



Fundamentos de medicina basada en la evidencia

Revisión sistemática y metanálisis (I): conceptos básicos

Javier González de Dios. Departamento de Pediatría. Hospital de Torrevieja. Universidad Miguel Hernández. Alicante (España). Correo electrónico: jgonzalez@torrevieja-salud.com
Albert Balaguer Santamaría. Departamento de Pediatría. Hospital Sant Joan de Reus. Universidad Rovira i Virgili. Tarragona (España). Correo electrónico: albert.balaguer@urv.es

Términos clave en inglés: evidence-based medicine; systematic review; meta-analysis

Términos clave en español: medicina basada en pruebas; revisión sistemática; metanálisis

Fecha de recepción: 10 de noviembre de 2007
Fecha de aceptación: 19 de noviembre de 2007

Fecha de publicación: 1 de diciembre de 2007

Evid Pediatr. 2007; 3: 107 doi: [vol3/2007_numero_4/2007_vol3_numero4.23.htm](https://doi.org/10.4238/2007_vol3_numero4.23.htm)

Cómo citar este artículo

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

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín por medio del ETOC <http://www.aepap.org/EvidPediatr/etoc.htm>

Este artículo está disponible en: http://www.aepap.org/EvidPediatr/numeros/vol3/2007_numero_4/2007_vol3_numero4.23.htm
EVIDENCIAS EN PEDIATRÍA es la revista oficial del Grupo de Pediatría Basada en la Evidencia de la Asociación Española de Pediatría de Atención Primaria. © 2005-06. Todos los derechos reservados

Revisión sistemática y metanálisis (I): conceptos básicos

Javier González de Dios. Departamento de Pediatría. Hospital de Torrevieja. Universidad Miguel Hernández. Alicante (España). Correo electrónico: jgonzalez@torrevieja-salud.com

Albert Balaguer Santamaría. Departamento de Pediatría. Hospital Sant Joan de Reus. Universidad Rovira i Virgili. Tarragona (España). Correo electrónico: albert.balaguer@urv.es

La necesidad de luchar contra la explosión de la información científica que se genera en biomedicina y hacer de esta información algo útil y aplicable en la práctica clínica es un hecho constatado principalmente desde finales del siglo XX. Y ésta es la función que cumplen los distintos métodos de síntesis de la bibliografía científica: servir de instrumentos que resumen y analizan los hallazgos más relevantes en las distintas áreas de conocimiento para, posteriormente, devolverlos a modo de síntesis a esa misma comunidad con el fin de que se haga uso de ellos¹. A diferencia de los movimientos clásicos de la bibliografía científica (entre ellos la revisión clásica o de autor), los procedimientos contemporáneos de síntesis adoptan la lógica y la estructura de un proceso de investigación sobre un problema, conocido como investigaciones secundarias (basadas en el análisis y en la síntesis de publicaciones originales) y que se publican en las revistas de biomedicina en forma, por ejemplo, de revisión sistemática (RS) y metanálisis (MA). Todos estos ejemplos de fuentes de información secundaria están en relación con el movimiento en ciencias de la salud conocido como Medicina basada en la evidencia (MBE) o en pruebas²⁻⁴.

Existen diversos motivos para gestionar correctamente el conocimiento científico, principalmente ante el exceso de información científica, lo que es un problema tanto cuantitativo (es imposible acceder y revisar a fondo todo lo que se publica sobre un determinado tema) como cualitativo (es difícil analizar críticamente la evidencia científica existente y discernir la utilidad de lo nuevo en relación al conocimiento previo). El acceso ordenado, sistemático y sin sesgos de la información derivada de los trabajos científicos es muy complejo, pese a la accesibilidad que proporciona Internet a las bases de datos bibliográficas y revistas biomédicas. Por tanto, necesitamos disponer de herramientas que nos permitan acceder a la información adecuada en términos de cantidad, calidad y actualidad⁵. Y, como no disponemos de mucho tiempo, no sólo aspiramos a información veraz y adecuada a nuestro problema sino que deseamos que esté presentada de modo práctico, que sea fácil de interpretar, además de fiable y rigurosa. El médico actual ha dejado de ser un acumulador de información para convertirse en un buscador de fuentes de información eficientes y de calidad. El flujo de ideas en biomedicina se ha hecho internacional, masivo y de gran velocidad de renovación⁶.

Revisión sistemática frente a revisión narrativa (o de autor)

Las revisiones nos ofrecen la posibilidad de estar informados sin necesidad de invertir un tiempo que no disponemos. Precisamente, una de las maneras más

eficientes de encontrar la evidencia científica a la que hace referencia la MBE es buscar una revisión sistemática de los ensayos clínicos (EC) aleatorizados sobre los efectos de la atención sanitaria. El EC se considera el patrón oro de las mejores evidencias externas de la literatura si el tema que nos preocupa es un problema terapéutico; una metodología más reciente denominada RS (que cuando utiliza procedimientos estadísticos se conoce como MA) ha permitido extraer conclusiones más firmes, al combinar todos los EC aleatorizados (y, con menor frecuencia, otros estudios con menor evidencia científica) llevados a cabo sobre un problema de salud específico. El uso de métodos explícitos y sistemáticos en las RS limita el sesgo (esto es, el error sistemático) y reduce los efectos encontrados por azar, de modo que proporciona resultados más fiables sobre los cuales sacar conclusiones y tomar decisiones⁷⁻⁹.

Consecuentemente las RS pretenden ser¹⁰:

- Rigurosas: en cuanto a los estudios incluidos (criterios de inclusión, criterios de calidad, etc.).
- Informativas: enfocadas hacia problemas reales, tratando de contestar a una pregunta clínica claramente delimitada e, idealmente, analizando y presentando los datos de la forma que mejor ayude a la toma de decisiones.
- Exhaustivas: utilizar la mayor y mejor información pertinentes, sin introducir sesgos (de selección, de publicación, etc.).
- Explícitas: todos los métodos utilizados en la revisión deben describirse con suficiente detalle.

Los problemas de la revisión de autor son dos: primero, los autores no especifican el proceso que han seguido para buscar, reunir y evaluar la información y, en segundo lugar, y ante la ausencia de dicha información, no serán capaces de repetir y verificar los resultados y conclusiones de la revisión^{6,10}. Las RS buscan "toda la verdad" e intentan responder en profundidad a cuestiones clínicas generalmente muy concretas y específicas. Sin embargo, las revisiones narrativas abarcan un amplio abanico de asuntos relacionados con un tema, más que uno en profundidad; son útiles más para adquirir una perspectiva amplia sobre un tema, pero no tanto para obtener respuestas concretas y mensurables a cuestiones clínicas específicas. Las diferencias entre ambos tipos de revisiones se exponen en la tabla 1, destacando que en las narrativas están basadas frecuentemente en opiniones (además de, por supuesto, en algunos datos objetivos), mientras que las RS están basadas fundamentalmente en pruebas^{6,10-12}. La Colaboración Cochrane se considera el prototipo de fuente de información secundaria y uno de los recursos bibliográficos más importantes en términos de validez científica e importancia clínica. En la Cochrane cada revisor es miembro de un Grupo

Colaborador de Revisión (Collaborative Review Group), el cual está constituido por individuos que comparten un interés específico sobre un tema determinado. Cada Grupo Colaborador de Revisión está coordinado por un equipo editorial, quien difunde sus resultados principalmente a través de The Cochrane Library, que incluye diferentes bases de datos: Base de Datos Cochrane de Revisiones Sistemáticas (The Cochrane Database Systematic Reviews-CDSR-, la de mayor interés), la Base de Datos de Resúmenes de Efectividad (The Database of Abstracts of Reviews of Effectiveness –DARE-), la Base de Datos de Evaluación Económica (NHS Economic Evaluation Database –NHS-EED-), el Registro Cochrane de Ensayos Controlados (The Cochrane Controlled Trials Register –CCTR-), etc^{13,14}.

Actualmente existen Grupos Colaboradores de Revisión en más de 50 áreas clínicas, algunas íntimamente ligadas a la Pediatría (ej. Neonatal Group, Pregnancy and Childbirth Group, Airways Group, Acute Respiratory Infection Group, Infection Disease Group, Cystic Fibrosis and Genetic Disorder Group, Cochrane Ear, Nose and Throat Disorders Group). La Pediatría (y concretamente la medicina perinatal-neonatal) han tenido un lugar preponderante en los orígenes de la CC; de hecho, la publicación en 1989 del libro *Effective Care in Pregnancy and Childbirth* y *The Oxford Database of Perinatal Trials*, fueron los precursores de la CC y The Cochrane Library. Las revisiones Cochrane tiene un formato estándar, perfectamente definido, que incorpora los apartados expuestos en la tabla 2^{8,10}.

No todas las RS (y MA) se encuentran volcadas en la Colaboración Cochrane, pues muchas otras se encuentran publicadas en revistas biomédicas¹⁵. Si se ha constatado que las RS (y MA) que aparecen en la Cochrane suelen tener mayor rigor metodológico y mayor grado de actualización que las publicadas en revistas biomédicas¹⁶.

Metodología para la preparación de una revisión sistemática (y metanálisis)

Basado en el Manual de revisores de la CC (versión española de Cochrane reviewer's Handbook del Centro Cochrane Iberoamericano)^{17,18}, y fundamentado en revisiones previas⁹⁻¹⁰.

Primer paso: formulación del problema

a. Justificación de las preguntas bien formuladas:

Es necesario que el artículo exponga claramente de qué trata la RS, y es preferible que trate de responder a una pregunta y no a muchas. La pregunta principal, formulada a priori, determinará el enfoque principal de la RS y ayuda a combatir las denominadas "expediciones de pesca". Si la pregunta que se quiere responder no está bien formulada es probable que la revisión no sea pertinente. Exponer claramente el objetivo de la RS es importante por dos razones fundamentales:

1.- Ayuda al autor a determinar las estrategias para la selección de la información (ejemplos: las RS referentes a eficacia terapéutica podrían limitarse a datos de EC, las RS referentes a etiología deben hacer especial hincapié en los estudios de casos y controles,...), y

2.- Permite que el lector decida si la revisión trata de un tema relevante para el interés de su práctica clínica^{19,20}. Dos tareas resultan prioritarias en la elaboración de la pregunta que define el marco conceptual: definir cuál es la medida de efectividad que va a medir las virtudes de la intervención estudiada y determinar los factores (cofactores, confusores y predictores) que pueden influir en los resultados del análisis.

b. Componentes clave de una pregunta:

Cualquier pregunta clínica estructurada (sea de etiología, diagnóstico, tratamiento, prevención o pronóstico) se compone de los siguientes componentes: el paciente o problema de interés (definiendo las enfermedades, la población y el ámbito de interés), la intervención que se va a estudiar, la intervención que se va a comparar (en el caso de que sea pertinente disponer de un grupo control) y la(s) variable(s) o resultados que se valoran. Habitualmente utilizamos el acrónimo PICO (Patient, Intervention, Comparison, Outcome) para recordar las partes fundamentales de una pregunta clínica estructurada.

c. Identificación y selección de los estudios mediante los componentes clave de la pregunta:

El tipo de pregunta condiciona el tipo de estudio o diseño preferencial que responderá a la cuestión planteada: si la pregunta es cuándo, cuánto, dónde y quién se valorarán estudios descriptivos; si la pregunta es qué exposición produce un daño se valorarán estudios de casos y controles; si la pregunta es qué daño produce una exposición se valorarán estudios de cohortes; si la pregunta es cuánto disminuye el daño la aplicación de un tratamiento se valorarán ensayos clínicos; si la pregunta es valorar la operatividad de una prueba diagnóstica se valorarán estudios sobre pruebas diagnósticas²⁰.

d. Amplitud de las preguntas:

Son varias las ventajas y desventajas que se derivan de formular preguntas amplias o restringidas. Se prefiere en las RS el enfoque hacia preguntas restringidas, si bien presenta el inconveniente de que quizá no sean generalizables a entornos diversos.

e. Modificación de las preguntas:

Las preguntas deben plantearse en el protocolo antes de iniciar la revisión completa, pero no debe convertirse en una coraza que impida la exploración de aspectos inesperados. Sin embargo, las preguntas a posteriori son más proclives a sesgos.

Segundo paso: localización y selección de los estudios

a. Fuentes de información:

El lector necesita tener la seguridad de que todas las publicaciones pertinentes e importantes se han incluido en la RS. Cuanto más selectivo o anárquico sea el método de búsqueda, más probable será que haya sesgos en la revisión. Lo ideal es que dichos métodos incluyan una búsqueda en varias bases de datos bibliográficos (especificando las palabras clave, los años de revisión y otros

aspectos para mejorar el rendimiento de las estrategias de búsqueda), una investigación de los informes que sean citados en artículos importantes, las referencias de todos los artículos pertinentes encontrados y la comunicación personal con investigadores sobre el tema que se está revisando. Es importante garantizar que no se omitan artículos importantes publicados y, en particular, detectar artículos metodológicamente adecuados que no hayan sido publicados. Dado que el estudio que ofrece resultados positivos tiene más posibilidades de ser publicado que el que ofrece resultados negativos, se corre el riesgo de cometer sesgo de publicación en la RS²¹.

Ya hemos comentado la clasificación, desde un punto de vista didáctico, de las fuentes de información bibliográfica en dos grandes grupos: fuentes de información secundarias y primarias⁵.

b. Desarrollo de una estrategia de búsqueda:

Internet constituye la herramienta fundamental en la búsqueda de información bibliográfica. La búsqueda electrónica o informatizada (Medline, Embase, Excerpta Médica, Índice Médico Español, Cochrane Library, etc) es la fuente más utilizada para identificar los trabajos más importantes publicados sobre un tema, pero si confiamos sólo en este método de búsqueda perderemos parte de información relevante. Otras posibilidades de búsqueda para identificar información relevante son: 1) la búsqueda manual de bases bibliográficas (Index Medicus, Excerpta Medica, Current Contents, etc); 2) la revisión de las listas de referencia de los artículos encontrados, revisiones, libros de texto,...; 3) la búsqueda de la bibliografía "gris" u "opaca", como tesis doctorales, libros de actas a congresos, informes de instituciones,...; 4) el intercambio directo de información con expertos en el tema que se investiga; 5) la búsqueda de registros internacionales de los estudios (básicamente EC) realizados sobre un tema concreto, tanto los publicados como los no publicados; 6) agencias financiadoras (FIS, CICYT, fundaciones, etc.); 7) industria farmacéutica; etc. Deben usarse exhaustivamente todas las fuentes de información a nuestro alcance para minimizar la probabilidad de omitir estudios relevantes, y detallar siempre los procedimientos de búsqueda empleados⁵.

Aumentar la exhaustividad en una búsqueda implica reducir su precisión e identificar más artículos no relevantes. Por ello siempre es necesario mantener un equilibrio entre el grado de exhaustividad (búsquedas sensibles) y precisión (búsquedas específicas). Pese a la necesidad de elaborar estrategias eficientes de búsqueda de información bibliográfica se producen numerosas pérdidas de información por la inexperiencia del investigador en la búsqueda en bases de datos mecanizadas y por la ingente cantidad de investigación no registrada en ellas (bibliografía "gris" u "opaca").

c. Selección de estudios:

Los autores deben especificar los criterios de inclusión y exclusión a la hora de seleccionar los artículos válidos para la RS, así como seleccionaron los artículos haciendo referencia a los tres elementos básicos de su objetivo, bajo el acrónimo PICO (población, exposición o intervención, comparación y resultado). Del conjunto de artículos que

se extraen de la literatura para una revisión, muchos no serán pertinentes al tema que se estudia y otros serán deficientes desde el punto de vista metodológico.

d. Documentación de una estrategia de búsqueda:

La exposición explícita de la metodología utilizada para la RS permite que el lector pueda evaluar con conocimiento de causa el rigor científico de la revisión y la solidez de los datos en que se apoyan las inferencias. Los lectores que necesiten respuestas a problemas clínicos específicos deben buscar RS que indiquen claramente los métodos utilizados. Sin conocer los métodos de la RS el lector no puede distinguir las afirmaciones basadas en pruebas (nuevo paradigma, más objetivo) de aquéllas apoyadas únicamente en la opinión de quien escribió la revisión (antiguo paradigma, más subjetivo).

Tercer paso: evaluación de la calidad de los estudios

a. Validez. Sesgos en los EC:

La validez de un estudio se refiere a la capacidad de evitar los errores sistemáticos o sesgos mediante el correcto diseño y realización del estudio. El término sesgo hace referencia a los problemas de validez interna de una investigación, esto es, se plantea si los resultados obtenidos son ciertos en la población de referencia a la que se pretende aplicar los resultados. En los estudios sobre los efectos de la atención sanitaria pueden detectarse cuatro tipos de sesgos^{10,22}:

- Sesgo de selección (selection bias): conviene instaurar un programa de asignación aleatoria implantado por alguien no implicado en el reclutamiento (ocultación de la asignación).
- Sesgo de realización (performance bias): diferencias sistemáticas en la atención sanitaria proporcionada a los participantes en los grupos de comparación.
- Sesgo de desgaste o pérdida (attrition bias): diferencias sistemáticas en la pérdida de participantes durante el estudio entre los grupos de comparación;
- Sesgo de detección (detection bias): diferencias sistemáticas en la evaluación de resultados entre los grupos de comparación; el cegado es particularmente importante en aquellas investigaciones que utilizan medidas subjetivas de resultado.

b. Evaluación de la validez y calidad de los estudios.

Una RS llegará a conclusiones adecuadas solo si valora con mucha meticulosidad la validez de los estudios primarios en que se basa la revisión. Si la mayoría de los estudios primarios tienen deficiencias básicas, sus conclusiones pueden ser cuestionables, incluso cuando sus resultados son comparables. Existen guías metodológicas para los distintos tipos de investigaciones sobre etiología, diagnóstico, tratamiento, prevención o pronóstico, usando los mismos criterios de validez que para las publicaciones originales: diseño del estudio, aleatorización, enmascaramiento, análisis por intención de tratar, etc.²³. Basarse en que el artículo ha sido sometido al proceso de revisión por expertos (peer review) de las revistas no garantiza de forma absoluta la calidad de la investigación que se publica en ella. La competencia científica del autor y el prestigio de

la revista biomédica son criterios importantes, pero no suficientes para la credibilidad de una investigación²⁴. Lo ideal sería que todos los estudios primarios potencialmente apropiados (en base a los criterios de inclusión y exclusión de la RS) fueran evaluados al menos por dos autores, cada uno desconocedor de la decisión del otro. El grado de concordancia debería hacerse constar. Aparte del problema de reproducibilidad, también deben tenerse en cuenta las causas principales de sesgo, que se relacionan con los autores, sus instituciones y los resultados del estudio, tal como se ha demostrado en algunos estudios: los trabajos procedentes de instituciones respetadas y que exponen resultados positivos tienen mayor posibilidad de ser aceptados. Lo ideal sería que se evaluara el contenido y calidad de los estudios incluidos en una RS a partir de su metodología, sin que los revisores conocieran a los autores, instituciones y resultados. Sin duda esta precaución adicional reforzaría las conclusiones de la RS.

Los problemas de reproducibilidad y sesgo pueden afectar a dos etapas del proceso de RS: la decisión final sobre qué estudios se incluirán y la valoración de la calidad de los estudios incluidos. Estos problemas pueden minimizarse si se observan los criterios explícitos previamente expuestos. Se han identificado muchas escalas para valorar la validez y calidad de los EC; una de las más empleadas es la propuesta por Jadad y cols²⁵, en la que se puntúan por separado tres componentes esenciales: la aleatorización, el doble ciego y la descripción de las pérdidas durante el seguimiento.

El análisis de calidad metodológica es importante por dos razones: el grado de "evidencia" es más débil en presencia de problemas de calidad metodológica, y las variaciones en calidad pueden explicar diferencias encontradas en los estudios. También es posible ponderar los estudios según la validez, de manera que los estudios más válidos tengan más influencia en la medida resumen de efecto (no obstante, no existe una base empírica para determinar qué pesos debe asignarse a los distintos criterios de validez, por lo que este enfoque es de dudosa utilidad).

Cuarto paso: proceso de extracción de datos

El formulario de recogida de datos es un puente entre lo que han informado los investigadores de los estudios primarios y lo que, en último término, informa un revisor. El formulario de extracción de datos debe incluir características del acrónimo PICO (población, exposición o intervención, comparación y resultado –sin olvidar el factor tiempo–).

No existe un modelo universalmente válido para codificar la información de los estudios a combinar. Las principales variables que deben extraerse de los estudios originales son:

- Identificación del estudio: autores y país, fuente de financiación
- Fecha de publicación del estudio
- Fecha de inicio y final del trabajo
- Problema estudiado o hipótesis puesta a prueba
- Características de los individuos: edad, sexo, diagnósticos,

duración del seguimiento, evolución, etc.

- Diseño del estudio: experimental (con aleatorización o no), observacional (prospectivo o retrospectivo).
- Exposición o tratamiento estudiado: dosis, pauta, duración, etc.
- Co-exposición o co-tratamiento
- Resultados que permiten medir los efectos principales y secundarios: resultados brutos y medida de efecto ajustada con sus intervalos de confianza o valores p asociados e interacciones representativas
- Valoración de la calidad o validez de los estudios
- Susceptibilidad del estudio a posibles sesgos, aspectos destacables y limitaciones del estudio, y explicaciones que los autores presentan de los resultados.

Esta etapa de recogida de la información puede tener una considerable influencia sobre la conclusión final, en el que pueden surgir dos problemas: 1) diferencias en el proceso y calidad de recogida de la información, lo que explica en parte que existan RS sobre un mismo tema que obtienen conclusiones opuestas; para evitar estos sesgos de observación, se intenta que la extracción de datos la realicen al menos dos personas; y 2) la falta de información relevante que se obtiene a veces en las publicaciones originales; de ahí lo importante que puede resultar que el autor del RS contacte con los autores de los estudios primarios, por si fuera preciso solicitarles información adicional.

Quinto paso: análisis y presentación de resultados. Metanálisis.

Al realizar una RS se encontrarán variaciones en los resultados de los estudios en los que se investiga la cuestión de interés. La tarea del revisor es precisamente explicar dicha variabilidad. Las posibles fuentes de variación son el diseño de la investigación, la pura casualidad y las diferencias en los tres aspectos básicos del estudio (población, exposición o intervención y resultado medido). Así, los lectores de una revisión deben verificar si se han considerado estas cinco explicaciones de la diferencia de resultados de los estudios y deben ser escépticos cuando las diferencias se atribuyen a una explicación sin dar la debida consideración a las demás. La síntesis final de información debe implicar una integración sistemática y no selectiva, en la que pueden asignarse valores relativos a los datos, en función de los resultados de la valoración estandarizada. Los revisores deben estar dispuestos a usar técnicas de síntesis cuantitativa (MA) para complementar y suplementar las técnicas cualitativas. Hay que ser precavidos contra las RS que concluyen la inexistencia de un efecto sin haber considerado la potencia estadística de los estudios para detectar un efecto clínicamente importante, lo que resulta especialmente ventajoso en condiciones de baja prevalencia o cuando se evalúen efectos de pequeña magnitud.

a. Justificación para realizar un MA en una revisión:

Un paso más allá de la RS se encuentra el MA, que es un conjunto de procedimientos estadísticos cuyo objetivo es hacer una síntesis cualitativa y cuantitativa de los resultados de diferentes estudios independientes que tratan de un mismo tema, bajo la idea de que la

combinación de estudios individuales (preferentemente EC, y con menos frecuencia estudios observacionales) permitiría obtener una información más completa.

La razón de realizar RS es asegurar la validez de los resultados. Por eso debe tenerse en cuenta que el papel del análisis estadístico en las revisiones puede estar menos claro y, a veces, el MA puede parecer más un obstáculo que una ayuda. Además, cabe recordar dos razones importantes para no realizar un MA en una revisión: la principal es la falta de datos válidos y relevantes, y la segunda es que a veces el estudio estadístico no tiene sentido. Los objetivos que se persiguen con la realización de un MA son los siguientes^{9,26,27}:

- 1.- Obtener un mayor tamaño muestral, lo que permite aumentar la potencia o poder estadístico, y así se pueden detectar diferencias que pasaron inadvertidas en los estudios individuales.
- 2.- Obtener conclusiones cuando los resultados de los estudios individuales son contradictorios.
- 3.- Aumentar la generalización de las conclusiones, dada la heterogeneidad de las poblaciones de los diferentes estudios.
- 4.- Responder a preguntas no planteadas en los estudios originales.

Así pues, el MA es una estrategia sistemática de revisión que utiliza procedimientos estadísticos que resulta especialmente útil cuando los tamaños muestrales individuales son demasiado pequeños para detectar un efecto y calificarlo de estadísticamente significativo, cuando los resultados de varios estudios son contradictorios en cuanto a la dirección o magnitud del estudio, o cuando una investigación a gran escala exige demasiado tiempo o dinero. Los MA son proyectos de investigación por sí mismos, en los que las unidades de observación son los estudios originales. Los MA requieren una adecuada planificación y una dedicación considerable de recursos humanos, en la que se debería constituir un equipo investigador formado por investigadores con experiencia en el tema objeto de estudio, epidemiólogos y/o bioestadísticos con experiencia en la realización de MA y documentalistas expertos en la realización de búsquedas bibliográficas exhaustivas.

b. Resumen de los efectos de los estudios de un MA y métodos estadísticos:

El primer paso es calcular en cada uno de los estudios combinados las medidas del efecto que nos interesa (RAR, RRR, NNT,...) de una variable (tratamiento, exposición,...) sobre otra (mejoría, curación, supervivencia,...), controlando si es necesario los potenciales factores confusores (sexo, edad, gravedad,...).

Estas medidas de efecto pueden obtenerse a partir de los datos individuales, si se detallan en los estudios, o tomarse los efectos presentados. También es posible estimar el efecto a partir de los estadígrafos *t* de Student, *F* de Snedecor, "ji" cuadrado, coeficiente de correlación de rangos de Spearman, o incluso de los valores de *p*. Una exposición detallada de los diferentes métodos utilizables para combinar resultados de diferentes estudios está

fuera de nuestro objetivo, pero pueden clasificarse en dos grandes grupos: los modelos de efectos fijos (fixed effects) y los modelos de efectos aleatorios (random effects).

Los modelos de efectos fijos asumen que todos los estudios estiman el mismo efecto, es decir, que las diferencias que se observan entre los estudios se deben únicamente a variabilidad aleatoria. Depende de la variabilidad intraestudios. Los principales métodos de efectos fijos son el Mantel-Haenszel (el más atractivo y de elección en los MA realizados a partir de EC con variables de efecto dicotómica), el de Woolf (menos eficiente que el anterior, pero tiene la ventaja que puede aplicarse a los MA de EC en los que se combinan riesgos relativos, odds ratios, diferencias de medias, etc) y el de Yusuf-Peto (menos utilizado, dado que es bastante intuitivo y puede llevar a resultados paradójicos).

Los modelos de efectos aleatorios se suelen aplicar cuando los estudios tienen un grado de heterogeneidad superior al esperado por simple variabilidad aleatoria. Depende de la variabilidad intra e interestudios. Se utiliza el método de DerSimonian y Laird que incorpora la variabilidad interestudio en el estimador combinado, y cuyo uso es controvertido. Ciertos autores argumentan que deberían ser habituales en los MA, ya que es frecuente la existencia de heterogeneidad residual no explicada. Es frecuente, por otro lado, encontrar MA que presentan tanto los resultados de modelos de efectos fijos como los de efectos aleatorios^{9,26,27}.

Siempre habrá opiniones distintas sobre cuál es el método más apropiado para hacer un MA particular, de manera que el investigador debe preguntarse en todos los casos en qué medida sus resultados metanalíticos son sensibles a la manera en que se realizó el MA. Nunca está absolutamente claro si usar modelos de efectos fijos o de efectos aleatorios: en el caso de estudios homogéneos se obtienen resultados similares con ambos modelos (varía algo el intervalo de confianza), en el caso de estudios heterogéneos se obtienen resultados diferentes con ambos modelos, y en el caso de estudios con gran heterogeneidad, conviene no realizar MA.

Los métodos estadísticos utilizados en el MA son múltiples y pueden encontrarse integrados en programas informáticos que pueden facilitar mucho la realización de un MA o una RS, como DESCARTES, EASY MA, FAST*PRO, META-ANALYST, TRUE EPISTAT,... pero uno de los más conocidos es el programa REVMAN utilizado por la Colaboración Cochrane.

c. Evaluación de la heterogeneidad:

El supuesto que subyace a la combinación de resultados de estudios individuales en una estadística descriptiva es que sus diferencias son solo casuales (variación de muestreo) y que, por tanto, los resultados de todos los estudios son homogéneos, es decir, que reflejan el mismo efecto "verdadero". La falta de uniformidad de los resultados medidos en los estudios puede reflejar el hecho de que el efecto del tratamiento varía en función de las características particulares de los estudios, por ejemplo, las dosis y/o la edad de los pacientes. No solo es importante saber la efectividad terapéutica, también

es importante la cuestión de en qué pacientes y bajo qué circunstancias el tratamiento funciona mejor. En la tabla 3 se presentan posibles causas de heterogeneidad entre los estudios originales.

Un cierto grado de heterogeneidad entre los estudios combinados puede ser saludable para aumentar la capacidad de generalización (validez externa), pero, no obstante, una notable heterogeneidad estadística de efectos no es deseable. El límite de homogeneidad en los efectos que debemos aceptar nos vendrá dado por los procedimientos estadísticos habituales para probar que las diferencias entre las medidas de los efectos en los estudios primarios no son mayores de lo que cabría esperar por la influencia del azar. También será útil una representación gráfica de los efectos de los estudios primarios (que nos permitirá detectar los valores extremos y/o los patrones relacionados con las características clínicas) y sobre todo el uso de la lógica clínica, que deberá prevalecer sobre unos criterios exclusivamente matemáticos. Desde un punto de vista global cabe recordar que el test de homogeneidad aplicado a un MA puede tener escaso poder cuando se comparan menos de diez EC.

Cuando se juzga que hay una relativa homogeneidad, puede obtenerse una medida sumaria del resultado final combinando los resultados de los estudios primarios. Todos los lectores deberían buscar en la publicación de un MA una discusión del tema de la homogeneidad (mediante pruebas gráficas y estadísticas) previa al uso de procedimientos de combinación que la dan por supuesta. Se suele estimar el punto de corte de la "ji" cuadrado de la homogeneidad en 0,1: si $p > 0,1 =$ heterogeneidad, si $p < 0,1 =$ homogeneidad²⁷.

El gran problema del MA es la heterogeneidad; ¿cómo combatirla?: 1) restringir objetivos; 2) plantear análisis de subgrupos; 3) y si no es posible evitarla, se deben investigar los resultados para identificar las fuentes de heterogeneidad: análisis de la sensibilidad (estudios de mayor calidad frente estudios de menor calidad, estudios enmascarados frente no enmascarados, etc.), meta-regresión... Si hay heterogeneidad significativa, los revisores deben tener precaución al interpretar la estimación global derivada de un MA, así como al atribuir las diferencias entre los estudios a cualquier factor. Cuando hay heterogeneidad sustancial en los resultados hay desacuerdo en la forma de resumirlos y en si estos deben combinarse; sin embargo, existe un acuerdo en que los revisores deben alertar al lector cuando haya heterogeneidad sustancial y, por tanto, deben recomendar una interpretación cautelosa de los resultados.

Se debe tener en cuenta que los test estadísticos no pueden compensar la falta de sentido común, la perspicacia clínica y la plausibilidad biológica. De este modo, existen revisiones en que la síntesis de datos no es posible (o mejor, no es deseable) y se quedan en honrosas RS cualitativas^{9,10}.

d. Análisis de subgrupos:

Se refiere a las agrupaciones particulares de pacientes entre los estudios, no a subgrupos de estudios. Con frecuencia se abusa de los análisis de subgrupos, por

lo que cabe plantearse las siguientes preguntas para decidir si debe realizarse: 1) ¿hay evidencia indirecta que apoye la existencia de una diferencia en algún subgrupo en particular?, 2) ¿las hipótesis acerca de la diferencia preceden al análisis y no surgen después de éste?, 3) ¿la diferencia se insinúa a partir de comparaciones dentro de los estudios y no entre los estudios?, 4) ¿la diferencia es consistente a lo largo de los estudios?, 5) ¿la magnitud de la diferencia es clínicamente relevante?, 6) ¿la diferencia es estadísticamente significativa?¹⁰.

e. Análisis de sensibilidad:

Puede utilizarse como herramienta para analizar por qué dos MA sobre un mismo tema obtienen conclusiones diferentes; con este análisis se intenta responder a la pregunta ¿serían diferentes los resultados finales de la técnica estadística si se modifican los datos dudosos que se han utilizado en el análisis? El análisis de sensibilidad se realiza repitiendo la técnica estadística realizada con diferentes valores para determinados datos y observando si la modificación cambia sustancialmente el efecto final; si ocurre así, los resultados no tienen la solidez suficiente y, por tanto, es necesario ser muy cauto en la valoración de los datos y de las conclusiones.

f. Presentación de los resultados del MA:

Con el objetivo de mejorar y homogeneizar la calidad de la presentación de los resultados de las RS y MA, el grupo QUOROM ha publicado recientemente un documento donde se destacan los puntos que deberían existir ineludiblemente^{28,29}, y que incluye una lista-guía (checklist) y un diagrama de flujo (flow diagram) de los pacientes para evaluar correctamente su presentación. La presentación gráfica de los resultados es un método complementario de análisis de los resultados combinados y del resultado global del MA que tiene la ventaja de ser intuitivo y fácil de visualizar, pero en el que debe cuidarse la elección de las escalas para conseguir una impresión visual no distorsionada. Cada línea horizontal representa el resultado de un EC (cuanto más corta es la línea, más cierto es el resultado, pues menor es el intervalo de confianza), cuya autoría y año de realización del EC se indica al margen. También se representa en una línea horizontal el resultado global de la combinación de todos los EC. La línea vertical indica la posición alrededor de la cual las líneas horizontales se concentrarán si los dos tratamientos comparados en los EC tuviesen efectos similares. La posición del punto a la izquierda de la línea vertical indica que el tratamiento estudiado es beneficioso, y su posición a la derecha que es perjudicial. Si una línea horizontal toca la línea vertical, significa que aquel EC concreto no encontró diferencias significativas entre los tratamientos.

La virtud de esta presentación gráfica es que consigue hacer visualmente sencillo toda la complejidad metodológica del MA.

g. Interpretaciones erróneas y limitaciones de los MA:

Después de un notable entusiasmo inicial de los MA, las discrepancias entre las conclusiones extraídas de algunos MA y los hallazgos encontrados en macroensayos

aleatorizados³⁰ han llevado a un planteamiento más crítico del MA. Se profundiza más en el análisis de la heterogeneidad entre estudios y sus posibles causas, en la influencia de las características de los pacientes incluidos y en el análisis de sensibilidad de los resultados.

El MA es complementario, no sustituto, del EC aleatorizado de gran tamaño y representa una estrategia coste-efectiva porque reduce el número de estudios primarios innecesarios.

Sexto paso: interpretación de los resultados

La síntesis final de información debe implicar una integración sistemática y no selectiva, en la que pueden asignarse valores relativos a los datos, en función de los resultados de la valoración estandarizada. Los revisores deben estar dispuestos a usar técnicas de síntesis cuantitativa (MA) para complementar y suplementar las técnicas cualitativas (RS). Hay que ser precavidos contra las RS que concluyen la inexistencia de un efecto sin haber considerado la potencia estadística de los estudios para detectar un efecto clínicamente importante, lo que resulta especialmente ventajoso en condiciones de baja prevalencia o cuando se evalúen efectos de pequeña magnitud. Los resultados de los estudios primarios individuales, se haya utilizado una síntesis cualitativa y/o cuantitativa, deben indicarse con suficiente detalle para que los lectores sean capaces de evaluar críticamente el fundamento de las conclusiones a las que se llega en la RS. El método de presentación de resúmenes de los estudios individuales dependerá de la cuestión estudiada^{9,10}:

- En estudios de etiología se deben evaluar los criterios de inferencia causal, en el que los aspectos más comunes son la magnitud y consistencia de la asociación, y la necesidad de demostrar la relación temporal apropiada.
- En estudios de métodos diagnósticos se debe indicar la sensibilidad, especificidad y cocientes de probabilidad (con sus intervalos de confianza).
- En estudios de efectividad del tratamiento y prevención se debe indicar la magnitud del efecto (con su intervalo de confianza).
- En estudios de pronóstico son importantes las curvas de supervivencia.

Las conclusiones están justificadas solo cuando el proceso de recogida, análisis e integración de la información se aplica de forma completa y sistemática. Se debe aprovechar el esfuerzo de la RS para identificar los huecos del conocimiento actual en ese tema y sugerir recomendaciones sobre futuras iniciativas. De esta manera, el lector de la RS tiene una visión de conjunto de lo que se sabe y no se sabe sobre el tema.

En la RS se realiza una revisión estructurada, explícita y sistemática de los diferentes estudios independientes que tratan de un mismo tema (síntesis cualitativa), pero además, el MA utiliza procedimientos estadísticos para combinar los resultados de los estudios originales (síntesis cuantitativa). Dada la elevada calidad de la evidencia científica que se desprende de la RS y MA, éstos se han convertido en una herramienta clave en la toma de decisiones en salud y en la evaluación de las necesidades

de investigación y, a menudo, se convierte en punto de partida para el trabajo de grupos de consenso, paneles de expertos o comisiones con responsabilidades reguladoras y de alto impacto sanitario.

Bibliografía:

- 1.- Gálvez Toro A. Métodos contemporáneos de síntesis científica: una propuesta de homogeneidad. *Med Clin (Barc)*. 2003;121:256-63.
- 2.- 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.
- 3.- Jovell AJ, Navarro Rubio MD. Evaluación de la evidencia científica. *Med Clin (Barc)*. 1995; 105:740-3.
- 4.- González de Dios J. De la Medicina basada en la evidencia a la Evidencia basada en la medicina. *An Esp Pediatr*. 2001;55:429-39.
- 5.- González de Dios J, Buñuel Álvarez JC. Búsqueda eficiente de las mejores pruebas científicas disponibles en la literatura: fuentes de información primarias y secundarias. *Evid Pediatr*. 2006;2:12.
- 6.- Balaguer A, González de Dios J. Digiriendo la información científica. Excelencias y limitaciones de las revisiones sistemáticas y metanálisis. *Acta Pediatr Esp*. 2004;62:4-10.
- 7.- González de Dios J. Las mejores evidencias (o pruebas) científicas (I): el ensayo clínico. *Rev Esp Pediatr*. 1999;55:157-65.
- 8.- González de Dios J. Las mejores evidencias (o pruebas) científicas (II): la revisión sistemática. *Rev Esp Pediatr*. 1999;55:250-6.
- 9.- González de Dios J. Las mejores evidencias (o pruebas) científicas (III): el metaanálisis. *Rev Esp Pediatr*. 1999;55:368-76.
- 10.- Gisbert JP, Bonfill X. ¿Cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis?. *Gastroenterol Hepatol*. 2004;27:129-49.
- 11.- Siwek J, Gourlay ML, Slawson DC, Shaughnessy AF. How to writ and evidence-based clinical review article. *Am Fam Physician*. 2002;65:251-8.
- 12.- Oxman AD, Cook DJ, Guyatt GH. User's guides to the medical literature. VI. How to use an overview. *Evidence-Based Medicine Working Group*. *JAMA*. 1994;272:1367-71.
- 13.- López Arrieta JM, Qizilbash N. La medicina basada en pruebas: revisiones sistemáticas. *La Colaboración Cochrane*. *Med Clin (Barc)*. 1996;107:581-5.
- 14.- González de Dios J. Fuentes de información en Pediatría: ¿dónde encontrar las mejores evidencias de la bibliografía? (III) *Colaboración Cochrane y revistas con resúmenes estructurados*. *Pediatr Integral*. 2000;5:657-62.
- 15.- Piehl JH, Green S, McDonald S. Converting systematic reviews to Cochrane format: a cross-sectional survey of australian authors of systematic reviews. *BMC Health Service Research*. [en línea] [fecha de consulta: 8-XI-2007]. URL disponible en: <http://www.biomedcentral.com/1472-6963/3/2>
- 16.- Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, et al. Methodology and reports of systematic reviews and meta-analyses. A comparison of Cochrane Reviews with articles published in paper-based journals. *JAMA*. 1998;280: 278-80.
- 17.- Clarke M, Oxman AD, editors. *Cochrane Reviewers' Handbook 4.2* [update March 2003]. En: *Review Manager (RevMan)* [computer program]. Version 4.2. Oxford, England: The Cochrane Collaboration, 2003. Disponible en: <http://www.cochrane.org>

chrane.de/cochrane/hbook.htm

18.- Centro Cochrane Iberoamericano. Manual de revisores (versión española de "Cochrane Reviewers' Handbook") [actualización de junio 2000]. Barcelona: Centro Cochrane Iberoamericano, 2001. Disponible en: <http://www.cochrane.es/?q=es/handbook>

19.- Buñuel Álvarez JC, Ruiz-Canela Cáceres J. Cómo elaborar una pregunta clínica. Evid Pediatr. 2005;1:10.

20.- Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med. 1997;127:380-7.

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

22.- Delgado Rodríguez M, Sillero Arenas M. Sesgos en el metaanálisis. Med Clin (Barc). 1999; 112(Supl 1):43-50.

23.- González de Dios J, Buñuel Álvarez JC, Ochoa Sangrador C. La valoración crítica de documentos científicos y su aplicabilidad a la práctica clínica: aspecto clave en la toma de decisiones basada en las mejores pruebas científicas. Evid Pediatr. 2006; 2:37.

24.- Gervas J, Pérez Fernández M. La revisión por pares en las revistas científicas. Aten Primaria. 2001;27:432-9.

25.- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary?. Control Clin Trials. 1996;17:1-12.

26.- Fernández Ballart JD, Vobecky J, Martí Henneberg C. Metaanálisis: síntesis e integración de los resultados de estudios independientes en medicina. Med Clin (Barc). 1991;96:382-7.

27.- Molinero LM. Interpretación, utilidad y limitaciones del metaanálisis. Gastroenterol Hepatol Continuada. 2002;1:311-8.

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

29.- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Mejora de la calidad de los informes de metaanálisis de ensayos clínicos controlados: el acuerdo QUOROM. Rev Esp Salud Pública. 2000;74:107-18.

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

Tabla 1: Características diferenciales de los distintos tipos de revisiones		
	Revisión narrativa	Revisión sistemática
Pregunta	Frecuentemente amplia y poco definida	Pregunta clara y concreta sobre una cuestión definida
Fuentes-Búsqueda	No especificadas. Alta probabilidad de sesgo	Estrategia especificada, sistemática y explícita
Selección de estudios	No especificadas. Alta probabilidad de sesgo	Criterios especificados y aplicados de forma uniforme
Calidad de estudios	Variable pero generalmente no evaluada	Crítica y rigurosa (según criterios explícitos)
Síntesis de datos	Frecuentemente cualitativa, subjetivo y sin un estimador estadístico	Cuantitativa, si es posible, a través de un estimador estadístico
Inferencia	Variable, en ocasiones basada en la evidencia, pero más frecuente basado en opiniones personales	Basadas en "evidencias" y con la identificación de las lagunas de conocimiento que persisten

Tabla 2: Estructura de una revisión sistemática en la Colaboración Cochrane

- Carátula
 - o Título
 - o Nombre de los revisores
 - o Fuentes de financiación
 - o Qué contenido es nuevo
- Texto de la revisión
 - o Sinopsis
 - o Resumen
 - o Antecedentes
 - o Objetivos
 - o Criterios para la selección de los estudios de la revisión:
 - Tipos de estudios
 - Tipos de participantes
 - Tipos de intervenciones
 - Tipos de medidas de resultados (outcomes)
 - o Estrategia de búsqueda para la identificación de estudios
 - o Métodos de revisión
 - o Descripción de los estudios
 - o Calidad metodológica
 - o Resultados
 - o Discusión
 - o Conclusiones de los revisores:
 - Implicaciones para la práctica
 - Implicaciones para la investigación
 - o Agradecimientos
 - o Conflictos de interés
- Referencias
 - o Referencias de los estudios
 - Estudios incluidos
 - Estudios excluidos
 - Estudios pendientes de evaluación
 - Estudios en curso
 - o Otras referencias
 - Referencias adicionales
 - Otras versiones publicadas de esta revisión
- Tablas y figuras
 - o Características de los estudios incluidos
 - o Características de los estudios excluidos
 - o Características de los estudios en curso
 - o Comparaciones, datos y gráficos
 - o Tablas adicionales

Tabla 3: Fuentes de heterogeneidad entre los artículos originales

- Características del diseño de los estudios:
 - o Tipo de diseño
 - o Metodología en la recogida de la información (forma de medir las variables)
 - o Tipo de análisis estadístico empleado
- Características de las poblaciones de los estudios:
 - o Procedencia del grupo de expuestos (o casos)
 - o Procedencia del grupo de no expuestos (o controles)
 - o Criterios de inclusión / exclusión de los estudios
 - o Modificaciones de la población durante el seguimiento (entradas-salidas)
- Problemas de error sistemático:
 - o Control del fenómeno de confusión
 - o Control de los problemas de selección y seguimiento
 - o Control de los problemas de clasificación