

## Enfoque: Métodos contemporáneos en bioestadística (III)

# Revisiones sistemáticas y metaanálisis: bases conceptuales e interpretación

Ignacio Ferreira González<sup>a,b,\*</sup>, Gerard Urrútia<sup>b,c</sup> y Pablo Alonso-Coello<sup>b,c</sup>

<sup>a</sup> Unidad de Epidemiología, Servicio de Cardiología, Área del Cor, Hospital Vall d'Hebron, Barcelona, España

<sup>b</sup> CIBER de Epidemiología y Salud Pública (CIBERESP), España

<sup>c</sup> Centro Cochrane Iberoamericano-Servei d'Epidemiologia Clínica i Salut Pública, Institut d'Investigació Biomèdica Sant Pau, Barcelona, España

Historia del artículo:

On-line el 30 de junio de 2011

Palabras clave:

Revisión sistemática

Metaanálisis

Medicina basada en la evidencia

Keywords:

Systematic review

Meta-analysis

Evidence-based medicine

## RESUMEN

Las revisiones sistemáticas son investigaciones científicas en las cuales la unidad de análisis son los estudios originales primarios. Constituyen una herramienta esencial para sintetizar la información científica disponible, incrementar la validez de las conclusiones de estudios individuales e identificar áreas de incertidumbre donde sea necesario realizar investigación. Además, son imprescindibles para la práctica de una medicina basada en la evidencia y una herramienta fundamental en la toma de decisiones médicas. Sin embargo, la realización de una revisión sistemática de calidad no es una tarea sencilla, como en ocasiones tampoco lo es su interpretación. En este artículo especial se presentan las bases conceptuales para la realización y la interpretación de revisiones sistemáticas, poniendo especial énfasis en los puntos clave durante su ejecución mediante un ejemplo hipotético.

© 2011 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L. Todos los derechos reservados.

## Systematic Reviews and Meta-Analysis: Scientific Rationale and Interpretation

## ABSTRACT

Systematic reviews represent a specific type of medical research in which the units of analysis are the original primary studies. They are essential tools in synthesizing available scientific information, increasing the validity of the conclusions of primary studies, and identifying areas for future research. They are also indispensable for the practice of evidence-based medicine and the medical decision-making process. However, conducting high quality systematic reviews is not easy and they can sometimes be difficult to interpret. This special article presents the rationale for carrying out and interpreting systematic reviews and uses a hypothetical example to draw attention to key-points.

Full English text available from: [www.revespcardiol.org](http://www.revespcardiol.org)

© 2011 Sociedad Española de Cardiología. Published by Elsevier España, S.L. All rights reserved.

## Abreviaturas

ECA: ensayo clínico aleatorizado

MA: metaanálisis

RS: revisión sistemática

## INTRODUCCIÓN

Has finalizado una semana agotadora. Al final de la semana, te sientas tranquilo y reflexionas sobre las decisiones que has tenido que tomar. Entre otras, has indicado una intervención quirúrgica a un paciente con enfermedad de tres vasos, tuviste que decidir si se fibrinolizaba o se trasladaba a tu centro para angioplastia primaria a un paciente de 82 años con infarto inferior de 70 min de duración, y en consultas externas decidiste anticoagular a una paciente con fibrilación auricular.

\* Autor para correspondencia: Unidad de Epidemiología, Servicio de Cardiología, Hospital Vall d'Hebron, Pg. Vall d'Hebron 119-129, 08035 Barcelona, España.

Correo electrónico: [nacho@ferreiragonzalez.com](mailto:nacho@ferreiragonzalez.com) (I. Ferreira González).

Aunque estás razonablemente convencido de que tus decisiones se basaron en la mejor evidencia disponible, la duda te asalta. ¿Se habrá publicado algún estudio que ponga en cuestión mis decisiones?, ¿diferentes estudios sobre la misma intervención pueden discrepar en los resultados? Lo cierto es que en los últimos meses no has tenido mucho tiempo para leer. Te das cuenta de que para despejar tus dudas necesitas rápidamente una síntesis concisa, actualizada y rigurosa sobre la mejor evidencia disponible en las decisiones que tuviste que tomar. Dicho de otra manera, necesitas una revisión sistemática (RS)<sup>1</sup>.

Se considera que las RS son la fuente más fiable para informar la toma de decisiones médicas<sup>2</sup>. Quizá por ello, cada vez son más populares, y se han incrementado de forma muy importante las RS publicadas en los últimos años<sup>2</sup>. Sin embargo, la realización de una RS de calidad no es una tarea sencilla. De hecho, existen normas para su elaboración y, al igual que otros diseños, recomendaciones para la forma de presentación de sus resultados acorde con unos estándares de calidad, que han sido desarrollados por grupos internacionales multidisciplinares de expertos, que incluyen a autores de RS, metodólogos, clínicos y editores<sup>2-4</sup>. En este artículo se presentan las bases conceptuales para la realización y la interpretación de RS, poniendo especial énfasis en los puntos clave durante su ejecución mediante un ejemplo hipotético.

## CONCEPTO Y NOMENCLATURA

Las RS son investigaciones científicas en las que la unidad de análisis son los estudios originales primarios, a partir de los cuales se pretende contestar a una pregunta de investigación claramente formulada mediante un proceso sistemático y explícito. Por eso se las considera investigación secundaria («investigación sobre lo investigado»). Por el contrario, a las revisiones que no siguen un proceso sistemático, denominadas revisiones narrativas, no se las puede considerar un proceso formal de investigación, sino simplemente un formato de literatura científica basada sobre todo en opinión.

Desde un punto de vista formal, las RS sintetizan los resultados de investigaciones primarias mediante estrategias que limitan el sesgo y el error aleatorio<sup>5</sup>. Estas estrategias incluyen:

- La búsqueda sistemática y exhaustiva de todos los artículos potencialmente relevantes.
- La selección, mediante criterios explícitos y reproducibles, de los artículos que serán incluidos finalmente en la revisión<sup>1</sup>.
- La descripción del diseño y la ejecución de los estudios originales, la síntesis de los datos obtenidos y la interpretación de los resultados.

Aunque la RS es una herramienta de síntesis de información, no siempre es posible presentar resumidamente los resultados de los estudios primarios. Cuando estos no se combinan estadísticamente, la revisión se denomina RS cualitativa. Por el contrario, una RS cuantitativa, o metaanálisis (MA), es una RS que usa métodos estadísticos para combinar los resultados de dos o más estudios<sup>1</sup>.

No debe confundirse una RS con un MA. La primera siempre es posible, mientras que el segundo, sólo a veces. Sin embargo, cuando se cumplen las condiciones para realizarlo, el MA aporta información muy útil y manejable para facilitar la comprensión acerca del efecto de un tratamiento o intervención, tanto en general como en grupos específicos de pacientes. Además, permite aumentar la precisión en la estimación del efecto detectando efectos de magnitud moderada pero clínicamente importantes que podrían haber pasado inadvertidos en los estudios primarios. Habitualmente, el MA combina datos agregados procedentes de los estudios publicados, pero en ocasiones puede combinar datos individualizados de los pacientes que han participado en varios estudios. Es el llamado MA con datos individuales de pacientes (*individual patient data meta-analysis*), considerado patrón de referencia de las RS<sup>6</sup>.

Nótese que, a diferencia de las revisiones narrativas, en las RS se emplea un método sistemático para la búsqueda de todos los estudios potencialmente relevantes junto con unos criterios explícitos y reproducibles, fijados de antemano en la selección. Eso es lo que les confiere el carácter científico, a diferencia de las revisiones narrativas. La [tabla 1](#) presenta las diferencias entre ambas.

Al igual que en el caso de los ensayos clínicos, se recomienda contar con un protocolo previo a la realización de una RS<sup>7</sup>. Esto ayudará a reflexionar y fijar los métodos más adecuados que luego

se aplicarán en la revisión y, además, evitará las decisiones tomadas *a posteriori* en función de los resultados. Actualmente, el primer registro internacional de protocolos de revisiones sistemáticas, aparte de las RS Cochrane, se ha publicado recientemente con el nombre de PROSPERO (<http://www.crd.york.ac.uk/prosperto/>).

## ETAPAS DE UNA REVISIÓN SISTEMÁTICA

De forma resumida, una RS sigue las siguientes etapas:

- Definición de la pregunta clínica de interés y los criterios de inclusión y exclusión de los estudios.
- Localización y selección de los estudios relevantes.
- Extracción de datos de los estudios primarios.
- Análisis y presentación de los resultados.
- Interpretación de los resultados.

### Definición de la pregunta clínica de interés

El primer paso es la correcta formulación de dicha pregunta. De forma general, esta ha de ser explícita y estructurada incluyendo los siguientes componentes clave<sup>8</sup>:

- La población específica y el contexto. Por ejemplo, pacientes añosos (mayores de 75 años) ingresados por infarto agudo de miocardio con elevación del ST.
- La exposición de interés. Podría ser un factor de riesgo, un factor pronóstico, una intervención o tratamiento, o un test diagnóstico. En el caso de una intervención, tratamiento o test diagnóstico, es habitual definir al mismo tiempo una exposición control. Por ejemplo, angioplastia primaria (intervención) frente a fibrinólisis (control).
- Eventos de interés. Por ejemplo, mortalidad total, mortalidad cardiovascular, nuevos ingresos por síndrome coronario, nuevas revascularizaciones, etc.

Así, a partir de estos elementos, se podría formular la pregunta: ¿en pacientes mayores de 75 años, la angioplastia primaria comparada con la fibrinólisis reduce la mortalidad y el infarto de miocardio? Una vez definida y delimitada la pregunta de interés, es más fácil establecer criterios de inclusión y exclusión de los estudios primarios. Por el contrario, una pregunta mal delimitada llevará a decisiones confusas acerca de los estudios que resultan relevantes para darle respuesta.

En muchas ocasiones no resulta fácil decidir lo específica que debe ser la pregunta de interés. Lo que está claro es que debe ser clínicamente relevante. Una pregunta demasiado poco específica (p. ej., ¿es útil la angioplastia primaria en el infarto agudo de miocardio?) será de poca ayuda al clínico para tomar una decisión en un paciente concreto. Además, debería tener en cuenta elementos de la exposición o de los pacientes que se piensa que

**Tabla 1**  
Diferencias entre revisiones narrativas y sistemáticas

Característica	Revisión narrativa	Revisión sistemática
Pregunta de interés	No estructurada, no específica	Pregunta estructurada, problema clínico bien delimitado
Búsqueda de artículos y sus fuentes	No detallada y no sistemática	Búsqueda estructurada y explícita
Selección de artículos de interés	No detallada y no reproducible	Selección basada en criterios explícitos uniformemente aplicados a todos los artículos
Evaluación de la calidad de la información	Ausente	Estructurada y explícita
Síntesis	A menudo resumen cualitativo	Resumen cualitativo y cuantitativo
Inferencias	A veces basadas en la evidencia	Normalmente basadas en la evidencia

pueden afectar al evento de interés. Por ejemplo, no es infrecuente que los pacientes mayores de 75 años estén en tratamiento con anticoagulación oral, el cual previsiblemente puede afectar al evento de interés. Así pues, podríamos restringir nuestra población de estudio a los pacientes sin anticoagulación oral. Sin embargo, unos criterios de inclusión excesivamente específicos limitan la aplicabilidad de los resultados. Por el contrario, se podría definir una pregunta poco restrictiva que tenga sentido clínico, y a partir de ella, explorar preguntas más específicas. Por ejemplo, incluir a todos los pacientes con infarto agudo de miocardio, y después realizar análisis exploratorios en aquellos con y sin anticoagulación oral. Sin embargo, dicha estrategia puede dar problemas semejantes al análisis de subgrupos<sup>9</sup>. Por último, la elección de unos criterios de inclusión excesivamente amplios conlleva el riesgo de no tener sentido clínico ni biológico<sup>10</sup>.

Es importante en esta fase decidir qué diseños de estudio vamos a considerar para su inclusión en nuestra revisión. Ello dependerá del tipo de cuestión que se intenta abordar. Si queremos evaluar la eficacia de una intervención, como en el ejemplo anterior, está claro que deberemos incluir ensayos clínicos aleatorizados (ECA), si estos están disponibles. Lo mismo ocurre en el caso de la evaluación de la fiabilidad y la seguridad de una prueba diagnóstica. En el caso de una RS para la evaluación de intervenciones comunitarias o en salud pública o la evaluación de resultados de una intervención a largo plazo, sobre todo en aspectos de seguridad, los estudios observacionales son más relevantes. Sin embargo, en ocasiones no habrá ECA sobre una intervención concreta, y habrá que analizar estudios observacionales.

En la **figura 1** se presenta, de forma simplificada, los elementos de la pregunta de investigación de interés de nuestro ejemplo y los criterios de inclusión y exclusión para la RS que se derivan de ella.

### Localización y selección de los estudios relevantes en relación con la pregunta de interés

Esta etapa consta de varios componentes:

#### 1. Identificación de artículos potenciales

- Decidir sobre restricciones en relación con el idioma de publicación.
- Decidir sobre las fuentes de obtención de los estudios primarios.
- Obtención de los títulos y resúmenes de los potenciales estudios primarios.

#### 2. Selección de artículos potenciales

- Aplicar los criterios de inclusión y exclusión a los títulos y resúmenes obtenidos.

- Obtención de los artículos potenciales a partir de los títulos y resúmenes elegibles y aplicar los criterios de inclusión y exclusión.
- Evaluar la concordancia en la selección de los estudios.

### Identificación de artículos potenciales

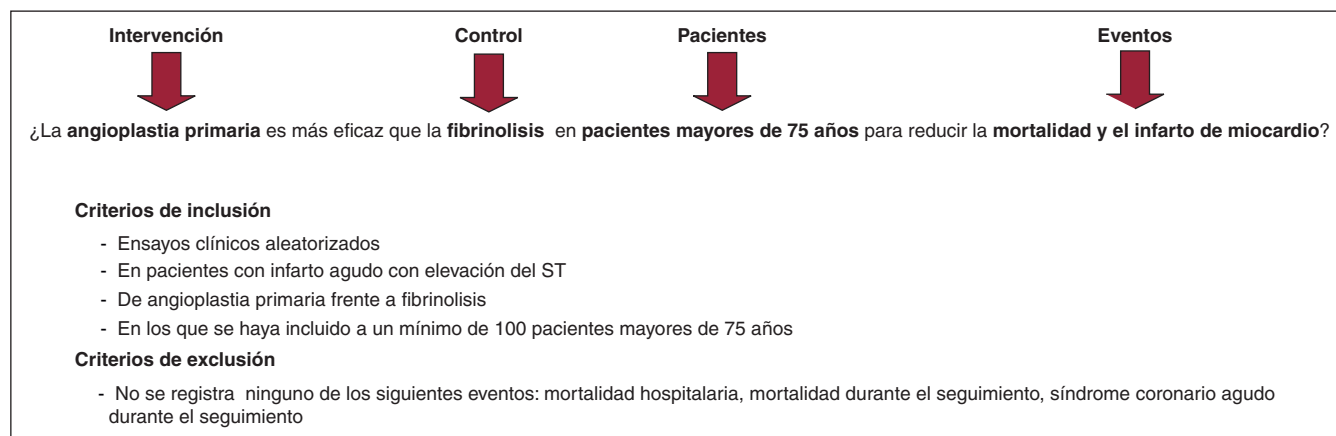
Como ocurre en cualquier estudio de investigación, la presencia de errores en la extracción de datos puede invalidar los resultados de una RS. Es de suma importancia conseguir tantos estudios primarios sobre la pregunta de interés como sea posible. Con ello se persigue minimizar el error aleatorio y el sesgo. Si se omiten estudios, se puede introducir sesgo si la muestra finalmente seleccionada no es representativa. Hay que tomar dos decisiones en este punto: en relación con la restricción o no del idioma de publicación y en relación con la inclusión o exclusión de estudios no publicados en revistas médicas.

Con respecto al idioma, lo más frecuente por razones pragmáticas es incluir únicamente publicaciones en inglés y en el idioma nativo del autor de la RS. Sin embargo, los datos disponibles indican que la calidad de la investigación no está necesariamente en relación con el idioma de publicación<sup>11</sup>. Por el contrario, las restricciones en el idioma pueden sesgar los resultados de la RS al excluir estudios que pueden ser relevantes<sup>12</sup>.

Por razones de factibilidad, parece razonable incluir sólo los estudios publicados en revistas médicas. Se podría argumentar que estos, al haber superado un proceso de revisión por pares, son los más fiables<sup>13</sup>. Lo cierto es que, independientemente de su calidad, es menos probable que se publiquen estudios con resultados no concluyentes o negativos<sup>14,15</sup>, por lo que su exclusión puede sesgar los resultados de la RS. Es el llamado sesgo de publicación, que da lugar a que es más probable que las RS que excluyen estudios no publicados sobrestimen la relación entre la exposición y el evento de interés. En situaciones extremas, podrían revelar como eficaces tratamientos totalmente fútiles<sup>16,17</sup>.

Una vez tomada la decisión sobre las dos cuestiones anteriores el siguiente paso es crucial: ¿dónde buscar los estudios primarios? Existen varias estrategias:

- Bases de datos electrónicas: MEDLINE, EMBASE, CENTRAL.
- Bases de datos no indexadas: AMED, CINAHL, BIOSIS, etc.
- Búsqueda manual en sumarios de revistas, actas y sumarios de reuniones científicas y libros.
- Listas de referencias y citaciones: *Science Citation Index* y similares.
- Registros de estudios en curso (p. ej., clinicaltrials.gov).
- Contacto con compañías farmacéuticas.
- Contacto con colegas expertos en el tema de interés.



**Figura 1.** Ejemplo de los elementos de una pregunta de interés de una revisión sistemática y de los criterios de inclusión y exclusión de los estudios primarios.

Evidentemente, la estrategia más utilizada hoy es la búsqueda en bases de datos electrónicas. Sin embargo, no es una estrategia simple, dado que, aunque hay un solapamiento entre bases de datos, muchas revistas incluidas en una base de datos específica no se consideran en otras. MEDLINE, por ejemplo, indexa tan sólo unas 5.600 de entre las más de 16.000 revistas biomédicas, la mayor parte en inglés. EMBASE indexa más de 1.000 revistas que no están incluidas en MEDLINE, muchas de ellas europeas. Otras bases de datos son complementarias a las anteriores, pues se ha realizado esfuerzos para registrar literatura sobre estudios no publicados (literatura gris)<sup>18</sup>.

La identificación de potenciales artículos elegibles se suele realizar, por operatividad, identificando títulos y resúmenes. Sin embargo, cada base de datos tiene su estructura particular y utiliza unos criterios de indexación y palabras clave más o menos específicas. Por ejemplo, en MEDLINE, el índice específico de vocabulario controlado se denomina MeSH (*Medical Subject Headings*). La ayuda de un documentalista experto en RS en esta fase es crucial. Además, hay que tener en cuenta que la estrategia de búsqueda finalmente utilizada tiene que aparecer en la publicación de la RS, con objeto de comprobar su reproducibilidad.

La búsqueda de estudios exclusivamente en bases de datos electrónicas podría no ser óptima según el tema de interés. En ocasiones, resulta conveniente incluir también una estrategia complementaria con objeto de identificar estudios no publicados. Esta fase es una de las más laboriosas, ya que suele requerir la búsqueda manual en sumarios de revistas o actas de congresos, contactos con expertos en el tema, con compañías farmacéuticas, etc. Puede comprenderse que todo ello conlleva una inversión importante en tiempo y recursos económicos. Consciente de ello, la Colaboración Cochrane ha impulsado una iniciativa internacional

para desarrollar un registro de ensayos clínicos controlados, antiguamente conocido como *Cochrane Controlled Trials Register* y actualmente denominado CENTRAL<sup>19</sup>. Este recurso, que cuenta con cientos de miles de registros o citas de estudios publicados en revistas indexadas, así como en suplementos (habitualmente resúmenes de congresos), y se actualiza constantemente, es de indudable interés para la localización de ECA controlados<sup>20</sup>.

La aplicación de las estrategias de búsqueda en las diversas bases de datos electrónicas proporcionará un número habitualmente elevado de referencias bibliográficas. Además, es muy probable que un número elevado de ellas estén duplicadas entre las bases de datos. Por ello, resulta de gran utilidad utilizar en esta fase un *software* para la gestión automatizada de las citas bibliográficas como, por ejemplo, ProCite o *Reference Manager*.

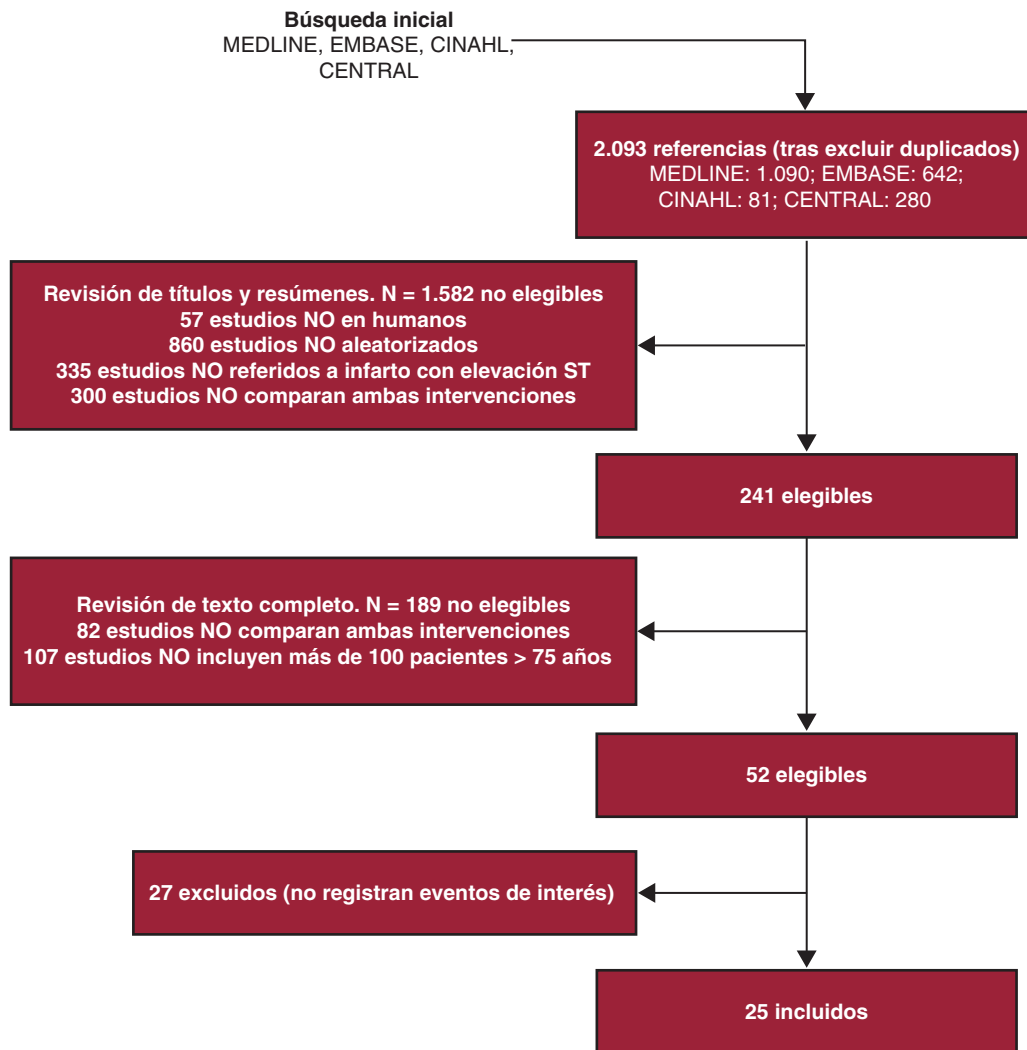
#### Selección de artículos potenciales

A partir de los títulos y resúmenes identificados, se ha de realizar una primera selección o cribado de los potenciales artículos elegibles. Para ello es conveniente diseñar una hoja de selección de estudios que sea operativa incluyendo criterios explícitos y comprensibles<sup>21</sup>. Un ejemplo simple, en relación con la pregunta de interés antes referida, sería el expuesto en la [figura 2](#). Normalmente se comienza el proceso de selección mediante la revisión de títulos y resúmenes y, en caso de que haya dudas, se deberá revisar el artículo a texto completo.

La selección de los estudios se ha de realizar por dos revisores de forma independiente, con objeto de aumentar la fiabilidad y la seguridad del proceso. Igualmente importante es medir el grado de acuerdo entre los revisores mediante el cálculo del estadístico

<p><b>Referencia del estudio:</b> _____</p> <p><b>Nivel de selección (marcar donde aplique):</b></p> <p>Título ___    Resumen ___    Texto ___</p> <p><b>Criterios de selección:</b></p> <p><b>Pacientes</b></p> <p>¿Se estudia a pacientes con infarto agudo con elevación del ST? Sí/No</p> <p>¿Se incluye a pacientes mayores de 75 años? Sí/No</p> <p><b>Intervención</b></p> <p>¿Recibió al menos uno de los grupos de tratamiento angioplasti a primaria? Sí/No</p> <p><b>Control</b></p> <p>¿Recibió al menos uno de los grupos de tratamiento fibrinólisis? Sí/No</p> <p><b>Eventos</b></p> <p>¿Se incluye alguno de los siguientes eventos como variables de resultado individuales o combinadas: mortalidad total, mortalidad cardiovascular, infarto de miocardio no mortal, nuevas revascularizaciones? Sí/No</p> <p><b>Diseño</b></p> <p>¿Es un ensayo clínico aleatorizado? Sí/No</p> <p>ACCIÓN (Incluir SÓLO en el que caso de que la respuesta a TODAS las anteriores haya sido «sí»):</p> <p>Incluir ___    Excluido ___    Dudoso ___</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Figura 2.** Ejemplo de una hipotética hoja de selección de artículos potenciales para una revisión sistemática.



**Figura 3.** Ejemplo del proceso de selección de estudios.

kappa para cada uno de los ítems de la hoja de selección. Dicho estadístico, expresado de forma simple, mide el grado de acuerdo entre los revisores por encima de lo esperable por el azar<sup>22,23</sup>. Para casos en que haya discrepancias entre los dos revisores respecto a la decisión de incluir o no un artículo, se suele nombrar a un tercer investigador sénior que arbitra las discrepancias y finalmente es quien toma la decisión.

Por último, es importante tener en cuenta que todo el proceso de localización y selección de estudios se ha de reportar correctamente, indicando en un diagrama de flujo los artículos identificados en cada fase, así como los eliminados y las causas de su eliminación (fig. 3).

### Extracción de datos de los estudios primarios

Esta fase requiere la máxima fiabilidad de la información que se recoge de cada estudio seleccionado por lo que, de nuevo, es conveniente que la extracción de datos se haga en duplicado. Si no es factible, una alternativa es la realización de una auditoría por un revisor independiente sobre una muestra de estudios escogida aleatoriamente.

La información a extraer de los estudios primarios se debe haber consensuado durante la fase de diseño del estudio. En general, debe ser aquella que nos permita aceptar o rechazar la hipótesis del estudio. De forma resumida, la hoja de extracción de datos suele incluir<sup>24</sup>:

- Información sobre los pacientes, intervención de interés, intervención control y diseño de estudio.
- Información sobre los resultados.
- Información sobre la calidad metodológica del estudio.

El apartado A corresponde a toda información que pensamos que puede ser relevante por su potencial impacto en el resultado y puede variar entre estudios. En nuestro ejemplo, no será lo mismo un ECA en el que la edad media de los pacientes incluidos fuera de 65 años y con un tiempo medio puerta-balón de 63 min que otro con edad media de 75 años y tiempo puerta-balón de 96 min. Todas estas diferencias podrían explicar que la magnitud del efecto de la intervención varíe de un estudio a otro. Es decir, podrían ayudar a explicar la heterogeneidad del efecto. En todo caso, el reto consiste en conseguir un equilibrio en la exhaustividad de la información a recoger evitando, al mismo tiempo, un exceso de información innecesario que podría sobrecargar el trabajo de revisión.

La información del apartado B corresponde a la extracción de los resultados. El formato que elijamos dependerá de cómo estén definidos los eventos de interés: como variable dicotómica o como variable continua. En el primer caso, a veces será fácil obtener el número y el porcentaje de pacientes en los que sucedió el evento de interés en cada rama de la intervención. Otras veces los resultados vendrán expresados en forma de alguna medida de asociación o impacto: riesgo relativo, reducción de riesgo relativo, reducción absoluta de riesgo, razón de ventajas (*odds ratio*) o tasa de riesgo



Referencia de estudio: \_\_\_\_\_

	Angioplastia primaria		Fibrinólisis		Medidas de asociación o de impacto			
	N total aleatorizados: _____ N total analizados: _____		N total aleatorizados: _____ N total analizados: _____		RR (IC del 95%)	RRR (IC del 95%)	RAR (IC del 95%)	HR (IC del 95%)
	N	%	N	%				
Mortalidad total								
Mortalidad CV								
Reingresos por SCA								
Nuevas revascularizaciones								

**Figura 4.** Ejemplo de una hipotética hoja de extracción de resultados de estudios para una revisión sistemática. CV: cardiovascular; HR: hazard ratio; IC: intervalo de confianza; RAR: reducción absoluta de riesgo; RR: riesgo relativo; RRR: reducción de riesgo relativo; SCA: síndrome coronario agudo.

(hazard ratio). En la figura 4 se presenta un ejemplo con una hipotética hoja de extracción de resultados. Afortunadamente, a partir de cualquiera de las medidas habituales de asociación o impacto es fácil llegar a obtener el número y porcentaje aproximado de pacientes con el evento de interés. En el caso de que la variable de resultado sea continua, por ejemplo fracción de eyección, la información de interés será la media y la desviación estándar en cada grupo de tratamiento.

Por último, la hoja de extracción de datos debe incluir información sobre la calidad metodológica de cada estudio incluido, pues está estrechamente relacionada con la magnitud del efecto. Hay controversia sobre cuál es la mejor forma de reflejar la calidad metodológica de un estudio<sup>25</sup>. Hay quien defiende el uso de escalas de puntuación de calidad. Se han desarrollado varias<sup>26</sup>, la mayor parte de ellas para ECA. Algunas son genéricas y otras específicas para determinadas áreas clínicas. Sin embargo, se ha demostrado que el uso de una escala u otra puede variar sustancialmente los resultados de un MA<sup>27</sup>, por lo que ninguna es totalmente fiable. Recientemente, se ha desarrollado un nuevo sistema para cuantificar la calidad metodológica de los estudios incluidos llamado GRADE (*Grading of Recommendations Assessment, Development, and Evaluation*)<sup>28</sup>. Este sistema, desarrollado y consensuado por un grupo de líderes internacionales en realización de guías de práctica clínica, ofrece algunas ventajas interesantes frente a otros. Fundamentalmente:

- La calidad de la evidencia, clasificada como alta, moderada, baja y muy baja, se reporta separada del grado de recomendación (recomendación fuerte o débil).
- Se reconoce y se pondera por los valores y preferencias de los pacientes.
- Ofrece una interpretación clara y pragmática del grado de recomendación (fuerte o débil) para clínicos, pacientes y gestores.
- Evalúa explícitamente la importancia para los pacientes de las variables de resultado de las alternativas terapéuticas consideradas.

Algunos autores abogan por la recogida y valoración de elementos metodológicos individuales de cada estudio, en lugar de usar las controvertidas escalas<sup>29</sup>. La información dependerá del diseño de estudio seleccionado. En el caso concreto de los ECA, diseño más común en las RS sobre intervenciones, los elementos o dimensiones del diseño y la ejecución que están más relacionados con el riesgo de sesgo son: el tipo de ocultación de la asignación aleatoria, el tipo de enmascaramiento de las intervenciones, las pérdidas de seguimiento, el tipo de análisis o la interrupción precoz del ensayo por un beneficio aparente. En la figura 5 se presenta un ejemplo hipotético de una hoja de recogida de elementos metodológicos de ECA.

### Análisis y presentación de los resultados

La presentación resumida de los resultados de los estudios primarios, obtenidos mediante una metodología sistemática y reproducible, constituye en sí una RS cualitativa. El paso siguiente consiste en la combinación mediante métodos estadísticos de los resultados de los estudios primarios, es decir, el MA propiamente dicho.

Conceptualmente, el MA combina los resultados de dos o más estudios similares sobre una intervención particular, siempre que se hayan medido las mismas variables de resultado. El MA no consiste en una simple media aritmética de los resultados de los diferentes estudios, sino en una media ponderada. En otras palabras, el MA concede un mayor peso relativo a los estudios con mayor carga informativa, es decir, que tienen mayor tamaño y/o que presentan mayor número de eventos. Así, al combinar los resultados, se asigna un peso distinto a cada estudio, y se obtiene una media ponderada. Además, la combinación de los resultados tiene en cuenta tanto la variabilidad en un estudio como entre los estudios con el objeto de mejorar la validez de las conclusiones. Esto significa que si hay mucha variación entre los resultados de los estudios incluidos (heterogeneidad) podría no ser apropiado combinarlos estadísticamente. En este caso, se debería presentar únicamente los resultados (puede ser útil presentarlos de forma gráfica —plot en inglés—, pero sin combinarlos) y las características de los estudios individuales (en forma de tablas) sin más.

Básicamente existen dos modelos para combinar estadísticamente los resultados: el modelo de efectos fijos y el modelo de efectos aleatorios. El primero asume que el efecto del tratamiento es constante en todos los estudios, mientras que el segundo asume que sigue una distribución al azar entre los distintos estudios. En otras palabras, el modelo de efectos fijos asume que sólo hay una fuente de variabilidad en los resultados (la del estudio), mientras que el modelo de efectos aleatorios introduce una segunda fuente de variación entre los estudios. La consecuencia práctica de esto es que el modelo de efectos aleatorios suele producir estimaciones más conservadoras (intervalos de confianza más amplios) del efecto combinado. Emplear uno u otro modelo dependerá del juicio que realicemos sobre las similitudes y diferencias de los estudios que vamos a combinar, aunque normalmente se suele emplear los dos.

Ahora bien, ¿qué quiere decir que hay «heterogeneidad» entre los estudios y cómo se mide? Básicamente, se refiere al hecho de que, una vez ponderados, los resultados de los estudios individuales (el efecto de la intervención) difieren entre sí más de lo que cabría esperar por el azar. En otras palabras, ya sea por diferencias en el tipo de diseño empleado, en los métodos empleados para la recogida de información, en el tipo de análisis utilizado y/o en las

<p><b>1. Tipo de ocultación de la asignación aleatoria:</b></p> <ul style="list-style-type: none"> <li>• Aleatorización central (mediante llamada telefónica, página <i>web</i>, investigador independiente o realizada por farmacia)</li> <li>• Contenedores codificados con la medicación en estudio y el control</li> <li>• Sobres codificados consecutivamente, opacos y sellados</li> <li>• Sobres, sin especificar los tres elementos anteriores</li> <li>• Seudoaleatorización (lanzar una moneda, fecha de nacimiento, día de la semana, etc.)</li> <li>• Se menciona que hubo enmascaramiento de aleatorización, pero no se describe</li> <li>• No declarado</li> <li>• Se declara que no hubo enmascaramiento de la aleatorización</li> </ul> <p><b>2. Enmascaramiento o «cegado» (<i>blinding</i>) de la intervención (marcar donde aplique):</b></p> <ul style="list-style-type: none"> <li>• Pacientes</li> <li>• Personal sanitario que administra la medicación</li> <li>• Investigadores que recogen los datos</li> <li>• Investigadores que evalúan los eventos</li> <li>• Investigadores que realizan el análisis estadístico</li> </ul> <p><b>3. Participantes perdidos durante el seguimiento:</b></p> <p>N = ____ % = ____</p> <ul style="list-style-type: none"> <li>• No se declara</li> </ul> <p><b>4. Análisis de los datos</b></p> <ul style="list-style-type: none"> <li>• Por intención de tratar (se analiza a los pacientes en el brazo de la intervención al que fueron aleatorizados)</li> <li>• Por protocolo (se analiza a los pacientes según la intervención que finalmente recibieron)</li> <li>• Ambos</li> <li>• No se declara</li> </ul> <p><b>5. Se interrumpió el estudio precozmente por beneficio</b></p> <ul style="list-style-type: none"> <li>• Sí</li> <li>• No</li> </ul>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Figura 5.** Ejemplo de una hipotética hoja sobre elementos calidad de ensayos clínicos aleatorizados incluidos en una revisión sistemática.

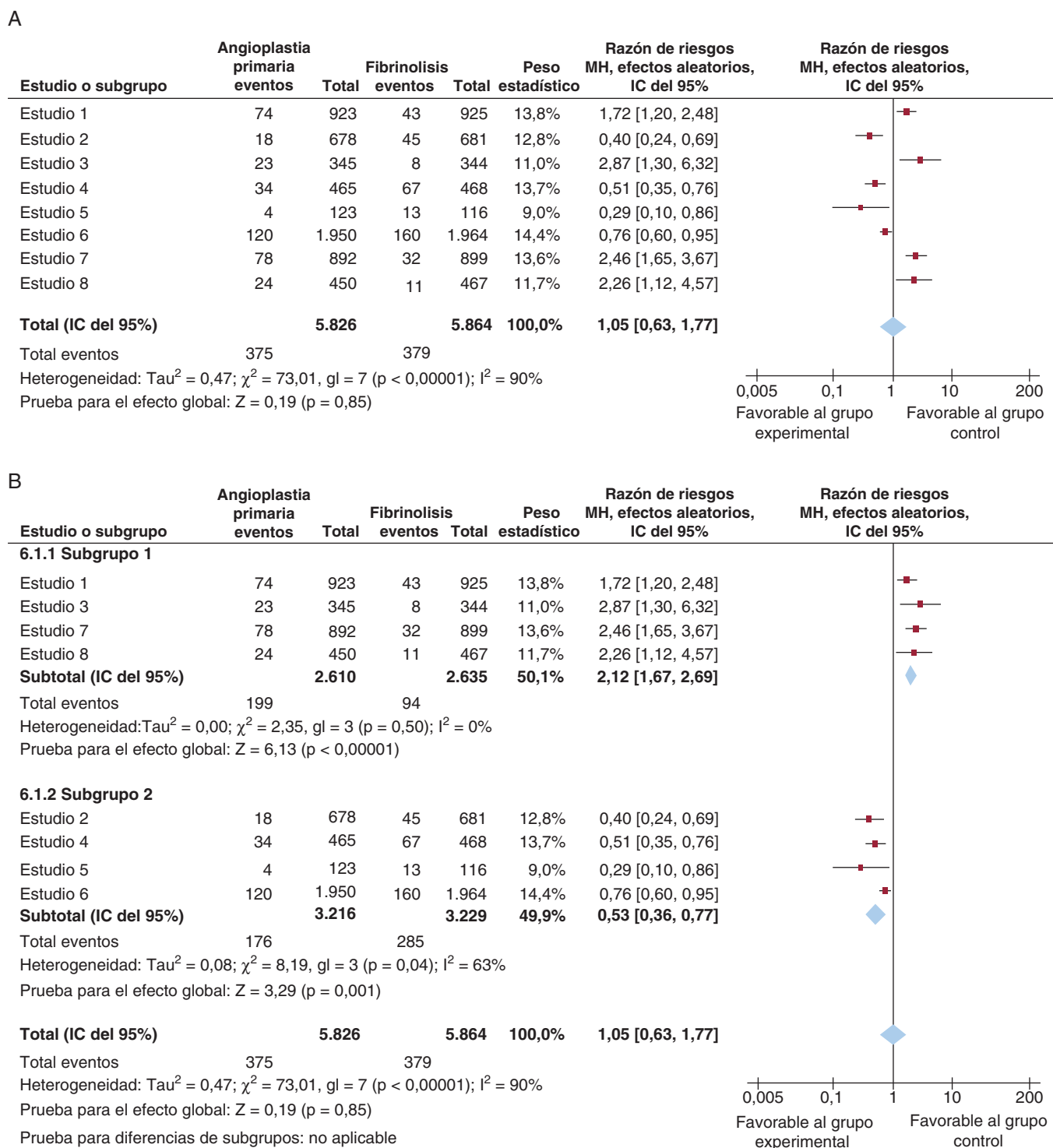
características de la población de estudio, el efecto de la intervención fue «diferente» en cada uno de ellos. Por ejemplo, imaginemos que un ensayo clínico sobre un antihipertensivo incluyó un 70% de población afroamericana y otro estudio, con el mismo fármaco, incluyó sólo un 10%. Y ahora imaginemos que el efecto de dicho antihipertensivo es muy potente en la población afroamericana y es nulo en el resto de la población. Evidentemente, el efecto de la intervención en el primer estudio será positivo y nulo en el segundo. Combinar ambos estudios no será apropiado, pues obtendremos una estimación del efecto «medio» del antihipertensivo que nos oculta una realidad mucho más rica y compleja, y nos confunde.

Existen varios estadísticos para cuantificar la heterogeneidad. Los más comunes son el estadístico Q, el H y el  $I^2$ . El más fácil de interpretar es el  $I^2$ . Indica la proporción de la variabilidad observada en el efecto de la intervención (entre estudios) que se debe a heterogeneidad entre los estudios y no al azar. Se suele considerar que, si es del 25%, hay poca heterogeneidad; del 50%, moderada, y del 75%, alta<sup>30</sup>.

Para la presentación gráfica de los resultados del MA se emplea el diagrama de árbol (*forest plot*). Este tipo de gráfico muestra los

datos de los estudios individuales junto con una representación del peso estadístico de cada estudio en relación con los intervalos de confianza y el error estándar de la media. Por ejemplo, supongamos que en nuestro ejemplo combinamos ocho estudios con objeto de analizar el efecto de la angioplastia respecto a la fibrinólisis de mayores de 75 años en la mortalidad cardiovascular. La **figura 6A** presenta el análisis combinado de los ocho estudios en forma de diagrama de árbol. Puede observarse que el efecto total no es concluyente y que, además, hay una importante heterogeneidad entre los estudios ( $I^2 = 90\%$ ).

Una interpretación simplista sería que la angioplastia primaria no es superior respecto a la fibrinólisis para reducir la muerte cardiovascular en pacientes mayores de 75 años. Sin embargo, la alta heterogeneidad entre estudios nos debe hacer sospechar que ese resultado es poco realista. De hecho, ya visualmente se comprueba que el efecto de la angioplastia en los estudios 2, 4, 5 y 6 fue muy superior a la fibrinólisis, mientras que en los estudios 1, 3, 7 y 8 fue inferior. Una lectura detallada de esos estudios revela que el tiempo medio puerta-balón en los que la angioplastia superó a la fibrinólisis fue < 90 min en todos ellos, mientras que el tiempo medio puerta-balón en los que la angioplastia fue peor que la fibrinólisis superó los



**Figura 6.** Metaanálisis del efecto de la angioplastia primaria frente a la fibrinólisis en la mortalidad total de pacientes con infarto de miocardio. A: considerando integralmente todos los estudios. B: análisis de subgrupos en función del tiempo puerta-balón; en la parte superior, estudios en que el tiempo medio puerta-balón excedió de 140 min, y en la parte inferior, aquellos en que el tiempo medio puerta balón fue < 90 min. IC: intervalo de confianza.

140 min en todos ellos. Entonces, estaría justificado realizar un análisis por subgrupos, analizando por separado los estudios con tiempo puerta-balón < 90 min y aquellos con tiempos más largos. La figura 6B presenta dichos resultados. Se puede observar cómo se reduce de forma importante la heterogeneidad en cada subgrupo y que los resultados son coherentes: importante beneficio de la angioplastia primaria con tiempos de puerta balón reducidos (parte inferior de la fig. 6B) y efecto inferior al de la fibrinólisis con tiempos superiores (parte superior de la fig. 6B).

Por último, hay que recalcar que existen guías y recomendaciones para la presentación y publicación de RS, cada una en forma de listado o lista de comprobación donde se enumeran los ítems que hay que declarar en la publicación. La guía MOOSE detalla los puntos específicos que ha de incluir una RS de estudios observacionales<sup>4</sup>, mientras que las guías PRISMA (que sustituyen a las QUORUM) se refieren a MA de ECA<sup>3</sup>. El lector interesado puede consultarlas por internet (<http://www.consort-statement.org>).



## Interpretación de los resultados

Finalmente, la RS concluye con la interpretación de los resultados. Ello incluye una discusión sobre las limitaciones del estudio (de la revisión), como potenciales sesgos de los estudios originales, así como potenciales sesgos que podrían afectar a la RS en sí misma. También es importante una discusión sobre la consistencia de los hallazgos y su aplicabilidad, así como proponer recomendaciones para futuras investigaciones sobre el tema de interés.

## REVISIONES SISTEMÁTICAS Y PRÁCTICA CLÍNICA

Por mucho que los hallazgos de una RS sean consistentes y convincentes, al final, es el clínico quien tiene que tomar una decisión sobre un paciente en particular. No deben tomarse los hallazgos de la RS como normas fijas e invariables fruto de la «ortodoxia evidencialista». Dicho de otra manera, hay que adaptar los hallazgos de una RS al paciente, y no al revés. En este sentido, antes de tomar una decisión sobre un paciente basándose en una RS, se recomienda la siguiente reflexión<sup>31</sup>:

### ¿Son aplicables los hallazgos a mi paciente?

La RS podría haber mostrado que la angioplastia primaria es superior a la fibrinólisis en pacientes ancianos. Pero imaginemos que mi paciente en particular presenta una condición que fue criterio de exclusión en todos los ensayos clínicos incluidos en la RS, como podría ser un aclaramiento de creatinina < 30 ml/min. En este caso, nuestro paciente en concreto no estaría representado por los ensayos clínicos de la RS.

### ¿Es factible la intervención en mi paciente?

Puede haber diferencias regionales en la disponibilidad y/o la experiencia de aplicación de una técnica concreta, que hay que tener en cuenta a la hora de aplicar la intervención a un paciente individual.

### ¿Cuál es el riesgo-beneficio para mi paciente?

Aunque la intervención sea factible y aplicable, se deben considerar en cada paciente concreto los riesgos específicos, generalmente poco representados en los ensayos clínicos.

### ¿Cuáles son los valores y las preferencias de mi paciente en particular?

Acostumbrados a tomar decisiones en función de marcadores de necrosis miocárdica, electrocardiogramas y demás pruebas complementarias, podemos caer en una medicina excesivamente paternalista, en la que tratamos al paciente con la mejor intención, según lo que creemos que es mejor para él, pero sin contar con él.

## CONCLUSIONES

Las RS constituyen una herramienta esencial para sintetizar la información científica disponible, incrementar la validez de las conclusiones de estudios individuales e identificar áreas de incertidumbre donde sea necesario realizar investigación. Además, son imprescindibles para la práctica de una medicina basada en la evidencia. Sin embargo, la realización de una RS se debe llevar a cabo siguiendo una estricta metodología y un control de calidad para evitar conclusiones sesgadas. En última instancia, es el clínico quien tiene que tomar la decisión sobre un paciente en particular, y la RS es una herramienta más que se ha de utilizar de forma juiciosa en la toma de decisiones.

## CONFLICTO DE INTERESES

Ninguno.

## BIBLIOGRAFÍA

1. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997;126:376-80.
2. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med.* 2007;4:e78.
3. 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. *J Clin Epidemiol.* 2009;62:e1-34.
4. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA.* 2000;283:2008-12.
5. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol.* 1995;48:167-71.
6. Stewart LA, Tierney JF, Clarke M. Reviews of individual patient data. En: Higgins JP, Green S, editores. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (actualizada sept 2008).* The Cochrane Collaboration; 2008.
7. Straus S, Moher D. Registering systematic reviews. *CMAJ.* 2010;182:13-4.
8. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club.* 1995;123:A12-3.
9. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ.* 2010;340:c117.
10. Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. *Epidemiol Rev.* 1992;14:154-76.
11. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet.* 1996;347:363-6.
12. Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol.* 1995;48:159-63.
13. Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med.* 1987;6:315-28.
14. Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci.* 1993;703:135-46.
15. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet.* 1991;337:867-72.
16. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med.* 1987;6:11-29.
17. Egger M, Smith GD. Misleading meta-analysis. *BMJ.* 1995;311:753-4.
18. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ.* 1994;309:1286-91.
19. Dickersin K, Larsen K. Section V: Establishing and maintaining an International register of RCTs. *The Cochrane Collaboration Handbook.* London: BMJ Publishing Group; 2010.
20. Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *Int J Technol Assess Health Care.* 2003;19:591-603.
21. Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. *Ann Intern Med.* 1997;127:531-7.
22. Streiner DL, Norman GR. Reliability. En: *Health measurement scales: a practical guide to their development and use.* New York: Oxford University Press; 1995.
23. Altman DG. Measuring agreement. En: *Practical statistics for medical research.* London: Chapman and Hall; 1991.
24. Guyatt GH, Rennie D, Meade MO. Summarizing the evidence. *Users' guides to the medical literature.* New York: McGraw Hill; 2009.
25. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care.* 1996;12:195-208.
26. Moher D, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials.* 1995;16:62-73.
27. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA.* 1999;282:1054-60.
28. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336:924-6.
29. Juni P, Altman DG, Egger M. Assessing the quality of randomized controlled trials. En: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care: metaanalysis in context.* London: BMJ Publishing Group; 2001.
30. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21:1539-58.
31. McAlister FA. Applying the results of systematic reviews at the bedside. En: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context.* London: BMJ Publishing Group; 2001.