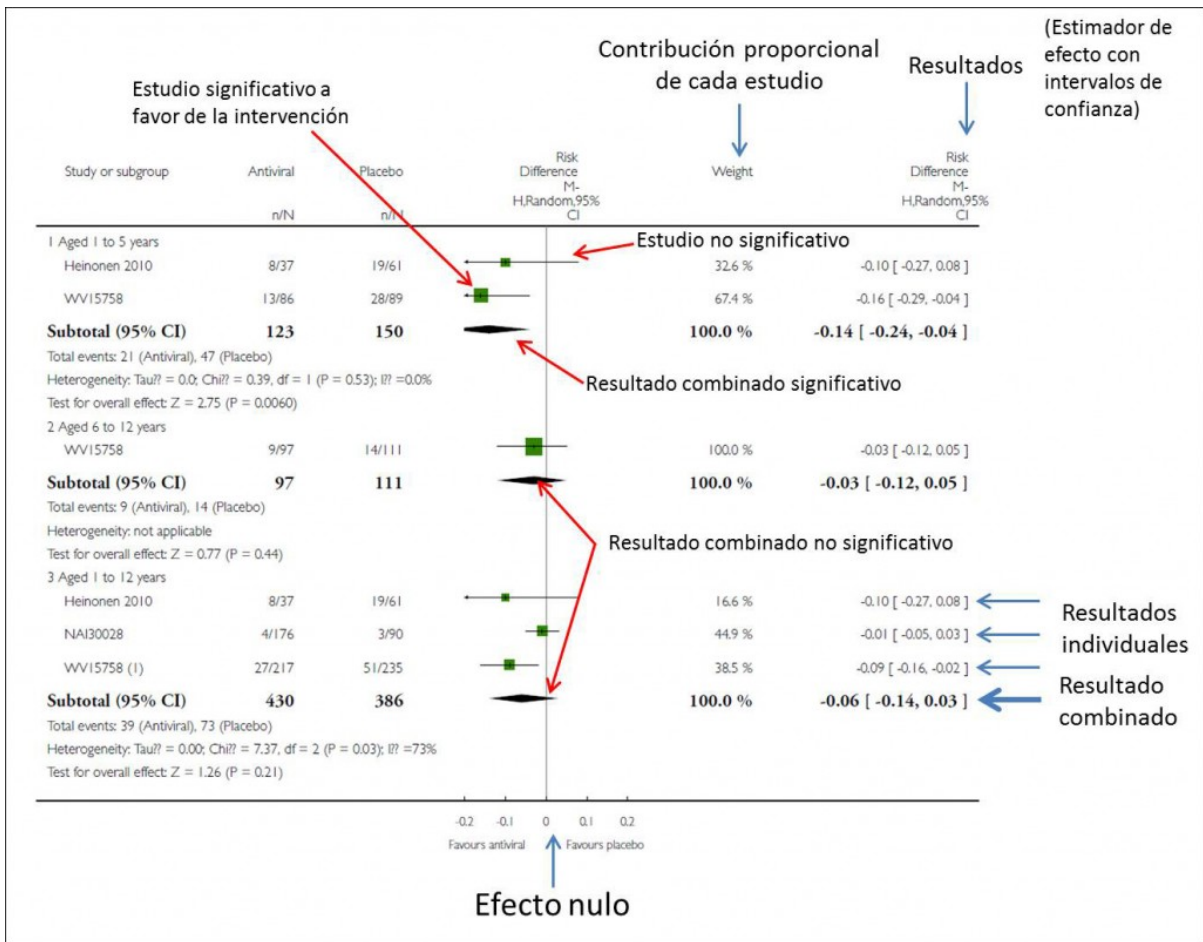
Primero debemos informar sobre el proceso de selección de estudios: cuántos hemos encontrado y evaluado, cuántos hemos seleccionado y cuántos rechazado, explicando además las razones para hacerlo. Para esto resulta muy útil el diagrama de flujo que debe incluir la revisión sistemática de la que procede el metanálisis si se acoge a la declaración PRISMA.

En segundo lugar deben especificarse las características de los estudios primarios, detallando qué datos sacamos de cada uno de ellos y sus correspondientes citas bibliográficas para facilitar que cualquier lector del trabajo pueda comprobar los datos si no se fía de nosotros. En este sentido va también el tercer apartado, que se refiere a la evaluación del riesgo de sesgos de los estudios y su validez interna.

Cuarto, debemos presentar los resultados de cada estudio individual con un dato resumen de cada grupo de intervención analizado junto con los estimadores calculados y sus intervalos de confianza. Estos datos nos servirán para confeccionar la información que PRISMA nos pide en su quinto punto referente a la presentación de resultados y no es otro que la síntesis de todos los estudios del metanálisis, sus intervalos de confianza, resultados del estudio de homogeneidad, etc.

Esto suele hacerse de forma gráfica mediante un diagrama de efectos, una herramienta gráfica popularmente más conocida por su nombre en inglés: el *forest plot*. Este gráfico es una especie de bosque donde los árboles serían los estudios primarios del metanálisis y donde se resumen todos los resultados relevantes de la síntesis cuantitativa.

La [Cochrane Collaboration](#) recomienda estructurar el *forest plot* en cinco columnas bien diferenciadas. En la columna 1 se listan los estudios primarios o los grupos o subgrupos de pacientes incluidos en el metanálisis. Habitualmente se representan por un identificador compuesto por el nombre del primer autor y la fecha de publicación.



La columna 2 nos muestra los resultados de las medidas de efecto de cada estudio tal como las refieren sus respectivos autores.

La columna 3 es el *forest plot* propiamente dicho, la parte gráfica del asunto. En él se representan las medidas de efecto de cada estudio a ambos lados de la línea de efecto nulo, que ya sabemos que es el cero para diferencias de media y el uno para odds ratios, riesgos relativos, *hazard ratios*, etc. Cada estudio se representa por un cuadrado cuya área suele ser proporcional a la contribución de cada uno al resultado global. Además, el cuadrado está dentro de un segmento que representa los extremos de su intervalo de confianza.

Estos intervalos de confianza nos informan sobre la precisión de los estudios y nos dicen cuáles son estadísticamente significativos: aquellos cuyo intervalo no cruza la línea de efecto nulo. De todas formas, no olvidéis que, aunque crucen la línea de efecto nulo y no sean estadísticamente significativos, los límites del intervalo pueden darnos mucha información sobre la importancia clínica de los resultados de cada estudio. Por último, en el fondo del gráfico encontraremos un diamante que representa el resultado global del metanálisis. Su posición respecto a la línea de efecto nulo nos informará sobre la significación estadística del resultado global, mientras que su anchura nos dará una idea de su precisión.

(su intervalo de confianza). Además, en la parte superior de esta columna encontraremos el tipo de medida de efecto, el modelo de análisis de datos que se ha utilizado (efectos fijos o efectos aleatorios) y el valor de significación de los intervalos de confianza (habitualmente 95%).

Suele completar este gráfico una cuarta columna con la estimación del peso de cada estudio en tantos por cien y una quinta columna con las estimaciones del efecto ponderado de cada uno. Y en algún rincón de todo este bosque estará la medida de heterogeneidad que se ha utilizado, junto con su significación estadística en los casos en que sea pertinente.

Para finalizar la exposición de los resultados, PRISMA recomienda un sexto apartado con la evaluación que se haya hecho de los riesgos de sesgo del estudio y un séptimo con todos los análisis adicionales que haya sido necesario realizar: estratificación, análisis de sensibilidad, metarregresión, etc.

Como veis, nada es fácil en esto de los metanálisis. Por eso, la Cochrane recomienda seguir una serie de pasos para interpretar correctamente los resultados. A saber:

1. Verificar qué variable se compara y cómo. Suele verse en la parte superior del *forest plot*.
2. Localizar la medida de efecto utilizada. Esto es lógico y necesario para saber interpretar los resultados. No es lo mismo una *hazard ratio* que una diferencia de medias o lo que sea que se haya usado.
3. Localizar el diamante, su posición y su amplitud. Conviene también fijarse en el valor numérico del estimador global y en su intervalo de confianza.
4. Comprobar que se ha estudiado la heterogeneidad. Esto puede verse a ojo mirando si los segmentos que representan los estudios primarios están o no muy dispersos y si se solapan o no. En cualquier caso, siempre habrá un estadístico que valore el grado de heterogeneidad. Si vemos que existe heterogeneidad, lo siguiente será buscar qué explicación dan los autores sobre su existencia.
5. Sacar nuestras conclusiones. Nos fijaremos en qué lado de la línea de efecto nulo están el efecto global y su intervalo de confianza. Ya sabéis que, aunque sea significativo, el límite inferior del intervalo conviene que esté lo más lejos posible de la línea, por aquello de la importancia clínica, que no siempre coincide con la significación estadística. Por último, volved a mirar el estudio de homogeneidad. Si hay mucha heterogeneidad los resultados no serán tan fiables.

Y con esto damos por finalizado el tema del metanálisis. En realidad, el *forest plot* no es exclusivo de los metanálisis y puede usarse siempre que queramos comparar estudios para dilucidar su significación estadística o clínica, o en casos como los estudios de equivalencia, en los que a la

línea de efecto nulo se le unen las de los umbrales de equivalencia. Pero aún tiene una utilidad más. Una variante del *forest plot* sirve también para valorar si existe sesgo de publicación en la revisión sistemática aunque, como ya sabemos, en estos casos le cambiamos el nombre por el de gráfico en embudo. Pero esa es otra historia...
