

¿Cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis?

J.P. Gisbert^{a,b} y X. Bonfill^c

^aServicio de Aparato Digestivo. Hospital Universitario de La Princesa. Universidad Autónoma de Madrid. Madrid. España.

^bEditor del *Cochrane Upper Gastrointestinal and Pancreatic Diseases Group*.

^cCentro Cochrane Iberoamericano. Servicio de Epidemiología Clínica y Salud Pública. Hospital de la Santa Creu i Sant Pau. Universidad Autónoma de Barcelona. Barcelona. España.

INTRODUCCIÓN

Una de las definiciones más populares de la tan nombrada medicina basada en la evidencia es la de «la utilización consciente, explícita y juiciosa de la mejor evidencia clínica disponible para tomar decisiones sobre el cuidado de los pacientes». La práctica de la medicina basada en la evidencia significa, por tanto, la integración de la maestría clínica individual con las mejores evidencias científicas externas disponibles a partir de una investigación clínica. Aunque este último componente parece, en teoría, relativamente fácil de adquirir, en la práctica los profesionales sanitarios se ven desbordados por una cantidad de información imposible de manejar.

Así, hoy más que nunca existen diversos motivos que hacen necesario gestionar correctamente el conocimiento¹. En primer lugar, sufrimos actualmente un exceso de información; se ha pasado en tan sólo unos pocos años de no disponer de fuentes de información a nadar en una sobreabundancia que nos satura y puede ser ciertamente contraproducente. Se calcula que se habrán realizado cientos de miles de ensayos clínicos aleatorizados sobre diferentes aspectos de la asistencia sanitaria desde 1948, año en el que se publicó el primero de estos estudios². Y el futuro se promete imparable en cuanto a la expansión continua de la información obtenida mediante la investigación. A modo de ejemplo del exceso de información al que nos vemos sometidos, si introducimos el término

«enfermedad de Crohn» o «*Crohn's disease*» en cualquiera de los buscadores de Internet actualmente disponibles (Altavista, Yahoo, Google, etc.), encontraremos más de 300.000 direcciones con información relacionada, lo que evidentemente supone una cifra desorbitada.

En segundo lugar, disponemos en la actualidad de un número inagotable de fuentes¹; así, se publican anualmente más de 25.000 revistas y más de 10 millones de referencias en MEDLINE. Por ejemplo, si introducimos de nuevo el término «*Crohn's disease*» en la casilla de PubMed (el motor de búsqueda más conocido de MEDLINE), identificaremos casi 20.000 referencias bibliográficas relacionadas con esa enfermedad, lo que de nuevo supone una cifra realmente difícil de digerir. En este sentido, un médico que intente mantenerse al día sólo en su especialidad debería leer, teóricamente, más de 100 artículos completos cada semana, o más de 500 el día de vuelta de sus vacaciones de verano³. Por otro lado, el tiempo disponible para la búsqueda de información y el estudio es cada vez más limitado. Como consecuencia de lo anterior, los conocimientos del médico se deterioran rápidamente tras finalizar la carrera de medicina (o el programa de formación continuada que posteriormente se haya llevado a cabo). Este hecho está magníficamente expresado en una figura ya clásica, donde el eje de ordenadas representa los conocimientos actualizados sobre el mejor tratamiento y el eje de abscisas, los años desde la graduación; y en dicho gráfico se observa cómo existe lo que se ha denominado una peligrosa «pendiente resbaladiza».

En tercer lugar, la calidad de las diversas fuentes de información es desigual, por lo que el problema al que nos enfrentamos no es sólo el exceso de información, sino el que la calidad de la mayoría de las fuentes es deficiente¹. Por último, los contenidos de las fuentes de información se quedan obsoletos rápidamente, lo que supone una limitación adicional de los sistemas actuales. Por tanto, necesitamos disponer de herramientas que nos permitan acceder a la información adecuada en términos de cantidad, calidad y actualidad.

Esta revisión ha sido realizada en parte gracias a la concesión de las becas del Instituto de Salud Carlos III (C03/02 y G03/090) y de la Red Temática de «Medicina basada en la evidencia: la Colaboración Cochrane Iberoamericana» (exp. 2003/XT/00087), Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya.

Correspondencia: Dr. J.P. Gisbert.
Playa de Mojácar, 29. Urb. Bonanza.
28669 Boadilla del Monte. Madrid. España.
Correo electrónico: gisbert@meditex.es

Recibido el 14-10-2003; aceptado para su publicación el 15-10-2003.

Las revisiones nos ofrecen la posibilidad de estar informados sin necesidad de invertir la enormidad de tiempo antes mencionada³. Precisamente, una de las maneras más eficientes de encontrar la evidencia científica a la que hace referencia la medicina basada en la evidencia es buscar una revisión sistemática de los ensayos clínicos aleatorizados sobre los efectos de la atención sanitaria. Por tanto, necesitamos imperiosamente revisiones sistemáticas para integrar de forma eficiente toda la información válida y proporcionar una base para tomar decisiones de manera racional. Las revisiones sistemáticas establecen cuándo los efectos de la atención sanitaria son consistentes y cuándo pueden aplicarse los resultados de la investigación a los diferentes grupos de población. El uso de métodos explícitos y sistemáticos en las revisiones 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 decisiones. El metaanálisis, es decir, el uso de los métodos estadísticos para resumir los resultados de estudios independientes entre sí, puede proporcionar estimaciones más precisas sobre los efectos de la atención sanitaria que las derivadas de los estudios individuales incluidos en una revisión.

El presente trabajo versará sobre cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis, en concreto de ensayos clínicos aleatorizados (*randomised controlled trials*), ya que éstos tienen mayores posibilidades de proporcionar información más fiable que otras fuentes de evidencia sobre los efectos diferenciales de las distintas opciones de atención en salud. A lo largo de ese artículo se citarán numerosos ejemplos prácticos, con la intención de hacer más ameno y didáctico el enfoque de cada uno de los aspectos de las revisiones sistemáticas.

DEFINICIÓN DE REVISIÓN SISTEMÁTICA

Las revisiones sistemáticas son investigaciones científicas en sí mismas, con métodos prefigurados y un ensamblaje de los estudios originales, que sintetizan los resultados de éstos. Las características básicas que mejor definen a una revisión sistemática se podrían resumir en las siguientes: *a)* son síntesis y análisis de la información con un enfoque práctico; *b)* se basan en la mejor evidencia científica disponible; *c)* formulan preguntas claramente definidas, y *d)* utilizan métodos sistemáticos y explícitos para identificar y seleccionar estudios, evaluar éstos críticamente, extraer los datos de interés y analizarlos.

Consecuentemente, las revisiones sistemáticas pretenden ser: *a)* rigurosas, en cuanto a los estudios incluidos (con criterios de calidad, etc.); *b)* informativas, esto es, enfocadas hacia problemas reales, tratando de contestar una pregunta claramente delimitada o específica, e idealmente analizando y presentando los datos de la forma que mejor ayude a la toma de decisiones; *c)* exhaustivas: su objetivo es identificar y utilizar la mayor cantidad posible de información pertinente, sin introducir sesgos (de publicación, de selección, etc.), y *d)* explícitas, ya que todos los métodos utilizados en la revisión deben describirse con suficiente detalle.

REVISIONES SISTEMÁTICAS FRENTE A NO SISTEMÁTICAS

Llegados a este punto, cabría preguntarse: ¿es que no se han empleado desde siempre este tipo de revisiones «sistemáticas»? ¿es esto de las revisiones «sistemáticas» algo realmente novedoso? Tradicionalmente se ha realizado un tipo de revisiones que podríamos denominar «narrativas». El inconveniente de este tipo de revisiones es que utilizan métodos «informales», no explícitos, y a menudo personales y subjetivos³. En consecuencia, los lectores de una revisión narrativa encararán 2 problemas: primero, los autores no especifican el proceso que han seguido para buscar, reunir y evaluar la información que les ha llevado a ofrecernos sus conclusiones, y en segundo lugar, y ante la ausencia de dicha información, no serán capaces de repetir y verificar los resultados y las conclusiones de la revisión³.

En resumen, los principales defectos de las revisiones tradicionales o narrativas son la dependencia de la subjetividad del revisor y la falta de explicitación de los métodos de búsqueda y selección de las fuentes⁴. Las limitaciones antes mencionadas están magníficamente expresadas en la frase de D. Sackett, al referirse a las revisiones narrativas: «Creo que la razón de esta deficiencia en las revisiones se debe a la tradición de llamar a un experto en el tema para hacer la revisión... Esos autores comienzan su apartado de objetivos con una conclusión... Desafortunadamente los resultados están ya anunciados». Además, lamentablemente, la calidad de la mayoría de las revisiones médicas tradicionales deja mucho que desear, debido a que las labores de recogida, análisis y publicación de los resultados no se realizan con el mismo cuidado y detalle que cuando se trata de un trabajo de investigación original. Esta falta de método lleva a la tendencia a extraer conclusiones que, en el mejor de los casos, están sesgadas y, en el peor de los casos, son erróneas².

En contraposición, como se ha señalado anteriormente, las revisiones sistemáticas emplean métodos explícitos y sistemáticos que limitan el sesgo, y de este modo proporcionan resultados más fiables^{5,6}. Como se deduce de su nombre, las revisiones sistemáticas, no satisfechas con averiguar parte de la verdad, buscan «toda la verdad», es decir, recogen y ensamblan todas las evidencias de calidad disponibles sobre una cuestión clínica en concreto. Además, esta forma de revisión da al lector una importante ventaja sobre otras revisiones: la posibilidad de reproducirla y verificar si se llega a la misma conclusión³.

Otra diferencia importante es que las revisiones sistemáticas se generan para responder en profundidad a cuestiones clínicas generalmente muy concretas y específicas. Estas cuestiones pueden y deben formularse con precisión (véase más adelante el apartado «Componentes clave de una pregunta»). Si la pregunta que trata la revisión no está clara en su enunciado, en el resumen o en la introducción, o si no se incluye un apartado de «Métodos», lo más probable es que el artículo sea una revisión narrativa, no sistemática.

En contraposición, la mayor parte de los artículos de revisión de tipo narrativo abarcan un amplio abanico de asun-

TABLA I. Revisiones sistemáticas frente a revisiones narrativas

Característica	Revisión narrativa	Revisión sistemática
Pregunta de investigación	Con frecuencia amplia y poco definida	Clara, concreta, centrada en una cuestión clínica bien definida
Fuentes de información y revisión de la bibliografía	No especificada. No dirigida a localizar todos los estudios. Alta probabilidad de sesgo	Estrategia de búsqueda sistemática y explícita de todas las fuentes y artículos potencialmente relevantes
Selección de estudios	Generalmente sin criterios de selección. Alta probabilidad de sesgo	Descripción explícita de criterios de selección. Selección aplicada de manera uniforme
Calidad de los estudios	Generalmente no evaluada	Evaluación crítica de la calidad metodológica de los estudios revisados
Síntesis de datos	Resumen subjetivo, generalmente cualitativo, sin un estimador estadístico	Basada en la calidad metodológica de los estudios, a menudo resumida (cuantificada) por un estimador estadístico
Interpretación	En ocasiones basada en la evidencia, pero con frecuencia basada en los estudios que se han seleccionado subjetivamente y en opiniones personales	Generalmente basada en la evidencia científica, con identificación de las lagunas de conocimiento que persisten

tos relacionados con un tema, más que uno en profundidad. Por ejemplo, una revisión narrativa sobre la pancreatitis aguda, como la que puede encontrarse en el capítulo de un libro de texto, incluiría apartados sobre anatomía del páncreas, fisiopatología de la función pancreática, clínica, diagnóstico, complicaciones y tratamiento. Por lo tanto, las revisiones narrativas pueden ser más útiles para adquirir una perspectiva amplia sobre un tema, pero no lo son tanto para obtener respuestas concretas y mensurables ante cuestiones clínicas específicas. En este sentido, un ejemplo de una revisión sistemática de la Colaboración Cochrane relacionada con este tema plantea concretamente la pregunta de si la nutrición parenteral total, en comparación con la nutrición enteral, disminuye la mortalidad, la morbilidad o la estancia hospitalaria en los pacientes con pancreatitis aguda.

Por último, las recomendaciones halladas en las revisiones narrativas publicadas en las revistas y los libros de texto suelen diferir de las que se encuentran en las revisiones sistemáticas⁷. Así, las revisiones narrativas, por ejemplo, pueden retrasar más de una década la recomendación de un nuevo tratamiento de probada eficacia o continuar defendiendo un tratamiento mucho tiempo después de que se haya demostrado su escasa eficacia o incluso su potencial iatrogénico. Asimismo, las revisiones sistemáticas que incorporan técnicas cuantitativas pueden detectar mejor que las revisiones narrativas pequeños, aunque significativos, efectos del tratamiento.

En resumen, se podría concluir que, de entre las múltiples diferencias que existen entre las revisiones narrativas y las sistemáticas –resumidas en la tabla I–, destaca que las

primeras están basadas frecuentemente en opiniones (además de, por supuesto, en algunos datos objetivos), mientras que las revisiones sistemáticas están basadas fundamentalmente en pruebas⁸.

DEFINICIÓN DE METAANÁLISIS

El término «metaanálisis» fue introducido por Glass en 1976 (es decir, hace relativamente poco tiempo), que lo definía en los siguientes términos: «un análisis estadístico de una amplia serie de análisis de resultados de estudios individuales con el objetivo de integrar sus hallazgos»⁴. De entre las múltiples definiciones que se han propuesto posteriormente, una de las más difundidas quizá sea la de Jenicek: «Una integración estructurada, con una revisión cualitativa y cuantitativa, de los resultados de diversos estudios independientes acerca de un mismo tema»⁴.

Aunque a veces se utilicen indistintamente los 2 términos, revisión sistemática no es igual a metaanálisis³. La primera es el proceso que nos lleva a obtener los estudios cuyos resultados pueden combinarse –o no– matemáticamente para poder ofrecer conclusiones. Cuando los resultados de los estudios primarios se resumen pero no se combinan con métodos estadísticos, el resultado puede denominarse revisión sistemática «cualitativa». Así, el método matemático llamado metaanálisis (que daría como resultado una revisión sistemática «cuantitativa») es sólo una parte, aunque muy importante, de la revisión sistemática. Dicho de otro modo, un metaanálisis es la combinación estadística de al menos 2 estudios para obtener una estimación o suma única del efecto de la intervención en salud que estamos evaluando.

EVOLUCIÓN DE LAS REVISIONES SISTEMÁTICAS Y LOS METAANÁLISIS

Las revisiones sistemáticas en general, y los metaanálisis en particular, han experimentado en la última década un crecimiento exponencial que se ha calificado de «epidemia de metaanálisis». Baste el dato de que durante el período 1981-1986 se encontraron sólo 21 referencias con ese encabezamiento en MEDLINE, mientras que una década después (período 1991-1996) esta cifra había aumentado hasta 1.459⁴. En el año 2003, una búsqueda en esta base de datos que incluyera como descriptor o palabra clave el término «*meta-analysis*» recuperó nada menos que 14.000 referencias bibliográficas. La aplicación del metaanálisis se ha generalizado, habiéndose empleado en la mayoría de las especialidades, por no decir en todas. En aparato digestivo, por ejemplo, no se había publicado ningún metaanálisis hasta el año 1986; entre los años 1987 y 1991 se publicaron tan sólo 15, y esta cifra ascendió hasta 65 en el período 1992-1998⁹. Además, dentro de las diversas especialidades, las técnicas metaanalíticas se han utilizado en casi todas las áreas temáticas. A modo de ejemplo, hasta el año 2003 se llevan publicados aproximadamente 130 metaanálisis relacionados con la infección por *Helicobacter pylori*, una bacteria que se descubrió hace relativamente poco tiempo.

TABLA II. Esquema de una revisión Cochrane

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

TABLA III. Pasos para la preparación de una revisión sistemática

1. Formulación del problema
2. Localización y selección de los estudios
3. Evaluación de la calidad de los estudios
4. Extracción de datos. Metaanálisis
5. Análisis y presentación de los resultados
6. Interpretación de los resultados

Las causas más probables de esta eclosión son fundamentalmente 2: en primer lugar, el aumento de información acaecido a partir de 1940, reflejado en el incremento del número de revistas biomédicas, que pasó de 2.300 a más de 25.000, con más de 9 millones de artículos publicados por año¹⁰; y en segundo lugar, el auge del movimiento denominado «medicina basada en la evidencia», que, basándose precisamente en la revisión sistemática de los trabajos realizados en diversas áreas, busca dotar a la práctica de la medicina de los argumentos más objetivos posibles¹⁰.

Esta «fiebre de metaanálisis» ha suscitado vivas polémicas entre diversos autores, con posturas encontradas entre los que afirman que «cualquiera que se inicie en una nueva área de investigación, solicite una beca o realice su tesis doctoral debería antes realizar un metaanálisis del tema, con el objeto de establecer lo que ya se conoce...» y los más heterodoxos, que manifiestan todo lo contrario: «prefiero fiarme de un buen estudio individual que de un metaanálisis»⁴.

REVISIONES SISTEMÁTICAS, METAANÁLISIS Y GRADO DE EVIDENCIA CIENTÍFICA

Para evaluar el grado de evidencia científica existen diversas escalas que puntúan la calidad de la información en función del tipo de estudio que la soporta. Una de las más utilizadas es la propuesta por los National Institutes of Health de EE.UU., que se emplea para graduar las recomendaciones en las guías de práctica clínica basadas en la evidencia y en la que las revisiones sistemáticas y los metaanálisis ocupan una posición privilegiada⁴.

REVISIONES SISTEMÁTICAS Y METAANÁLISIS DE ENSAYOS CLÍNICOS

Archie Cochrane señalaba no hace mucho la necesidad de un mayor número de revisiones de conjunto sistemáticas cuando escribió: «Seguramente es una crítica grave a nuestra profesión el que no hayamos organizado un resumen crítico, por especialidades o subespecialidades, adaptado periódicamente, de todos los ensayos clínicos de asignación aleatoria relevantes»⁸. Precisamente, la respuesta a ese desafío la constituyen las revisiones llevadas a cabo por la Colaboración que lleva su nombre, una organización internacional sin ánimo de lucro que intenta ayudar a tomar decisiones clínicas y sanitarias bien fundamentadas preparando, manteniendo y divulgando revisiones sistemáticas sobre los efectos de la atención sanitaria. Las revisiones Cochrane tienen un formato estándar, perfectamente definido, que debe constar de los siguientes apartados (tabla II): un resumen en un formato estructurado; el texto de la revisión, que consta de una introducción (antecedentes y objetivos), materiales (criterios de selección y estrategia de búsqueda), métodos, resultados (descripción de los estudios, calidad metodológica y resultados propiamente dichos), discusión y conclusiones de los revisores. Además, incorpora tablas y figuras estándar para mostrar las características de los estudios incluidos, la especificación de las intervenciones que se han comparado, los resultados propiamente dichos y una lista de los estudios que se han excluido. El programa Review Manager (RevMan) está diseñado para ayudar a los revisores a construir las revisiones según el formato estructurado descrito y se puede adquirir gratuitamente en la web de la Colaboración Cochrane (www.Cochrane-net.org/revman).

PASOS NECESARIOS PARA REALIZAR, EVALUAR Y UTILIZAR UNA REVISIÓN SISTEMÁTICA

En la tabla III se resumen los pasos necesarios para preparar correctamente una revisión sistemática. La mayoría de lo que se expone en este apartado está basado en el «Manual de revisores» de la Colaboración Cochrane (versión española de «*Cochrane reviewers' handbook*» del Centro Cochrane Iberoamericano)^{11,12}, texto de obligada consulta para todo aquel que quiera embarcarse, sin miedo al naufragio, en la preparación de una revisión sistemática. Las referencias bibliográficas utilizadas en este

artículo de revisión podrían ser innumerables; puesto que gran parte de éstas se encuentran recogidas en el «Manual de revisores», en este manuscrito citaremos únicamente algunas de las más relevantes, remitiendo al lector interesado por un detalle bibliográfico más exhaustivo a la consulta del citado manual.

Primer paso: formulación del problema

Justificación de las preguntas bien formuladas

Las preguntas mal enfocadas conducen a decisiones poco claras acerca de qué estudios conviene incluir en la revisión y cómo resumir sus datos. Así, una pregunta confusa conduce, con mucha probabilidad, a una respuesta también confusa, hecho que está magistralmente expresado en la afirmación de Counsell¹³: «*Ask a poor question and you'll get a poor review*». Desde la perspectiva del lector, si el principal problema que aborda la revisión no está claro a partir del título o del resumen, probablemente no merece la pena perder el tiempo leyendo el artículo de revisión en su totalidad⁸. El desarrollo de un objetivo principal claramente definido *a priori* ayuda a luchar contra las denominadas «expediciones de pesca»³; éstas se organizan cuando la pregunta principal es confusa y se opta por unir hallazgos atrayentes entre todos los estudios, hasta que finalmente alguna combinación de datos ofrece un resultado positivo. Al igual que en cualquier proyecto de investigación, la primera y más importante decisión a la hora de preparar una revisión es determinar su enfoque, lo que se consigue planteando preguntas claramente delimitadas. Dichas preguntas y los objetivos de una revisión sirven también a los lectores para valorar inicialmente su relevancia. Así, los lectores pueden utilizar las preguntas y los objetivos planteados para juzgar si la revisión les puede resultar de interés y si es pertinente para las cuestiones que ellos se plantean.

Componentes clave de una pregunta

Existen varios componentes clave que son necesarios para formular correctamente una pregunta y que deben especificarse en la sección «Criterios de selección de los estudios» de la revisión. Una pregunta claramente definida debe especificar el tipo de personas (participantes), el tipo de intervenciones y el tipo de resultados que son de interés para la revisión. A modo de ejemplo, una pregunta correctamente planteada sería: «En la población asintomática de más de 50 años (los participantes), ¿el cribado con colonoscopia (la intervención) se asocia con una reducción de la mortalidad debida al cáncer de colon (el resultado)?».

Con respecto al tipo de participantes, resulta de utilidad considerar en 2 fases el tipo de personas que son objeto de interés para la revisión. En primer lugar, hay que definir las enfermedades o condiciones que pretendemos evaluar, debiéndose establecer criterios que sean suficientemente explícitos para estas definiciones. En segundo lugar, hay que identificar a la población y el ámbito de in-

terés. Esto implica tener que decidir si se está interesado en un grupo específico de la población, en función de factores como la edad, el sexo, la etnia, etc. Cualquier restricción basada en características específicas de la población o del ámbito de estudio debe estar justificada por evidencia ya reconocida. Por ejemplo, enfocar una revisión sobre la efectividad del cribado (*screening*) del cáncer de colon mediante colonoscopia dirigida únicamente a individuos mayores de 50 años de edad parece estar justificado por diversas razones, entre las que destaca que la incidencia de este tipo de cáncer en individuos menores de 50 años es muy baja.

El segundo componente clave de una pregunta bien formulada es especificar las intervenciones que son de interés. También es importante definir las intervenciones con las que éstas se compararán, es decir, el tipo de grupo control que es aceptable para la revisión. Así, debe considerarse con detenimiento si las personas en el grupo control pueden estar recibiendo otras intervenciones (y cuáles) distintas del placebo.

El tercer componente clave de una pregunta bien formulada es la especificación —con criterios explícitos— de aquellos resultados concretos que sean de interés. Los revisores deben evitar abrumar a los lectores con datos que tienen poca o ninguna trascendencia pero, al mismo tiempo, han de tener cuidado en no omitir datos importantes.

Un componente adicional al plantear una pregunta es el tipo de diseño de los estudios, pues algunos resultan superiores a otros para tratar de responder a determinadas cuestiones. Muchos autores consideran que los ensayos clínicos controlados y aleatorizados son una condición *sine qua non* para tratar asuntos referentes a la eficacia terapéutica. Ya que las revisiones Cochrane plantean cuestiones acerca de los efectos de la atención sanitaria, éstas se centran principalmente en ensayos clínicos aleatorizados. Existen 2 razones fundamentales por las que se debe ser cauteloso a la hora de incluir estudios no aleatorizados en una revisión sobre los efectos de la atención sanitaria, ambos relacionados con los sesgos. En primer lugar, la aleatorización es el único modo de controlar aquellos que no se conocen o no están cuantificados¹⁴. Datos empíricos indican que, por lo general, los estudios no aleatorizados tienden a sobreestimar los efectos de las intervenciones sanitarias. A pesar de las consideraciones anteriores, a veces resulta apropiado realizar una revisión sistemática de los estudios no aleatorizados sobre los efectos de un determinado tratamiento. Por ejemplo, en ocasiones el curso de una enfermedad es tan uniforme o los efectos de una intervención son tan evidentes que resulta innecesario, e incluso dudosamente ético, realizar un ensayo clínico aleatorizado. Por último, otros aspectos relevantes acerca del diseño de los estudios que merecen considerarse desde el inicio son si se revisarán únicamente los estudios que tienen un seguimiento de una cierta duración. No obstante, es preciso tener en cuenta que cuanto más restrictivos sean los criterios exigidos acerca del diseño de los estudios para responder a la pregunta planteada, menos probabilidades habrá de encontrar datos que respondan de forma específica a preguntas tan restringidas.

Identificación y selección de los estudios mediante los componentes clave de la pregunta

Una vez que se ha planteado correctamente una pregunta, deberían determinarse qué componentes clave se tendrán en cuenta para definir la estrategia de búsqueda inicial¹⁵. Para las revisiones Cochrane, la búsqueda de estudios se ve enormemente facilitada por la existencia de los registros especializados de ensayos clínicos elaborados por los Grupos Colaboradores de Revisión. Las búsquedas que requieran de la presencia simultánea de varios de los componentes de la pregunta –o bien de formulaciones muy específicas de ciertos componentes– suelen ser demasiado específicas y tienden a pasar por alto información relevante. Por ejemplo, si uno busca estudios que aborden los efectos a largo plazo de la erradicación de *H. pylori* sobre la prevención de la recidiva ulcerosa duodenal y lo hace buscando los que están indexados mediante los términos «*Helicobacter pylori*», «tratamiento erradicador», «úlcera duodenal», «prevención», «recidiva ulcerosa» y «largo plazo», es probable que no se identifiquen estudios que pueden ser relevantes. Por otro lado, si sólo se emplean los términos «*Helicobacter pylori*» y «recidiva ulcerosa» en la búsqueda, se identificarán cientos de artículos irrelevantes. En general, los componentes clave que resultan de utilidad cuando se realiza una búsqueda incluyen la condición o enfermedad de interés y la intervención o exposición que se está evaluando. Aunque se pueda estar específicamente interesado en un ámbito determinado, los estudios no suelen estar indexados según esta variable dentro de las bases de datos electrónicas. Además, los estudios pueden evaluar múltiples resultados, algunos de los cuales pueden ser relevantes para la revisión, pero es posible que no formen parte de la indexación del artículo.

Amplitud de las preguntas

El alcance de las preguntas que aborda una revisión puede ser amplio o restringido. Por ejemplo, un revisor podría plantear una pregunta amplia al considerar si los aminosalicilatos son eficaces en la prevención de la recidiva de la enfermedad inflamatoria intestinal crónica (en general, sin particularizar a ninguna subpoblación de pacientes). Por el contrario, otro revisor podría preguntarse si estos fármacos son efectivos para reducir el riesgo de recidiva únicamente en los pacientes con enfermedad de Crohn que han sido sometidos a cirugía de resección intestinal. Son varias las ventajas y desventajas que se derivan del hecho de formular preguntas amplias o restringidas. Las revisiones enfocadas de manera restringida quizá no sean generalizables a entornos diversos, a distintas poblaciones o a formas alternativas de una misma intervención. Además, un enfoque restringido comporta un elevado riesgo de conclusiones sesgadas cuando el revisor conoce la bibliografía sobre un determinado tema y restringe los criterios de inclusión de tal manera que queden excluidos algunos o varios estudios con resultados que estén en conflicto con la postura que defiende. Por su par-

te, la validez de las revisiones definidas en términos muy amplios puede no ser apropiada cuando existe evidencia sólida que induce a pensar que las diversas definiciones de la condición de interés se asocian con efectos de la intervención que son marcadamente diferentes.

Modificación de la pregunta

La pregunta debe plantearse en el protocolo *antes* de iniciar la revisión completa, pero no debe convertirse en una «camisa de fuerza» que impida la exploración de aspectos inesperados. Aunque es de esperar que se obtenga una cierta fluidez y refinamiento al formular las preguntas de revisión conforme se va adquiriendo una mayor comprensión del problema, es importante protegerse del riesgo de sesgo al modificarlas. Las preguntas *a posteriori* son más proclives al sesgo que aquellas que han sido formuladas previamente, y las preguntas modificadas por los propios datos pueden generar conclusiones falsas basadas en resultados espurios. Por tanto, en todo caso, debe documentarse y justificarse cualquier cambio que sufra el protocolo derivado de la modificación de la pregunta inicial de revisión.

Segundo paso: localización y selección de los estudios

Como se ha mencionado previamente, que la búsqueda de los estudios sea exhaustiva y no sesgada es una de las diferencias clave entre las revisiones sistemáticas y las tradicionales. Si bien las bases de datos electrónicas, como MEDLINE, son potentes herramientas para la localización de estudios, sólo pueden identificar entre un 30 y un 80% de todos los ensayos clínicos controlados aleatorizados que se han publicado. Por una parte, esta base de datos no incluye la totalidad de la bibliografía médica publicada y, por otra, algunos estudios nunca llegan a publicarse. Las referencias de estudios escritos en lengua no inglesa tienen una baja representación en MEDLINE, por lo que si se utiliza esta base de datos para identificar estudios la revisión puede estar afectada potencialmente por un sesgo de publicación, que se origina cuando lo que se publica no es representativo de lo que se investiga¹⁶⁻¹⁸. Desde otra perspectiva, el sesgo de publicación puede deberse a la tendencia existente entre autores, revisores y editores de revistas a permitir que la dirección y la significación estadística de los resultados de la investigación influyan en las decisiones de enviar y aceptar para publicación los resultados de la investigación, al seleccionarse prioritariamente estudios con resultados positivos. Aunque en un principio se pensó que las revistas médicas podían ser las responsables fundamentales de este tipo de sesgo, ya que tenderían a publicar únicamente estudios con resultados significativos, en la actualidad, y a partir de las entrevistas efectuadas directamente con investigadores, se sabe que los principales responsables son los propios autores al considerar que, si no hay resultados significativos, los estudios no son interesantes para su pu-

blicación⁴. En el sesgo de publicación intervienen diversos factores: además de que los trabajos con resultados estadísticamente significativos tienen mayor probabilidad de ser publicados que los que no arrojan diferencias, una vez aceptados se publican más rápidamente, lo hacen en revistas de más impacto y son citados con más frecuencia¹⁰.

Para evitar estos sesgos y asegurar que, en la medida de lo posible, todos los datos relevantes sean incluidos en una revisión, es necesario utilizar varias fuentes de información para identificar los estudios. Debe también tenerse en cuenta la posibilidad de que haya un sesgo de referencias (una tendencia a citar preferentemente los estudios que apoyan el propio punto de vista) y evitarlo mediante el empleo de múltiples estrategias de búsqueda. Posibles fuentes de información para la localización de estudios incluyen⁴: *a*) bases de datos bibliográficas computarizadas de estudios publicados en revistas indexadas (MEDLINE, EMBASE, Excerpta Medica, Índice Médico Español, Cochrane Library); *b*) artículos de revisión y metaanálisis; *c*) artículos originales (publicaciones preliminares o completas) en publicaciones primarias; *d*) publicaciones secundarias: resúmenes de artículos publicados en otras revistas como originales (*Evidence Based Medicine*, *ACP Journal Club*, etc.); *e*) resúmenes de congresos; *f*) tesis doctorales (registro de publicaciones de facultades); *g*) libros de texto sobre el tema; *h*) contactos personales con expertos, sociedades científicas, grupos de trabajo, etc.; *i*) agencias financiadoras (FIS, CICYT, fundaciones, etc.); *j*) registros de ensayos clínicos (Ministerio de Sanidad, Colaboración Cochrane, registros monográficos, entre otros); *k*) industria farmacéutica, y *l*) búsquedas manuales a partir de las citas bibliográficas de los artículos originales. A continuación se revisan con más detalle las fuentes de información más relevantes para la localización de los estudios.

Registros especializados de la Colaboración Cochrane

La principal fuente de estudios para las revisiones Cochrane es el registro especializado de los Grupos Colaboradores de Revisión. Al desarrollar una estrategia de búsqueda es importante consultar al coordinador de Búsqueda de Ensayos o al coordinador del grupo de revisión del Grupo de Colaboradores de Revisión para evitar la duplicación innecesaria de esfuerzos y garantizar que se está haciendo un uso eficiente de otras bases de datos adicionales.

Control de las listas de referencia

Los revisores deben controlar las listas de referencias de todos los artículos relevantes que hayan identificado. En general, éste es un medio eficiente, aunque lento, de identificar estudios. Algunas de las fuentes de referencia de ensayos más obvias y convenientes son las revisiones ya existentes. La Cochrane Library incluye la Base de Datos de Resúmenes de Revisiones de Efectividad (Database of Abstracts of Reviews of Effectiveness-DARE), producida

por el NHS Centre for Reviews and Dissemination (CRD) en York (Reino Unido). MEDLINE, EMBASE y otras bases de datos bibliográficas también pueden usarse para identificar artículos de revisión (en MEDLINE, por ejemplo, los artículos de revisión aparecen indexados como «Review» en el campo «Publication type»).

Comunicación personal

Los colegas pueden ser una importante fuente de información de ensayos recientes que todavía no se han publicado o de ensayos inéditos más antiguos. A veces estos canales informales de comunicación pueden ser los únicos medios para identificar datos no publicados. La «literatura gris u opaca» (*grey literature*), compuesta por informes técnicos, actas de congresos y tesis doctorales, entre otros, también puede resultar útil, ya que muchos estudios referenciados de esta manera no son publicados en su totalidad^{2,19}.

Bases de datos electrónicas

Desde octubre de 1996, la Colaboración Cochrane ha venido desarrollando su propio registro de estudios, denominado «CENTRAL». Este registro incluye principalmente referencias de los ensayos clínicos controlados en atención sanitaria, aportadas por grupos de revisión y otras entidades que han identificado los ensayos a través de diversos mecanismos de búsqueda. Por su parte, Index Medicus y Excerpta Medica son repertorios de revistas de ciencias de la salud que están accesibles en formato electrónico bajo la denominación de MEDLINE y EMBASE, respectivamente. Estas bases de datos referenciales pueden ser interrogadas mediante descriptores temáticos (*subject terms*) asignados por indexadores contratados por las empresas productoras. Esta sencilla estrategia de búsqueda funciona bien si el objetivo es recuperar algún buen artículo sobre un tema o simplemente identificar un artículo específico. Sin embargo, cuando se realiza una revisión sistemática debe considerarse con cierto escepticismo la precisión con la que se aplican los descriptores a las referencias, dado que los autores quizá no describan bien sus métodos u objetivos; además, los indexadores no siempre son expertos en el área temática del artículo que están indexando. Por otra parte, ninguna de las bases de datos mencionadas cubre todas las revistas. Así, se ha calculado que el solapamiento entre MEDLINE y EMBASE es de tan sólo un tercio aproximadamente, por lo que una búsqueda más exhaustiva requiere que se interroguen, como mínimo, ambas bases de datos.

Publicaciones secundarias

Las publicaciones secundarias seleccionan estudios originales con la exigencia de que sean metodológicamente correctos y además clínicamente relevantes. Presentan un resumen estructurado de dichos estudios, junto con un breve comentario de un experto que aporta una perspectiva clínica a los resultados. Entre las publicaciones secun-

darias se encuentran el *ACP Journal Club* (<http://www.acpjc.org/>) y el *EBM online* (<http://ebm.bmjournals.com/>). Estas 2 publicaciones se encuentran reunidas en formato CD-ROM bajo el título *Best Evidence*. Algunas especialidades disponen de publicaciones secundarias específicas, como es el caso de *Evidence-Based Gastroenterology* (www.evidence-based-gastro.com/) en la especialidad de aparato digestivo.

Desarrollo de una estrategia de búsqueda

Debe describirse el proceso mediante el que se seleccionan los estudios para su inclusión en una revisión, es decir, debe explicitarse la estrategia de búsqueda. La primera etapa de revisión de los resultados de una búsqueda electrónica implica la evaluación detallada de los títulos y resúmenes para determinar si cada artículo reúne los requisitos predeterminados para ser escogido. Si al examinar la información disponible se puede determinar que un artículo no cumple los criterios de inclusión, éste puede rechazarse. Si, a criterio del revisor, el título o el resumen del artículo ofrecen dudas que impiden rechazarlo definitivamente, debería obtenerse el texto completo del artículo. La revisión del texto completo puede llevar, finalmente, a la exclusión del estudio porque no reúne los requisitos de inclusión. En cada etapa del proceso de selección es preferible errar por exceso que por defecto, ya que una vez que un estudio ha quedado excluido del proceso de selección es improbable que se vuelva a considerar su inclusión. Debido a la importancia que tienen estas decisiones, es aconsejable que más de un revisor examine cada registro obtenido antes de que sea rechazado.

Aumentar la exhaustividad en una búsqueda implica reducir su precisión e identificar más artículos no relevantes. Aunque es importante que la búsqueda sea lo más completa posible, siempre es necesario mantener un equilibrio entre el grado de exhaustividad y su precisión. A menudo, los bibliotecarios o documentalistas que se han especializado en búsquedas electrónicas pueden ser de gran ayuda en el desarrollo y ejecución de éstas. El desarrollo de una estrategia de búsqueda es un proceso iterativo en el cual los términos que se utilizan se modifican según los resultados que se vayan obteniendo. Por lo tanto, es importante trabajar estrechamente con un documentalista en el momento de decidir en qué bases de datos buscar y qué estrategias de búsqueda utilizar en cada una de ellas. Idealmente, los revisores deben estar junto al documentalista cuando se ejecute la estrategia de búsqueda.

Un buen abordaje para desarrollar una estrategia de búsqueda exhaustiva es empezar utilizando términos múltiples que describan la condición o enfermedad de interés y combinarlos con el operador booleano «OR». Esto significa que se identificarán los artículos indexados que incluyan por lo menos uno de los términos utilizados. Los resultados pueden reducirse utilizando posteriormente el operador «AND». En este paso, se identifican los artículos que utilizan los términos que describen las intervenciones que se están evaluando (agrupados con el operador «OR») si *también* abordan la condición de estudio (es de-

cir, si cumplen la intervención «Y» la condición). Por ejemplo, una estrategia de búsqueda para evaluar la posible relación entre la infección por el virus C de la hepatitis y el linfoma no hodgkiniano sería: («hepatitis C» OR «hepatitis C virus» OR HCV) AND («lymphoma» OR «non-Hodgkin lymphoma» OR «lymphoproliferative disease» OR «lymphoproliferative disorder»).

Selección de estudios

Todas las referencias de estudios identificados como potencialmente elegibles deben evaluarse con posterioridad para confirmar si cumplen los criterios de inclusión para la revisión. Los revisores deben decidir: *a)* cuándo se necesitará que más de un revisor evalúe la relevancia de cada artículo (actitud, en general, recomendable); *b)* si las decisiones en cuanto a la selección las tomarán especialistas en el tema, no especialistas o ambos (los especialistas en un área específica con frecuencia tienen opiniones formadas previamente que pueden sesgar sus evaluaciones, tanto sobre la relevancia como sobre la validez de los artículos; por lo tanto, aunque es importante que al menos un revisor esté versado en el tema objeto de revisión, quizá sea una ventaja disponer de un segundo revisor que no sea un experto en esa área), y *c)* cómo se resolverán las discrepancias si existe más de un revisor (generalmente mediante discusión, pues la causa del desacuerdo es a menudo un simple descuido por parte de uno de los revisores; cuando la discrepancia se deba a una diferencia de interpretación más profunda, el tema debe resolverse por consenso).

Documentación de una estrategia de búsqueda

La estrategia de búsqueda debe describirse con suficiente detalle en la revisión como para que el proceso pueda reproducirse. La documentación debe incluir las principales fuentes de información utilizadas para identificar los estudios, la estrategia de búsqueda para cada base de datos consultada y la fecha en la que se efectuó dicha búsqueda, las limitaciones temporales (p. ej., de 1966 a 2003) y las limitaciones de idioma o de otro tipo (aunque éstas deben evitarse en la medida de lo posible).

Tercer paso: evaluación de la calidad de los estudios

Como expresa muy gráficamente el refrán inglés «*garbage in, garbage out*», si se introduce basura en una revisión sistemática se obtendrá también basura^{3,20}. De este modo, si la calidad de los estudios incluidos es defectuosa, la calidad de la revisión sistemática que los ha incluido será asimismo deficiente. En este caso, dicha revisión no sólo no nos aportará beneficio alguno sino que incluso podrá confundirnos en nuestra práctica clínica. Por tanto, la evaluación de la calidad de los estudios individuales que se incluyen en las revisiones sistemáticas es necesaria para limitar los sesgos y guiar la interpretación de resultados. Incluso si la revisión incluye sólo ensayos clínicos

aleatorizados, es importante conocer si éstos eran de buena calidad. Los factores que justifican una evaluación detallada de la calidad son los relacionados con la aplicabilidad de los resultados, la validez de los estudios individuales y ciertas características del diseño que afectan a la interpretación de los resultados. La aplicabilidad, también llamada validez externa o generalización, está relacionada con la definición de los componentes clave de las preguntas correctamente formuladas, aspecto ya comentado en el apartado correspondiente. La interpretación de los resultados depende de la validez de los estudios incluidos y de otras características. Por ejemplo, una revisión quizá resuma 20 ensayos clínicos válidos que evalúan los efectos del tratamiento erradicador de *H. pylori* sobre los síntomas de la dispepsia funcional; sin embargo, los ensayos clínicos quizá examinen diferentes tipos de pacientes (con dispepsia de tipo ulceroso o, por el contrario, de tipo dismotilidad) y es posible que tengan períodos de seguimiento variables (desde unos pocos meses hasta más de un año). Estos aspectos afectarían a la interpretación aunque quizá no estén relacionados directamente con la validez interna de los ensayos. Dado que la mayoría de las revisiones Cochrane incluyen ensayos clínicos aleatorizados, nos centraremos a continuación en cómo evaluar, precisamente, la validez de dichos estudios.

Validez. Sesgos en los ensayos clínicos

Antes de fijarnos en las conclusiones de una revisión sistemática debemos valorar si nos las vamos a creer³. Cuanta mayor sea la validez de los estudios que sustentan dicha revisión, más nos fiaremos de las conclusiones. En el contexto de una revisión sistemática, 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. En una revisión sistemática, las diferencias observadas en la validez de los estudios incluidos podrían ayudarnos a explicar la variabilidad de sus resultados. De este modo, los estudios más rigurosos tienen mayor probabilidad de dar resultados más próximos a la «verdad», mientras que los estudios menos rigurosos suelen estar sesgados hacia la sobreestimación de la eficacia de la intervención.

En resumen, 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 generalizar los resultados¹⁶. Por ejemplo, si es cierta la menor incidencia de hemorragia digestiva por varices esofágicas en los pacientes cirróticos que reciben profilaxis con bloqueadores beta. Por el contrario, la validez externa de una investigación se plantea a la hora de generalizar los resultados obtenidos, es decir, si los datos encontrados en una población son aplicables a otra¹⁶. En los estudios sobre los efectos de la atención sanitaria pueden existir 4 tipos de sesgos o errores sistemáticos –de selección, de realización (*performance bias*), de desgaste o pérdida (*attrition bias*) y de detección– que se comentan a continuación:

– *Sesgo de selección*. Uno de los factores más importantes que puede sesgar y distorsionar las comparaciones entre tratamientos está relacionado con el proceso de obtención de los grupos a comparar. Así, en el diseño de ensayos clínicos es muy importante utilizar un método apropiado para prevenir que se conozca de antemano la asignación de los tratamientos²¹. Mientras se evalúa la elegibilidad potencial de un participante en un ensayo clínico, las personas que reclutan a los participantes deben desconocer el tratamiento que corresponde a la próxima asignación. Lo ideal es que el proceso de asignación no pueda estar influido por las personas que lo llevan a cabo, para lo que debería instaurarse un programa de asignación aleatoria implantado por alguien no implicado en el reclutamiento (p. ej., alguien ubicado en una oficina central del ensayo). Si no puede organizarse tal aleatorización central, entonces es necesario tomar otras precauciones para evitar la manipulación del proceso de asignación por las personas que participan en el reclutamiento. El proceso de ocultar la asignación hasta que se asigna el tratamiento se denomina «ocultación de la asignación» (en inglés, *allocation concealment*), que no debe confundirse con el término «cegar», que se reserva para las medidas tomadas para reducir el sesgo una vez que la intervención ya ha sido asignada.

– *Sesgo de realización*. Hace referencia a las diferencias sistemáticas en la atención sanitaria proporcionada a los participantes en los grupos de comparación, aparte, claro está, de la intervención objeto de estudio. Para proteger contra las diferencias en la atención y contra el efecto placebo, puede «cegar» a aquellos que dan y reciben el tratamiento para que no conozcan el grupo al que están asignados quienes lo reciben.

– *Sesgo de desgaste o pérdida*. Se refiere a las diferencias sistemáticas entre los grupos de comparación en cuanto a la pérdida de participantes durante el estudio. La manera de tratar las pérdidas puede sesgar los resultados, y la información incompleta sobre cómo se tratan las pérdidas puede ocultar este problema.

– *Sesgo de detección*. El sesgo de detección se refiere a las diferencias sistemáticas entre los grupos de comparación en la evaluación de resultados. Los ensayos clínicos que ciegan la asignación de las intervenciones a quienes evalúan los resultados son, lógicamente, menos propensos a estar sesgados que los ensayos clínicos que no lo hacen. El cegamiento es particularmente importante en aquellas investigaciones que utilizan medidas subjetivas de resultado, como el dolor.

Evaluación de la validez y calidad de los estudios

Si se han utilizado, como es aconsejable, varios criterios explícitos para valorar la validez, es recomendable resumirlos para obtener una evaluación global de los resultados de cada estudio^{22,23}. Se han identificado al menos 25 escalas y 9 listas que se han utilizado para evaluar la validez y la «calidad» de los ensayos clínicos controlados y aleatorizados, pero ninguna de ellas puede recomendarse sin reservas^{24,25}. Una de las más empleadas es la escala de calidad propuesta

por Jadad et al²⁶, en la que se puntúan por separado 3 componentes esenciales: la aleatorización, el doble ciego y la descripción de las pérdidas durante el seguimiento.

Hay distintas formas de utilizar las evaluaciones de la validez en una revisión: como un umbral para la inclusión de estudios, como una posible explicación de las diferencias en los resultados entre los estudios, en los análisis de sensibilidad y como pesos en el análisis estadístico de los resultados de los estudios. En algunas revisiones, si el riesgo de sesgo es elevado, se justifica la exclusión de los estudios que no cumplen uno o varios criterios de validez. Por ejemplo, para medidas de resultado muy subjetivas, como la dispepsia, podría decidirse incluir sólo los estudios que previenen el sesgo de realización mediante el cegamiento de los participantes.

Hay diversos métodos para explorar si la validez puede explicar las diferencias entre los resultados de los estudios incluidos. Se pueden utilizar gráficos visuales de los resultados ordenando los estudios según su grado de validez. Un segundo enfoque consiste en analizar subgrupos de estudios situados por encima del punto de corte metodológico, el cual preferiblemente debe especificarse *a priori* en el protocolo de la revisión. Este enfoque puede utilizarse mediante un análisis de sensibilidad, a fin de determinar si los resultados globales coinciden con los que se obtienen cuando sólo se incluyen en el análisis los estudios con riesgo de sesgo bajo. 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 peso debe asignarse a los distintos criterios de validez, por lo que este último 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. Aunque la búsqueda y el proceso de selección deberían haber descartado la mayoría de estudios no elegibles, es conveniente comprobar la elegibilidad de éstos en el momento del resumen o de la recogida de datos. Así, las revisiones Cochrane incluyen una tabla de estudios excluidos en el caso de aquellos que inicialmente parecían reunir los criterios de inclusión pero que, tras un examen más minucioso, fueron finalmente excluidos.

Algunas de las características de los métodos de los estudios que pueden ser relevantes incluyen la duración del estudio, el tipo de ensayo (diseños paralelos, cruzados), el cegamiento (de pacientes, proveedores y evaluadores de resultados), el número de pérdidas y de cambios de tratamiento, las cointervenciones y otros factores potenciales de confusión. Además, el apartado de métodos del formulario de recogida de datos debe incluir los criterios de validez utilizados (véase el apartado anterior).

Los criterios diagnósticos utilizados para definir el problema de salud de interés pueden ser una fuente importante de variabilidad clínica y deberían describirse con

detalle. Por ejemplo, en una revisión sobre terapia farmacológica con infliximab en la enfermedad de Crohn es importante conocer cómo se estableció en cada estudio el fenotipo (inflamatorio o fistulizante) y cómo se definió la refractariedad al tratamiento convencional (corticorresistencia o corticodependencia).

Cómo medir el resultado de los estudios puede ser una de las secciones más difíciles de diseñar del formulario de extracción de datos. Los informes de los estudios suelen incluir más de una medida de resultado (mortalidad, morbilidad, calidad de vida, etc.); además, pueden presentar la misma medida de resultado usando diferentes escalas de cuantificación, pueden incluir medidas de resultado para diferentes subgrupos de pacientes o pueden señalar medidas de resultado en diferentes momentos del tiempo. Por ejemplo, la valoración del efecto del tratamiento erradicador de *H. pylori* sobre la dispepsia funcional es un tema muy controvertido²⁷. En parte, ello se debe a la falta de acuerdo entre los diversos estudios realizados sobre cómo se deben cuantificar los síntomas dispépticos y por lo tanto su mejoría: los síntomas podrían considerarse una variable cuantitativa, y el beneficio, un descenso determinado en el valor medio de todos los pacientes tras el tratamiento; o también podría compararse el porcentaje de pacientes con síntomas dispépticos de una determinada gravedad antes y después del tratamiento; por último, otra opción más sencilla consiste en definir como éxito terapéutico sencillamente la ausencia de síntomas dispépticos al final del estudio. Por otra parte, el momento preciso en que se evalúan los síntomas tras haber administrado tratamiento erradicador de *H. pylori* puede ser también relevante, pues algún estudio ha señalado que existe una cierta latencia desde que desaparece la infección hasta que se comprueba la mejoría sintomática (en caso de que esto realmente ocurra).

Quinto paso: análisis y presentación de resultados. Metaanálisis

La razón fundamental para realizar las revisiones de forma sistemática es asegurar la validez de los resultados, pero el papel del análisis estadístico en las revisiones puede estar menos claro. A veces, el uso de la estadística (metaanálisis) puede parecer, y quizá lo sea, más un obstáculo que una ayuda para los que no están familiarizados con estas técnicas. Existe también el riesgo de prestar demasiada atención a la línea estadística inferior o al rombo-diamante situado al final del gráfico que resume los resultados de estudios individuales en un metaanálisis. En ocasiones, los revisores y los lectores, con las prisas de llegar al rombo-diamante, descuidan la importancia de la reflexión y de las consideraciones hechas en los pasos anteriores al análisis o en el análisis mismo.

Justificación para realizar un metaanálisis en una revisión

Para preparar una revisión se deben obtener datos de los estudios individuales y, posteriormente, se pueden usar

métodos estadísticos para analizar y resumir estos datos. Si se usan apropiadamente, dichos métodos proporcionan una herramienta potente para llegar a unas conclusiones que tengan sentido y ayudan a evitar errores de interpretación. Cuando en una revisión no se usan métodos estadísticos, es fácil cometer el error de comparar el número de estudios «positivos» con el número de estudios «negativos». Esta especie de «recuento de votos», en el que se confiere el mismo valor a los estudios de pequeño y gran tamaño es poco fiable y debe evitarse, porque un estudio puede considerarse positivo o negativo dependiendo de cómo interpreten los resultados los revisores, y además porque no tiene en cuenta el peso relativo de la evidencia fiable proporcionada por cada estudio⁸. Al «contar los votos» también se tiende a pasar por alto los efectos pequeños pero clínicamente relevantes, en particular al considerar negativos los estudios con resultados que no son estadísticamente significativos (debido, con probabilidad, a un error beta consecuencia de la escasa potencia estadística). Además, un investigador puede interpretar un estudio como positivo mientras que otro lo considera negativo; por ejemplo, un efecto importante que no sea estadísticamente significativo podría interpretarse como positivo a la luz de la importancia clínica y como negativo desde la perspectiva de la significación estadística⁸. Sin embargo, los metaanálisis valoran los estudios de acuerdo, entre otras cosas, con su tamaño, de modo que conceden un mayor peso a los más grandes; por este motivo, los resultados globales representan un promedio ponderado de los resultados de los estudios individuales⁸. Por supuesto, el uso de métodos estadísticos –metaanálisis– no garantiza que los resultados de una revisión sean válidos, como ocurre también en un estudio primario. Así, como cualquier herramienta, matemática o no, los métodos estadísticos pueden utilizarse de modo inapropiado.

Objetivos de los metaanálisis

Entre los propósitos generales comunes a cualquier metaanálisis se encuentran los siguientes⁴:

1. Resumir los resultados de la bibliografía recogida y hacer de ella una valoración cualitativa y cuantitativa.
2. Resolver la incertidumbre (p. ej., si un determinado fármaco tiene efecto beneficioso o no).
3. Mejorar la precisión del estimador del efecto testado (esto es, aumentar el poder estadístico).
4. Comparar y contrastar los resultados divergentes de los estudios individuales (resolver discrepancias).
5. Responder preguntas nuevas que los estudios originales no habían planteado o no podían responder por no tener suficiente poder estadístico (análisis de subgrupos).
6. Orientar futuros trabajos de investigación (generar hipótesis).

Indicaciones para la realización de un metaanálisis

Existen, además, ciertas situaciones concretas en las que estaría específicamente indicado realizar un metaanálisis⁴:

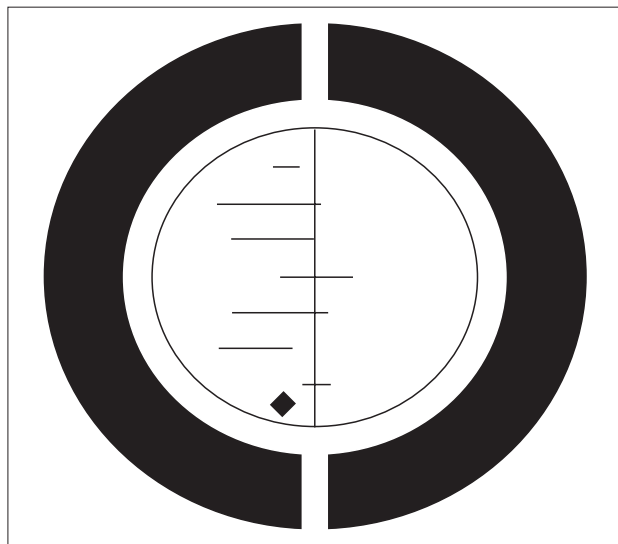


Fig. 1. Logotipo de la Colaboración Cochrane.

1. Cuando hay que tomar decisiones (p. ej., terapéuticas) y no se dispone de tiempo suficiente para realizar nuevos estudios. La revisión sistemática es una técnica científica eficiente, pues, aunque es laboriosa y ocupa mucho tiempo, este procedimiento es generalmente más rápido y menos costoso que embarcarse en un nuevo estudio².

2. Cuando existen muchos estudios no concluyentes sobre los efectos de un tratamiento, bien porque el efecto esperable es pequeño –aunque relevante desde el punto de vista clínico–, bien porque el estudio que se tendría que llevar a cabo para obtener una respuesta definitiva es irrealizable. Así, los metaanálisis aumentan el poder estadístico de los estudios originales, lo que es especialmente ventajoso en condiciones de baja prevalencia o cuando se evalúan efectos de pequeña magnitud². Un bonito ejemplo es el representado por el logotipo de la Colaboración Cochrane (fig. 1), que ilustra la capacidad de los corticoides para reducir la mortalidad perinatal en los recién nacidos de mujeres con riesgo de parto prematuro, mediante un metaanálisis que incluye 7 ensayos clínicos aleatorizados: el efecto beneficioso es estadísticamente significativo, a pesar de que sólo 2 de los estudios mostraban un claro efecto a favor de los corticoides².

3. Cuando los resultados entre estudios son divergentes o incluso contradictorios. En estos casos el metaanálisis puede contribuir a explicar el porqué de las discrepancias entre estudios.

4. Cuando se desea establecer con cierta precisión la magnitud de la eficacia de una intervención sanitaria.

Contraindicaciones para la realización de un metaanálisis

Por otro lado, hay también una serie de situaciones en las que no debería llevarse a cabo un metaanálisis⁴:

1. Para decidir la aprobación o rechazo de un nuevo fármaco (en las fases iniciales del desarrollo de un fármaco

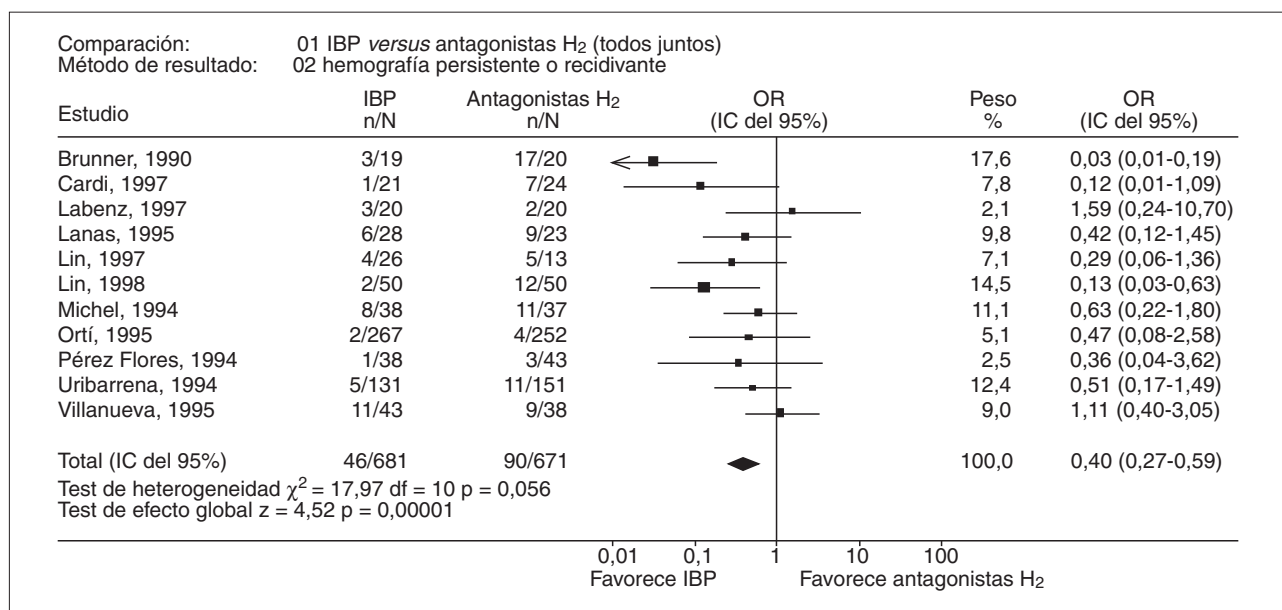


Fig. 2. Gráfico de efectos en el que se representa la odds ratio de cada estudio (cuadrado), sus intervalos de confianza del 95% (líneas horizontales) y la odds ratio global (diamante). Se trata de un metaanálisis que compara los inhibidores de la bomba de protones con los antagonistas H₂ en el tratamiento de la hemorragia digestiva por úlcera péptica. La medida de resultado evaluada es la persistencia o la recidiva hemorrágica. IBP: inhibidores de la bomba de protones; H₂ antagonistas: antagonistas H₂ de la histamina; OR: odds ratio; IC: intervalo de confianza.

los ensayos clínicos aleatorizados constituyen la herramienta más apropiada para valorar la eficacia terapéutica).

2. En situaciones en las que existe suficiente evidencia y se conoce con precisión el efecto de una intervención sanitaria (es obvio que no tiene sentido efectuar metaanálisis sobre aspectos en los que existe una certeza casi absoluta).
3. Para intentar convertir un efecto clínicamente pequeño en estadísticamente significativo combinando muchos estudios irrelevantes que muestran efectos insignificantes.
4. Para suplir la necesidad de llevar a cabo un estudio original, con el único objetivo de aumentar el número de publicaciones (aunque esta situación pueda parecer absurda, no es excepcional encontrar en la bibliografía metaanálisis sin preguntas de interés científico o sanitario).

Resumen de los efectos de los estudios en un metaanálisis

Un objetivo de la mayoría de los metaanálisis es proporcionar una estimación fiable de los efectos de una intervención, basada en una media ponderada de los resultados de todos los estudios relevantes. Así, la síntesis cuantitativa o metaanálisis ocurre cuando al menos una medida de resultados procedente de como mínimo 2 estudios diferentes se combina estadísticamente para obtener un resultado llamado «efecto sumatorio global» (*overall summary effect* en la terminología anglosajona)³. Generalmente, a cada estudio se le asigna como peso el inverso de su variancia; es decir, a las estimaciones más precisas, provenientes de estudios más amplios y con más sucesos, se les da más peso. También es posible dar más o menos peso a los estudios basándose en otros factores como su calidad metodológica, pero esto raramente se hace.

Para el cálculo de un efecto global, la situación ideal es aquella en la que se cuenta con bastantes estudios de pequeño tamaño, muy homogéneos tanto en su diseño como en sus resultados, lo que permitirá obtener una sólida estimación global del efecto¹⁰. La combinación de los hallazgos de un grupo de estudios tiene sentido si las diferencias observadas entre sus resultados no son estadísticamente significativas o clínicamente relevantes. En tal caso, la combinación resulta relativamente sencilla y cada estudio se resume usando una medida de efecto (riesgo relativo, *odds ratio*, etc.) que representa la comparación del grupo de intervención y de control. De este modo, los participantes en cada estudio se comparan únicamente con los demás participantes en el mismo estudio.

Para la presentación de los resultados se suelen emplear gráficos, el más habitual de los cuales es el denominado *forest plot* o gráfico de efectos, en el que se ordenan en vertical, según un cierto criterio (p. ej., el nombre del autor o la fecha de publicación), los resultados obtenidos en cada estudio, con su intervalo de confianza del 95% (líneas horizontales), así como el efecto global del metaanálisis. Se suele representar el valor medio de cada estudio mediante un símbolo (cuadrado o rombo) cuya área es proporcional al peso con el que interviene en el cálculo global (inversamente proporcional a la variancia), para así resaltar de algún modo los estudios más precisos¹⁰.

En la figura 2 se puede observar el gráfico de efectos de un metaanálisis que compara los inhibidores de la bomba de protones con los antagonistas H₂ en el tratamiento de la hemorragia digestiva por úlcera péptica, siendo la medida de resultado evaluada la persistencia o la recidiva hemorrágica²⁸. Se observa cómo en la mayoría de los casos el cuadrado que representa el efecto resumen de cada

estudio se encuentra situado a la izquierda de la línea media (la cual señala una *odds ratio* igual a la unidad, es decir, la ausencia de efecto) y, por tanto, indica un mayor efecto preventivo (sobre la persistencia o recidiva hemorrágica) de los inhibidores de la bomba de protones. Sin embargo, en esta misma figura se puede comprobar cómo en todos menos en 2 de los estudios la línea horizontal, que representa el intervalo de confianza del 95%, cruza la línea media, lo que indica que, aunque exista una tendencia a un mayor efecto con los inhibidores de la bomba de protones, las diferencias no alcanzan significación estadística en la mayoría de los estudios individuales (probablemente por el reducido tamaño muestral de éstos). Y es aquí donde la potencia del metaanálisis, que aglutina los resultados de todos los estudios, nos proporciona una estimación más precisa sobre el efecto del tratamiento: una *odds ratio* global de 0,4 (representada por el rombo-diamante situado en la parte inferior de la figura) con un intervalo de confianza del 95% (representado por los extremos izquierdo y derecho de dicho rombo-diamante) que no incluye a la unidad (de 0,27 a 0,59) y, por consiguiente, indica un efecto estadísticamente significativo.

Métodos estadísticos de los metaanálisis

El análisis estadístico es, quizá, la parte más sencilla en la ejecución de un metaanálisis. Si los otros pasos se han seguido con rigor, la combinación estadística de los valores estimados por cada estudio es de naturaleza meramente mecánica⁴. Los metaanálisis utilizan diversas técnicas estadísticas pero, lamentablemente, no hay una única correcta, pues la elección de ésta depende del tipo de datos analizados. Para datos dicotómicos (o binarios) hay varios pros y contras en cada uno de los estadísticos de resumen disponibles: *odds ratio*, riesgo relativo y diferencia de riesgo. La interpretación de la *odds ratio* como aproximación del riesgo relativo sólo es válida para tasas pequeñas, mientras que para sucesos frecuentes el primer estadístico no es una buena aproximación del segundo¹⁰.

El «número necesario a tratar» (NNT), que puede obtenerse fácilmente a partir de la diferencia de riesgos (es su inversa), es una forma útil e intuitiva, con mayor significación clínica, de expresar de nuevo los resultados: nos dice cuántos pacientes es necesario tratar con el nuevo tratamiento para producir un evento de interés, en comparación con el placebo o el tratamiento control. Por ejemplo, en un reciente metaanálisis que compara la incidencia de recidiva hemorrágica por úlcera péptica en los pacientes que reciben tratamiento erradicador de *H. pylori* frente a los tratados únicamente con antisecretores (sin recibir posteriormente tratamiento antisecretores de mantenimiento), se calcula una *odds ratio* de 0,18 (con un intervalo de confianza del 95% que no incluye la unidad)²⁹. Esta cifra, que evidentemente indica un efecto protector del tratamiento erradicador sobre la recidiva hemorrágica, no es fácil de «trasladar» a la práctica clínica. Sin embargo, si expresamos la magnitud de dicho efecto protector diciendo que el NNT correspondiente es de 5, lo que indica que es necesario administrar tratamiento erradicador

(en comparación con el tratamiento antisecretores) a tan sólo 5 pacientes para conseguir prevenir un episodio de recidiva hemorrágica, la interpretación es más directa e intuitiva.

Aunque no existe regla alguna para clasificar un NNT como alto o bajo, parece evidente que un valor de 5 traduce un efecto terapéutico importante. Por el contrario, un reciente metaanálisis que evalúa el efecto del tratamiento erradicador de *H. pylori* sobre la mejoría de los síntomas de la dispepsia funcional demuestra que el NNT con terapia erradicadora para conseguir la mejoría sintomática de uno de estos pacientes (en comparación con el placebo) es de 25, e incluso de 50 cuando se excluye un estudio responsable de heterogeneidad en los resultados²⁷. Es obvio que esta última cifra (uno entre 50) supone un beneficio más que dudoso, que deberá sopesarse con los posibles perjuicios que podría causar la administración de tratamiento erradicador a los 49 pacientes restantes que no se beneficiarán de él.

Cuando se presentan medidas absolutas –como el NNT– es útil dar la información para un intervalo de riesgos basales (p. ej., para el intervalo observado en los estudios incluidos). Por otra parte, los NNT son específicos de un período de seguimiento y solamente deben calcularse cuando se combinan ensayos en los que esta variable tiene un valor similar. Por último, conviene recordar que el NNT es sumamente sensible a los sesgos de todo tipo y en particular a la heterogeneidad entre los estudios, por lo que su uso a partir de los resultados globales de un metaanálisis debe ser siempre juicioso³.

Las medidas absolutas pueden ser más informativas que las relativas porque reflejan el riesgo basal y el cambio de riesgo con la intervención. No obstante, si la tasa en el grupo control no es similar en todos los estudios, no es correcto emplear una medida, como la reducción absoluta de riesgo, para estimar el efecto global y habrá que utilizar una medida relativa, como puede ser la *odds ratio* o el riesgo relativo¹⁰. Como resumen, podríamos recomendar que siempre que sea posible deberían describirse los resultados con una medida relativa (*odds ratio*, riesgo relativo o reducción del riesgo relativo) y también mediante un estimador absoluto (NNT o diferencia de riesgos)^{4,30}.

Evaluación de la heterogeneidad

En la actualidad constituye una de las motivaciones fundamentales para realizar un metaanálisis, incluso mayor que el mero cálculo de una estimación del efecto global¹⁰. Es importante preguntarse si los resultados de los estudios son similares dentro de cada comparación. Una manera de responder a esta pregunta es examinar una presentación gráfica de los resultados. Si los intervalos de confianza para los resultados de cada estudio (representados típicamente mediante líneas horizontales) no se superponen, esto indica que es probable que las diferencias entre los estudios sean estadísticamente significativas. Las pruebas de homogeneidad son análisis estadísticos formales para examinar si la variación observada en los resultados de los estudios es compatible con la variación

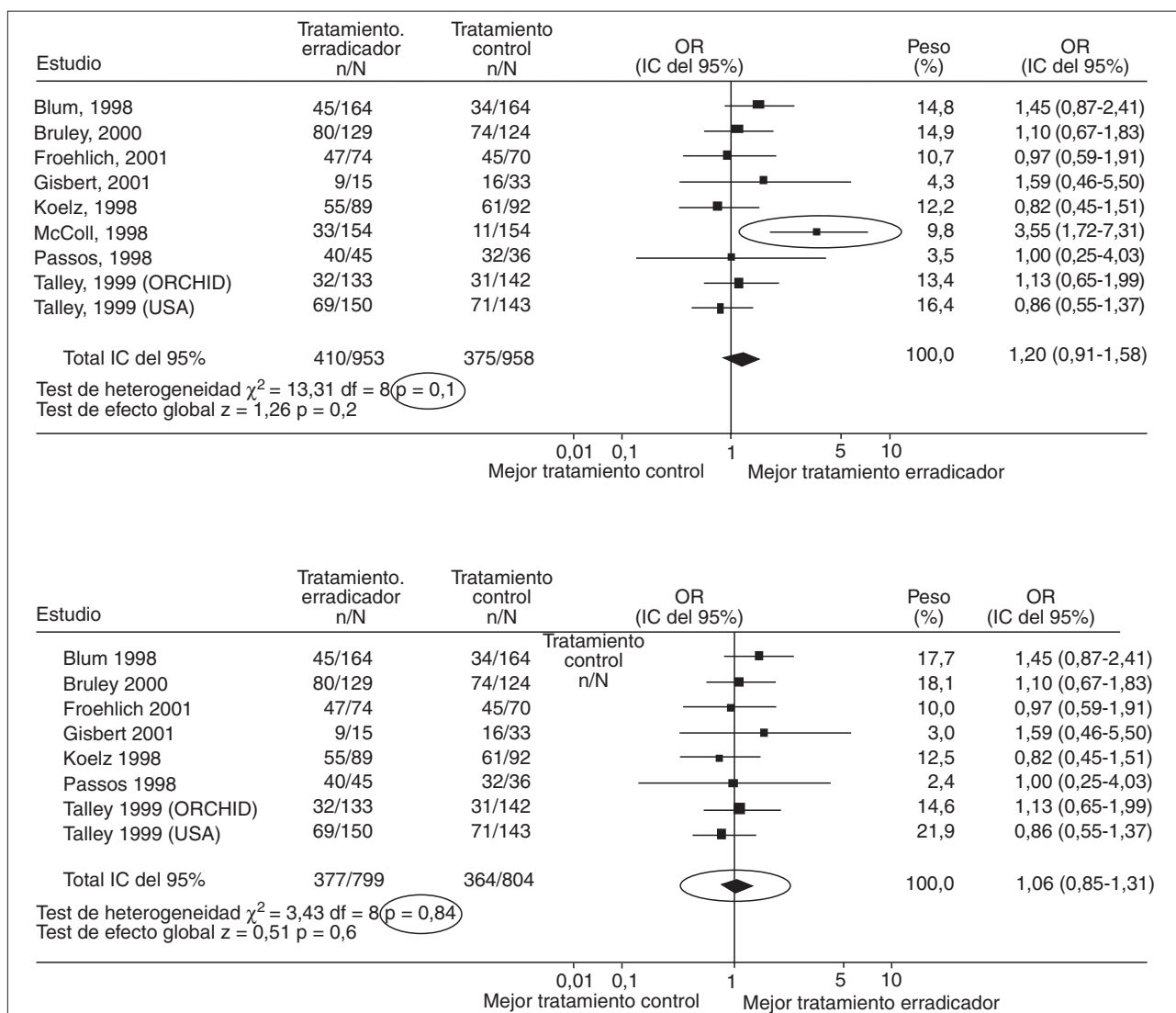


Fig. 3. Metaanálisis de la eficacia del tratamiento erradicador de Helicobacter pylori en la dispepsia funcional²⁷. Panel superior: presencia de heterogeneidad, debida a los resultados discordantes del estudio de McColl et al; panel inferior: resultados homogéneos tras la exclusión de dicho estudio. OR: odds ratio; IC: intervalo de confianza.

esperada únicamente por azar. Cuanto más significativos sean los resultados de la prueba (valor de p más pequeño), más probable es que las diferencias observadas no sean debidas únicamente al azar. En general, las pruebas estadísticas que valoran la presencia de heterogeneidad suelen tener una baja potencia estadística, por lo que es habitual situar el punto de corte para la p¹⁰ en 0,1 en lugar del habitual 0,05. Además, una prueba no significativa no descarta necesariamente la existencia de heterogeneidad. Por lo tanto, las diferencias clínicamente importantes entre los resultados de los estudios siguen imponiendo precaución en la interpretación de los hallazgos globales, a pesar de que la prueba estadística de heterogeneidad pueda no ser significativa⁸.

En el ejemplo del metaanálisis anteriormente mencionado sobre el efecto del tratamiento erradicador de H. pylori en la dispepsia funcional²⁷, la mera observación de la representación gráfica nos indica que uno de los estudios

(McCull, 1988) ofrece resultados claramente discordantes (fig. 3, panel superior). De hecho, la p del test de heterogeneidad, con un valor de 0,1, confirma que nos encontramos en el límite estadístico que la define. Los resultados contradictorios entre varios estudios podrían explicarse, al menos en parte, por la diferente prevalencia de úlcera gastroduodenal en las distintas áreas geográficas. Así, es posible que el efecto favorable de la terapia erradicadora evidenciado en los ensayos clínicos positivos sea debido al tratamiento de la enfermedad ulcerosa no diagnosticada, que, como es sabido, tiene una excelente respuesta a la erradicación de H. pylori. Por tanto, la terapia frente a H. pylori podría estar tratando únicamente al reducido subgrupo de pacientes con úlcera péptica a los que se ha clasificado erróneamente como dispepticos funcionales al no objetivarse durante la gastroscopia una úlcera gastroduodenal activa. La elevada incidencia de úlcera gastroduodenal en Escocia, donde se llevó a cabo el

único estudio con resultados positivos, apoyaría esta posibilidad. Precisamente, la heterogeneidad encontrada al principio en el metaanálisis desaparecía al excluir el estudio de McColl et al, situación en la que el valor de la *p* del test de heterogeneidad ascendía hasta 0,84 (fig. 3, panel inferior).

Como se ha comentado antes, la existencia de heterogeneidad estadísticamente significativa indica que las diferencias observadas en los resultados de los estudios individuales podrían estar causadas por otros factores distintos del azar. En este caso, los revisores deben alertar a los lectores de su revisión y recomendar cautela en la interpretación de los resultados combinados. Incluso podrían llegar a decidir que la heterogeneidad es de tal magnitud que es preferible no combinar los resultados de los estudios individuales. Se debe tener siempre presente que los tests 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 revisiones sistemáticas cualitativas³. Los revisores pueden comentar, en ese caso, que les fue imposible combinar los resultados porque los estudios que encontraron no ofrecían datos combinables. Por último, los revisores deben intentar explorar las causas de la heterogeneidad y tratar de explicarla, pero cualquier interpretación debe ser cautelosa, puesto que estos análisis se hacen habitualmente *a posteriori*. Idealmente, deben especificarse *a priori*, dentro del protocolo inicial de la revisión, los factores que podrían conducir a diferencias entre los resultados de estudios individuales. En resumen, es importante recalcar que un adecuado análisis de los factores que pueden ser responsables de la heterogeneidad entre los estudios es de mayor importancia que la idea simplista de buscar una medida de efecto global cuando existen resultados contradictorios entre los diversos estudios¹⁰.

Análisis de subgrupos

A menudo resulta interesante examinar una categoría particular de participantes en una revisión (como mujeres, un grupo de edad en particular o los sujetos con un subtipo específico o gravedad de la enfermedad). Es preciso recalcar aquí que el término «subgrupos» se refiere a las agrupaciones particulares de pacientes entre los estudios, no a subgrupos de estudios³. Con frecuencia se abusa de los análisis de subgrupos, por lo que los revisores deberán ser cuidadosos al efectuarlas y los lectores al interpretarlas. Las preguntas incluidas en la tabla IV pueden utilizarse para ayudar a decidir si se lleva o no a cabo este tipo de análisis. De entre ellas destaca la primera, que plantea si es plausible esperar que haya una diferencia basándose en lo que se conoce acerca de la intervención y del problema, ya que si no es así probablemente no debería emplearse el análisis de subgrupos. En la medida de lo posible, los revisores deben intentar reducir al mínimo el número de estos análisis, pues cuanto más grande sea el número de hipótesis a evaluar mayor será el número de las diferencias que podrán encontrarse simplemente por

TABLA IV. Preguntas que ayudan a decidir si debe o no hacerse un análisis de subgrupos

- | |
|--|
| <ol style="list-style-type: none"> 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 <i>dentro</i> de los estudios y no <i>entre</i> 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? |
|--|

azar. Es más probable que las decisiones acerca de qué análisis realizar estén reorientadas por los datos cuando se planean *a posteriori* y, en consecuencia, es más probable que sean espurias. Se debe ser extremadamente cauteloso al examinar y plantear conclusiones acerca de la existencia de diferencias en la respuesta de los subgrupos basándose en diferencias en los resultados *entre* los estudios, pues puede haber muchos factores que lo justifiquen (p. ej., diferencias en los fármacos, en las dosis, en los esquemas de tratamiento específicos que se han utilizado, etc.). Incluso si la revisión incluye únicamente ensayos aleatorizados, debe recordarse que no se aleatorizó a los pacientes a un estudio o a otro, sino que, dentro de un mismo estudio, se les aleatorizó a recibir uno u otro tratamiento.

Análisis de sensibilidad

Dado que hay varios enfoques para realizar una revisión sistemática, los revisores deben preguntarse cuán sensibles son los resultados del análisis a los cambios en el modo en que éste fue realizado. Esto proporciona a los revisores un mecanismo para probar cuán robustos son los resultados de la revisión en relación con las decisiones y con las asunciones clave surgidas en el proceso de realizarla. Si los análisis de sensibilidad que se efectúan no cambian sustancialmente los resultados, la confianza que tendremos en éstos queda fortalecida. Por el contrario, si los resultados cambian de manera que conducen a conclusiones diferentes, se deberá ser más cauteloso en la interpretación de los datos y al plantear las conclusiones.

Presentación de los resultados del metaanálisis

Con el ánimo de mejorar y homogeneizar la calidad de la presentación de los resultados de los metaanálisis, el grupo QUOROM (Quality Of Reporting Of Meta-analysis) ha publicado recientemente un documento donde se destacaban los puntos que deberían exigirse ineludiblemente^{31,32}. Según este grupo, existen 6 áreas temáticas generales importantes para la realización y presentación de los resultados de cualquier metaanálisis de ensayos clínicos aleatorizados: diseño del estudio, combinación adecuada de diversos estudios, control del sesgo, análisis estadístico, análisis de sensibilidad y problemas de aplicabilidad. Esta propuesta incluye una lista (fig. 4) que «obliga» a los autores del metaanálisis a ofrecer a los lectores información explícita sobre la estrategia de búsqueda, los criterios para la se-

Categoría	Subcategoría	Ítem	¿Incluido en informe? (S/N)	N.º páginas
Título		Identificar el trabajo como un metaanálisis [o revisión sistemática] de ECC ²⁶		
Resumen		Utilizar un formato estructurado ²⁷		
	Objetivos	Describir La cuestión clínica explícitamente		
	Fuentes de datos	Las bases de datos (es decir, una relación) y otras fuentes de información		
	Métodos de revisión	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		
	Resultados	Caraterísticas de los ECC incluidos y excluidos: hallagos cualitativos y cuantitativos (estimaciones puntuales e intervalos de confianza), y análisis de subgrupo		
	Conclusión	Resultados principales		
		Describir		
Introducción		El problema clínico concreto; los fundamentos biológicos de la intervención y los fundamentos de la revisión		
Métodos	Búsqueda	Las fuentes de información, con detalle ²⁸ (p.e. bases de datos, registros, ficheros personales, expertos, agencias, búsqueda manual) y cualquier restricción (años considerados, situación de publicación ²⁹ , idioma de la publicación ^{30,31})		
	Selección	Los criterios de inclusión y exclusión (relativos a la población, la intervención, los resultados principales y el diseño del estudio ³²)		
	Evaluación de la validez	Criterios y proceso utilizados (p.e. condiciones encubiertas, valoración de la calidad y sus hallazgos ³³⁻³⁶)		

Fig. 4. Lista (checklist) para evaluar la correcta presentación de los resultados de un metaanálisis. Propuesta QUOROM (Quality Of Reporting Of Meta-analysis)³². ECC: ensayos clínicos controlados.

lección 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. Además, la propuesta QUOROM incluye un diagrama de flujo (fig. 5) en el que los autores deberán aportar información detallada sobre el número de ensayos clínicos aleatorizados identificados, incluidos y excluidos, y los motivos por los que estos últimos lo fueron.

Interpretaciones erróneas y limitaciones de los metaanálisis

Después de un notable entusiasmo inicial, las discrepancias entre las conclusiones extraídas de algunos metaanálisis y los hallazgos encontrados en macroensayos aleatorizados^{33,34} han llevado a un planteamiento mucho más crítico del papel del metaanálisis¹⁰. No se debe caer en el error de considerar el metaanálisis una mera combinación

de múltiples estudios, como si fuera, por tanto, «una revisión ordinaria, pero más grande»²⁰. Así, actualmente se profundiza mucho 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 resultados¹⁰. Por otra parte, no se debe olvidar que, en todo caso, el metaanálisis es un estudio observacional, con las limitaciones que ello comporta¹⁰. Desde sus orígenes el metaanálisis ha sido y sigue siendo una técnica controvertida, sobre todo cuando se ha limitado a un ingenuo cálculo del efecto global, dado que los métodos estadísticos no pueden «corregir» o «ajustar» las discrepancias entre estudios, y mucho menos lograr que resultados contradictorios se combinen en una única medida global¹⁰. La aparente precisión con la que se efectúa el cálculo de una estimación global en un metaanálisis puede disuadir de la realización de estudios posteriores sobre ese asunto, lo que en ocasiones puede llegar a ser

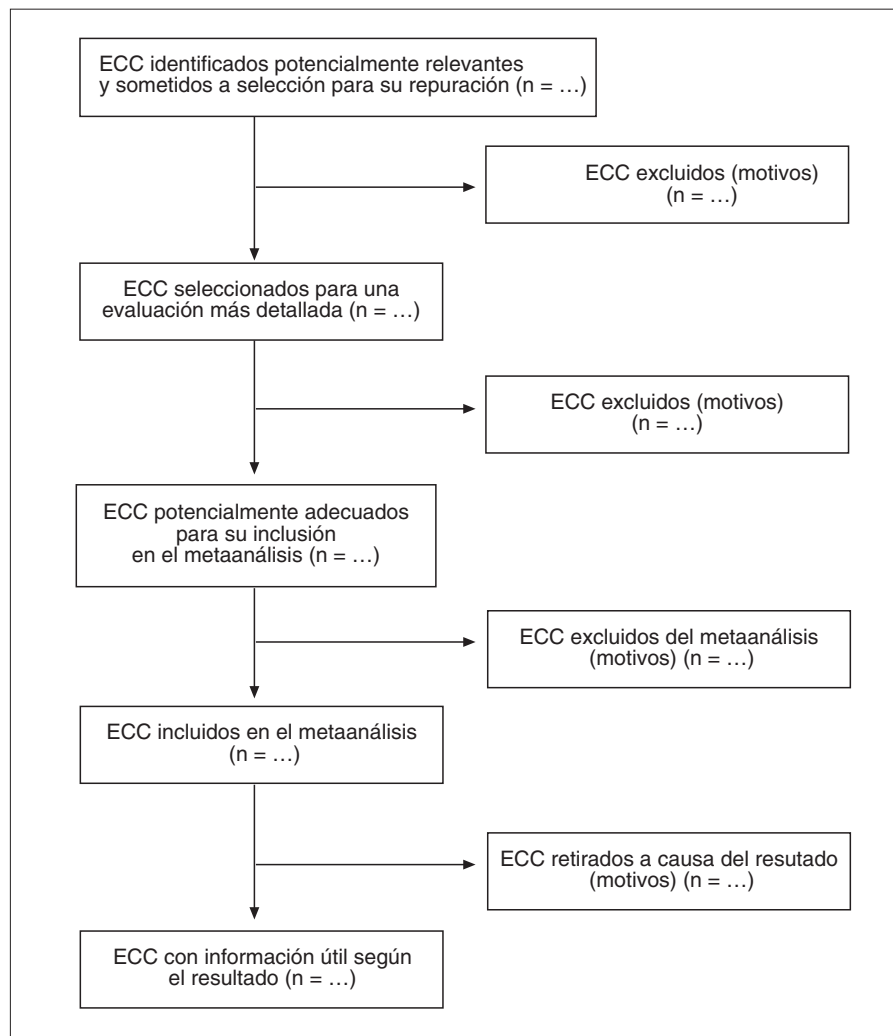


Fig. 5. Diagrama de flujo (flow diagram) para evaluar la correcta presentación de los resultados de un metaanálisis. Propuesta QUOROM (Quality Of Reporting Of Meta-analysis)³². ECC: ensayos clínicos controlados.

un inconveniente¹⁰. En este sentido, el metaanálisis no debe sustituir a los ensayos clínicos de gran tamaño; sin embargo, cuando éstos no existen, el metaanálisis constituye una guía para la práctica clínica e incluso para el diseño de futuros ensayos¹⁰. No se debe caer en el error de considerar el metaanálisis un sustituto de los grandes ensayos clínicos de elevada calidad. Así, el metaanálisis debe ser una herramienta para la generación de hipótesis más que para la toma de decisiones¹⁰. Idealmente deberá localizar dónde se encuentran las lagunas de conocimiento, para así orientar las líneas de investigación futuras. Dicho de otro modo, el metaanálisis es complementario, no sustituto, del ensayo clínico aleatorizado de gran tamaño y representa una estrategia coste-efectiva porque reduce el número de estudios primarios innecesarios. Como se ha comentado previamente, el metaanálisis no es la panacea que convierte estudios primarios metodológicamente deficientes en una combinación brillante de ensayos clínicos, por lo que hay que tener especial cuidado en no combinar de forma indiscriminada (dando igual valor) las investigaciones de buena y mala calidad. Mezclar «manzanas y naranjas» en la misma cesta es también

un error que se comete frecuentemente al realizar un metaanálisis, al ignorar las diferencias fundamentales entre los distintos estudios y combinarlos sin más²⁰. Por ejemplo, no sería apropiado evaluar el efecto de los bloqueadores beta en la prevención de la hemorragia digestiva por varices esofágicas mezclando, en el mismo metaanálisis, estudios dirigidos a la profilaxis primaria (del primer episodio hemorrágico) y secundaria (de la recurrencia de la hemorragia). Ambas poblaciones de pacientes cirróticos tienen un riesgo de base diferente, del que depende directamente el valor de la variable que se pretende medir, en este caso, la incidencia de hemorragia digestiva por varices esofágicas.

En resumen, el metaanálisis, si se utiliza correctamente, constituye la mejor herramienta posible para combinar los resultados de diversos estudios en un mundo cada vez más desbordado por el volumen de información que precisa ser sintetizada¹⁰. Sin embargo, debido a su elevada potencia, es también un arma de doble filo que puede llevarnos a sacar conclusiones sesgadas a partir de estudios imperfectos que aisladamente no habrían sido capaces de «confundirnos»²⁰.

TABLA V. Lista de verificación (checklist) que permite la detección de errores importantes en una revisión sistemática

1. Formulación de la pregunta	¿Las preguntas de la revisión están bien formuladas y contienen los componentes clave?
2. Identificación de estudios	¿Se realiza una búsqueda minuciosa de datos pertinentes usando fuentes apropiadas? ¿Las estrategias de búsqueda son adecuadas para la pregunta que se ha planteado?
3. Selección de estudios	¿Se aplican criterios de inclusión y de exclusión apropiados para seleccionar los estudios? ¿Se aplican los criterios de selección de manera que limiten los sesgos?
4. Evaluación de los estudios	¿Se valora de manera fiable la validez de los estudios individuales? ¿Se abordan de modo sistemático los parámetros importantes que pueden afectar a los resultados del estudio?
5. Obtención de 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 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 informa de la precisión de los resultados?
7. Discusión	¿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?
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?

Sexto paso: interpretación de los resultados

Existen algunos temas que deben ser objeto de discusión en una revisión sistemática y que pueden ser de ayuda en la toma de decisiones, como la solidez de la evidencia, la aplicabilidad de los resultados o el balance entre los beneficios esperados y los posibles daños y costes de la intervención de interés. Con respecto a la solidez de la evidencia, un buen punto de partida o de destino para la sección de discusión de una revisión es abordar cualquier limitación metodológica importante de los estudios incluidos y de los métodos usados en la revisión que pueda afectar a las decisiones prácticas a tomar.

En relación con la aplicabilidad, una de las ventajas de una revisión de conjunto es que, dado que incluye numerosos estudios, los resultados proceden de una variedad muy diversa de pacientes⁸. Si los resultados son consistentes a través de los estudios, se aplicarán por tanto a esta amplia variedad de pacientes. Incluso así, se puede seguir teniendo dudas sobre la aplicabilidad de los resultados. Por ejemplo, una revisión sobre la estrategia *test-and-treat* en la dispepsia, basada en el diagnóstico y el consiguiente tratamiento de la infección por *H. pylori* sin necesidad de practicar una

gastroscoopia, se fundamenta en una serie de estudios que evalúan dicha estrategia en pacientes dispépticos jóvenes (p. ej., menores de 50 años), pues en este grupo de edad la incidencia de cáncer gástrico es mínima. Sin embargo, el lector puede que proceda de o resida en una zona de alta incidencia de adenocarcinoma gástrico como Japón o Latinoamérica, donde probablemente no sea aplicable este punto de corte para la edad y sea prudente reducirlo. Por tanto, los usuarios de cualquier revisión sistemática deben decidir si la evidencia que encuentran en ella es o no aplicable a sus circunstancias específicas. Para hacerlo, deben determinar si los participantes y los entornos en los estudios incluidos son razonablemente similares a su propia situación. En general, en lugar de aplicar rígidamente los criterios de inclusión y de exclusión de los estudios a situaciones específicas, es mejor preguntarse si hay razones evidentes por las cuales los ensayos no deben aplicarse en ciertas circunstancias.

Un tercer aspecto es si los beneficios compensan los riesgos de la intervención. Por ejemplo, en una reciente revisión sistemática y metaanálisis, llevada a cabo por la Colaboración Cochrane sobre el efecto del tratamiento erradicador de *H. pylori* en la dispepsia funcional, se concluye que la erradicación del microorganismo tiene un pequeño pero estadísticamente significativo efecto beneficioso³⁵; los autores plantean un modelo económico que apunta finalmente a que este modesto beneficio es, probablemente, coste-efectivo.

Finalmente, un error común que se comete cuando la evidencia disponible no es concluyente, tanto al describir resultados como al sacar conclusiones, es confundir «la falta de evidencia de efecto» con «la evidencia de que no hay ningún efecto». Dicho de otro modo, no es lo mismo decir que un estudio «no ha demostrado diferencias estadísticamente significativas» entre 2 tratamientos que afirmar que realmente «no existen diferencias» y que ambos tratamientos son equivalentes. Los autores y los lectores deben ser cuidadosos para no caer en esta trampa, que afecta típicamente a los estudios (y también a los metaanálisis) que incluyen un reducido número de pacientes, lo que se traduce en una escasa potencia estadística y, cuando no se demuestran diferencias entre los 2 grupos de tratamiento, una elevada probabilidad de cometer un error beta o de tipo II. Un ejemplo que destaca este problema de «equivalencia» es el siguiente: el tamaño muestral necesario para demostrar que existen diferencias estadísticamente significativas entre un tratamiento A (con una tasa de éxito terapéutico del 80%) y un tratamiento B (con una diferencia clínicamente significativa de un 10%, esto es, una eficacia menor del 70% o mayor del 90%) es de nada menos que 250 pacientes por grupo aproximadamente (500 en total) si aplicamos un contraste bilateral.

Una segunda equivocación común consiste en sacar conclusiones que van más allá de la evidencia que se analiza. Por ejemplo, si un metaanálisis ha demostrado que un determinado tratamiento es más eficaz que el placebo para controlar la hemorragia digestiva por varices en los pacientes con cirrosis hepática, esto no implica que dicho tratamiento se siga necesariamente de una disminución de la mortalidad, ya que es posible que ésta dependa, como

a) ¿Los resultados de la revisión son válidos?

Preguntas «de eliminación»

<p>1 ¿Se hizo la revisión sobre un tema claramente definido?</p> <p>Pista: un tema debe ser definido en términos de:</p> <ul style="list-style-type: none"> - La población de estudio - La intervención realizada - Los resultados ("anteriores") considerados 	<p><input type="checkbox"/> Sí <input type="checkbox"/> No sé <input type="checkbox"/> No</p>	<p>3 ¿Crees que estaban incluidos los estudios importantes y relevantes?</p> <p>Pista: busca:</p> <ul style="list-style-type: none"> - ¿Qué bases de datos bibliográficas se han usado? - ¿Seguimientos de la referencias? - ¿Contacto personal con expertos? - búsqueda de estudios no publicados? - ¿Búsqueda de estudios en idiomas distintos del inglés? 	<p><input type="checkbox"/> Sí <input type="checkbox"/> No sé <input type="checkbox"/> No</p>
<p>2 ¿Buscaron los autores el tipo de artículos decuado?</p> <p>Pista: el mejor "tipo de estudio" es el que:</p> <ul style="list-style-type: none"> - Se dirige a la pregunta objeto de la revisión - Tiene un diseño apropiado para la pregunta 	<p><input type="checkbox"/> Sí <input type="checkbox"/> No sé <input type="checkbox"/> No</p>	<p>4 ¿Crees que los autores de la revisión han hecho suficiente esfuerzo para valorar la calidad de los estudios incluidos?</p> <p>Pista: los autores necesitan considerar el rigor de los estudios que han identificado. La falta de rigor puede afectar al resultado de los estudios ("No es oro todo lo que reluce", <i>El mercader de Venecia</i>, Acto II)</p>	<p><input type="checkbox"/> Sí <input checked="" type="checkbox"/> No sé <input type="checkbox"/> No</p>
<p>5 Si los resultados de los diferentes estudios han sido mezclados para obtener un resultado "combinado", ¿era razonable hacer eso?</p> <p>Pista: considera si:</p> <ul style="list-style-type: none"> - Los resultados de los estudios eran similares entre sí - Los resultados de todos los estudios incluidos están claramente presentados - Están discutidos los motivos de cualquier variación de los resultados 		<p><input type="checkbox"/> Sí <input type="checkbox"/> No sé <input type="checkbox"/> No</p>	

Fig. 6. Preguntas propuestas por el grupo CASP (Critical Appraisal Skills Programme) para la lectura crítica de una revisión sistemática: a) ¿los resultados de la revisión son válidos?⁸.

en realidad ocurre, de otros factores que no se ven modificados por el tratamiento en cuestión. Por último, incluso afirmaciones generales del tipo «se necesitan más estudios...», tan empleadas en las últimas líneas de muchos manuscritos, deberían evitarse. En todo caso, los revisores deben especificar exactamente qué investigación se necesita y su justificación precisa.

EVALUACIÓN CRÍTICA DE UNA REVISIÓN SISTEMÁTICA

Al plantearse llevar a cabo una revisión sistemática, o para revisar una ya realizada por otros autores, es útil emplear una lista de verificación (*checklist*) como guía para descubrir errores importantes, como la propuesta en la tabla V. Otro formato alternativo para la evaluación crítica de una revisión sistemática por parte de los lectores es la propuesta por el grupo CASP (Critical Appraisal Skills Programme, o programa de habilidades en lectura crítica), que, como su nombre indica, tiene como objetivo ayudar a desarrollar habilidades para hacer una lectura crítica de la evidencia científica. La lectura crítica sería el proceso de evaluar e interpretar la evidencia aportada por la bibliografía científica, considerando sistemáticamente los

resultados que se presentan, su validez y su relevancia para el trabajo propio. En España existe un grupo CASP (CASPe), en cuya página *web* (www.redcaspe.org) se pueden encontrar herramientas muy útiles para estas actividades. Siguiendo las «guías para usuarios de la bibliografía médica» publicadas en la revista *JAMA* (en concreto, el artículo escrito por Oxman et al⁸ sobre «cómo utilizar una revisión»), existen 3 preguntas fundamentales que deben plantearse cuando se hace la lectura crítica de este tipo de artículos: a) ¿son válidos los resultados? (fig. 6); b) ¿cuáles son los resultados? (fig. 7), y c) ¿son aplicables en el medio particular en que nos encontramos? (fig. 8). Las 2 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.

A MODO DE CONCLUSIONES

En las páginas anteriores se han aportado los argumentos necesarios para poder considerar las revisiones sistemáti-

b) ¿Cuáles son los resultados?

<p>6 ¿Cuál es el resultado global de la revisión?</p> <p>Pista: considera:</p> <ul style="list-style-type: none"> - Si tienes claro los resultados últimos de la revisión - ¿Cuáles son? (numéricamente, si es apropiado) - Cómo están expresados los resultados? (NNT, odds ratio, etc.) 	
<p>7 ¿Cuán precisos son los resultados?</p> <p>Pista: busca los intervalos de confianza de los estimadores</p>	

Fig. 7. Preguntas propuestas por el grupo CASP (Critical Appraisal Skills Programme) para la lectura crítica de una revisión sistemática: b) ¿cuáles son los resultados??. NNT: número necesario a tratar.

c) ¿Son los resultados aplicables en tu medio?

<p>8 ¿Se pueden aplicar los resultados en tu medio?</p> <p>Pista: considera si:</p> <ul style="list-style-type: none"> - Los pacientes cubiertos por la revisión pueden ser suficientemente diferentes de los de tu área - Tu medio parece ser muy diferente al del estudio 	<input type="checkbox"/> Sí <input type="checkbox"/> No sé <input type="checkbox"/> No
<p>9 ¿Se han considerado todos los resultados importantes para tomar la decisión?</p>	<input type="checkbox"/> Sí <input type="checkbox"/> No sé <input type="checkbox"/> No
<p>10 ¿Los beneficios merecen la pena frente a los perjuicios y costes?</p> <p>Aunque no esté planteado explícitamente en la revisión, ¿qué opinas?</p>	<input type="checkbox"/> Sí <input type="checkbox"/> No

Fig. 8. Preguntas propuestas por el grupo CASP (Critical Appraisal Skills Programme) para la lectura crítica de una revisión sistemática: c) ¿son los resultados aplicables en el medio particular en que nos encontramos?.

cas como piezas fundamentales del complejo engranaje que constituye la producción y divulgación de los conocimientos científicos en medicina. Asimismo, se han detallado las diversas etapas metodológicas que conviene tener en cuenta para desarrollar con solvencia y validez una revisión sistemática, según los criterios de la Colaboración Cochrane. Precisamente, uno de los rasgos distintivos de esta organización internacional, independiente y

sin ánimo de lucro, es que las revisiones sistemáticas son elaboradas por las personas interesadas por un tema particular. Es decir, no se llevan a cabo de manera centralizada por parte de un instituto especializado (aunque existen diversos organismos que producen revisiones sistemáticas de calidad), sino de una red de profesionales sanitarios (y también de pacientes y ciudadanos) de todo el mundo interesados en revisar críticamente la base científica que sustenta las diferentes decisiones clínicas y sanitarias. Para ello, la Colaboración Cochrane proporciona formación y apoyo metodológico a todos aquellos que se muestran interesados en participar activamente en esta labor. El Centro Cochrane Iberoamericano es el que en España e Iberoamérica se responsabiliza de coordinar este apoyo y facilitar la información necesaria^{2,36}, y desde aquí animamos a cualquiera que esté interesado en colaborar con esta organización a visitar su página web (<http://www.cochrane.es>).

Además, desde octubre de 2003 y gracias a una suscripción realizada por el Ministerio de Sanidad y Consumo, cualquier ciudadano puede consultar gratuitamente la revista electrónica Cochrane Library Plus, la cual contiene la versión en castellano de la Cochrane Library y otras publicaciones de interés (<http://www.update-software.com/clibplus>).

BIBLIOGRAFÍA

1. Gisbert JP. Fuentes y centros de gestión del conocimiento. Enfermedad inflamatoria intestinal al día 2003;2:32-8.
2. López Arrieta JM, Qizilbash N. La medicina basada en pruebas: revisiones sistemáticas. La Colaboración Cochrane. Med Clin (Barc) 1996;107:581-5.
3. Sáenz A, Ausejo M. Guía para leer e interpretar una revisión sistemática. Rev Soc Mad Med Fam Comunit 2000;2:29-36.
4. Gabriel-Sánchez R, Pladevall-Vila M. Evaluación de la evidencia en medicina: revisiones sistemáticas y metaanálisis. Medicina 1998;7:4845-51.
5. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126:376-80.
6. Siwek J, Gourlay ML, Slawson DC, Shaughnessy AF. How to write an evidence-based clinical review article. Am Fam Physician 2002;65:251-8.
7. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 1992;268:240-8.
8. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994;272:1367-71.
9. Ceballos C, Valdizan JR, Artal A, Almarcegui C, Allepuz C, García Campayo J, et al. ¿Por qué la medicina basada en la evidencia? 20 años de metaanálisis. An Med Interna 2000;17: 521-6.
10. Molinero LM. Interpretación, utilidad y limitaciones del metaanálisis. Gastroenterol Hepatol Continuada 2002;1:311-8.
11. Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2 [updated March 2003]. En: Review Manager (RevMan) [computer program]. Versión 4.2. Oxford, England: The Cochrane Collaboration, 2003. Disponible en: <http://www.cochrane.de/cochrane/hbook.htm>
12. 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/Castellano/CC_Handbook

13. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997; 127:380-7.
14. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998; 317:1185-90.
15. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;309:1286-91.
16. Delgado-Rodríguez M, Sillero-Arenas M. Sesgos en el metaanálisis. *Med Clin (Barc)* 1999;112(Supl 1):43-50.
17. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;350:326-9.
18. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;320:1574-7.
19. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000;356:1228-31.
20. Huang JQ, Hunt RH. Uses and abuses of meta-analysis in gastroenterology. *Chin J Dig Dis* 2002;3:39-42.
21. Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309:1358-61.
22. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352:609-13.
23. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001; 323:42-6.
24. Oxman AD. Checklists for review articles. *BMJ* 1994;309: 648-51.
25. 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. *Control Clin Trials* 1995;16:62-73.
26. 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.
27. Gisbert JP, Calvet X, Gabriel R, Pajares JM. Infección por *Helicobacter pylori* y dispepsia funcional. Metaanálisis de la eficacia del tratamiento erradicador. *Med Clin (Barc)* 2002; 118:405-9.
28. Gisbert JP, González L, Calvet X, Roque M, Gabriel R, Pajares JM. Proton pump inhibitors versus H₂-antagonists: a meta-analysis of their efficacy in treating bleeding peptic ulcer. *Aliment Pharmacol Ther* 2001;15:917-26.
29. Gisbert JP, Khorrami S, Carballo F, Calvet X, Gené E, Domínguez-Muñoz E. *Helicobacter pylori* eradication therapy versus antisecretory non-eradication therapy (with or without long-term maintenance antisecretory therapy) for the prevention of recurrent bleeding from peptic ulcer (Cochrane Review). En: *The Cochrane Library*, Issue 4, 2003. Chichester: John Wiley & Sons, Ltd.
30. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann Intern Med* 1997;126: 712-20.
31. 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.
32. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Mejora de la calidad de los informes de los metaanálisis de ensayos clínicos controlados: el acuerdo QUOROM. *Rev Esp Salud Pública* 2000;74:107-18.
33. Cappelleri JC, Ioannidis JP, Schmid CH, De Ferranti SD, Aubert M, Chalmers TC, et al. Large trials versus meta-analysis of smaller trials: how do their results compare? *JAMA* 1996;276: 1332-8.
34. 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.
35. Moayyedi P, Soo S, Deeks J, Delaney B, Harris A, Innes M, et al. Eradication of *Helicobacter pylori* for non-ulcer dyspepsia (Cochrane Review). En: *The Cochrane Library*, Issue 1, 2003. Oxford: Update Software.
36. Bonfill X, Etcheverry C, Martí J, Glutting JP, Urrutia G, Pladevall M. El desarrollo de la Colaboración Cochrane española. *Med Clin (Barc)* 1999;112(Supl 1):17-20.