

Valoración crítica de artículos

SEGUNDA PARTE

Revisión sistemática y metaanálisis

J. GONZÁLEZ DE DIOS¹ Y A. BALAGUER SANTAMARÍA²

¹ Departamento de Pediatría. Hospital de Torrevieja. Universidad Miguel Hernández (Alicante)

² Pediatra. Unidad Neonatología y Servicio de Pediatría. Hospital Universitari St. Joan. Universidad Rovira i Virgili. Reus (Tarragona)

Adaptación del artículo González de Dios J, Balaguer Santamaría A. Revisión sistemática y metanálisis (II): evaluación crítica de la revisión sistemática y metanálisis. *Evid Pediatr* 2008;4:19.

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 a un procedimiento estructurado bien definido e intentan responder una pregunta clínica sobre un tema concreto. Deben definirse claramente la población y/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 o de un metaanálisis existen diferentes listados de verificación de su calidad. Todos ellos evalúan tres aspectos clave: la validez metodológica, la importancia clínica de sus resultados y la posible aplicabilidad de los mismos a la práctica clínica cotidiana del lector.

INTRODUCCIÓN

En Medicina, difícilmente un estudio aislado puede aportar la solución definitiva de un problema médico, aunque este es-

tudio tenga la calidad científica de un ensayo clínico (EC) 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^{1,2}. Para resolver tales discrepancias se ha recurrido a los métodos de síntesis de información sobre un mismo tema, bien cualitativa (RS) o cuantitativa (MA).

Sin embargo, todas las RS y/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 buena³ y, a veces, los resultados no coinciden con los de los grandes EC⁴. 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)^{4,5}.

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

VALOR DE LAS LISTAS-GUÍA EN LAS REVISIONES SISTEMÁTICAS Y METAANÁLISIS

Debido al creciente número de RS y/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

Tabla 1

LISTA DE VERIFICACIÓN (CHECK-LIST) QUE PERMITE LA DETECCIÓN DE ERRORES IMPORTANTES EN UNA REVISIÓN SISTEMÁTICA Y/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 y/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 y/o MA de forma objetiva e imparcial?

7. Discusión

- ¿Se discuten los resultados de la RS y/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.?

analiza globalmente cada apartado, lo que permite un análisis más matizado).

Al plantearse llevar a cabo una RS y/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 1**. En dicha tabla se presenta una lista de preguntas clave para valorar una RS y/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 y/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

Tabla 2
PROPUESTA QUOROM PARA LA PRESENTACIÓN DE UNA REVISIÓN SISTEMÁTICA Y METAANÁLISIS

Categoría	Subcategoría	Ítem
TÍTULO		<ul style="list-style-type: none"> • Identificar el trabajo como un MA (o RS) de ECC
Resumen	<ul style="list-style-type: none"> • Objetivos • Fuentes de datos • Métodos de revisión • Resultados • Conclusión 	<ul style="list-style-type: none"> • Utilizar un formato estructurado • DESCRIBIR • La cuestión clínica explícitamente • Las base de datos (es decir, una relación) y otras fuentes de información • Los criterios de selección (es decir, población, intervención, resultado y diseño del estudio); los métodos de valoración de la validez, análisis de datos y características del estudio, y síntesis de datos cuantitativos suficientemente detallada para permitir la repetición • Características de los EC incluidos y excluidos: hallazgos cualitativos y cuantitativos (estimaciones puntuales e intervalos de confianza), y análisis de subgrupo • Resultados principales
INTRODUCCIÓN		<ul style="list-style-type: none"> • El problema clínico concreto: los fundamentos biológicos de la intervención y los fundamentos de la revisión
Métodos	<ul style="list-style-type: none"> • Búsqueda • Selección • Evaluación de la validez • Análisis de datos • Características del estudio • Síntesis de datos cuantitativos 	<ul style="list-style-type: none"> • Las fuentes de información, con detalle (p. ej., bases de datos, registros, ficheros personales, expertos, agencias, búsqueda manual) y cualquier restricción (años considerados, situación de publicación, idioma) • Los criterios de inclusión y exclusión (relativos a la población, la intervención, los resultados principales y el diseño del estudio) • Criterios y proceso utilizados (p. ej., condiciones encubiertas, valoración de la calidad y sus hallazgos) • Proceso o procesos utilizados (p. ej., completados, independientemente, en duplicado) • Tipo de diseño del estudio, características de los participantes, detalle de la intervención, definiciones de los resultados, etc., y cómo se evaluó la heterogeneidad clínica • Principales medidas de efectos (p. ej., riesgo relativo), el método de combinación de resultados (pruebas estadísticas e intervalos de confianza), tratamiento de resultados perdidos; evaluación de la heterogeneidad clínica; fundamento para cualquier análisis de sensibilidad y subgrupo a priori; y cualquier sesgo de publicación
RESULTADOS	<ul style="list-style-type: none"> • Flujo de resultados 	<ul style="list-style-type: none"> • Ofrecer un perfil de MA que resuma el flujo de resultados (ver figura 1)
	<ul style="list-style-type: none"> • Características del estudio 	<ul style="list-style-type: none"> • Descripción de cada resultado (p. ej., edad, tamaño de la muestra, intervención, dosis, duración, período de seguimiento)
	<ul style="list-style-type: none"> • Síntesis de datos cuantitativos 	<ul style="list-style-type: none"> • Selección y evaluación de la validez; presentar un resultado sencillo de los resultados (para cada grupo en cada resultado, para cada resultado primario); presentar los datos necesarios para calcular los tamaños del efecto y los intervalos de confianza en los análisis (p. ej., tablas 2x2, medias y desviación estándar, proporciones)
Discusión		<ul style="list-style-type: none"> • Resumir los hallazgos clave; discutir las inferencias clínicas basadas en la validez interna y externa; interpretar los resultados a la luz de la totalidad de la evidencia disponible; describir los sesgos potenciales en el proceso de revisión (p. ej., sesgo de publicación); y sugerir un futuro plan de investigación

de mejorar y homogeneizar la calidad de la presentación de este tipo de estudios con alta «evidencia» científica, se destacan los puntos que deberían exigirse⁶. Si esta iniciativa tiene éxito y es asumida por las revistas científicas más importantes, tal como ocurrió con su antecedente para los EC (el conocido CONSORT)⁷, dichas recomendaciones pueden convertirse en norma para publicar nuevas RS y/o MA. Esta propuesta contiene una lista-guía que «obliga» a los autores a ofrecer información explícita sobre cada uno de los apartados, incluyendo la estrategia de búsqueda, los criterios para la selección de estudios, la evaluación de la validez de éstos, el modo en que se extrajeron los datos, las características de los diversos estudios y la síntesis cuantitativa de los datos (tabla 2).

Según este grupo, existen seis áreas temáticas importantes para la realización y presentación de los resultados de cualquier RS y/o MA de EC aleatorizados: diseño del estudio, combinación adecuada de diversos estudios, control del sesgo, análisis estadístico, análisis de sensibilidad y problemas de aplicabilidad. Además, la propuesta QUOROM destaca la necesidad de incluir un diagrama de flujo (figura 1) en el que se deberá aportar información detallada sobre el número de EC aleatorizados identificados, incluidos y excluidos, y los motivos por los que estos últimos lo fueron.

Los «Requisitos de uniformidad para manuscritos enviados a revistas biomédicas» en su última edición de 2007 recoge las recomendaciones para la publicación de estudios con diseños específicos (EC bajo la Declaración CONSORT, RS y/o MA bajo la Declaración QUOROM, estudios sobre pruebas diagnósticas bajo la Declaración STARD, guías de práctica clínica bajo la Declaración AGREE, etc.), y algunas revistas pueden exigir a los autores en sus instrucciones que atiendan dichos criterios⁸.

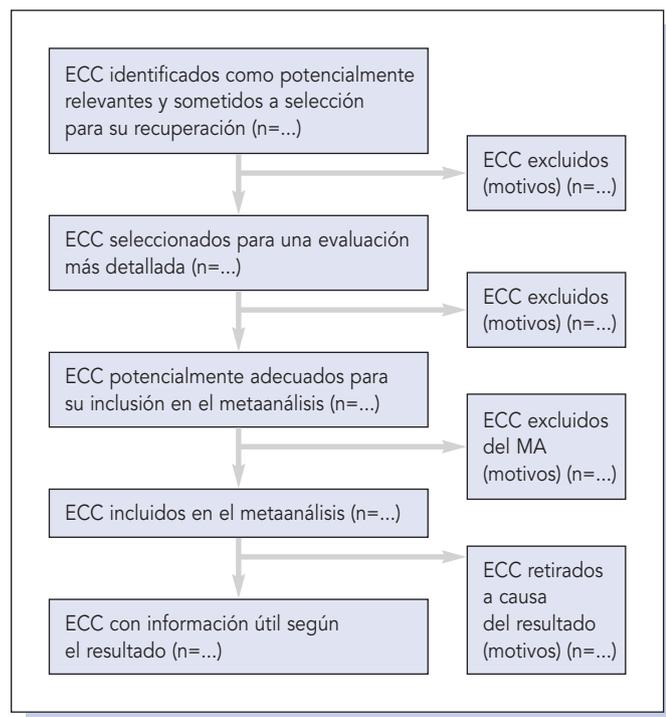


FIGURA 1. Diagrama de flujo de una revisión sistemática.

LECTURA CRÍTICA DE PUBLICACIONES SOBRE REVISIONES SISTEMÁTICAS Y/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 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 artículo⁹: 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 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 CASP (*Critical Appraisal Skills Programme*), 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 RS¹⁰.

Analizaremos a continuación los pasos a seguir para analizar la validez, la importancia y la aplicabilidad (tabla 3) de los artículos sobre pronóstico, planteando diversas cuestiones:

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

En base a estas fuentes, los aspectos fundamentales a valorar en una RS y/o MA, para que ésta sea considerada de calidad (libre de los sesgos más importantes), son los siguientes¹¹⁻¹⁴:

¿Son válidos los resultados de la revisión?

Estas preguntas definen los criterios que se utilizan para evaluar la validez interna (o rigor científico) de las RS y/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.

Tabla 3
PREGUNTAS PARA VALORACIÓN CRÍTICA DE ARTÍCULOS SOBRE REVISIONES SISTEMÁTICAS Y/O METAANÁLISIS

- A) ¿Son válidos los resultados del estudio?**
Preguntas de eliminación:
1. ¿Se hizo la revisión sobre un tema claramente definido?
 2. ¿Buscaron los autores el tipo de artículos adecuados?
- 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.

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 mucha 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 (p. ej., 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 investigación¹⁵ (tabla 4). 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 (entre otros motivos, porque la mayoría de las publicadas hasta una época muy reciente son narrativas y no sistemáticas); sólo si la respuesta es afirmativa en ambas merece la pena continuar con la tercera pregunta. Tal como sugieren Oxman et al.¹³, si el principal

Tabla 4
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 del mismo?
- ¿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 y/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?

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 bibliográficas 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 Cochrane Library, 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 estudios 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)¹⁶. El diagrama de flujo propuesto por el grupo QUOROM (figura 1) permite evaluar gráficamente cómo se han seleccionado los estudios⁶.

¿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.

Si los resultados de los diferentes estudios han sido mezclados para obtener un resultado «combinado», ¿era razonable 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. Un criterio para decidir combinar los resultados cuantitativamente es si los estudios parecen medir la misma magnitud del efecto. Los revisores deben de haber analizado estas posibles diferencias mediante las llamadas «pruebas de homogeneidad». Por homogeneidad estadística entendemos que los resultados de cada estudio individual son matemáticamente compatibles con los resultados de los otros estudios primarios. Cuanto más homogénea es una prueba, más probable es que las posibles diferencias sean fruto de la casualidad. Por contra, cuanto existe «heterogeneidad» estadísticamente significativa, hay una posibilidad muy elevada de que haya diferencias y podría dejar de tener sentido integrar los resultados.

¿Cuáles son los resultados de la revisión?

¿Cuál es el resultado global de la revisión? ¿Han sido interpretados con sentido común?

Una RS y/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 (IC) del 95% de la estimación de la medida del efecto (lo que indica la precisión del resultado), y en la que una se-

ñal 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 diferencias 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.

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. También es oportuno en la RS y/o MA evaluar si sus resultados son sensibles a la forma en que se han realizado, en función de distintos criterios, lo que se conoce como análisis de sensibilidad.

¿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 y/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 del 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^{17,18}. 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 del 95% incluye el valor 0, que es el valor correspondiente a la hipótesis nula estadística (H_0) de que no existe diferencia 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 del 95% incluye el valor 1, que el valor correspondiente a la H_0 de que no existe diferencia entre ambos grupos, se concluirá que el resultado no es estadísticamente significativo.

¿Son los resultados 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 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^{19,20}.

¿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 y/o MA ofrece la mejor base posible para cuantificar los resultados previstos, pero estos resultados deben seguir considerándose en el contexto de los valores y preocupaciones de sus pacientes por los resultados previstos de una decisión.

BIBLIOGRAFÍA

1. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med* 1987;82:498-510.
2. Bailar JC III. When research results are in conflict. *N Engl J Med* 1985;313:1080-1.
3. Jadad AR, Cook DJ, Browman G. A guide to interpreting discordant systematic reviews. *Can Med Assoc J* 1997;156:1411-6.
4. 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.

5. Rodríguez Artalejo F, Guallar Castellón P. Sobre los meta-análisis y el Quorum. *Rev Esp Salud Pública* 2000;74:103-5.
6. 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.
7. 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.
8. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication [en línea] [fecha de consulta: 26-01-2007]. Disponible en <http://www.icmje.org/index.html>.
9. González de Dios J. Lectura crítica de documentos científicos. *An Pediatr Contin* 2003;1:51-5.
10. 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] [fecha de consulta: 06-03-2005]. Disponible en: <http://www.redcaspe.org/herramientas/lectura/10revision.pdf>.
11. Greenhalgh T. How to read a paper. Papers that summarise others papers (systematic reviews and meta-analyses). *BMJ* 1997; 315:672-5.
12. Martí Ramis J. Lectura crítica de revisiones sistemáticas de ensayos clínicos [en línea] [fecha de consulta: 26-01-2007]. Disponible en http://www.atheneum.doyma.es/Socios/sala_/lec08mbe.htm.
13. Oxman AD, Sackett DL, Guyatt GH, for the Evidence-Based Medicine Working Group. User's guide to the medical literature, I: how to get started. *JAMA* 1993;270:2093-5.
14. González de Dios J. Revisión crítica de publicaciones sobre tratamiento. *Pediatr Integral* 2002;6:455-62.
15. González de Dios J. Revisión crítica de publicaciones sobre pruebas diagnósticas. *Pediatr Integral* 2002;6:551-6.
16. González de Dios J. El sesgo de publicación: en busca de la evidencia perdida. *Pediatr Integral* 2000;5:760-6.
17. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled trials. *N Engl J Med* 1983;309:1358-61.
18. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995;16:62-73.
19. Gardner MJ, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. *BMJ* 1986;292: 746-50.
20. Yusuf S, Liffes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93-8.