Valoración crítica de artículos científicos. Parte 2: revisiones sistemáticas y metaanálisis

J. González de Dios[1], A. Balaguer Santamaría[2]

[1]Servicio de Pediatría. Hospital General Universitario de Alicante. Departamento de Pediatría. Universidad Miguel Hernández. ISABIAL-Instituto de Investigación Biomédica. Alicante. España. Comité de Trabajo de Pediatría Basada en la Evidencia de la Asociación Española de Pediatría.

[2]Servicio de Pediatría. Hospital Universitari General de Catalunya. Universitat Internacional de Catalunya. Barcelona. España. Comité de Trabajo de Pediatría Basada en la Evidencia de la Asociación Española de Pediatría.

RESUMEN

En el presente artículo se presentan las nociones básicas de una revisión sistemática y de un metaanálisis, junto con los principales instrumentos elaborados para la evaluación de su calidad metodológica. Las revisiones sistemáticas se elaboran de acuerdo con un procedimiento estructurado bien definido e intentan responder una pregunta clínica sobre un tema concreto. Deben definirse claramente la población o la enfermedad de interés, la intervención a valorar y los resultados clínicos que se consideren de importancia clínica.

Cuando los resultados de los diferentes estudios que conforman una revisión sistemática pueden combinarse (en base a criterios de sentido común y del resultado de pruebas estadísticas de heterogeneidad) se habla de metaanálisis. Cuando es posible combinar los resultados de varios estudios, el estimador combinado global suele ser la *odds ratio*, el riesgo relativo o la reducción absoluta del riesgo. Estas medidas han de acompañarse de su correspondiente intervalo de confianza del 95% para evaluar su precisión.

Para la correcta valoración crítica de una revisión sistemática (con o sin metaanálisis) existen diferentes listados de verificación de su calidad. Porque una revisión sistemática no es buena per se y, para ello, debe describir de forma completa y transparente su metodología: disponemos de la declaración PRISMA, como una lista de comprobación para este tipo de estudios y que viene a sustituir a la previa declaración QUOROM. Y, además, en la lectura crítica de este tipo de documentos se debe evaluar los tres aspectos clave en medicina basada en pruebas que pone CASP a nuestra disposición: la validez metodológica, la importancia clínica de sus resultados y la posible aplicabilidad de estos a la práctica clínica cotidiana del lector.

LAS REVISIONES COMO FUENTE DE INFORMACIÓN ACTUALIZADA: REVISIONES DE AUTOR Y REVISIONES SISTEMÁTICAS

En Medicina, difícilmente un estudio aislado puede aportar la solución definitiva de un problema médico, aunque este estudio tenga la calidad científica de un ensayo clínico aleatorizado. Se acepta que el avance científico se consigue con la acumulación de los resultados procedentes de varios estudios, pero es habitual que los resultados sean discrepantes. Un instrumento excelente para adquirir y actualizar nuestros conocimientos son las revisiones de la literatura médica. Su consulta puede

Cómo citar este artículo: González de Dios J, González de Dios J. Valoración crítica de artículos científicos. Parte 2: revisiones sistemáticas y metaanálisis. FAPap Monogr. 2021;6:14-26.

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN 2444-1953

proporcionarnos buenos resúmenes, puestos al día, difíciles de obtener en libros de texto. Muchas veces estos pueden ser los instrumentos para resolver un problema clínico que nos preocupa o nuestra necesidad de formación continuada. Sin embargo, a pesar de sus muchas virtudes, a nadie se escapan las deficiencias que arrastran muchas de ellas[1].

La revisión de la bibliografía es necesaria para depurar toda la desmesurada información científica a la que nos vemos sometidos, de forma que el problema es tanto cuantitativo (es difícil acceder a todo lo que se publica sobre un mismo tema) como cualitativo (es difícil conocer la importancia clínica de lo novedoso respecto a lo ya conocido)[2]. Sin embargo, la calidad de los artículos de revisión habituales dejan mucho que desear con frecuencia, debido a que las labores de recogida, análisis y publicación de los resultados no se realizan de una forma estructurada, explícita y sistemática, por lo que se constituyen en revisiones de autor subjetivas y científicamente dudosas. Cada día se tiene más claro que, ni la competencia científica del autor ni el prestigio de la revista biomédica, son criterios suficientes para la credibilidad de una revisión. Esta falta de método en las revisiones tiende a extraer conclusiones que, en el mejor de los casos, están sesgadas, y, en el peor de los casos, pueden ser erróneas[3].

Años atrás los impulsores del movimiento de la medicina basada en la evidencia (MBE), entre otros, ya llamaron la atención sobre la falta de rigor inherente a la metodología usada en el desarrollo de estas revisiones[4]. Particularmente destacaron el peligro de aquellas efectuadas por "expertos", que por su aparente garantía, podrían tener mayor resonancia. Estos podían dejarse llevar más fácilmente por sus particulares experiencias, cuando no por prejuicios o preferencias, a la hora de valorar los resultados procedentes de pruebas científicas. En cualquier caso, estos sesgos que fácilmente imbuyen las revisiones tradicionales de la literatura médica, pueden tratar de minimizarse usando una metodología predefinida y explícita. Con este propósito se han ido estructurando las llamadas revisiones sistemáticas (o sistematizadas)[5-7]. En la Tabla 1 se esquematizan algunas de las diferencias entre revisiones tradicionales (narrativas o de autor o de "experto") y las revisiones sistemáticas[7,8].

Actualmente las revisiones sistemáticas de calidad, ya con una metodología bien definida, han sido consideradas una de las mejores fuentes de "evidencia" científica disponible[1,5-10]. Porque una de las maneras de resolver la discrepancia del exceso de información sobre un tema es recurrir a los métodos de síntesis de información sobre un mismo tema, bien cualitativa (revisión sistemática [RS]) o cuantitativa (metaanálisis [MA]).

Las RS consisten básicamente en resúmenes bibliográficos, idealmente efectuados por más de una persona, siguiendo una metodología estructurada (al estar definida en diferentes etapas), explícita (ya que determina los diferentes tipos de diseños en cada una de las fases) y sistemática (en cuanto que pretende acceder a toda la información disponible). La elaboración de una RS sigue una metodología bien establecida[1,5-10]: definir el objetivo, identificar exhaustivamente la metodología de búsqueda de la información, seleccionar los datos con criterios de inclusión y exclusión definidos, valorar la calidad y validez de los estudios, sintetizar la información y analizar los resultados, extraer conclusiones en base a los datos de la revisión.

Sin embargo, todas las RS o MA no son sinónimo de calidad y estos métodos de síntesis de información son mejorables: como toda investigación retrospectiva tienen limitaciones, principalmente por la calidad de la información disponible. Además, la reproducibilidad de algunas RS no es buena[11] y, a veces, los resultados no coinciden con los de los grandes EC[12]. Lo que es peor, a menudo no está claro qué es más cierto: el resultado de una RS de pequeños EC (que suele tener alguna validez externa, porque se han realizado con poblaciones y métodos variados) o el de un gran EC (posiblemente con mayor validez interna)[12,13].

Tabla 1. **Características diferenciales de los distintos tipos de revisiones.**

	Revisión narrativa	Revisión sistemática
Objetivo	Tema general	Pregunta clínica estructurada
Pregunta	Frecuentemente amplia	Pregunta concreta
Fuentes-Búsqueda	Estrategia no especificada	Estrategia especificada
Selección	Criterios no explícitos	Criterios explícitos y rigurosos
Evaluación	Variable	Crítica y rigurosa (según criterios explícitos)
Síntesis	Cualitativa	Cuantitativa (si es posible) o cualitativa
Inferencias	Variable	Basadas en "evidencias"

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

En el presente artículo se realiza una valoración de las listas-guías para la correcta presentación e interpretación de las RS o MA, así como la evaluación de la lectura crítica de estos documentos científicos.

VALORACIÓN DE LAS LISTAS-GUÍAS EN LAS REVISIONES SISTEMÁTICAS Y METAANÁLISIS: DECLARACIÓN PRISMA

Debido al creciente número de RS o MA que se publican, puede ser de una gran utilidad contar con métodos que ayuden a evaluar de forma rápida y sistemática su calidad: son las listas-guía, cuestionarios o *check-lists*. Además de su objetivo fundamental, que es valorar si los métodos utilizados y los resultados obtenidos son adecuados para producir información útil, las listas-guía cumplen también una importante labor informativa, tanto en la evaluación crítica de distintos tipos de estudios publicados como en la evaluación de protocolos. Existen dos tipos fundamentales de listas-guía: cuantitativas (se asigna una puntuación a cada apartado) y cualitativas (se analiza globalmente cada apartado, lo que permite un análisis más matizado).

Al plantearse llevar a cabo una RS o MA, o para revisar una ya realizada por otros autores, es útil emplear una lista de verificación como guía para descubrir errores importantes, como la propuesta en la Tabla 2. En dicha tabla se presenta una lista de preguntas clave para valorar una RS o MA, lo que puede ser útil tanto al posible autor como al lector. La evaluación sistemática de estas preguntas ayuda a interpretar de forma crítica y objetiva las RS o MA.

Para impulsar la mejora y uniformidad de las RS y MA un grupo de investigadores publicaron en 1999 en Lancet unas recomendaciones identificadas por el acrónimo QUOROM (*Quality Of Reporting Of Meta-analysis*). En ellas, con el ánimo de mejorar y homogeneizar la calidad de la presentación de este tipo de estudios con alta "evidencia" científica, se destacan los puntos que deberían exigirse[14,15]. QUOROM apareció como una lista de comprobación estructurada de 18 ítems (y que incluía un diagrama de flujo) cuyo objetivo era proporcionar toda aquella información que resulta esencial para escribir, interpretar y utilizar adecuadamente los resultados de un MA. La declaración QUOROM no logró la misma aceptación por parte de los editores de revistas que la declaración CONSORT para los ensayos clínicos[16]. Se comenta que una de las causas de esta menor implantación podría ser que no se han realizado suficientes estudios que demuestren su impacto en la mejoría de la calidad de los MA después de su inclusión.

Ya en el momento de su publicación, el grupo de trabajo de QUOROM estableció la necesidad de que debiera ser actualizado periódicamente. Actualizar QUOROM ha sido una labor que se ha llevado a cabo durante varios años por un heterogéneo grupo de de 29 profesionales de distintos países con el nombre de grupo *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA), y en el que tuvieron una labor fundamental distintas universidades (Otawa, Oxford, Módena, McMaster, Friburgo, Maastricht, San Francisco, etc.), centros Cochrane (Oxford, Italia, Noruega, etc.), así como la International Network of Agencies for Health Technology Assessment (INAHTA) o Guideline International Network (GIN). No en vano muchos de los autores del grupo de trabajo PRISMA estaban también involucrados como asesores metodológicos de la Colaboración Cochrane.

La declaración PRISMA es una lista de comprobación con 27 puntos ítems (frente a los 18 de QUOROM) y un diagrama de flujo de 4 fases[17,18]. PRISMA se convierte así en un avance significativo: a diferencia de QUOROM, la declaración PRISMA viene acompañada de un extenso documento donde se detalla la explicación o la justificación de cada uno de los 27 ítems propuestos, así como el proceso de elaboración de estas directrices a partir de siete cuadros explicativos sobre ciertos aspectos clave de las RS.

Los autores de PRISMA identifican cuatro aspectos conceptuales novedosos que conllevan la adición de nuevos ítems a la lista de comprobación[17,18]:

1. El carácter iterativo del proceso de desarrollo de una RS, dado que es muy importante que los métodos que utilice la revisión se establezcan y se documenten *a priori*.
2. La conducción y la publicación de un estudio de investigación son conceptos distintos.
3. Evaluación del riesgo de sesgo de los estudios o de los resultados.
4. Importancia de los sesgos de publicación, bien por omisión de estudios completos o por omisión de resultados dentro de un estudio individual.

De esta forma, y comparado con QUOROM, los cambios más relevantes introducidos en la lista de comprobación de PRISMA son [17,18]:

- En el título: PRISMA solicita la identificación de la publicación como RS, MA o ambos, mientras que QUOROM se refería exclusivamente a MA de ensayos clínicos.
- En el resumen: tanto QUOROM como PRISMA solicitan a los autores un resumen estructurado, pero PRISMA no especifica su formato, aunque sí hace recomendaciones.

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

Tabla 2. **Lista de verificación (*check-list*) que permite la detección de errores importantes en una revisión sistemática o metaanálisis.**

1. Formulación de la pregunta
• ¿Las preguntas de la revisión están bien formuladas y contienen los componentes clave: la población diana, la exposición o intervención en estudio y el resultado de valoración estudiado? • ¿Está suficientemente justificada la realización de la RS o MA? • ¿Están descritos los objetivos de forma clara y precisa? ¿Son relevantes los objetivos?
2. Identificación de estudios
• ¿Se realiza una búsqueda minuciosa y sistemática de estudios pertinentes y usando fuentes apropiadas? • ¿Las estrategias de búsqueda son adecuadas para la pregunta que se ha planteado? • ¿Se describen los métodos de búsqueda bibliográfica con suficiente detalle? • ¿Se han utilizado estrategias de búsqueda complementaria? • ¿Se han intentado localizar trabajos en literatura "opaca"? • ¿Se han intentado localizar trabajos no publicados para evitar el sesgo de publicación?
3. Selección de estudios
• ¿Se aplican los criterios de inclusión y de exclusión apropiados para seleccionar los estudios? • ¿Se aplican los criterios de selección de forma específica, detallada y reproducible, de manera que limiten los sesgos? ¿Se establecieron *a priori*?
4. Evaluación de los estudios
• ¿Se valora de manera fiable la validez y calidad de los estudios individuales? • ¿Se abordan de modo sistemático los parámetros importantes que pueden afectar a los resultados del estudio? • ¿Se siguieron criterios explícitos y reproducibles para evaluar la calidad de los estudios? ¿Se establecieron *a priori*?
5. Obtención de datos
• ¿Se describen los métodos para la extracción de los datos de los estudios originales? • ¿Se evaluó la reproducibilidad de la extracción de los datos? • ¿Hay una mínima cantidad de datos no disponibles en los resultados y en otras variables que se consideren clave para la interpretación de los resultados?
6. Síntesis de datos
• ¿Se tomaron decisiones razonables con respecto a la manera de combinar los datos? • ¿Se han considerado en la síntesis de factores importantes como el diseño de los estudios? • ¿Son sensibles los resultados a los cambios relacionados con el modo en que se hizo el análisis? • ¿Se describen los métodos utilizados para el análisis estadístico? • ¿Se describen los métodos utilizados para evaluar la heterogeneidad de los resultados y las fuentes de heterogeneidad? • ¿Con qué métodos se combinan los estudios originales? • ¿Se realizan análisis de sensibilidad de los resultados? • ¿Se presentan los resultados de los estudios originales y de la RS o MA de forma objetiva e imparcial?
7. Discusión
• ¿Se discuten los resultados de la RS o MA de forma equilibrada? • ¿Se mencionan las limitaciones de los estudios y del proceso de revisión? • ¿Se integran los resultados de la revisión en el contexto de otra evidencia indirecta? • ¿Se analizan las posibles discrepancias con otros trabajos publicados? • ¿Se discute la solidez de las inferencias según criterios de causalidad relevantes? • ¿Se concretan las consecuencias para la práctica clínica o de salud pública de los resultados?
8. Conclusiones de los revisores
• ¿Se apoyan las conclusiones en el contenido de la revisión? • ¿Se abordan explicaciones alternativas y verosímiles de los efectos observados? • ¿Hay interpretaciones adecuadas acerca de la evidencia que no es concluyente (sin evidencia de efecto frente a evidencia de que no hay efecto)? • ¿Se identifican las consideraciones que resultan importantes para las personas que deben adoptar decisiones? • ¿Existen "otros" factores que puedan influir en las conclusiones que alcanzan los investigadores, como conflictos de interés, fuentes de financiación, etc.?

- En la introducción: incluye el ítem 4 (objetivo), que se focaliza en la pregunta explícita que aborda la revisión utilizando el formato PICO (descripción de los participantes, las intervenciones, las comparaciones y las medidas de resultado de la revisión sistemática), así como el tipo de estudio (diseño).
- En el apartado de "Métodos" es donde encontramos mayores cambios: incluye el ítem 4 (protocolo) en el que se

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

solicita al autor que explique si un protocolo precedió la revisión y, en tal caso, cómo puede accederse a este. Así mismo, PRISMA desdobla el ítem sobre la "búsqueda" de QUOROM en dos ítems: 1) fuentes de información (ítem 7) y 2) búsqueda (ítem 8). Aunque la estrategia de búsqueda se reporta tanto en QUOROM como en PRISMA, este solicita a los autores que proporcionen la descripción completa de, al menos, una estrategia de búsqueda electrónica (ítem 8). También es importante la evaluación del riesgo de sesgo en los estudios incluidos: en QUOROM este ítem aparecía como "evaluación de la calidad", pero ahora este ítem (12) se focaliza en la evaluación del riesgo de sesgo dentro de cada estudio incluido en la RS y también se introduce el nuevo concepto de evaluación del sesgo al nivel de los resultados (ítem 19).

En la Tabla 3 se esquematizan estas diferencias entre PRISMA y QUOROM.

Como decíamos previamente, uno de los retos reales de PRISMA será poder de evaluar las ventajas (calidad mejorada de la publicación) y las posibles limitaciones (aumento de la extensión del texto) en la publicación de RS o MA. Durante el proceso de presentación del manuscrito, los editores pueden exigir a los autores completar la lista de comprobación de PRISMA e incluirla como parte del proceso de remisión del artículo. Como parte del proceso de revisión, los editores pueden sugerir a los revisores usar la lista para guiarlos durante la evaluación del artículo. Finalmente, los editores pueden utilizar PRISMA durante el proceso de revisión y decisión editorial.

Con esta lista de comprobación, las RS y MA no serán buenos per se, sino que "dependerán del PRISMA con que se miren". Disponemos ya del "prisma"; resta utilizarlo y demostrar que mejora la calidad de las RS o MA. En la Tabla 4 se exponen los 27 ítems de la declaración PRISMA[17-19]. Estos se clasifican en 7 secciones, que son las clásicas IMRD de cualquier artículo original (Introducción, Métodos, Resultados y Discusión), con dos secciones previas (Título y Resumen) y una posterior (Financiación).

Este es el "prisma" actual con el que hay que elaborar, revisar y valorar las RS y MA, dos tipos de artículos esenciales en el contexto de la "infoxicación" en el que nos movemos. Los autores ya hemos hablado de los conceptos básicos de las RS y MA[1,5] y de su valoración crítica[6,7,16]. Ahora es el momento de agrupar estos conocimientos y encuadrarlos en la declaración PRISMA. Una buena herramienta[7].

LECTURA CRÍTICA DE PUBLICACIONES SOBRE REVISIONES SISTEMÁTICAS O METAANÁLISIS

En las revisiones narrativas (tradicionales o de autor) alguien, normalmente un experto, expone los datos científicos existentes sobre un tema determinado, por lo que, en su mayor parte, constituyen simples resúmenes actualizados de la literatura. Las RS, a diferencia de las anteriores, utilizan de manera sistemática métodos para identificar, evaluar y sintetizar si es posible toda la información existente sobre un problema determinado. Cuando se integran resultados de los estudios incluidos en una revisión, se habla de MA. Pero un MA puede no estar basado en una revisión sistemática de la literatura; o puede que una RS de la evidencia no conlleve ninguna técnica estadística para integrar los resultados de los estudios primarios incluidos en ella.

La lectura crítica de de documentos científicos es un proceso fundamental en el contexto de la Medicina basada en la evidencia, en la que existen tres preguntas fundamentales a contestar en cualquier artículo[6,7,16]: juzgar si son válidas desde el punto de vista científico (próximas a la verdad y con rigor científico), decidir si son importantes (y, en consecuencia, valiosas y

Tabla 3. **Diferencias más relevantes entre declaración PRISMA y QUOROM.**

Sección/ítem	PRISMA (27 ítems)	QUORUM (18 ítems)
Título	RS, MA (o ambos) de distintos tipos de estudios	Solo MA de ensayos clínicos
Resumen	Se solicita No especifica formato	Se solicita Estructurado
Introducción/objetivo	PICO Tipo de estudios	Fundamentos biológicos y de la intervención
Métodos/protocolos	Si un protocolo precedió la revisión (acceso)	
Métodos/búsqueda	Desdobla en 2 ítems: fuentes y búsqueda	Búsqueda
Métodos/riesgo de sesgo en los estudios incluidos	Cambio de concepto: sesgo frente a calidad. Sesgo en resultados	Evaluación de la calidad

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

Tabla 4. **Declaración PRISMA para la presentación de una revisión sistemática y metaanálisis.**

Título y resumen	
Título	Identificar la publicación como RS, MA o ambos. Si se especifica que se trata de una revisión sobre ensayos clínicos, ayudará a priorizarla frente a otras revisiones
Resumen	
Resumen estructurado	Que incluya: antecedentes; objetivos; fuente de los datos; criterios de elegibilidad de los estudios, participantes e intervenciones; evaluación de los estudios y métodos de síntesis; resultados; limitaciones; conclusiones e implicaciones de los hallazgos principales; número de registro de la RS
Introducción	
Justificación	Implica describir lo que se conoce del tema y comentar la necesidad de la RS haciendo referencia a la importancia del problema y a la controversia existente sobre las intervenciones que se desea revisar
Objetivos	Plantear de forma explícita las preguntas que se desea contestar en relación con los participantes, las intervenciones, las comparaciones, los resultados y el diseño de los estudios (acrónimo PICOS)
Métodos	
Protocolo y registro.	Indicar si existe un protocolo de revisión al que se pueda acceder (por ej., dirección web) y, si está disponible, la información sobre el registro, incluyendo su número de registro
Criterios de elegibilidad.	Especificar las características de los estudios (por ej., cada uno de los elementos del acrónimo PICOS, duración del seguimiento) y de las características de la búsqueda (por ej., años abarcados, idiomas o estatus de publicación) utilizadas como criterios de elegibilidad y su justificación
Fuentes de información	Describir todas las fuentes de información (por ejemplo, bases de datos y periodos de búsqueda, contacto con los autores para identificar estudios adicionales, etc.) y la fecha de la última búsqueda realizada
Búsqueda	Presentar la estrategia completa de búsqueda electrónica en, al menos, una base de datos, incluyendo los límites utilizados, de tal forma que pueda ser reproducible
Selección de los estudios	Especificar el proceso de selección de los estudios, con los criterios de inclusión y exclusión. Definir la elegibilidad de los estudios incluidos en la RS y, cuando sea pertinente, en el MA
Proceso de extracción de datos	Describir los métodos para la extracción de datos de las publicaciones (por ejemplo, formularios pilotados, por duplicado y de forma independiente) y cualquier proceso para obtener y confirmar datos por parte de los investigadores
Lista de datos	Listar y definir todas las variables para las que se buscaron datos (por ejemplo, PICOS, fuente de financiación) y cualquier asunción y simplificación que se hayan hecho
Riesgo de sesgo en los estudios individuales	Describir los métodos utilizados para evaluar el riesgo de sesgo en los estudios individuales (especificar si se realizó al nivel de los estudios o de los resultados) y cómo esta información se ha utilizado en la síntesis de datos
Medidas de resumen	Especificar las principales medidas de resumen (por ejemplo, razón de riesgos o diferencia de medias)
Síntesis de resultados	Describir los métodos para manejar los datos y combinar resultados de los estudios, cuando esto es posible, incluyendo medidas de consistencia para cada MA
Riesgo de sesgo entre los estudios	Especificar cualquier evaluación del riesgo de sesgo que pueda afectar la evidencia acumulativa (por ejemplo, sesgo de publicación o comunicación selectiva)
Análisis adicionales	Describir los métodos adicionales de análisis (por ejemplo, análisis de sensibilidad o de subgrupos, metarregresión) y, en el caso de que se hiciera, indicar cuáles fueron preespecificados
Resultados	
Selección de estudios	Facilitar el número de estudios cribados, evaluados para su elegibilidad e incluidos en la RS, y detallar las razones para su exclusión en cada etapa, idealmente mediante un diagrama de flujo. Este se convierte en un elemento esencial en las RS, en donde se valora el número de registros en cada etapa: identificación, cribado, elegibilidad e inclusión (ver figura 1)
Características de los estudios	Para cada estudio, presentar las características para las que se extrajeron los datos (por ejemplo, tamaño, PICOS y duración del seguimiento) y proporcionar las citas bibliográficas
Riesgo de sesgo en los estudios	Presentar datos sobre el riesgo de sesgo en cada estudio y, si está disponible, cualquier evaluación del sesgo en los resultados (ver ítem 12)
Resultados de los estudios individuales	Para cada resultado considerado en cada estudio (beneficios o daños), presentar: a) el dato resumen para cada grupo de intervención, y b) la estimación del efecto con su intervalo de confianza, idealmente de forma gráfica mediante un diagrama de bosque (*forest plot*)

continúa

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

Tabla 4. **Declaración PRISMA para la presentación de una revisión sistemática y metaanálisis (*cont.*).**

Síntesis de los resultados	Presentar los resultados de todos los MA realizados, incluyendo los intervalos de confianza y las medidas de consistencia
Riesgo de sesgo entre los sujetos	Presentar los resultados de cualquier evaluación del riesgo de sesgo entre los estudios (ver ítem 15)
Análisis adicionales	Facilitar los resultados de cualquier análisis adicional, en el caso de que se hayan realizado (por ej., análisis de sensibilidad o de subgrupos, metarregresión (ver ítem 16)
Discusión	
Resumen de la evidencia	Resumir los hallazgos principales, incluyendo la fortaleza de las evidencias para cada resultado principal; considerar su relevancia para grupos clave (por ejemplo, proveedores cuidados, usuarios y decisores en salud)
Limitaciones	Discutir las limitaciones de los estudios y de los resultados (por ejemplo, riesgo de sesgo) y de la revisión (por ejemplo, obtención incompleta de los estudios identificados o comunicación selectiva
Conclusiones	Proporcionar una interpretación general de los resultados en el contexto de otras evidencias, así como las implicaciones para la futura investigación. En las RS/MA son esenciales los apartados finales de "implicaciones para la práctica clínica" e "implicaciones para la investigación"
Financiación	
Financiación	Describir las fuentes de financiación de la RS y otro tipo de apoyos (por ejemplo, aporte de los datos), así como el papel de los financiadores en la RS

relevantes en potencia para el lector en su condición de clínico) y aplicables a la práctica clínica habitual (trasladar los resultados de la evidencia científica a la práctica clínica en nuestro[s] paciente[s]). Una buena manera de abordar una lectura crítica de una RS es la propuesta por el grupo Critical Appraisal Skills Programme (CASP), que es un programa del servicio inglés de salud que ayuda a adquirir habilidades para hacer lectura crítica y obtener así la "evidencia" científica necesaria para las decisiones clínicas. En España existe un grupo CASP, que se denomina CASPe (Programa de habilidades en lectura crítica), y desde el que podemos descargar también las preguntas orientadas para realizar lectura crítica de documentos científicos, entre ellos la RS [20].

Analizaremos a continuación los pasos a seguir para analizar la validez, la importancia y la aplicabilidad (Tabla 5) de los artículos sobre RS o MA, planteando tres preguntas clave:

- ¿Son válidos los resultados?
- ¿Cuáles son los resultados?
- ¿Son aplicables en tu medio?

¿Son válidos los resultados de la RS o MA?

Estas preguntas definen los criterios que se utilizan para evaluar la validez interna (o rigor científico) de las RS o MA, lo que permitirá estar razonablemente convencidos de que sus resultados no están sesgados y proporcionan una respuesta correcta a la pregunta planteada en su objetivo.

Criterios primarios o preguntas "de eliminación"

¿Se hizo la revisión sobre un tema claramente definido? El primer criterio para valorar una RS es si aborda una pregunta clínica concreta e importante, que debe estar definida con mu-

Tabla 5. **Preguntas para la valoración crítica de artículos sobre revisiones sistemáticas o metaanálisis.**

A. ¿Son válidos los resultados del estudio?

A.1. Preguntas de eliminación:

1. ¿Se hizo la revisión sobre un tema claramente definido?
2. ¿Buscaron los autores el tipo de artículos adecuados?

A.2. Preguntas de detalle:

3. ¿Estaban incluidos los estudios importantes y relevantes?
4. ¿Se ha hecho lo suficiente para valorar la calidad de los estudios incluidos?
5. Si los resultados de los diferentes estudios han sido mezclados para obtener un resultado "combinado", ¿era razonable hacer esto?

B. ¿Cuáles son los resultados?

6. ¿Cuál es el resultado global de la revisión? ¿Han sido interpretados con sentido común?
7. ¿Cuán precisos son los resultados?

C. ¿Son los resultados aplicables al escenario?

8. ¿Se pueden aplicar los resultados en tu medio?
9. ¿Se han considerado todos los resultados importantes para tomar la decisión?
10. ¿Los beneficios merecen la pena frente a los perjuicios y costes?

Según orientación del programa CASPe. Disponible en http://www.redcaspe.org. En la mayoría de las preguntas se pueden dar tres tipos de respuesta: Sí, No, No se puede saber.

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

cha claridad, para poder evaluar si los estudios que se han incluido son relevantes. La mayor parte de problemas clínicos pueden formularse en términos de una simple relación entre condición de interés (patología de estudio), población, alguna exposición (a un tratamiento, prueba diagnóstica, potencial efecto adversos, etc.), y uno o más resultados de interés.

¿Buscaron los autores el tipo de artículos adecuados? Los criterios utilizados para la identificación y selección de los artículos candidatos a ser incluidos deben estar definidos explícitamente y ser apropiados a la pregunta planteada. El mejor tipo de estudio es el que se dirige a la pregunta objeto de la revisión (pacientes, intervenciones y resultados) y tiene un diseño apropiado. El lector debe conocer los criterios que los autores utilizaron para seleccionar los estudios incluidos en la revisión. Asimismo, es preciso que especifiquen los estándares metodológicos utilizados para seleccionar los estudios (por ejemplo, para RS de cuestiones terapéuticas deben utilizarse preferentemente EC controlados) y estos patrones deben ser similares a los criterios de validez primaria descritos para los artículos originales de investigación[21-24] (Tabla 6). Además del tipo de diseño, deben valorarse el tipo de personas y ámbito de las intervenciones evaluadas y las de comparación, los grupos control y los resultados clínicos que interesa estudiar, de forma que se incluyan estudios relativamente homogéneos en cuanto al problema de salud y al abordaje metodológico que realizan del mismo. Hay que tener en cuenta que diferencias en estos criterios podría llevar a diferentes resultados en estudios de revisión que inicialmente se planteaban una misma pregunta clínica. También debe tenerse presente que diferencias en los tres elementos esenciales de todo estudio clínico (pacientes, intervenciones y resultados) pueden conducir a conclusiones diferentes en revisiones que parecen abordar el mismo problema clínico.

Las dos primeras preguntas son de "eliminación" y se pueden responder rápidamente, con lo que es probable que desechemos la mayor parte de las revisiones; solo si la respuesta es afirmativa en ambas merece la pena continuar con la tercera pregunta. Tal como sugieren Oxman *et al*.[25], si el principal problema que aborda una revisión no está claro a partir del título o del resumen, y no se utiliza el tipo adecuado de estudios, probablemente es buena idea dejar de leer y pasar al siguiente artículo de revisión.

Criterios secundarios o preguntas "de detalle"

¿Estaban incluidos los estudios importantes y relevantes? La validez de una revisión depende de que se hayan evaluado todas las pruebas científicas disponibles y que no han quedado excluidos estudios relevantes. Una búsqueda global y no sesgada de la literatura biomédica es una de las diferencias clave entre una RS y una revisión narrativa. Es importante que los autores lleven a cabo una búsqueda completa de los estudios que cumplan sus criterios de inclusión. La recuperación de artículos de las bases de datos bibliográficos electrónicas puede variar mucho en función de la estrategia de búsqueda utilizada. Es preciso, por tanto, que los autores especifiquen cómo han identificado los artículos seleccionados, así como las bases de datos consultadas (idealmente no obviar PubMed/Medline y Embase). La búsqueda realizada exclusivamente por medios electrónicos con frecuencia es insuficiente, por lo que los autores deberían hacer un seguimiento de las referencias de los es-

Tabla 6. **Guía para la selección de artículos con mayores probabilidades de proporcionar resultados válidos.**

Tratamiento
• ¿Se ha realizado de forma aleatoria la asignación de los tratamientos a los pacientes? • ¿Se han tenido en cuenta adecuadamente todos los pacientes incluidos en el ensayo clínico y se los ha considerado a la conclusión de este? • ¿Se ha realizado un seguimiento completo? • ¿Se han analizados los pacientes en los grupos a los que fueron asignados aleatoriamente, según análisis por intención de tratar?
Diagnóstico
• ¿Existió una comparación ciega, independiente, con un patrón de referencia adecuado? • ¿Incluyó la muestra de pacientes a un espectro adecuado de pacientes a los cuales en la práctica clínica se aplicará el examen diagnóstico?
Efectos perjudiciales de intervenciones o exposiciones
• ¿Se han utilizado grupos de comparación similares en cuanto a otros factores determinantes del efecto diferentes a la exposición? • ¿Se han valorado de la misma forma en todos los grupos de comparación las exposiciones y los efectos? • ¿El seguimiento ha sido suficientemente largo y completo?
Pronóstico
• ¿La muestra de pacientes está bien definida y es representativa de pacientes en un momento similar del curso de la enfermedad? • ¿El seguimiento ha sido suficientemente largo y completo?

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

tudios identificados y, si fuera necesario, contactar con expertos para la identificación de estudios no publicados. La inclusión de estudios no publicados disminuye las posibilidades de "sesgo de publicación" (probabilidad más elevada de ser publicados los estudios con resultados positivos y el riesgo que esto comporta de sobrestimar la eficacia) [26]. El diagrama de flujo propuesto por el grupo PRISMA (Figura 1) permite evaluar gráficamente cómo se han seleccionado los estudios[17,18].

¿Se ha hecho lo suficiente para valorar la calidad de los estudios incluidos? Es importante conocer la calidad de los estudios primarios incluidos en la revisión, porque la revisión de expertos no garantiza la validez de la investigación publicada. Las diferencias metodológicas de los estudios pueden explicar importantes diferencias entre los resultados (los estudios menos rigurosos tienen tendencia a sobreestimar la efectividad de las intervenciones terapéuticas). No existe una única forma correcta de evaluar la validez de los estudios, sino que existen multitud de métodos y debe utilizarse alguno de ellos. Los criterios utilizados deben ser explícitos y referidos tanto a aspectos generales, comunes a todos los estudios, como específicos del tema de investigación. Dado que esta evaluación está sujeta a variabilidad, es importante que se haya realizado de forma independiente por más de un observador y que se haya evaluado su grado de concordancia en cuanto a qué estudios incluir y qué datos extraer de ellos. Cada una de estas decisiones requiere la opinión de los revisores y cada una de ellas está sujeta tanto a errores aleatorios como a sesgos o errores sistemáticos.

Se han diseñados muchos instrumentos para evaluar los diferentes aspectos de una revisión, pero pocos de ellos permiten una evaluación crítica integral. La herramienta Ameasurement Tool to Assess Systematic Reviews (AMSTAR) fue desarrollado para evaluar RS de ensayos aleatorizados [27] y, más adelante, AMSTAR-2 [28], que permite una evaluación más detallada de las RS que incluyen también estudios no aleatorizados de intervenciones sanitarias, que son cada vez más incorporados en las RS. AMSTAR-2 es un cuestionario que contiene 16 dominios con opciones de respuesta si, no y si parcial. De ellos, siete dominios son considerados críticos (entre ellos, el ítem 9, riesgo de sesgo de los estudios individuales incluidos), dado que pueden afectar sustancialmente la validez de una RS y sus conclusiones, y de donde surgen cuatro niveles de confianza: alta, moderada, baja y críticamente baja.

Si los resultados de los diferentes estudios han sido mezclados para obtener un resultado "combinado", ¿era razonable

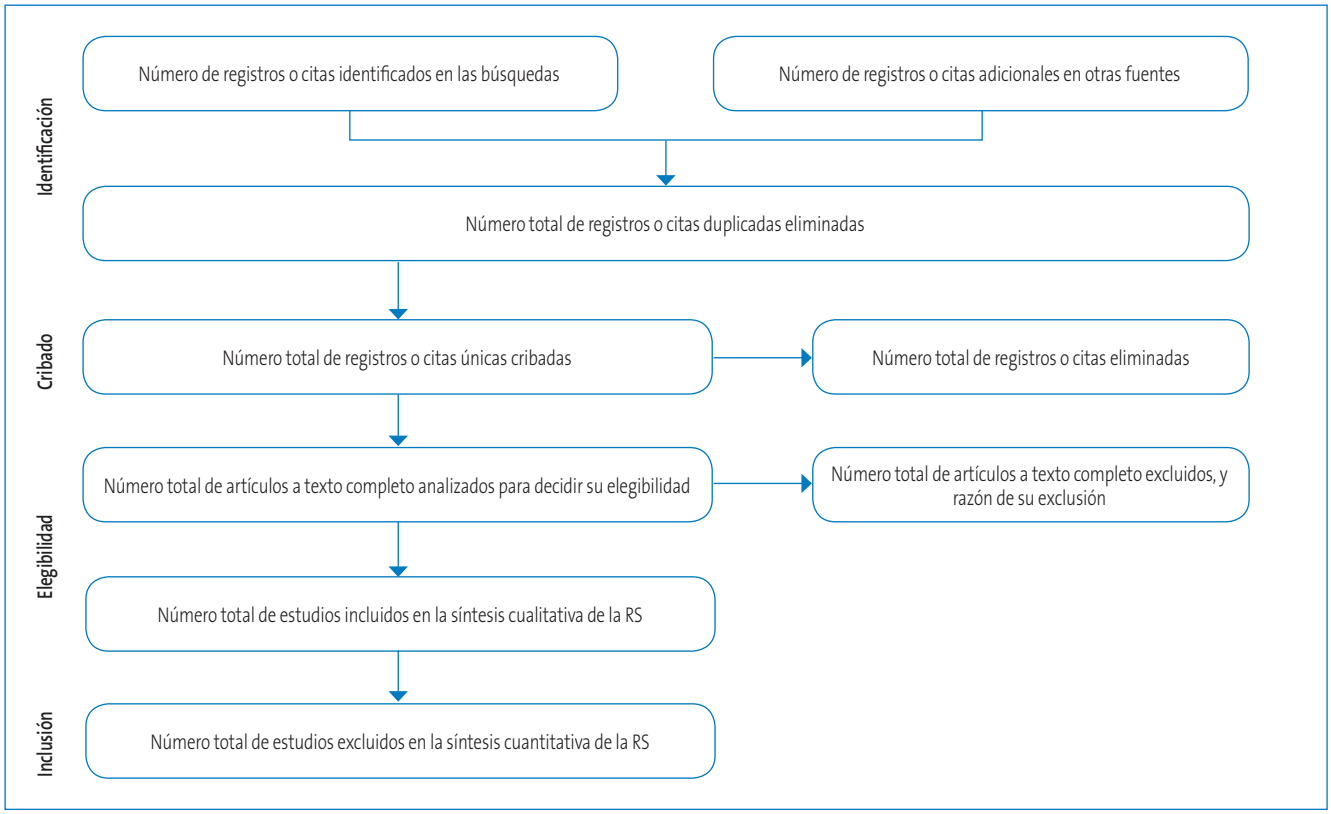

Figura 1. **Diagrama de flujo de la información a través de las diferentes fases de una revisión sistemática.**

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

hacer eso? Para poder obtener una estimación global del efecto de interés a partir de diversos estudios, debe poderse asumir que las diferencias entre ellos son debidas exclusivamente al azar, es decir, que los resultados de los diferentes estudios son homogéneos. La mayor parte de las revisiones documentan diferencias importantes de los pacientes, exposiciones, parámetros de resultados y métodos de investigación de un estudio a otro. Es necesario que el lector sepa cuándo estos factores son tan diferentes que deja de tener sentido integrar los resultados de los estudios para hacer un MA y por ello debemos valorar que se ha estudiado la heterogeneidad de los estudios primarios, de forma que: 1) si existe mucha heterogeneidad desde el punto de vista clínico, quizás lo más idóneo sea no hacer MA y limitarse a una análisis de síntesis cualitativa de los resultados de la revisión, y 2) si se considera que los estudios se parecen lo suficiente como para intentar combinarlos, se tratará de medir la heterogeneidad con algún método estadístico.

Entre los métodos estadísticos empleados para el estudio de heterogeneidad, quizás el más clásico sea la Q de Cochran, aunque el más utilizado en la actualidad es la I^2 (y existe un tercer método menos conocido, la H^2)[29]. Con I^2 se suelen usar los límites de 25, 50 y 75% para establecer si hay heterogeneidad baja, moderada y alta, respectivamente. En resumen, siempre que se muestre un indicador de homogeneidad que represente un porcentaje indicará la proporción de variabilidad que no es debida al azar y se considerará una heterogeneidad significativa cuando $p < 0{,}05$. Además, se han ideado una serie de gráficos para inspecciona los estudios y comprobar que no hay datos de heterogeneidad, siendo los más utilizados el gráfico de Galbraith y el gráfico de L'Abbé.

Si se considera que existe una heterogeneidad significativa entre los estudios primarios habrá varias posibilidades: 1) no combinar los estudios primarios y limitarse a una síntesis cualitativa de los resultados; 2) hacer un análisis estratificado según la variable que cause la heterogeneidad, siempre que sea posible identificarla, y 3) recurrir a una metarregresión, opción más compleja.

¿Cuáles son los resultados de la revisión?

¿Cuál es el resultado global de la revisión? ¿Han sido interpretados con sentido común? Una RS o MA busca proporcionar una estimación global del efecto de una intervención basada en un promedio ponderado de los resultados de todos los estudios de calidad disponibles. Generalmente, los resultados de cada estudio se ponderan por el inverso de su variancia, de forma que se concede un mayor peso a los estudios de mayor tamaño que proporcionan estimaciones más precisas. A veces, la ponderación también tiene en cuenta la calidad metodológica de los estudios.

Los resultados suelen expresarse como una medida relativa (*odds ratio* [OR], riesgo relativo [RR], reducción relativa del riesgo [RRR]), aunque sería conveniente complementarla con una medida absoluta (reducción absoluta del riesgo [RAR], número necesario de pacientes a tratar [NNT]). Un buen MA debe ser más fácil de interpretar que el conjunto de EC que contiene. Aparte de sintetizar los datos numéricos, la información relevante relativa a los estudios primarios incluidos debe estar correctamente tabulada (criterios de inclusión, tamaños de muestra, características de los pacientes, características fundamentales de los EC, resultados principales y secundarios). Actualmente los resultados de los MA tienden a presentarse de manera estandarizada. El resultado principal de cada uno de los estudios incluidos en la revisión se indica como una línea horizontal, cuya anchura representa el intervalo de confianza del 95% (IC 95) de la estimación de la medida del efecto (lo que indica la precisión del resultado), y en la que una señal indica la estimación puntual observada. La línea vertical en el centro del gráfico representa la "línea de efecto nulo", con una OR igual a 1 o una diferencia de medias igual a 0. En caso de que la línea horizontal (intervalo de confianza) se cruce con la línea vertical, esto puede significar que, o bien no existen diferencias significativas entre los tratamientos, o que el tamaño de la muestra es insuficiente para detectar dichas diferencias. Cada uno de los estudios individuales tiene su OR y su intervalo de confianza, pero el diamante de la parte inferior de las líneas horizontales representa la suma de los OR de todos los EC, con un nuevo intervalo de confianza mucho más estrecho o preciso. Si el diamante sobrepasara claramente la línea de efecto nulo, no podríamos afirmar con seguridad que un tratamiento es mejor que otro. Cuanto más estrecho sea el diamante, más preciso será el resultado final, y cuanto más alejado esté de la línea de efecto nulo, más clara será la diferencia entre los tratamientos comparados.

La Cochrane Collaboration recomienda estructurar el diagrama de efectos (*forest plot*) en cinco columnas bien diferenciadas (Figura 2)[30]: 1) se listan los estudios primarios (autor y año) con los grupos o subgrupos de pacientes incluidos en el MA; 2) resultados de cada estudios (tanto del grupo intervención como del grupo control); 3) es la representación propia del *forest plot* tal como se ha explicado previamente y alrededor de la línea vertical del efecto nulo y con el diamante como resultado combinado; 4) estimación del peso de cada estudio (en proporción), y 5) estimación del efecto ponderado de cada estudio. En este gráfico también se indica la medida de heterogeneidad.

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

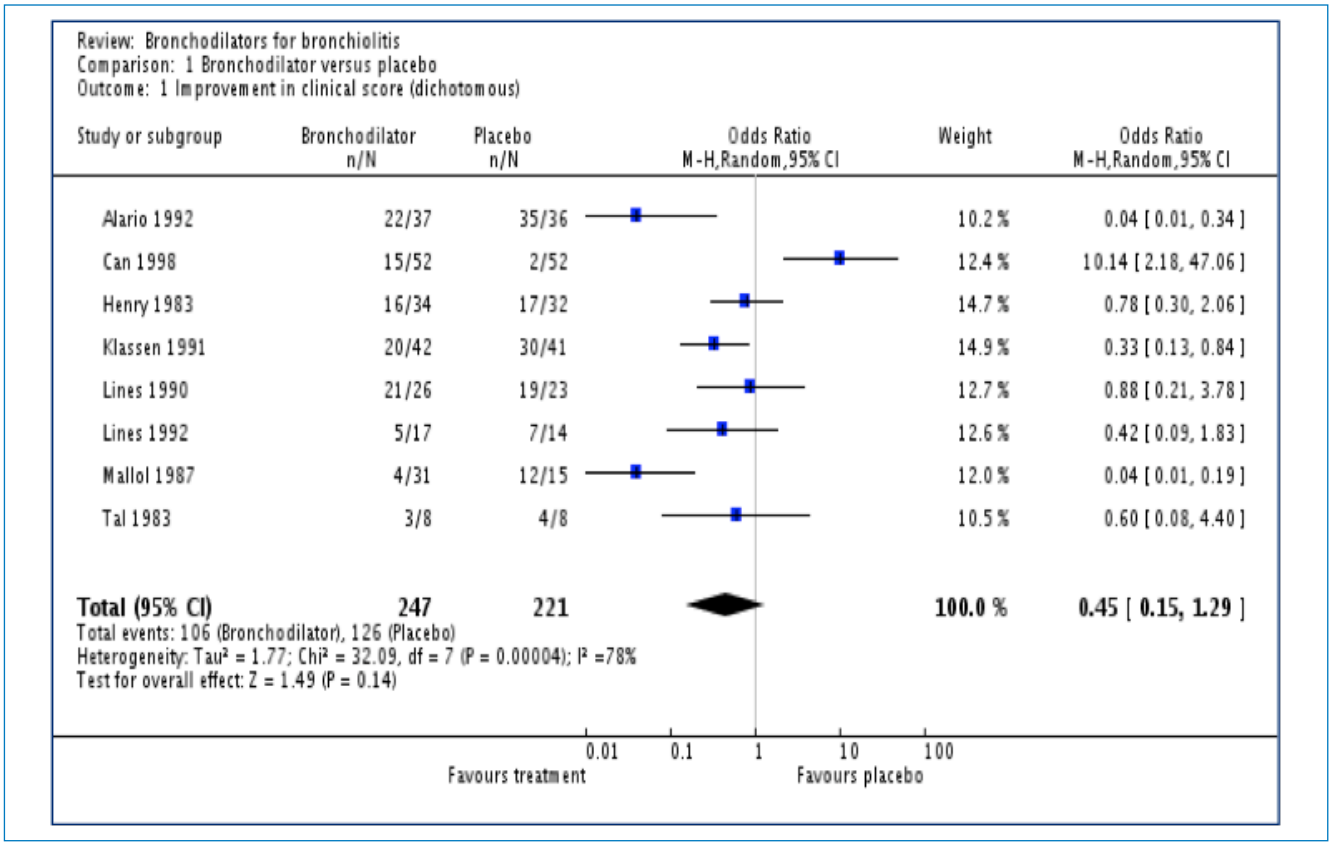

Figura 2. **Ejemplo de *forest plot* de un metaanálisis.**

Es importante señalar que, por muy favorable que sea el resultado final del MA, esto no significa que debamos ofrecer indiscriminadamente el tratamiento estudiado, pues el uso de la mejor de las evidencias no puede pasar por encima del conocimiento personal del paciente, es decir, es necesario conocer sus características individuales y preferencias.

La principal asunción de un MA para obtener una estimación conjunta es que las diferencias entre los diferentes estudios son debidas exclusivamente al azar, es decir, que son estudios homogéneos. Por lo tanto, en primer lugar, debe analizarse la homogeneidad tanto clínica como estadística de los estudios, tanto con métodos gráficos como con las pruebas estadísticas adecuadas referidas previamente. Si se decide que pueden combinarse de los estudios primarios, esto no puede hacerse uniendo sin más todos los resultados (riesgo de que se produzca la paradoja de Simpson), sino que se pueden utilizar dos modelos: modelo de efecto fijo (que incorpora solo un término de error por la variabilidad de cada estudio) y modelo de efectos aleatorios (que incorpora, además del anterior error, también el error debido a la variabilidad entre los estudios y, por tanto, es más conservador)[30]. El tipo de modelo a elegir tiene su importancia: si el análisis previo muestra que los estudios son homogéneos, podrá utilizarse el modelo de efecto fijo; pero si se detecta que existe heterogeneidad, dentro de los límites que permiten combinar estudios, será preferible utilizar el modelo de efectos aleatorios (y ante la duda, utilizar mejor este modelo)[30].

¿Cuán precisos son los resultados? La magnitud del resultado observado en un estudio no es más que una estimación puntual de la verdadera magnitud del efecto o asociación de interés. Si se hubiera estudiado una muestra diferente de sujetos (o de estudios en el caso de la RS o MA), podría haberse obtenido un resultado distinto, aunque es de esperar que no demasiado diferente. Por ello, es importante determinar la precisión con que se ha realizado la estimación mediante el IC 95, que es aquel intervalo entre cuyos límites se tiene el 95% de confianza de que se encuentre la verdadera magnitud del efecto. El IC presenta grandes ventajas respecto al grado de significación estadística. Un resultado estadísticamente significativo no implica de forma inequívoca que sea clínicamente relevante. Cuando se utiliza como medida del efecto una diferencia, si el IC 95 incluye el valor 0, que es el valor correspondiente a la hipótesis nula estadística (Ho) de que no existe diferencia

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

entre ambos grupos, se concluirá que el resultado no es estadísticamente significativo. Cuando se utiliza como medida del efecto una medida relativa basada en un cociente (RR, OR, etc.), si el IC 95 incluye el valor 1, que el valor correspondiente a la Ho de que no existe diferencia entre ambos grupos se concluirá que el resultado no es estadísticamente significativo.

¿Son los resultados de la RS o MA aplicables al escenario?

¿Se pueden aplicar los resultados en tu medio? Cualquier resultado numérico, por más preciso y "estadísticamente significativo" que sea, es decir, incontrovertible, debe situarse en el contexto práctico de la pregunta a la que la revisión pretende dar respuesta. El clínico debe decidir cómo va a influir este resultado numérico en la atención sanitaria que va a dispensarse a un paciente concreto. Un aspecto fundamental a considerar cuando se lee críticamente una RS es la validez externa de los EC (u otros estudios) que contiene. Es decir, ¿están descritas las características de los pacientes incluidos en los estudios primarios? ¿Son parecidos a los nuestros? Una de las ventajas de una RS es que, dado que incluye numerosos estudios, los resultados proceden de una variedad muy diversa de pacientes. Incluso así, el clínico puede seguir teniendo dudas sobre la aplicabilidad de los resultados a nuestro(s) paciente(s), con aspectos como edad, tipo de enfermedad, tipo de tratamiento, etc. Estas preguntas suscitan el problema del análisis de subgrupo; en cualquier caso, están disponibles guías detalladas para decidir si son creíbles los análisis de subgrupo.

¿Se han considerado todos los resultados importantes para tomar la decisión? Aunque es conveniente buscar RS focalizados a una pregunta clínica (variable principal), porque es más probable que ofrezca resultados válidos, esto no significa que no requiera la consideración de otros parámetros de interés (variables secundarias).

¿Los beneficios merecen la pena frente a los perjuicios y costes? Explícita o implícitamente, cuando se toma una decisión clínica es preciso sobrepasar los beneficios frente a los posibles riesgos y costes. Un artículo válido y relevante de RS o MA ofrece la mejor base posible para cuantificar los resultados previstos, pero estos resultados deben seguir considerándose en el contexto de los valorares y preocupaciones de sus pacientes por los resultados previstos de una decisión.

BIBLIOGRAFÍA

1. Balaguer A, González de Dios J. Digiriendo la información científica. Excelencias y limitaciones de las revisiones sistemáticas y metaanálisis. Acta Pediatr Esp. 2004;62:4-10.

2. González de Dios J. Búsqueda de información en Pediatría basada en la evidencia (I): "infoxicación" e Internet. Rev Esp Pediatr. 2003;59:246-58.

3. López Arrieta JM, Qizilbash N. La medicina basada en pruebas: revisiones sistemáticas. La Colaboración Cochrane. Med Clín (Barc). 1996;107:581-5.

4. Straus SE, McAlister FA. Evidence-based medicine: a commentary on common criticisms. CMAJ. 2000;163:837-41.

5. González de Dios J, Balaguer Santamaría A. Revisión sistemática y metaanálisis (I): conceptos básicos. Evid Pediatr. 2007;3:107.

6. González de Dios J, Balaguer Santamaría. Revisión sistemática y metaanálisis (II): evaluación crítica de la revisión sistemática y metánalisis. Evid Pediatr. 2007;4:19.

7. González de Dios J, González Muñoz M, Alonso A, Aleixandre R. Comunicación científica (XX). Conocimientos básicos para leer (y escribir) un artículo científico (7): Lista de comprobación de revisiones sistemáticas. Acta Pediatr Esp. 2015;73;47-51.

8. Molina Arias M. Lectura crítica en pequeña dosis. La revisión sistemática. Rev Pediatr Aten Primaria. 2013;15:283-5.

9. González de Dios J. Las mejores evidencias (o pruebas) científicas (III): el metaanálisis. Rev Esp Pediatr. 1999;55:368-76.

10. González de Dios J, Buñuel JC, González P, Aleixandre R. Fuentes de información bibliográfica (XIV). Sobre "fuentes", "pirámides" y "revoluciones" en la gestión del conocimiento en Pediatría. Acta Pediatr Esp. 2012;70:289-95.

11. Jadad AR, Cook DJ, Browman G. A guide to interpreting discordant systematic reviews. Can Med Assoc J. 1997;156:1411-6.

12. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med. 1997;337:536-42.

13. Rodríguez Artalejo F, Guallar Castellón P. Sobre los meta-análisis y el Quorom. Rev Esp Salud Pública. 2000;74:103-5.

14. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. Lancet. 1999;354: 1896-900.

15. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, *et al*. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA. 1996;276:637-9.

16. González de Dios J, González Muñoz M, Alonso A, Aleixandre R. Comunicación científica (XIX). Conocimientos básicos para leer (y escribir) un artículo científico (6): Lista de comprobación de estudios experimentales. Acta Pediatr Esp. 2015;73;19-24.

17. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009;6:e1000097.

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953

18. Urrutia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. Med Clin (Barc). 2010;135:507-11.

19. González de Dios J, Buñuel JC, Aparicio M. Listas guía de comprobación de revisiones sistemáticas y metaanálisis: declaración PRISMA. Evid Pediatr. 2011;7:97.

20. CASPe. Programa de habilidades en lectura crítica. Entendiendo la evidencia acerca de la eficacia clínica. 10 preguntas para ayudarte a entender una revisión [en línea] [consultado el 08/04/2021]. Disponible en: www.redcaspe.org/system/tdf/materiales/plantilla_revision.pdf?file=1&type=node&id=154&force=

21. González de Dios J, González P. Evaluación de artículos científicos sobre intervenciones terapéuticas. Evid Pediatr. 2006;2:90.

22. Ochoa C, Buñuel JC, González de Dios J. Evaluación de artículos sobre intervenciones diagnósticas. Evid Pediatr. 2007;3:24.

23. González de Dios J, Buñuel JC. Evaluación de artículos científicos sobre efectos perjudiciales de intervenciones y/o exposiciones. Evid Pediatr. 2007;3:54.

24. González de Dios J, Ibáñez V, Modesto V. Evaluación de artículos científicos sobre pronóstico. Evid Pediatr. 2007;3:81.

25. Oxman AD, Sackett DL, Guyatt GH, for the Evidence-Based Medicine Working Group. User's guide to the medical literatura, I. how to get started. JAMA. 1993;270:2093-5.

26. González de Dios J. El sesgo de publicación: en busca de la evidencia perdida. Pediatr Integral. 2000;5:760-6.

27. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, *et al.* Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.

28. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, *et al*. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.

29. Molina Arias M. Aspectos metodológicos del metaanálisis (1). Rev Pediatr Aten Primaria. 2018;20:297-302.

30. Molina Arias M. Aspectos metodológicos del metaanálisis (2). Rev Pediatr Aten Primaria. 2018;20:401-5.

©AEPap 2021 • Copia para uso personal, se prohíbe la reproducción y/o transmisión de este documento por cualquier medio o formato • http://monograficos.fapap.es • ISSN | 2444-1953