



Check for updates

ARTÍCULO DE REVISIÓN

Metaanálisis: bases conceptuales, análisis e interpretación estadística

Meta-analysis: Conceptual bases, statistical analysis and interpretation

Christian Fau* y Solange Nabzo

Fundación Oftalmológica 2020, Iberoamerican Cochrane Network, Santiago, Chile

Resumen

El término metaanálisis fue empleado por primera vez por G.V. Glass en 1976, en un artículo titulado Primaria, secundaria y metaanálisis de la investigación, para referirse al análisis estadístico del conjunto de resultados obtenidos en diferentes ensayos clínicos sobre una misma cuestión, con la finalidad de evaluarlos de manera conjunta. Este tipo de investigación fue inicialmente aplicado en las ciencias sociales y psicología, y en la década de 1980 el metaanálisis se hizo cada vez más popular en medicina, particularmente en los campos de enfermedad cardiovascular, oncología y perinatal. En la actualidad son muy frecuentes los artículos que describen resultados de metaanálisis en publicaciones médicas. Una de las razones para realizar un metaanálisis y combinar los resultados de los estudios es que al aumentar el tamaño de la muestra aumenta su potencia estadística, además, incorporar trabajos efectuados en centros y lugares diferentes conlleva que los resultados obtenidos pueden ser más fácilmente generalizables. Los metaanálisis no han estado ni están libres de controversia, entre otras cosas porque, quizás, como suele ser habitual con todo lo que se pone de moda, ha sido utilizado con una frecuencia excesiva sin suficiente rigor metodológico, ya que existen múltiples limitaciones a tener en cuenta a la hora de evaluar los resultados de un metaanálisis. En este artículo nos vamos a centrar fundamentalmente en los aspectos estadísticos del tema, desde el punto de vista de la descripción de los métodos, indicaciones e interpretaciones, sin entrar en otros detalles, como pueden ser los protocolos de realización de una revisión sistemática o fórmulas estadísticas. El enfoque va a ser en los métodos de análisis utilizados en metaanálisis de ensayos clínicos controlados que evalúan eficacia terapéutica o reacciones adversas.

Palabras clave: Revisión sistemática. Metaanálisis. MBE. Heterogeneidad. Sesgo de publicación. Medicina basada en la evidencia. Búsqueda de literatura. Análisis de sensibilidad. Análisis de subgrupos.

Abstract

The term meta-analysis was first used by G.V. Glass in 1976 in an article called "Primary, secondary and meta-analysis of investigations" ("Primaria, secundaria y meta-análisis de la investigación"). He used this term to refer to the statistical analysis of all the results obtained in different clinical studies regarding the same subject and that were to be analyzed together. At the beginning, this type of analysis was mainly used for the examination of social studies and psychology investigations, but later, during the 1980s, it became a popular method used in medicine; particularly in the cardiovascular, cancer and perinatal specialties. Nowadays it is not rare to find several medical articles using this method. For this study we decided to

Correspondencia:

*Christian Fau

Avda. Presidente Riesgo, 5157

Dep. 212, Las Condes, Santiago, Metropolitana, Chile

E-mail: chfauf@gmail.com

0187-4519/© 2020 Sociedad Mexicana de Oftalmología. Publicado por Permanyer. Este es un artículo open access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Fecha de recepción: 20-04-2020

Fecha de aceptación: 12-06-2020

DOI: 10.24875/RMO.M20000134

Disponible en internet: 09-11-2020

Rev Mex Oftalmol. 2020;94(6):260-273

www.rmo.com.mx

perform a meta-analysis and also combine the results of the studies because when the sample size increases, the statistical potential increases as well. Furthermore, when including studies and researches performed in different Health Centers, the results obtained can be easily generalized. Nevertheless, the meta-analysis method has its controversies, many of them due to an excessive use of the method together with a lack of methodological rigor; there are many limitations to be considered when evaluating the results of a meta-analysis. In this article we will focus mainly on the statistical aspects of the subject, from the point of view of the description of the methods, indications and interpretations, without specifying other details such as the protocols for carrying out a systematic review or statistical formulas. The focus will be on the analytical methods used in meta-analyses of controlled clinical trials evaluating therapeutic efficacy or adverse reactions.

Key words: Systematic Reviews. Meta-Analysis. EBM. Heterogeneity. Publication bias. Evidence-based medicine. Literature search. Sensitivity analyses. Subgroup analysis.

Introducción

El primer texto que combinó las evidencias de diferentes fuentes fue realizado en 1861 por el astrónomo inglés George Biddell Airy. En el área médica, el estadístico inglés Karl Pearson, en 1904, agrupó las estadísticas de instalaciones médicas y militares de Sudáfrica e India, y concluyó que la vacuna contra la fiebre intestinal no era eficaz, pero no fue hasta 1976 cuando el término *metaanálisis* fue empleado por primera vez por el psicólogo G.V. Glass, en un artículo titulado *Primaria, secundaria y metaanálisis de la investigación*. Él usó este término para referirse al análisis estadístico del conjunto de resultados obtenidos de diferentes ensayos clínicos con la finalidad de evaluarlos de manera conjunta. Este tipo de investigación fue inicialmente usada en las ciencias sociales y la psicología, y, en la década de 1980, el metaanálisis se empezó a hacer cada vez más popular en medicina, desarrollándose en 1989 el término *meta-analysis*, para indexar estos estudios en el Medical Subject Headings (MeSH) del Index Medicus.

Antiguamente, uno de los principales problemas en medicina eran las escasas investigaciones, además con cantidad insuficiente de pacientes, publicadas en un número limitado de revistas impresas y de acceso difícil. Existía, entonces, una falta de información. Esto llevó a que las razones iniciales para realizar un metaanálisis fueran combinar los resultados de los pocos estudios que existían, y con ello aumentar el tamaño de la muestra y su potencia estadística. Hoy en día, con el desarrollo de los sistemas digitales y el volumen de información biomédica, que está en una expansión exponencial a nivel mundial, en muchas áreas se ha convertido simplemente en imposible el poder leer, valorar críticamente y sintetizar el estado de los conocimientos actuales, y mucho menos mantener una actualización de forma regular. Actualmente, el problema ya no es la falta de información, sino un exceso

de esta. El Dr. Sackett, padre de la medicina basada en la evidencia, en una conferencia en diciembre de 1997 expresaba: «Un médico que pretenda diagnosticar y tratar a su paciente con aquello que verdaderamente ha demostrado su valor necesita leer todos los días del año 19 artículos científicos; algo que es claramente una utopía». Por lo tanto, las revisiones se han convertido en herramientas esenciales para poder mantenerse al día, sin embargo, inicialmente, esto era suplido por las revisiones narrativas, desde Mulrow en 1987, y Oxman y Guyatt en 1988, se llamó la atención a nivel mundial sobre la mala calidad de estas, muchas de ellas incluso con una clara intencionalidad de inducir a conductas específicas, direccionadas por empresas farmacéuticas y otros actores del mercado, por lo tanto, surgen las revisiones sistemáticas y los metaanálisis como una forma de estructurar sistemáticamente la información, para, así, poder reducir los sesgos. Pero los metaanálisis no han estado ni estarán libres de controversia, entre otras cosas porque, quizás, como suele ser habitual con todo lo que se pone de moda, han sido utilizados de forma excesiva sin el suficiente rigor metodológico. Esto es una contradicción con lo que es precisamente su finalidad. Los metaanálisis presentan múltiples limitaciones a tener en cuenta a la hora de evaluar los resultados, por otro lado, hay quienes reverencian los resultados numéricos sin analizar críticamente los procedimientos con los que estos se obtienen, por lo tanto, hoy, uno debe tener siempre presente la nemotecnia GIGO (*garbage-in, garbage-out*) al evaluar la información de un metaanálisis¹⁻⁶.

En este artículo nos vamos a centrar fundamentalmente en los aspectos estadísticos del tema, desde el punto de vista de la descripción de los métodos, indicaciones e interpretaciones, sin entrar en otros detalles, como pueden ser los protocolos de realización de una revisión sistemática o fórmulas estadísticas, que, si bien son de una gran importancia, han sido

previamente analizados en múltiples publicaciones. Para profundizar en la realización de la protocolización de una revisión sistemática, se sugiere consultar el manual Cochrane de revisiones sistemáticas y la declaración PRISMA⁷⁻¹⁰, ambos artículos son de libre acceso en Internet. El enfoque del artículo va a ser sobre los métodos de análisis utilizados en un metaanálisis de ensayos clínicos controlados que evalúan eficacia o reacciones adversas (terapéutica). No se analizarán en este artículo los métodos más complejos que se utilizan para un metaanálisis de pruebas diagnósticas, de análisis de sobrevida o *network* metaanálisis¹¹, por lo que el objetivo del estudio es analizar los distintos métodos estadístico y gráficos que se pueden utilizar en la realización de un metaanálisis de terapéutica, y que permiten una mejor comprensión e interpretación de los resultados. Para facilitar la comprensión, los ejemplos serán realizados utilizando los datos de la publicación Cochrane Patching for corneal abrasion (Review)¹², que corresponde a un metaanálisis de 11 ensayos clínicos controlados y aleatorizados que evalúan el efecto del uso del parche ocular en la reparación o cicatrización de las erosiones superficiales corneales, menores de 10 mm². De este análisis solo se tomaron los datos de las tasas de reparación corneal con parche y sin parche al primer día de seguimiento.

¿Qué es un metaanálisis?

Una revisión sistemática se define como un tipo de investigación secundaria, investigación sobre lo investigado, que ha sido preparada con un enfoque sistemático para minimizar los sesgos y errores aleatorios, en la cual la unidad de análisis son los estudios originales primarios y no los pacientes. Una revisión sistemática, que es cualitativa, puede o no incluir un metaanálisis que es un análisis cuantitativo de los resultados de diferentes estudios independientes con un diseño similar, que generalmente tiene como objetivo producir una estimación ponderada del efecto, para lo cual se usan métodos estadísticos que permiten combinar los resultados de dos o más estudios. Esto, además, aumenta la precisión en la estimación del efecto. Cuando en diferentes estudios se analiza un mismo problema, es previsible encontrar por puro azar diferencias entre los resultados, que pueden maximizarse si se trata de estudios de pequeño tamaño, ya que la estimación es menos precisa. El objetivo de utilizar un metaanálisis para combinar los resultados de los estudios es que, al aumentar el

tamaño de la muestra, aumenta su potencia estadística, además, incorporar trabajos efectuados en centros y lugares diferentes lleva que los resultados obtenidos pueden ser más fácilmente generalizables. Al obtener una muestra mayor logramos disminuir la probabilidad de cometer un error de tipo II. Este error es la probabilidad de no detectar una diferencia de una determinada magnitud entre los grupos cuando esta sí existe o, lo que es lo mismo, la probabilidad de concluir en nuestro estudio que no existen diferencias estadísticamente significativas cuando en realidad sí las hay. Incluso cuando hay efectos de magnitud moderada o pequeña pero clínicamente importantes, estos podrían pasar inadvertidos en los estudios primarios y ser apreciados más claramente en una medida ponderada.

Los metaanálisis constituyen una herramienta esencial para sintetizar la información científica disponible, que permite la estimación cuantitativa de la relación entre dos variables, estimando una magnitud de efecto ponderado y aumentando la precisión de los estimadores. En su proceso permiten valorar la calidad de la metodología empleada, valorar la consistencia entre las investigaciones, obtener una mejor estimación global e identificar subgrupos con efectos diferentes. Esto posibilita incrementar la validez de las conclusiones de los estudios individuales e identificar áreas de incertidumbre donde sea necesario realizar una nueva investigación.

Sujetos, materiales y métodos

Las fases de un metaanálisis bien realizado son muy similares a las de una investigación primaria, con la salvedad de que, ahora, la unidad de observación en lugar de ser los pacientes la constituyen los estudios. Lo primero es definir un buen protocolo de investigación que describa claramente el motivo para realizar el metaanálisis, los objetivos que se persiguen y la hipótesis que se investiga. Primero se realiza la búsqueda sistemática y exhaustiva de todos los artículos potencialmente relevantes, y luego la selección de los artículos que serán incluidos en la revisión mediante criterios explícitos de inclusión y exclusión. Finalmente, habrá que proceder a resumir los datos a partir de los diferentes trabajos de forma bien estructurada, que evite errores y facilite su posterior análisis. Se realizará una descripción del diseño de los estudios originales, la síntesis de los datos obtenidos y la interpretación de los resultados. Todo este proceso científico utilizado debe quedar plasmado en la sección de *materiales y métodos* del documento que se publique⁷.

En 1999 se publicó la Declaración QUORUM (*Quality Of Reporting Of Meta-analysis*)⁸, con el objetivo de mejorar la calidad de las revisiones sistemáticas y los metaanálisis. Esta ofrece las primeras normas de publicación de dichos trabajos, las cuales se han ido actualizando hasta llegar al 2010, momento en que se publica la Declaración PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analysis*)⁹⁻¹⁰. En esta, a diferencia de la anterior, se detalla y justifica cada ítem que debe tenerse en cuenta para la elaboración de estos estudios. Para profundizar en la elaboración de una revisión sistemática se sugiere revisar estas normas internacionales y el manual Cochrane de revisiones sistemáticas⁷⁻⁹. Todos estos artículos son de acceso libre en Internet.

Análisis

Existen múltiples limitaciones que hay que tener en cuenta a la hora de evaluar los resultados de un metaanálisis, por tal motivo vamos a dividir el análisis en cuatro puntos, donde discutiremos los problemas y las posibles soluciones estadísticas:

1. La calidad de los estudios que se incluyen.
2. El sesgo propio de publicación.
3. El sesgo de elección de los artículos que se incluirán en el metaanálisis.
4. La variabilidad o heterogeneidad en la estimación de los estudios.

La calidad de los estudios

Considere la posibilidad de una serie de ensayos clínicos de alta calidad metodológica que examinaran la misma intervención en una población de pacientes. Cada ensayo proporcionará una estimación sin sesgo del mismo efecto del tratamiento subyacente, la variabilidad que se observa entre los ensayos con confianza puede ser atribuida a la variación aleatoria, y el metaanálisis debería proporcionar una estimación sin sesgo del tratamiento, con un aumento en la precisión de esta estimación. Una situación fundamentalmente diferente surge en el caso de los estudios epidemiológicos, estudios de casos y controles, estudios transversales y estudios de cohorte. En estos, debido a los efectos de la confusión y los sesgos, se pueden generar estimaciones que muestran asociaciones que se desvían del efecto en forma sistemática, y así proporcionar una estimación falsamente precisa. El examen a fondo de la heterogeneidad entre los resultados de los estudios observacionales, en particular de las posibles fuentes

de confusión y sesgo, proporcionará más ideas acerca del tema en revisión que el cálculo mecánico de una medida de efecto ponderado, para lo cual puede ser utilizado un análisis de sensibilidad que vamos a revisar más adelante. La diferencia fundamental que hay entre los estudios observacionales y los ensayos clínicos controlados aleatorizados no significa que estos últimos sean inmunes al sesgo. El sesgo también se puede introducir si la calidad metodológica de los ensayos clínicos controlados es insuficiente. La calidad de los estudios puede determinarse mediante técnicas de análisis de medicina basada en la evidencia y pautas publicadas para tales fines^{1-4,7}. Actualmente, las guías más frecuentemente utilizadas para evaluar la calidad de la evidencia de los artículos son las de GRADE (*Grading of Recommendations Assessment, Development and Evaluation* - <https://www.gradeworKinggroup.org>)

El sesgo de publicación

Los metaanálisis han permitido hacer tangible el sesgo de publicación en la interpretación de la evidencia. El sesgo de publicación alude a que los artículos científicos publicados no reflejan fielmente la totalidad de los ensayos clínicos realizados. El no incluir todos los estudios es una de las principales críticas que han recibido los metaanálisis, y que podemos resumir señalando que los trabajos con resultados positivos estadísticamente significativos tienen mayor probabilidad de ser enviados a publicación, de ser aceptados para su publicación y de ser publicados en revistas con mayor impacto que los que no arrojan diferencias, y una vez aceptados por las revistas tardan menos tiempo en ser publicados. También son citados con más frecuencia, lo que aumenta considerablemente la probabilidad de que aparezcan en una base de datos al realizar la búsqueda bibliográfica. Asimismo, los ensayos con gran número de pacientes tienen mayor probabilidad de ser publicados, aunque sean de tipo negativo, es decir, que no hayan encontrado diferencias entre el grupo de estudio y el grupo control. Por otro lado, muchas veces, los ensayos clínicos que dan resultado no significativo o en sentido contrario a la hipótesis planteada no llegan a hacerse públicos, no solo porque las revistas no los publican, sino también porque los autores no los envían a publicación, o debido a que gran parte de la investigación médica es financiada por las empresas farmacéuticas, que inevitablemente serán reacias a publicar trabajos en los que

una nueva terapia no se revela como más eficaz que la actual.

También puede ocurrir que un mismo ensayo haya sido publicado más de una vez, en forma aparentemente diferente. La publicación duplicada también puede introducir un sesgo. Hay artículos que se han publicado hasta 5 veces con los mismos pacientes pero con títulos diferentes. Algunos autores realizan esto con el fin de aumentar su currículum. Esto sesga los resultados al incluir más de una vez los registros de un mismo artículo. Por lo que no cabe duda, e incluso se han hecho estudios que lo confirman, que hay un sesgo favorecedor de determinados tipos de publicaciones, un importante sesgo en la selección de idiomas de publicación y un sesgo en el número de citas, lo que conlleva a que la probabilidad de que determinados resultados intervengan o no en un metaanálisis esté sesgada. Eso sí, hay que tener claro que el sesgo principalmente es en relación con la falta de artículos de pequeño tamaño muestral de efecto menor, nulo u opuesto al resultado del análisis.

Un método para evaluar este tipo de sesgo es utilizar un gráfico de *funnel plot* o gráfico de embudo propuesto por Vandenbroucke en 1988 (Fig. 1). El gráfico en embudo se basa en el hecho de que la precisión en la estimación del efecto del tratamiento subyacente aumentará a medida que el tamaño de la muestra de los estudios aumente: en el eje Y se registra el error estándar del logaritmo natural de la medida del efecto del riesgo relativo (RR) o razón de momios (OR) (o el tamaño de la muestra) y en el eje de las X, el logaritmo natural de la magnitud del efecto del RR u OR. Los resultados de los estudios pequeños se dispersan ampliamente como una nube de puntos en la parte inferior de la gráfica en torno al estimador central, generando un estrechamiento progresivo hacia el vértice del gráfico a medida que los estudios presentan un mayor tamaño muestral. En ausencia de sesgo, la trama se parecerá a un embudo invertido simétrico. Si existe sesgo de publicación de los estudios más pequeños, solo se encontrarán los artículos que muestran el beneficio del tratamiento explorado, lo que hace asimétrica la distribución de los puntos, por lo tanto, si hay sesgo, el gráfico de embudo será a menudo desigual y asimétrico.

Un análisis simple de los gráficos en embudo proporciona una prueba útil para la probable presencia de sesgo de publicación en un metaanálisis, pero esta capacidad de detectar el sesgo será limitada cuando los metaanálisis se basan en un número pequeño de ensayos clínicos, ya que la nube de puntos será

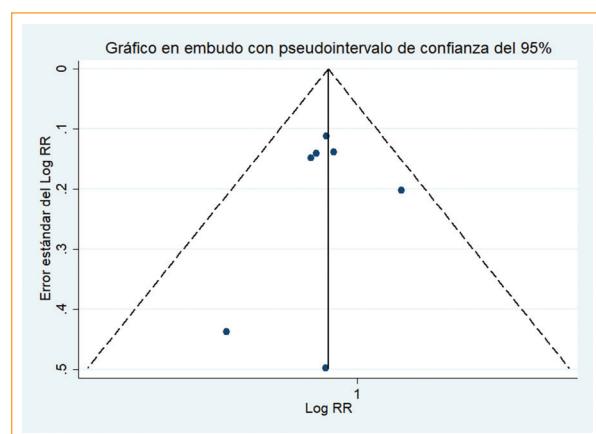


Figura 1. Se muestra un gráfico de funnel plot del riesgo relativo (RR) que evalúa el efecto del uso del parche ocular en la reparación o cicatrización al primer día de las erosiones superficiales corneales de menos de 10 mm². La gráfica impresiona como simétrica en torno a la línea negra, que es el Log del RR ponderado de todos los estudios. Pudieran faltar algunos estudios pequeños en torno al Log RR entre 0 y 0.5. Las líneas negras segmentadas de los bordes corresponden a un pseudointervalo de confianza $\pm 1.96^*$ error estándar, la diferencia está representada como el logaritmo del RR = 0.

pequeña, por lo tanto, los resultados deben analizarse con cautela. Es más, incluso si hay un sesgo de publicación, este puede no dar lugar a un gráfico en embudo asimétrico, por ejemplo, cuando la trama de la nube de puntos se ahueca.

Un gráfico en embudo no es un método muy fiable para investigar el sesgo de publicación, aunque sí nos da una idea de si los resultados del estudio son espaciados simétricamente en torno a un efecto central más preciso. Un gráfico de embudo asimétrico puede deberse a un sesgo de publicación, pero también puede resultar de la heterogeneidad clínica entre los estudios (por ejemplo, diferentes tasas de eventos) o heterogeneidad metodológica entre los estudios (por ejemplo, fallo para ocultar la asignación), por lo tanto, más que sesgo de publicación, esto tiende a considerarse como si el resultado del metaanálisis es influido o no por el efecto de estudios de pequeño tamaño muestral.

La simetría o la asimetría generalmente se ha definido de manera informal, a través de un examen visual en estos gráficos. Como es de esperar, los gráficos en embudo han sido interpretados de manera diferente por diferentes observadores, por lo tanto, se ha medido la asimetría del gráfico en embudo numéricamente

mediante una simple prueba de asimetría, prediciendo la discordancia de los resultados cuando los metaanálisis se comparan con los ensayos clínicos grandes. Esto ha generado algunas pruebas estadísticas para detectar la asimetría del gráfico en embudo, por lo tanto, es aconsejable complementar siempre el estudio de sesgo gráfico con pruebas estadísticas más objetivas, como pueden ser la prueba de Egger (1997) y la prueba de Begg y Mazumdar (1994). Estas contrastan la hipótesis nula de ausencia de sesgo de publicación (Fig. 2). La prueba de Begg y Mazumdar comprueba si existe una correlación entre la magnitud estandarizada del efecto y su variabilidad. Si a mayor efecto estandarizado existe una mayor varianza, significa que los estudios pequeños (de varianza grande) se publican o fueron seleccionados cuando el efecto es diferente del efecto de los estudios grandes (de varianza pequeña). La correlación se calcula con el coeficiente de correlación ordinal tau de Kendall, y la ausencia de significación estadística sugiere que no hay sesgo de publicación. La prueba de Egger¹³ es más específica que la de Begg y puede dar falsos positivos. El método consiste en ajustar una recta de regresión a los puntos con abscisa igual a la precisión, que es el inverso del error estándar $1/ES$ (variable independiente), y ordenada igual al efecto estandarizado $\ln RR/ES$ (variable dependiente). Si no hay sesgo de publicación, la ordenada en el origen será igual a cero, y cuanto más alejada de cero esté, más pronunciada será la asimetría y mayor la evidencia de sesgo (Fig. 2). Cuando la prueba de Begg y Mazumdar, que es una prueba de correlación de rango, no arroja un resultado significativo ($p > 0.10$), se aconseja precaución al interpretarla, ya que tiene muy poca potencia cuando el número de estudios es pequeño, como ocurre frecuentemente.

Macaskill, et al. (2001) proponen otra medida de asimetría. Esta se calcula ajustando un modelo de regresión lineal entre el efecto ponderado ($\ln OR$ o RR) y el tamaño de la muestra. En el año 2006, Peters, et al. comprobaron que este método funciona mejor si la variable independiente es el recíproco del tamaño de la muestra ($1/n$). La interpretación es la misma que con las otras pruebas, pero este método es menos sensible que el de Egger para detectar un sesgo de publicación, pero es más específico. El método de Begg y Mazumdar es un intermedio entre ambos.

Otra herramienta para poder evaluar el riesgo de sesgo es el *Trim and Fill*, método desarrollado por Duval y Tweedie (2000). Es un método no paramétrico de cuantificación del riesgo de sesgo de publicación en el metaanálisis. El método estima el número y los

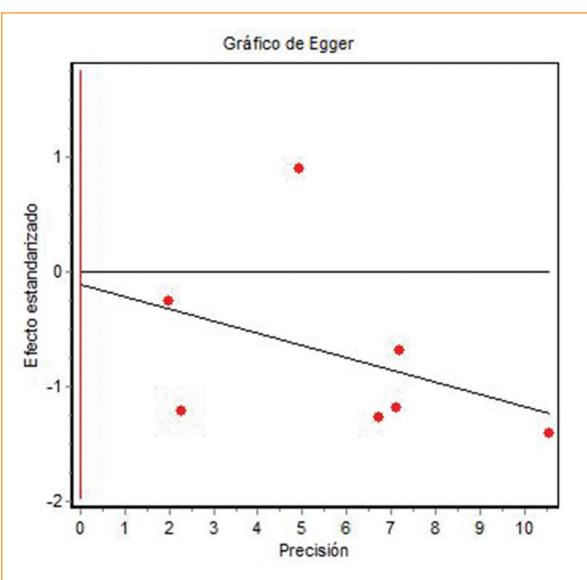


Figura 2. Muestra la dispersión de puntos de los diferentes estudios según la precisión de estos. La línea negra corresponde a la regresión de la prueba de Egger, la cual llega a 0, por lo que no hay evidencia de sesgo de estudios de pequeño tamaño.

resultados de los estudios que podrían faltar y ajusta el metaanálisis para incorporar los datos imputados desaparecidos. Los autores afirman que el método es eficaz y coherente con otras técnicas de ajuste. Este proporciona un gráfico en embudo con los datos marcados, en este caso con un cuadrado, a diferencia del círculo de los artículos o con el mismo símbolo, pero de distinto color, además permite calcular la medida ponderada global (OR o RR) para el modelo de efecto fijo o aleatorio, con y sin los estudios calculados como perdidos para evaluar cuánto se modifican los resultados.

En el ejemplo antes planteado en el gráfico de la figura 1, la prueba de Begg fue $p = 0.7639$ y la prueba de Egger fue $p = 0.8785$. No se estimó que faltaran estudios en este análisis mediante el método de *Trim and Fill*.

El sesgo de elección o selección

El sesgo de selección constituye otra de las limitaciones del metaanálisis, y es muy importante definir con claridad los criterios de inclusión y exclusión de los artículos siendo lo más objetivos posible. El sesgo de selección surgirá al incluir en el metaanálisis solo algunos de los artículos de investigación efectuados, como consecuencia de no haber llevado a cabo una exhaustiva búsqueda bibliográfica o haberla limitado

tendenciosamente o al haber eliminado inadecuadamente artículos por falta de validez. El propio autor del metaanálisis es quien sesga los resultados, ya que su criterio para incluir o excluir un estudio del análisis puede venir influenciado. Tanto en la incorporación como en la exclusión de los artículos hay que ser muy cuidadosos, lo ideal es que dos evaluadores independientes decidan y, en caso de haber una discrepancia, un tercer evaluador dirima el problema. Este tipo de sesgo solo puede ser manejado por el diseño del estudio y no por la estadística del análisis de los datos. Pautas como PRISMA, STROBE, CONCORT o el Manual Cochrane de Revisiones Sistemáticas de Intervenciones pueden ayudar a minimizar estos sesgos metodológicos^{1-4,7-10}.

Estimación del efecto y análisis de homogeneidad

Al ser uno de los objetivos del metaanálisis unir la información aportada por los diferentes trabajos, un asunto crucial es determinar una medida estandarizada del efecto que permita precisamente esa combinación. Antes de realizar el metaanálisis, el investigador ha de establecer las medidas que se van a utilizar para describir y representar los efectos, y así poderlos agrregar. Obviamente, estas medidas van a depender tanto del tipo de respuesta (binaria o continua) como del diseño del estudio (experimental u observacional) utilizados en los estudios primarios:

- Respuesta binaria: este tipo de respuesta solamente presenta dos categorías de resultado (muerte o supervivencia, enfermos o sanos, éxito o fracaso, mejora la visión o no mejora, presión intraocular controlada o sin control, progresión del glaucoma o sin progresión, etc.). En estudios con este tipo de respuesta, las medidas de efecto más comúnmente utilizadas son la tasa de riesgos (riesgo relativo) y la razón de momios (*odds ratio*).
- Respuesta continua: en la literatura médica existen multitud de estudios en los cuales la variable respuesta está medida en escala continua (presión intraocular, recuento de células endoteliales, grosor de retina mediante tomografía por coherencia óptica, paquimetría, etc.). Si las respuestas de todos los estudios están medidas en la misma escala, entonces, la medida del efecto o tamaño del efecto será simplemente la diferencia entre las medias de los grupos. Si no fuese así, es decir, si los estudios miden la respuesta en diferente escala, habrá que considerar en su lugar una diferencia estandarizada de medias (**Tabla 1**).

Para presentar esta medida de efecto ponderado se construye un gráfico, que en inglés se conoce como *forest plot*, en el cual, en el eje de las abscisas (eje X) se representa la medida del efecto (OR o RR), y a lo largo del eje de las ordenadas (eje Y) se sitúan los diferentes estudios, generalmente ordenados por el año de publicación o cualquier otro criterio de ordenación (**Fig. 3**). Para cada estudio y para la estimación global del efecto, se representa su estimación puntual y también el intervalo de confianza (IC), habitualmente del 95%, asumiendo un modelo de efectos fijos y/o efectos aleatorios. En la **figura 3**, en la zona C se muestra la parte gráfica y en la zona D, los valores numéricos. También se traza una línea vertical, que se ubica en el punto en el que no existe diferencia entre los eventos estudiados, que corresponde a la línea continua negra trazada en el $RR = 1$, la línea segmentada roja corresponde al valor ponderado del RR. A los lados del estudio se coloca el nombre del primer autor y el año del estudio y a veces se identifica el número de eventos existentes para cada situación estudiada (zonas A y B).

El problema de este tipo de gráfico radica en que los estudios con peor precisión, los que tienen menor peso en el cálculo del efecto global, son los que presentan un mayor impacto visual, ya que su IC es el más amplio. Para remediar esto, de alguna forma, se acude a representar el valor de su efecto medio con un símbolo cuya área sea proporcional a la precisión, de tal manera que los símbolos más grandes y, por tanto, más llamativos, corresponden a los estudios más precisos y con mayor peso en la estimación del efecto ponderado (cuadrado gris sobre el estudio en zona C). Por último, se coloca el valor de cuánto pesa cada estudio en la estimación ponderada (zona E). En la parte inferior del gráfico se coloca la medida del efecto ponderado, en este caso el RR con su IC del 95%, y se grafica mediante un diamante horizontal, donde los extremos corresponden al IC del 95%, además se colocan las pruebas estadísticas de homogeneidad que se verán más adelante.

Es crucial entender las limitaciones de los metaanálisis y la importancia de explorar fuentes de heterogeneidad y sesgo que permitan asegurar que las conclusiones son válidas. Cuando hablamos de heterogeneidad podemos distinguir dos aspectos, por un lado, las diferencias existentes entre los estudios en cuanto a las características de los pacientes incluidos, la metodología utilizada, el tiempo de seguimiento, las dosis empleadas, la localización geográfica, etc.; y, por otro lado, el concepto de heterogeneidad estadística,

Tabla 1. Resumen de los principales métodos estadísticos utilizados según tipo de datos y modelo

Tipo de variable			
Tipo de datos	Medida estadística	Modelo	Método
Dicotómico	Razón de momios (OR)	Efecto fijo	Peto
			Mantel-Haenszel
			Wolf
	Riesgo relativo (RR)	Efectos aleatorios	DerSimonian and Laird
		Efecto fijo	Mantel-Haenszel
			Wolf
Continuo	Diferencia de riesgo	Efectos aleatorios	DerSimonian and Laird
		Efecto fijo	Mantel-Haenszel
	Diferencia de medias		Wolf
		Efecto fijo	Wolf
	Diferencia de medias estandarizadas	Efectos aleatorios	DerSimonian and Laird
		Efecto fijo	Wolf
		Efectos aleatorios	DerSimonian and Laird

que únicamente cuantifica la variabilidad entre los resultados de los estudios, y que puede deberse a las diferencias reales de diseño y ejecución entre los estudios incluidos u otras causas. Uno de los aspectos de la heterogeneidad, el relativo a la homogeneidad clínica, es un problema metodológico, ya que habrá que decidir si las diferencias entre los estudios, que siempre existen, permiten o no combinarlos, independientemente de los resultados que en ellos se hayan obtenido y, por lo tanto, es previo a la ejecución del metaanálisis^{2,3}. Los estudios deben ser compatibles para poder ser combinados y deben tener por lo menos tres aspectos similares: los tratamientos que se emplean, los pacientes que se incluyen y las variables de resultado utilizadas. La heterogeneidad estadística trata de cuantificar la variabilidad del resultado medido con respecto al resultado del efecto ponderado, y determinar si dicha variabilidad es superior a la que sería esperable por puro azar. La situación ideal para la aplicación del metaanálisis en cuanto a la determinación de un efecto ponderado se da cuando los estudios que se combinan son ensayos clínicos aleatorizados y sus resultados son homogéneos, aunque el tamaño de la muestra original fuera insuficiente para obtener resultados concluyentes. Cuando hay estudios discrepantes, la investigación de

la presencia y causas de la heterogeneidad en los resultados puede pasar a convertirse en el principal objetivo del metaanálisis y desplazar al objetivo de obtener un efecto ponderado.

Se han desarrollado dos modelos estadísticos para realizar inferencias sobre el efecto ponderado de un grupo de estudios: el modelo de efectos fijos y el modelo de efectos aleatorios. El modelo de efectos fijos se conoce como *método ponderado por la inversa de la varianza (inverse variance weighted method)*, y fue descrito en 1930 por Birge y también por Cochran. Asume que el tamaño del efecto es una constante fija y desconocida que debemos estimar, da por supuesto que los estudios individuales son homogéneos y provienen de una misma población, con lo cual las diferencias entre las estimaciones del tamaño del efecto únicamente se deben a la variabilidad intraestudios (error de muestreo), habitualmente se calcula con el método de Mantel-Haenszel (M-H). El modelo de efecto aleatorio considera que el tamaño del efecto proviene de una distribución de parámetros con una media y una varianza interestudios. En este caso, las diferencias en las estimaciones del tamaño del efecto se deben a la variabilidad intraestudios y a la variabilidad interestudios, que puede ser estimada por diferentes

métodos, los más frecuentes son el de DerSimonian-Laird (D+L) y la máxima verosimilitud. Las diferencias entre ambos métodos son mínimas, ya que el primero se basa en el método de los momentos y el segundo en la función de máxima verosimilitud, ambos métodos suelen dar estimadores muy precisos y similares, siendo el primero el más usado (**Tabla 1**).

El modelo de efectos aleatorios, al tener en cuenta la variabilidad no solo intraestudios sino también interestudios, nos permite hacer una inferencia que va más allá del grupo de estudios con el que estamos trabajando. En cambio, el modelo de efectos fijos solo nos permite llegar a inferencias condicionadas a la muestra de los estudios.

El modelo de efectos aleatorios no asume un único tamaño del efecto paramétrico, sino una distribución normal de parámetros del tamaño del efecto, por lo que incluye una varianza interestudios, además de la intraestudios, para las estimaciones del tamaño del efecto de cada estudio. La varianza paramétrica de los efectos aleatorios (T^2) es una medida de la heterogeneidad interestudios. Desde este punto de vista, el modelo de efectos fijos es un caso especial del modelo de efectos aleatorios donde $T^2 = 0$. Para el modelo de efectos aleatorios, el tamaño del efecto medio es también una medida ponderada por la inversa de la varianza, solo que en este caso la varianza tiene dos componentes, varianza intrasujetos e intersujetos. Esto quiere decir que, en presencia de heterogeneidad, debería hacerse uso del modelo de efectos aleatorios. El modelo de efectos aleatorios, al tener en cuenta más variabilidad, produce IC más amplios que en el modelo de efectos fijos y, además, aumenta la importancia de los estudios pequeños, en los que es más frecuente el sesgo de publicación, a diferencia de los efectos fijos, donde los estudios de mayor tamaño muestral tendrán más peso.

Los modelos de efectos aleatorios deben utilizarse siempre que se determine, tras un análisis de heterogeneidad^{14,15}, que los estudios no son homogéneos y no puedan identificarse claramente las causas, lo que suele ser habitual, especialmente en estudios epidemiológicos. Para probar la heterogeneidad estadística normalmente se usa un estadístico definido por Cochran en 1954. Este se basa en calcular la suma de las desviaciones cuadráticas entre el resultado individual de cada estudio y el resultado global, ponderadas por el mismo peso con el que cada resultado interviene en el cálculo global: $Q_w = \sum w_i(d_i - \bar{d})^2$. La hipótesis nula de homogeneidad es $H_0: T^2 = 0$. Si la prueba Q_w nos da un valor significativo, rechazamos la hipótesis nula de homogeneidad entre los estudios y aplicamos el

modelo de efectos aleatorios. Para ver si Q_w es significativa, se contrasta frente a una χ^2 con $n-1$ grados de libertad. El problema es que la prueba Q_w adolece de baja potencia, por lo que en muchos casos estamos asumiendo homogeneidad y aplicando efectos fijos cuando en realidad deberíamos aplicar efectos aleatorios. Si el número de estudios es pequeño, su capacidad para detectar heterogeneidad es muy baja (poca potencia de contraste), mientras que, por el contrario, cuando el metaanálisis combina un gran número de estudios, el resultado puede ser estadísticamente significativo, incluso cuando la magnitud de la heterogeneidad no sea de relevancia clínica. Esto no es nada nuevo, son los problemas inherentes a la metodología de las pruebas de contraste estadístico. Puesto que la situación más habitual en los metaanálisis es la de un reducido número de estudios, se recomienda utilizar un nivel de rechazo en esta prueba al menos de $p > 0.1$, en lugar del tradicional $p > 0.05$.

Las limitaciones de la prueba Q_w anteriormente comentadas, así como el hecho de que su resultado no sirva para comparar diferentes metaanálisis en los que intervienen diferentes números de estudios, ha llevado a buscar otros índices que permitan cuantificar la heterogeneidad de manera que el parámetro calculado sirva para comparar diferentes metaanálisis, y así, de alguna manera, ponderar la validez del efecto medio calculado en ellos. En esa línea se ha propuesto un índice denominado I^2 , partiendo de la idea de que la pregunta de interés no es si existe o no heterogeneidad estadísticamente significativa, sino cómo afecta esta a las conclusiones del metaanálisis. Los autores trataron de buscar un índice que permita la comparación entre diferentes metaanálisis y, por lo tanto, que no dependa ni de la unidad de medida utilizada para cuantificar el efecto, ni del número de estudios incluidos, y que, además, sea fácilmente interpretable por personas no muy expertas en estadística. El parámetro I^2 indica la proporción de la variación entre estudios respecto de la variación total, es decir, la proporción de la variación total que es atribuible a la heterogeneidad, y se expresa en porcentajes (%). Los autores que lo desarrollaron proponen los valores de 25, 50 y 75% como posibles marcas para indicar una heterogeneidad baja, moderada o alta, respectivamente. Una vez más hay que resaltar que la ausencia de heterogeneidad estadística no garantiza que los estudios sean homogéneos en cuanto a sus características clínicas, biológicas o metodológicas, y, por lo tanto, no garantiza la validez de proceder a combinarlos (**Fig. 3**).

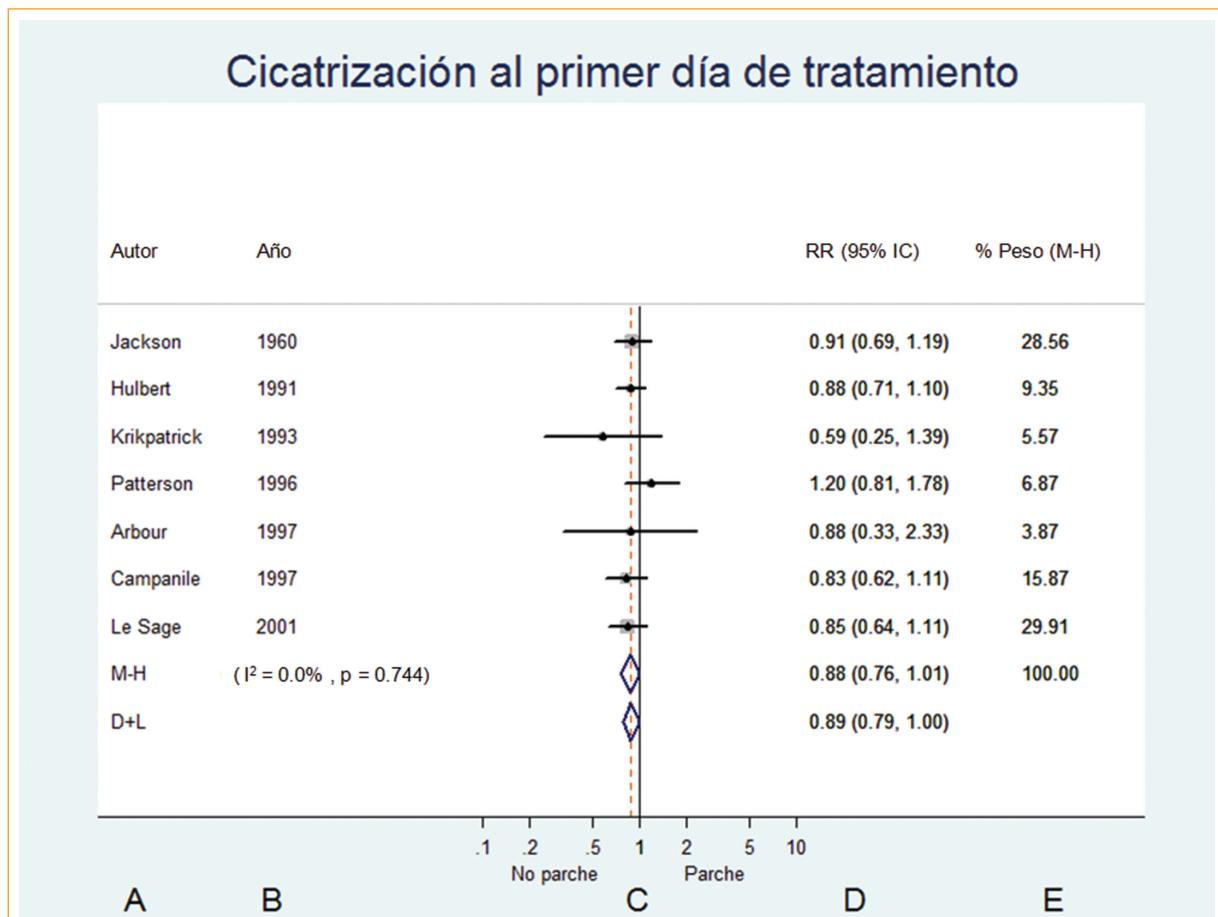


Figura 3. Se aprecia cómo cambia el valor del riesgo relativo (RR) en la zona de abajo de 0.88 en efecto fijo (Mantel-Haenszel [M-H]) a 0.89 en efecto aleatorio (DerSimonian and Laird [D+L]). Se vuelve más conservadora la estimación, luego, al lado de cada estimación se aprecia como varía el intervalo de confianza (IC) del 95%, siendo más amplio, por lo tanto, menos preciso, el de efecto aleatorio que el de efecto fijo. Entre paréntesis, al lado de M-H se presenta en este caso el cálculo de $I^2 = 0,0\%$ y la $p = 0.744$ del χ^2 del análisis de M-H, lo cual establece que no hay heterogeneidad estadística en este caso.

Debido a la poca potencia de las pruebas estadísticas existentes, estas pueden complementarse con algunas representaciones gráficas que permitan la inspección visual de la magnitud de la variabilidad entre los estudios. Las representaciones más utilizadas son el gráfico de Galbraith (Fig. 4), recomendado por ser aplicable a cualquier tipo de estudio (observacional y experimental), y el gráfico de L'Abbé (Fig. 5), más restrictivo, aplicable únicamente a metaanálisis de ensayos clínicos. El gráfico de Galbraith representa la precisión de cada estudio (el inverso del error estándar) frente al efecto estandarizado; asimismo, representa la línea de regresión ajustada a estos puntos y una banda de confianza. Los estudios que caen fuera de esta banda son los que más contribuyen a la heterogeneidad. Por otra parte, la posición de los estudios

en el eje de abscisas permite identificar visualmente aquellos que tienen un mayor peso en el metaanálisis. Este gráfico también se puede utilizar para detectar fuentes de heterogeneidad al etiquetar los estudios por diferentes variables, tales como el año de publicación.

Otro enfoque para determinar la presencia de heterogeneidad entre los estudios es tratar de localizar la posible causa de esta. Una pregunta que resulta natural en este contexto es si existe relación entre la tasa de respuesta y el riesgo base. Una buena herramienta gráfica para verificarlo la constituye el denominado gráfico de L'Abbé (Fig. 5), en el que para cada estudio se representa, en el eje de las Y, la frecuencia de eventos en el grupo de tratamiento frente a la frecuencia de eventos en el grupo de control, en el eje de las

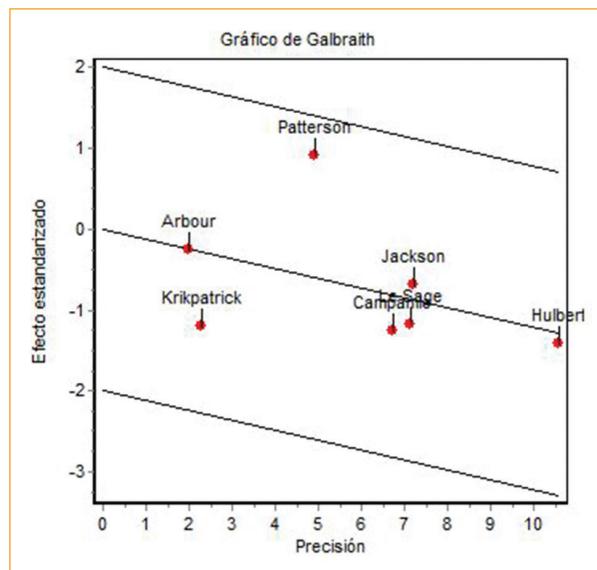


Figura 4. El gráfico de Galbraith representa la precisión de cada estudio frente al efecto estandarizado. Se aprecia como, en este caso, todos los estudios se encuentran entre las bandas de confianza y la mayoría de los estudios en torno a la línea de regresión ajustada, por lo tanto, no hay mucha heterogeneidad entre los estudios.

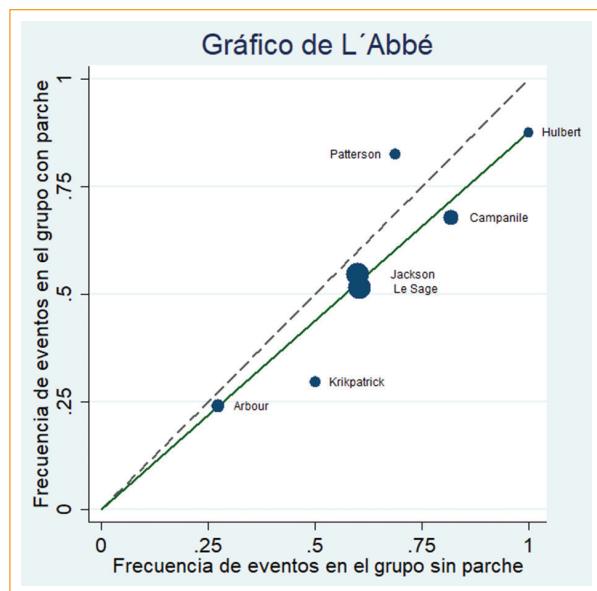


Figura 5. El gráfico de L'Abbé permite visualmente y de forma muy fácil evaluar los estudios que favorecen uno u otro tratamiento en relación con la tasa de eventos en cada grupo.

X. Este gráfico solo se presenta cuando la variable respuesta es binaria. Se traza una línea a 45° que divide el gráfico en dos secciones, a un lado de esta

línea quedarán los estudios que fueron favorables al grupo de tratamiento y al otro lado los estudios que fueron favorables al grupo de control, además, se traza una línea segmentada con el valor de la medida del efecto ponderado, en este caso el RR = 0.88. Se suelen representar los puntos de tamaño variable con un área inversamente proporcional a la varianza de la estimación del efecto en cada estudio (directamente proporcional a la precisión de la estimación), de tal manera que, cuanto más grande es un punto, mayor precisión tiene el estudio que representa. La presencia de puntos dispersos en el gráfico de L'Abbé indica heterogeneidad, mientras que, si se agrupan en una zona estrecha, es señal de resultados más homogéneos. La presencia de patrones en este gráfico puede orientar sobre lo que está ocurriendo. Si la eficacia depende del riesgo base, podría ocurrir que, para riesgos base bajos, los puntos estén muy próximos a la línea divisoria o incluso del lado de la no eficacia, y a medida que nos vamos moviendo hacia riesgos base mayores, los puntos se vayan alejando de esa línea hacia valores de mayor eficacia.

Podemos llevar más lejos la idea de que alguna característica subyacente en los estudios puede estar afectando a la heterogeneidad de los resultados, no solo el riesgo base, sino cualquier otra característica. Esto puede ser utilizado para explorar fuentes de heterogeneidad y probar hipótesis nuevas que no fueron planteadas en estudios individuales, mediante el uso de las técnicas de metarregresión. El término *metarregresión* describe el análisis en el que las características de los estudios o de los sujetos incluidos en cada estudio se usan como variables explicativas en un modelo de regresión multivariante, en el que la variable dependiente es el efecto medido o alguna medida de la desviación del efecto de cada estudio respecto del efecto ponderado. Aquí, la unidad de observación es cada estudio o cada subgrupo de los estudios, y a la hora de estimar los coeficientes de la ecuación de regresión se emplean métodos que ponderan la intervención de cada estudio en esa estimación, de acuerdo a su precisión (proporcional al tamaño del estudio, inversamente proporcional a la variabilidad del estudio). Puesto que la muestra es pequeña (número de estudios), el número de variables independientes que intervienen en la ecuación de regresión debe ser pequeño. Al igual que en el análisis de regresión tradicional, se usa la regla de 5 o 10 casos por cada variable, y dado que el número de estudios en un metaanálisis suele ser pequeño, no se deben utilizar muchas variables en la ecuación de regresión. Un ejemplo típico de

metarregresión, ya clásico, es el que presenta el *In odds ratio* de la eficacia de la vacuna BCG (*Bacillus Calmette Guérin*) contra la tuberculosis frente a la latitud del país en el que se efectuó el estudio. En el estudio que se realizó al respecto, se veía que probablemente el riesgo de ineficacia de la vacuna tiende a aumentar cuando nos acercamos al ecuador del planeta.

Por último, en caso de que se sospeche que existen motivos que puedan explicar la heterogeneidad de los resultados entre los estudios, la opción más recomendable es realizar un análisis de subgrupos que solo combine los estudios que cumplen determinada condición o característica, de modo que estos sean más homogéneos. La idea del análisis de subgrupos o estratificación consiste en efectuar el metaanálisis en diferentes grupos de estudios, reunidos según sus características (estudios publicados frente a no publicados, estudios de gran tamaño frente a pequeño tamaño, estudios según nivel de calidad, según fecha de publicación, etc.) o según las características de los pacientes incluidos, y comprobar cómo varía el resultado al estimarlo de nuevo en cada ocasión.

El análisis de sensibilidad pretende estudiar la influencia de cada uno de los estudios en la estimación del efecto ponderado, y, por lo tanto, la robustez o estabilidad de la medida final obtenida. Este análisis consiste en la repetición del metaanálisis tantas veces como la cantidad de estudios, de forma que cada vez que se calcula el efecto ponderado se omite un estudio y se utilizan todos los restantes. Si los resultados de los distintos metaanálisis son similares, y, por lo tanto, el efecto tiene una misma dirección, magnitud y significación estadística, se puede concluir que los resultados son robustos. En caso contrario no se tendría un estimador robusto, lo cual exigiría cierta precaución en la interpretación de los resultados o podría ser motivo para generar nuevas hipótesis. Una apreciación a tener en cuenta es que, cuando se retiran estudios de buena calidad y precisos, la magnitud del resultado final se amplía, lo que permite identificar este tipo de estudios. Se puede representar gráficamente mediante un *forest plot*, denominado *gráfico de influencias*, en el cual se grafica el efecto ponderado para cada análisis, excluyendo siempre un artículo. El análisis de sensibilidad también puede utilizarse para estudiar la influencia en los resultados del metaanálisis de ciertos aspectos relacionados con la validez de los estudios, como por ejemplo, la exclusión de estudios que no superan un determinado umbral de calidad; trabajos no publicados; o bien estudios en los que hay incertidumbre sobre el cumplimiento de los criterios de selección.

Además, se puede calcular, entonces, el número de estudios de efecto cero, realizados y no publicados, que deberían existir para modificar el sentido de una eventual conclusión «positiva» obtenida con un metaanálisis. Esto se denomina *Fail-Safe N*, y se utilizan los métodos de Rosenthal (1979) y Orwin (1983). Si este número es muy elevado, se considera que la probabilidad de que el sesgo de publicación haya modificado sustancialmente los resultados es baja, y se acepta la existencia de las diferencias sugeridas por el metaanálisis.

Otro método de análisis importante es el metaanálisis acumulado. Este se define como aquel proceso en el cual se lleva a cabo un nuevo metaanálisis cada vez que aparece un estudio nuevo publicado (Fig. 6). Este tipo de metaanálisis no requiere de técnicas estadísticas especiales para combinar los estudios. La combinación estadística de los resultados de los estudios se realizará tantas veces como estudios haya, menos uno. De modo que inicialmente se combinan los dos primeros y, en los metaanálisis sucesivos, se van incorporando, uno a uno, cada nuevo estudio. Aunque en el metaanálisis acumulado el orden cronológico suele ser el más utilizado (el año de publicación), en ocasiones resulta interesante realizar de forma independiente o complementaria otros tipos de agregación de estudios, basados, por ejemplo, en el orden creciente por tamaños muestrales; diferencias entre tratamiento y control; puntuaciones de calidad del estudio, etc. Un metaanálisis acumulado también permite estudiar de forma retrospectiva el momento o año de publicación en el cual el efecto del tratamiento alcanzó por primera vez los niveles convencionales de significación estadística, cuando un nuevo tratamiento supera al control. Esta forma de presentación de los resultados pone de manifiesto lo difícil que es para cualquier estudio individual, una vez que se han alcanzado resultados relativamente estables, aportar información adicional. De hecho, un estudio tiene que ser suficientemente grande y obtener resultados suficientemente extremos como para modificar sustancialmente los resultados acumulados.

Finalmente comentaremos las discrepancias que se plantean en la literatura entre los grandes ensayos clínicos y los metaanálisis, LeLorier, et al., en su artículo, revisaron las discrepancias entre 12 ensayos clínicos de gran tamaño, de más de 1.000 pacientes y 19 metaanálisis que intentaban responder a las mismas preguntas de investigación¹⁶. El grado de acuerdo fue moderado entre ambos tipos de investigaciones, los metaanálisis presentaron valores predictivos positivos del 68% y valores predictivos negativos del 67%, con una sensibilidad del 65% y una especificidad del 70%. Aunque parte de

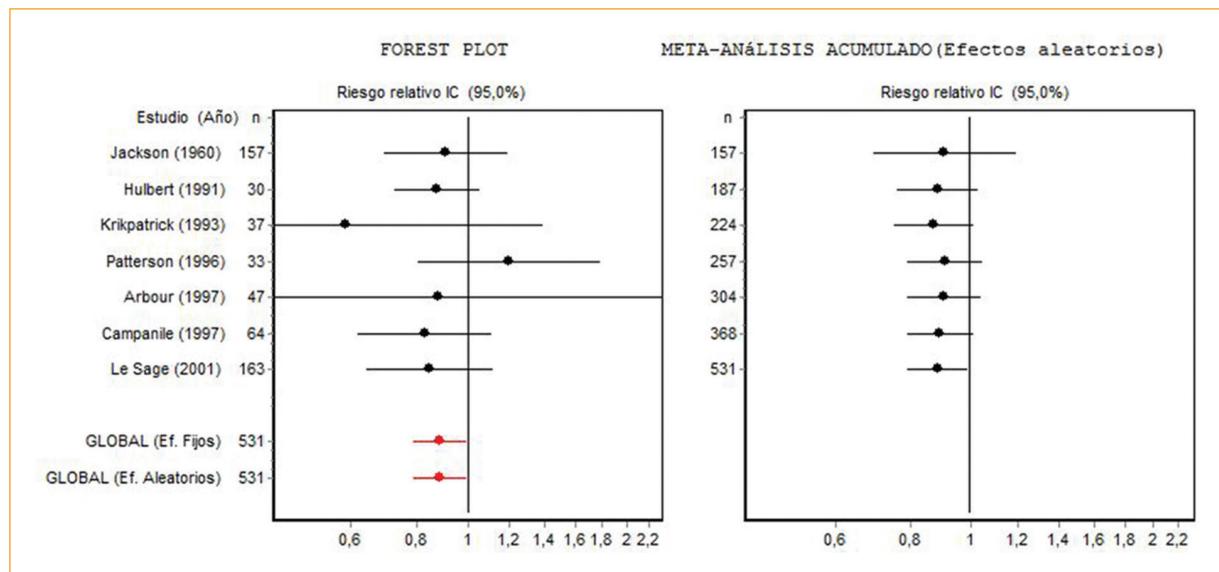


Figura 6. A la izquierda está el metaanálisis normal ordenado según el año de publicación. A la derecha se encuentran un análisis acumulado, donde se ve cómo cambia la estimación y el intervalo de confianza (IC) del 95% según se van agregando los diferentes estudios, y como en los últimos años casi no hay modificaciones en estos valores con la realización de nuevos estudios.

las discrepancias parecen deberse a una inadecuada consideración de la heterogeneidad de los estudios originales en los metaanálisis, los ensayos clínicos de gran tamaño y los metaanálisis desempeñan papeles complementarios en la evaluación de la evidencia clínica, diferenciados por la heterogeneidad de los pacientes involucrados, ya sean estos de mega ensayos clínicos o metaanálisis de ensayos clínicos de menor tamaño.

Análisis final del ejemplo

En relación con los ejemplos utilizados en el artículo de los datos extraídos de la revisión Cochrane que evalúa el uso del parche ocular, vamos a hacer un resumen del análisis completo secuencialmente. Primero, la variable utilizada es dicotómica, repara o no la córnea al primer día de ser o no ser parchado el ojo con un daño corneal menor de 10 mm^2 . Esto, según la tabla 1, nos obliga a decidir utilizar RR u OR mediante un modelo de efecto fijo o aleatorio. En este caso se utilizó el RR mediante un modelo de efecto fijo con el método de M-H. Esto es producto del tipo de estudios (ensayo clínico controlado) más el hecho de que al análisis de heterogeneidad, parte inferior izquierda de la figura 3, tanto el $I^2 = 0.0\%$ como la $p = 0.744$ del χ^2 del análisis de heterogeneidad mediante M-H mostraron muy baja heterogeneidad entre los artículos y la no existencia de diferencias significativas, respectivamente. De esto se

concluye que los datos analizados pueden ser metaanalizados, pero, ¿están todos los artículos o están sesgados los datos? Para responder esto vamos a la figura 1. En el funnel plot, los artículos se distribuyen más o menos homogéneamente excepto en la parte derecha inferior, donde podrían faltar algunos artículos, por lo que realizamos un gráfico de Egger (Fig. 2), donde la regresión tiende al 0, además calculamos el valor p de la prueba de Egger, el cual no fue significativo ($p = 0.8785$), por lo cual no parecen los resultados ser afectados por estudios de pequeño tamaño o sesgo de publicación. Para confirmar aún más esto, utilizamos el método de *Trim and Fill*, el cual demostró que no faltaban estudios en el análisis. Si todo lo anterior lo unimos a un protocolo de búsqueda y selección de artículos adecuado, entonces podemos estar tranquilos que los resultados pueden ser correctamente interpretados. Finalmente, los resultados los tomamos de la figura 3, los cuales son un $RR = 0.88$ con un IC del 95% de 0.76 a 1.01. Con esto, junto con la tendencia que muestra el metaanálisis acumulado de la figura 6, se puede concluir que el uso de un parche ocular en los pacientes con erosiones superficiales corneales de menos 10 mm^2 no reporta ningún beneficio en relación con la tasa de reparación al primer día. Es más, en el mejor de los escenarios, según el IC de 95%, la tasa de reparación sería igual a no parchar, por lo tanto, la conducta médica según este análisis sería finalmente no parchar a este

tipo de pacientes, ya que es muy improbable que un nuevo ensayo clínico tenga resultados tan rotundos en favor del parche para que al ser incorporado al análisis no solo desvíe el RR sobre el valor 1, sino que además desvíe completamente el IC del 95% sobre el valor 1, que sería lo único que podría generar un cambio de conducta.

Programas

Existen muchos programas que pueden ser utilizados para realizar un metaanálisis, la gran mayoría de ellos son para realizar un análisis frecuentista. Dentro de los dos mejores gratuitos están el RevMan (Review Manager) desarrollado por the Cochrane Collaboration, el cual además ayuda en el manejo del manuscrito y la bibliografía del estudio, y el paquete Metafor del programa R. Dentro de los programas comerciales que requieren de pago, uno de los mejores es el Stata, ya que tiene excelentes gráficos y permite actualizarse con las diferentes pruebas que se van desarrollando con el paso de los años, los ejemplos fueron desarrollados principalmente con este programa.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Confidencialidad de los datos. Los autores declaran que en este artículo no aparecen datos de pacientes.

Derecho a la privacidad y consentimiento informado. Los autores declaran que en este artículo no aparecen datos de pacientes.

Bibliografía

1. Oxman A, Cook D, Guyatt G. The Evidence Based Medicine Working Group. *JAMA*. 1994;272:1367-71.
2. Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. *BMJ*. 1995;310:1085-6.
3. Sackett D, Straus S, Richardson W, Rosenberg W, Haynes R. Evidence - Based Medicine. How to Practice and Teach EBM. 2nd ed. Edinburgh: Churchill Livingstone, 2000.
4. Guyatt G, Rennie D. Users' Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice. Chicago: AMA Press, 2002.
5. Egger M, Smith G, Altman D. Systematic Reviews in Health Care. Meta-analysis in context. London: BMJ; 2001.
6. Smith M, Glass G. Meta-analysis of psychotherapy outcome studies. *American Psychologist*. 1977;32:752-60.
7. Centro Cochrane Iberoamericano, traductores. Manual Cochrane de Revisiones Sistémáticas de Intervenciones, versión 5.1.0 [Internet]. Barcelona: Centro Cochrane Iberoamericano; 2012. Disponible en: https://es.cochrane.org/sites/es.cochrane.org/files/public/uploads/manual_cochrane_510_web.pdf
8. Moher D, Cook D, Eastwood S, Olkin I, Rennie D, Stroup D; for the QUOROM group. Improving the quality of reporting of meta-analysis of randomized controlled trials: The QUOROM statement. *Lancet*. 1999;354:1896-900.
9. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gotzsche P, Ioannidis J, et al., and the PRISMA Group. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med*. 2009;6: e1000100. Epub 2009 Jul 21.
10. Urrutia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Med Clin (Barc)*. 2010;135: 507-11.
11. Fau C, Nabzo S, Nasabun V. Metaanálisis en red. *Rev Mex Oftalmol*. 2018;92(3):153-9. Disponible en: https://www.rmo.com.mx/frame_esp.php?id=47.
12. Turner A, Rabiu M. Patching for corneal abrasion. *Cochrane Database of Systematic Reviews* 2006, Issue 2. Art. No.:CD004764. DOI: 10.1002/14651858.CD004764.pub2.
13. Egger M, Davey S, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-34.
14. Higgins J, Thompson S. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002; 21(11):1539-58.
15. Higgins J, Thompson S, Deeks J, Altman D. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-60.
16. LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337(8):536-42.