



Artículo especial

# Metaanálisis de ensayos clínicos aleatorizados, heterogeneidad e intervalos de predicción

## Meta-analysis of randomized trials, heterogeneity and prediction intervals

Ferrán Catalá-López<sup>a,b,c,\*</sup> y Aurelio Tobías<sup>d</sup><sup>a</sup> División de Farmacoepidemiología y Farmacovigilancia, Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), Madrid, España<sup>b</sup> Fundación Instituto de Investigación en Servicios de Salud, Valencia, España<sup>c</sup> Centro Superior de Investigación en Salud Pública (CSISP), Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Valencia, España<sup>d</sup> Instituto de Diagnóstico Ambiental y Estudios del Agua (IDAEA), Centro Superior de Investigaciones Científicas (CSIC), Barcelona, España

### INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 2 de mayo de 2013

Aceptado el 20 de junio de 2013

On-line el 7 de septiembre de 2013

### Introducción

Durante décadas, las revisiones sistemáticas y los metaanálisis se han consolidado como una herramienta metodológica que ofrece información con un elevado nivel de calidad y rigor científico<sup>1,2</sup>. La revisión sistemática integra toda la información empírica disponible sobre un tema de interés o pregunta de investigación, basándose en criterios predefinidos, explícitos y reproducibles, permitiendo reducir los sesgos generalmente presentes en las tradicionales revisiones narrativas. El metaanálisis es la técnica estadística que permite combinar resultados sobre los efectos procedentes de diversos estudios individuales identificados y valorados críticamente a través de una revisión sistemática previa<sup>3-6</sup>. Las principales ventajas de los metaanálisis son que permiten obtener una medida del efecto combinada con una mayor precisión que la de los estudios individuales incluidos en la revisión sistemática y, por lo tanto, una mayor potencia estadística, cuantificar la variabilidad de los resultados de los estudios y aproximarse a preguntas no planteadas en los análisis individuales explicando las diferencias entre sus resultados. Sin embargo, también presentan limitaciones, particularmente si no se utilizan los métodos estadísticos adecuados y si no interpretan cuidadosamente los resultados obtenidos cuando existen indicios de discrepancias entre los resultados de los estudios individuales. Para la obtención de estimadores combinados, los resultados de los estudios individuales son ponderados en función de su precisión a través de diversas técnicas y modelos. La elección del modelo empleado depende fundamentalmente de la consideración de la

hipótesis de homogeneidad de efectos entre los resultados de los estudios individuales incluidos en la revisión sistemática. Esto es importante, pues la interpretación de los resultados y su posterior aplicación a la toma de decisiones clínicas se verá influida directamente por ello. Del mismo modo, es importante conocer la consistencia de los resultados de un metaanálisis y si las conclusiones obtenidas son estables ante diferentes premisas.

El objetivo del artículo es introducir los aspectos metodológicos más importantes relacionados con los modelos estadísticos empleados habitualmente en el metaanálisis, y su relación con los conceptos de heterogeneidad e intervalo de predicción, aproximándose a su interpretación a través de diversos ejemplos.

### Modelo de efectos fijos frente a modelo de efectos aleatorios<sup>7,8</sup>

En la figura 1 se utiliza un escenario hipotético en el que el objetivo es estimar el efecto de un tratamiento sobre la reducción de la mortalidad cardiovascular a partir de 8 ensayos clínicos para un mismo tratamiento. La representación gráfica de los resultados del metaanálisis se hace en un *forest plot*, en el que en el eje de abscisas (eje X) se representa la medida del efecto del tratamiento y en el eje de coordenadas (eje Y) se sitúan los diferentes estudios. Para cada estudio se presenta el efecto del tratamiento de forma numérica y gráfica junto a su correspondiente intervalo de confianza. Cada estudio presenta el efecto del tratamiento hipotético en términos del riesgo relativo. También se presenta la estimación del efecto combinado resultado del metaanálisis. Los intervalos de confianza permiten comprobar si el efecto combinado es estadísticamente significativo, es decir, si, por ejemplo, el tratamiento presenta resultados estadísticamente diferentes (y supuestamente más beneficiosos) respecto a los del grupo de control o placebo. Para ello, se comprueba si el efecto neutro

\* Autor para correspondencia.

Correo electrónico: ferran\_catala@hotmail.com (F. Catalá-López).

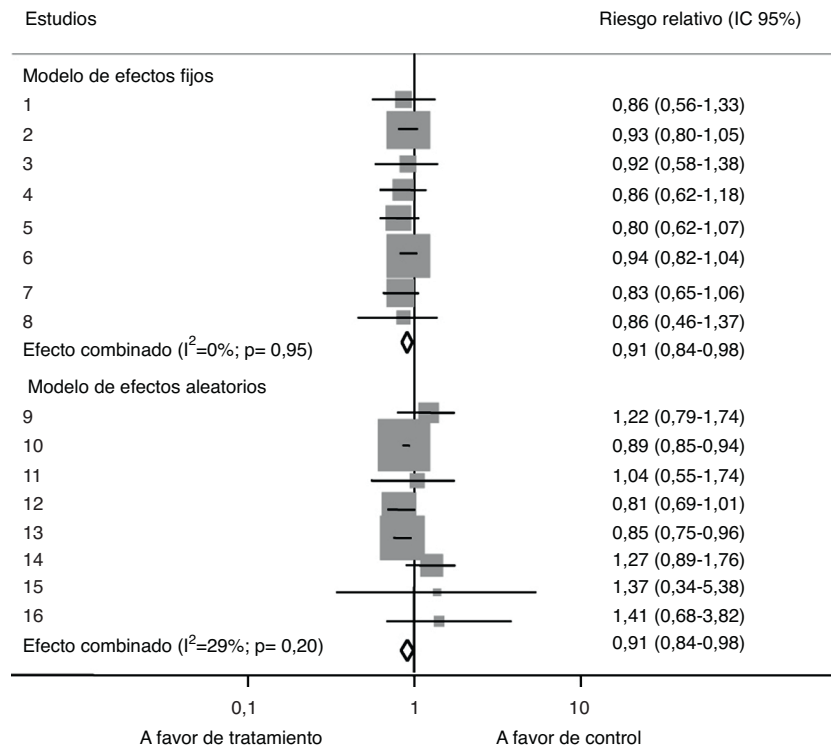


Figura 1. Ejemplo hipotético de metaanálisis.

(un riesgo relativo igual a 1) está incluido dentro del intervalo de confianza. Si dicho valor está dentro del intervalo, se concluye que el metaanálisis no ha sido capaz de demostrar que el tratamiento presenta un resultado estadísticamente diferente respecto al del control. En el ejemplo, los 2 metaanálisis estiman el mismo efecto combinado, una reducción de la mortalidad cardiovascular del 8% (riesgo relativo = 0,91; intervalo de confianza del 95% [IC 95%] 0,84-0,98), si bien utilizando modelos estadísticos distintos sobre distintas muestras de estudios.

En el modelo de efectos fijos se asume que los estudios incluidos en el metaanálisis están estimando a un único, y común, efecto poblacional, por lo que la única variabilidad presente es debida al azar, conocida como variabilidad intraestudio (al considerar el tamaño del estudio y su propia varianza). En la figura 1 del modelo de efectos fijos, el índice  $I^2$  cuantifica el porcentaje de la variabilidad en las estimaciones del efecto del tratamiento que se debe a la variabilidad entre estudios, que en este caso es del 0%.

El modelo de efectos aleatorios se fundamenta en que los estudios estiman una distribución de efectos en la población, asumiendo que los efectos subyacentes siguen una distribución normal con una media y varianza fijas. En otras palabras, el efecto individual de cada estudio estima a un efecto poblacional diferente. El modelo de efectos aleatorios considera, por lo tanto, 2 fuentes de variabilidad, la intraestudio (al igual que el modelo de efectos fijos) y la interestudios (que es la desviación de cada estudio respecto al tamaño del efecto promedio). En la figura 1 del modelo de efectos aleatorios, el  $I^2$  es del 29%, lo que evidencia que el 29% de la variabilidad en las estimaciones del efecto del tratamiento se debe a la variabilidad entre estudios. El modelo de efectos aleatorios se estima habitualmente con el método de DerSimonian y Laird<sup>9</sup>.

Independientemente de cuál sea la estimación del efecto combinado y su significación estadística, la interpretación clínica de los resultados es diferente según el modelo considerado. En el caso del modelo de efectos fijos, solo será posible generalizar los resultados del metaanálisis a una población de estudios de características idénticas a las de los incluidos. Por el contrario,

empleando el modelo de efectos aleatorios, los resultados pueden generalizarse a una población más amplia, y heterogénea, de estudios con características no necesariamente idénticas. El modelo de efectos aleatorios puede considerarse más realista que el modelo de efectos fijos porque no obliga a que la variabilidad entre estudios sea nula. Sin embargo, el modelo de efectos aleatorios termina por ponderar menos a los estudios más grandes y, por tanto, más precisos.

### Concepto de heterogeneidad<sup>10-13</sup>

Un aspecto importante a tener en cuenta en todo metaanálisis consiste en valorar si los efectos de los tratamientos de los estudios son homogéneos entre sí, es decir, si difieren solo por error de muestreo o aleatorio, o si, por el contrario, existe más heterogeneidad entre ellos de la que puede explicarse por el azar. De hecho, en cualquier metaanálisis es prácticamente inevitable que los estudios incluidos no difieran.

Del mismo modo, es importante distinguir diferentes tipos de heterogeneidad: 1) la variabilidad en las características de los pacientes o sujetos de estudio, los tratamientos y los resultados estudiados se puede describir como heterogeneidad clínica, y 2) la variabilidad en el diseño de los estudios (aleatorización, cegado, etc.) y algunos potenciales sesgos metodológicos se pueden describir como heterogeneidad metodológica o por diseño. La variabilidad en los efectos de los tratamientos que se evalúan en diferentes estudios se conoce como heterogeneidad estadística y puede ser consecuencia de los 2 tipos anteriores.

A su vez, atendiendo a diferentes razones, la heterogeneidad clínica puede diferenciarse en: 1) heterogeneidad «anónima», desconocida o aleatoria, en la que los efectos observados difieren, pero se desconoce su causa, y 2) heterogeneidad «clasificada» o conocida, que puede categorizarse y sus efectos pueden ser reproducidos (por ejemplo, evolución temporal de la práctica clínica habitual). Algunos autores las denominan heterogeneidad cuantitativa y cualitativa, respectivamente.

Se ha argumentado que puesto que prácticamente en todo metaanálisis siempre existirá heterogeneidad clínica y/o metodológica, la heterogeneidad estadística es inevitable<sup>13</sup>. Por lo tanto, una prueba estadística de heterogeneidad no debería condicionar totalmente la selección del modelo de análisis. La heterogeneidad (de cualquier tipo) siempre existirá, se tenga o no la capacidad de detectarla mediante una prueba estadística.

Se han desarrollado métodos para identificar y cuantificar la heterogeneidad estadística entre los estudios<sup>10</sup>. Entre ellos, la prueba de heterogeneidad utilizada con mayor frecuencia es la Q de Cochran<sup>14</sup>, aunque es conocido que puede llevar a no rechazar la hipótesis nula de homogeneidad de efectos debido a su baja potencia estadística cuando hay pocos estudios incluidos en la revisión sistemática. Por este motivo es recomendable complementar la información sobre la heterogeneidad estadística utilizando el índice I<sup>2</sup>. Este índice cuantifica el porcentaje de la variabilidad total en las estimaciones del efecto atribuible a la variabilidad interestudios<sup>10,12</sup>. El valor del índice I<sup>2</sup> se interpreta habitualmente en base a valores porcentuales del 25, 50 y 75% para identificar heterogeneidad baja, moderada, y alta, respectivamente.

En el caso de no identificarse heterogeneidad estadísticamente significativa, se podría obtener un efecto combinado a través del modelo de efectos fijos. Pero si el resultado de la prueba Q de heterogeneidad es que las diferencias observadas entre los resultados de los estudios incluidos en la revisión sistemática no pueden ser explicados por el azar, o bien si el índice I<sup>2</sup> nos proporciona indicios de un grado de heterogeneidad mayor del 50%, se concluye que hay heterogeneidad estadística. En ese caso, y siempre que fuera posible, habría que intentar explicar las características diferenciales de los estudios que pueden estar causando la heterogeneidad.

### Intervalo de confianza frente a intervalo de predicción

En un metaanálisis, el intervalo de confianza del efecto combinado permite comprobar si este es estadísticamente significativo, es decir, si el tratamiento presenta un resultado supuestamente más beneficioso respecto al del grupo de control<sup>15</sup>. Para ello, se comprueba si el efecto neutro (por ejemplo, un riesgo relativo igual a 1) no está incluido dentro del intervalo de confianza. Por otra parte, los límites superior e inferior del intervalo de confianza, así como su amplitud, nos informan del grado de precisión de la estimación del efecto promedio en la población. Cuanto más estrecho sea el intervalo de confianza obtenido, tanto mayor será la precisión de la estimación del efecto en la población. Generalmente, el intervalo de confianza que se obtiene en el modelo de efectos aleatorios es más amplio que el que se obtiene con el modelo de efectos fijos debido a que, como se mencionaba anteriormente, considera tanto la variabilidad intraestudio como la interestudios.

La presencia de heterogeneidad estadística es una situación muy común en los metaanálisis publicados. Si bien es deseable identificar las causas de la heterogeneidad (por ejemplo, a través de análisis de subgrupos o técnicas de regresión), en la práctica puede suceder que haya pocos estudios para examinarla, que no se conozca qué factores podrían causarla, o bien que no se disponga de la información necesaria. Aun cuando se identifica algún factor que puede explicar la causa de la heterogeneidad estadística, puede que no la explique en su totalidad y quede cierto grado de heterogeneidad residual por explicar.

Sin embargo, en ocasiones puede suceder que cuando se presenta un metaanálisis de efectos aleatorios, las miradas se centren excesivamente en la estimación del efecto combinado y en su intervalo de confianza, sin considerar cuál es el efecto potencial de la heterogeneidad en su estimación. De hecho, en un

metaanálisis de efectos aleatorios, un intervalo de confianza para un gran número de estudios que presenten una heterogeneidad moderada o alta podría ser de una amplitud similar a la de un intervalo de confianza procedente de un menor número de estudios más homogéneos, aunque en el primer caso será mucho menos probable que el intervalo de confianza refleje el intervalo de valores dentro del cual se encuentra el efecto del tratamiento para un nuevo estudio<sup>7,15</sup>. Por este motivo resulta de especial interés complementar el estimador combinado del modelo de efectos aleatorios con el cálculo de un intervalo de predicción<sup>7,8,16</sup>.

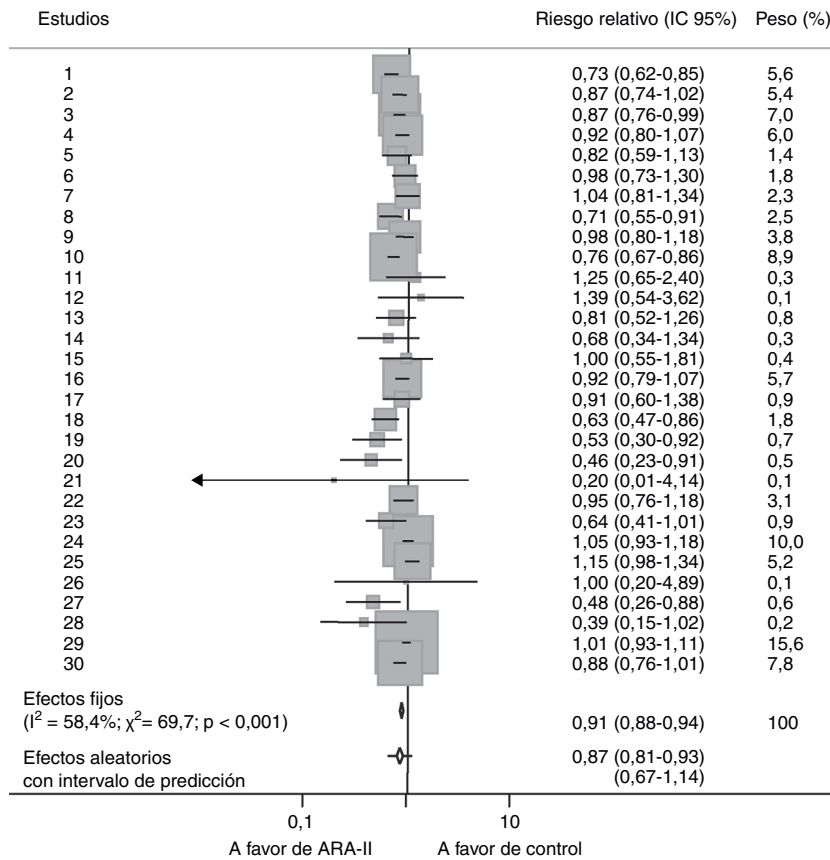
El intervalo de predicción permite dar información sobre un intervalo de efectos estimados en el estudio que puede ser más uniforme y certera, facilitando así la generalización de resultados a la práctica clínica. Por ejemplo, considerando el metaanálisis de efectos aleatorios de la figura 1, para el cual el intervalo de predicción del 95% oscila entre un intervalo de valores que va de 0,75 a 1,10, se observa que en algunas situaciones el tratamiento puede ser ineficaz. Este hallazgo queda enmascarado cuando en presencia de heterogeneidad exclusivamente nos centramos en el efecto combinado y su intervalo de confianza. En el *forest plot*, el intervalo de predicción se podría proporcionar de manera complementaria en la parte inferior del efecto combinado (figs. 2 y 3). El intervalo de predicción proporciona un conocimiento más completo sobre la incertidumbre que existe alrededor del efecto combinado, determinando un intervalo de valores sobre el verdadero efecto del tratamiento cuando se considera la variabilidad entre estudios.

Es importante diferenciar el intervalo de confianza del intervalo de predicción. Si bien ambos se centran en el efecto promedio, el intervalo de confianza solo considera la oscilación muestral en torno al valor estimado, mientras que el intervalo de predicción añade la oscilación de los estudios para pretender abarcar también el efecto observable en el 95% de los estudios.

### Ejemplos

En un reciente metaanálisis que incluyó 37 ensayos clínicos aleatorizados y controlados en 147.020 pacientes, Bangalore et al.<sup>17</sup> evaluaron el efecto de los antagonistas de los receptores de la angiotensina II (ARA-II) en comparación con los controles (placebo o tratamiento activo), llegando a concluir que el tratamiento no aumentó el riesgo de infarto de miocardio, observándose una reducción significativa del riesgo de accidente cerebrovascular, de insuficiencia cardíaca y diabetes. En las figuras 2 y 3 se presentan los resultados de los metaanálisis en los que se evalúa la eficacia de los ARA-II en la prevención de la insuficiencia cardíaca y de la diabetes.

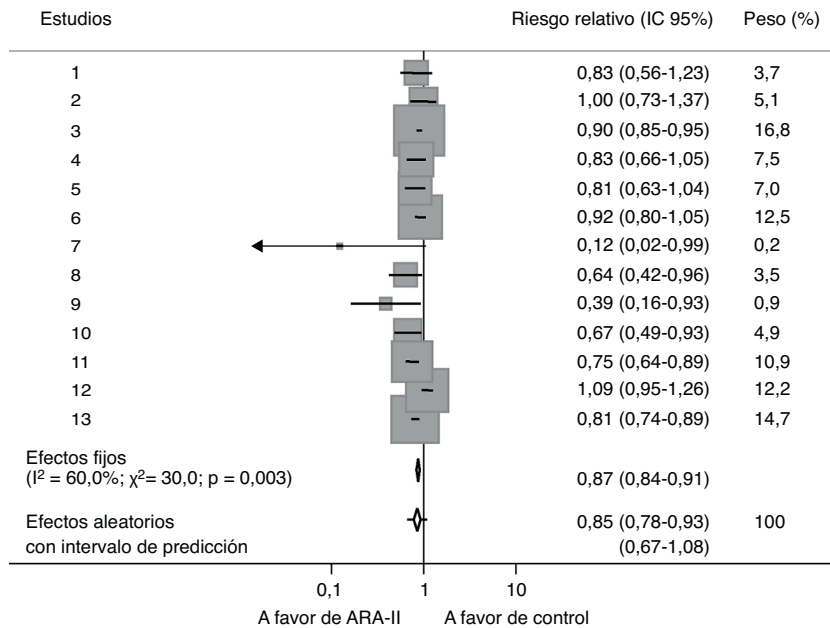
En la figura 2 se presentan los riesgos relativos del efecto de los ARA-II en el tratamiento de la insuficiencia cardíaca disponibles en 30 ensayos clínicos. Los valores inferiores a 1 indican un beneficio para los ARA-II. Los estudios utilizaron diferentes comparadores (10 estudios emplearon placebo, y 20, diferentes comparadores activos como betabloqueantes, inhibidores de la enzima convertidora de la angiotensina o bloqueantes de los canales de calcio), aunque también existían otras diferencias clínicas y metodológicas, lo que resultó en una gran heterogeneidad estadística entre los estudios (I<sup>2</sup> = 58,4%; p < 0,001). Por tanto, los autores utilizaron un modelo de efectos aleatorios y obtuvieron un resultado combinado de 0,87 (IC 95% 0,81-0,93), concluyendo que los ARA-II se asocian a reducciones del 13% de la insuficiencia cardíaca en comparación con los controles (de cualquier tipo). El resultado combinado presentado se relaciona con el efecto promedio de los ARA-II entre los ensayos incluidos en el metaanálisis. Si el intervalo de confianza es inferior a 1, se proporciona una fuerte evidencia de que el tratamiento antihipertensivo es beneficioso; sin embargo, no indica si los ARA-II son siempre eficaces. Los autores reconocen



**Figura 2.** Metaanálisis de ensayos clínicos aleatorizados y controlados que evalúan el efecto de los antagonistas de los receptores de la angiotensina II (ARA-II) en la reducción del riesgo de insuficiencia cardíaca. Datos de Bangalore et al. 2011<sup>17</sup>.

la heterogeneidad entre estudios, pero aunque los efectos de los tratamientos entre estudios podían diferir, concluyen que los resultados eran en su mayoría favorables al tratamiento. Se observa, entonces, un caso de heterogeneidad clasificada

(comentada en el apartado anterior), en la que como se presenta en el ejemplo original, conviene disponer de 2 medidas del efecto o subgrupos, una cuando se compara el tratamiento respecto a placebo, y otra cuando se considera el tratamiento en relación con



**Figura 3.** Metaanálisis de ensayos clínicos aleatorizados y controlados que evalúan el efecto de los antagonistas de los receptores de la angiotensina II (ARA-II) en la prevención del inicio de diabetes. Datos de Bangalore et al. 2011<sup>17</sup>.

un control activo. En función de la intervención estándar de la guía clínica de su entorno, de las características del paciente y de la relevancia clínica del efecto observado, el clínico ya sabrá cuál de los tratamientos aplica a sus pacientes. Sin embargo, en ocasiones esto también se puede cuantificar de manera más formal por un intervalo de predicción del 95%, que se ha calculado como 0,67 a 1,14. Este intervalo incluye valores superiores a 1, lo que indica que, en ocasiones, los ARA-II no serán beneficiosos cuando se utilizan en al menos el 95% de los efectos del estudio, un resultado que puede ser importante conocer para la práctica clínica.

En la *figura 3* se presentan los riesgos relativos del efecto de los ARA-II en la prevención de casos incidentes de diabetes disponibles en 13 ensayos clínicos. Los valores inferiores a 1 indican un beneficio para los ARA-II. El efecto combinado permitió obtener un riesgo relativo del 0,85 (IC 95% 0,78-0,93), lo que equivale a una reducción del riesgo del 15% en comparación con cualquier otro tratamiento. Como en el ejemplo anterior, el intervalo de confianza sugiere un fuerte nivel de evidencia a favor del tratamiento con ARA-II, ya que no incluye el 1. Sin embargo, nuevamente se presenta un grado considerable de heterogeneidad estadística ( $I^2 = 60,0\%$ ;  $p = 0,003$ ), que puede deberse a los comparadores utilizados, el período de seguimiento, la edad de los pacientes, el porcentaje de estos que son hipertensos o la presencia de otros factores de riesgo cardiovascular. De hecho, los autores en la discusión reconocen que «la reducción del riesgo de insuficiencia cardíaca y de diabetes fue mayor en los ensayos clínicos que incluían pacientes hipertensos en relación a los que no». Esta aclaración o advertencia, que puede pasar inadvertida en una primera lectura rápida del trabajo, puede cuantificarse con un intervalo de predicción del 95%, que hemos calculado en este trabajo igual a 0,67 a 1,08. Este intervalo incluye el 1 en el límite superior, lo que indica que, generalmente, los ARA-II serán beneficiosos, aunque en ocasiones no tendrán efecto. En este caso, quizás haya que realizar más estudios para identificar causas de la heterogeneidad, en particular, en las subpoblaciones de pacientes no hipertensos o/y en hipertensos controlados.

### Consideración final

La interpretación clínica de los resultados de un metaanálisis puede ser muy diferente según el modelo utilizado para combinar los resultados de los estudios incluidos en una revisión sistemática. Asimismo, en todo metaanálisis hay que tener en cuenta la descripción y valoración de la variabilidad entre los resultados de los estudios (heterogeneidad). Siguiendo las recomendaciones de la principal guía de publicación (*Preferred Reporting Items for Systematic reviews and Meta-Analyses* [PRISMA])<sup>5,6</sup>, sería deseable identificar las causas potenciales de esta heterogeneidad, si bien en la práctica es bastante frecuente que esto no se lleve a cabo. Una alternativa habitual es el uso del modelo de efectos aleatorios. En este caso sería deseable complementar el estimador combinado con su intervalo de predicción, ya que puede proporcionar información relevante sobre el intervalo de efectos estimados en

el estudio considerando la presencia de heterogeneidad, facilitando la generalización de resultados a la práctica clínica.

### Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Las opiniones expresadas en este trabajo son responsabilidad de los autores, por lo que no reflejan necesariamente el punto de vista de los organismos en los que trabajan.

### Agradecimientos

Los autores desean agradecer los comentarios recibidos por parte de los revisores anónimos de MEDICINA CLÍNICA por considerar que han contribuido a mejorar sustancialmente el contenido del presente manuscrito. Los autores somos responsables exclusivos del contenido, así como de cualquier error u omisión (no intencionado) que pudiera aparecer.

### Bibliografía

- Catalá-López F, Tobías A. Síntesis de la evidencia clínica y metaanálisis en red con comparaciones indirectas. *Med Clin (Barc)*. 2013;140:182-7.
- Ferreira González I, Urrútia G, Alonso-Coello P. Revisiones sistemáticas y metaanálisis: bases conceptuales e interpretación. *Rev Esp Cardiol*. 2011;64:688-96.
- Urrútia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Med Clin (Barc)*. 2010;135:507-11.
- Moher D, Liberati A. Revisiones sistemáticas y metaanálisis: la responsabilidad de los autores, revisores, editores y patrocinadores. *Med Clin (Barc)*. 2010;135:505-6.
- Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*. 2009;6:e1000097.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. 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.
- Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172:137-59.
- Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342:d549.
- DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177-88.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-60.
- Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ*. 1994;309:1351-5.
- Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I<sup>2</sup> index? *Psychol Methods*. 2006;11:193-206.
- Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol*. 2008;37:1158-60.
- Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10:101-29.
- Sánchez-Meca J, Marín-Martínez F. Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychol Methods*. 2008;13:31-48.
- Guddat C, Grouven U, Bender R, Skipka G. A note on the graphical presentation of prediction intervals in random-effects meta-analyses. *Syst Rev*. 2012;1:34.
- Bangalore S, Kumar S, Wetterslev J, Messerli FH. Angiotensin receptor blockers and risk of myocardial infarction: Meta-analyses and trial sequential analyses of 147 020 patients from randomised trials. *BMJ*. 2011;342:d2234.