



Apuntes para la lectura crítica de una revisión sistemática

F. García López* y C. Quereda**

Grupo de Nefrología Basada en la Evidencia de la Sociedad Española de Nefrología. *Unidad de Epidemiología Clínica. Hospital Universitario Puerta de Hierro. Madrid. **Servicio de Nefrología. Hospital Ramón y Cajal. Madrid.

INTRODUCCIÓN

Vamos a detallar los aspectos que consideramos más importantes en la lectura crítica de una revisión, desde la perspectiva de un clínico que tiene que tomar decisiones con respecto a su paciente. Nos centramos en las revisiones sistemáticas y dentro de ellas se incluyen los metanálisis, que son aquellas revisiones sistemáticas que emplean métodos cuantitativos para resumir los resultados. Las revisiones sistemáticas pueden abordar problemas de tratamiento, diagnóstico, causas o pronóstico. Estos apartados son un desarrollo de las conocidas guías de lectura crítica del Grupo de Medicina Basada en la Evidencia y toman mucha información de la declaración QUORUM acerca de la publicación de revisiones.

Un esquema de los conceptos que se desarrollan en este trabajo se resume en la tabla I.

A. ¿SON VÁLIDOS LOS RESULTADOS DEL ENSAYO?

A.1. Preguntas de «eliminación»

La respuesta negativa a estas preguntas supondría descartar, de entrada, la utilidad del artículo analizado, ya que estas preguntas constituyen elementos clave del diseño del trabajo sin los cuales no hay ninguna garantía de la validez de los resultados finales.

1. ¿Se hizo la revisión sobre un tema claramente definido?

Un tema bien definido significa que están claramente explicados tres aspectos importantes, la justifi-

ficación, el objetivo y los componentes básicos de la revisión:

1.1. En primer lugar, la justificación de la revisión. Se deben explicar los antecedentes y la necesidad de la revisión, es decir, los problemas que han dado lugar a ella.

Ante todo, una revisión sistemática debe referir y analizar las revisiones sistemáticas que se hayan realizado previamente sobre el tema, justificando la necesidad de realizar una nueva y, en su caso, comentando las coincidencias y desacuerdos que presenta la nueva revisión en relación a las previas. Las revisiones sistemáticas se buscan en el *Clinical Queries* del *PubMed* y también, específicamente en la base de datos de la colaboración Cochrane (*Cochrane Data Base of Systematic Reviews*).

1.2. En segundo lugar, el objetivo específico de la revisión, es decir, la expresión concreta de la pregunta de investigación que se pretende resolver.

Con frecuencia aparecen revisiones que abordan múltiples problemas, desde la etiología hasta el tratamiento, pasando por el diagnóstico y el pronóstico. Esto es habitual en los libros de texto o en las revisiones narrativas, no sistemáticas. Este tipo de revisiones son útiles como introducción a un tema, pero tienen un valor escaso pues sus conclusiones no se basan en una evaluación sistemática de la evidencia disponible. Las discrepancias habidas entre los resultados de los metanálisis y las recomendaciones efectuadas por expertos en revisiones no sistemáticas ponen de manifiesto las limitaciones de este tipo de revisiones.

Las revisiones sistemáticas se realizan para objetivar la respuesta que ofrece la literatura médica a preguntas clínicas muy concretas, analizando de modo riguroso la pertinencia de los artículos seleccionados y su calidad metodológica. Por tanto, debemos verificar que el objetivo de la revisión ha sido formulado de una forma clara y concisa, que no dé lugar a equívocos y que sea coherente con los elementos justificativos de la revisión que han sido comentados anteriormente.

Correspondencia: Carlos Quereda Rodríguez-Navarro
Hospital Ramón y Cajal
Ctra. de Colmenar Viejo, s/n
28032 Madrid
E-mail: cquereda.hrc@salud.madrid.org

Tabla I. Guía para la lectura crítica de una revisión sistemática

A) ¿Son válidos los resultados del ensayo?

A.1. Preguntas de «eliminación»:

1. *¿Se hizo la revisión sobre un tema claramente definido?*
 - Justificación de la revisión
 - Objetivo específico de la revisión
 - Componentes básicos del ensayo:
 - pacientes
 - * criterios de selección
 - intervención
 - comparación
 - desenlace o resultado
 - * variables de resultado primarias y secundarias
 - * definición de las variables de resultado
 - * ¿variables de resultado subrogadas?
 - * ¿variables de resultado combinadas?
2. *¿Fueron adecuados los criterios empleados para la selección de artículos?*
 - Los estudios deben abordar la cuestión objeto de la revisión
 - Criterios para la selección en una revisión de:
 - tratamiento
 - diagnóstico
 - etiología
 - pronóstico

A.2. Preguntas de detalle:

3. *¿Se incluyeron todos los estudios relevantes e importantes?*
 - Búsqueda en repertorios bibliográficos
 - Búsqueda manual de referencias, revistas y libros de resúmenes
 - Contacto con expertos
 - Recuperación de estudios no publicados
 - Estudios en idiomas distintos al inglés
 - Otras restricciones de la búsqueda
4. *¿Se evaluó la validez de los estudios incluidos?*
 - Evaluación de la calidad de los estudios
5. *¿Fueron reproducibles las evaluaciones de los estudios?*
 - Concordancia entre varios evaluadores
6. *¿Fueron similares los resultados en los distintos estudios de la revisión?*
 - Heterogeneidad entre los estudios
7. *¿Quién financió la revisión?, ¿se describió el papel del promotor en su diseño, ejecución, análisis y publicación?*

B) ¿Cuáles son los resultados?

8. *¿Cuáles fueron los resultados globales de la revisión?*
 - Características de los estudios
 - Diagrama de flujo de los estudios
 - Medidas de efecto
 - Modelos de efectos fijos y de efectos aleatorios
 - Análisis de sensibilidad
 - Análisis de subgrupos
 - Evaluación del sesgo de publicación
 - Descripción de acontecimientos adversos
 - Medida de beneficio o daño para un paciente concreto
9. *¿Fueron precisos los resultados?*
 - Intervalos de confianza
 - Razones de la imprecisión de los resultados
 - Valores de P frente a intervalos de confianza

C) ¿Pueden ayudarnos estos resultados?

10. *¿Se pueden aplicar estos resultados en nuestro medio?*
 - Tipo de pacientes y ámbito de los estudios de la revisión
 - Viabilidad de la intervención en nuestro medio
 11. *¿Se tuvieron en cuenta todos los resultados de importancia clínica?*
 12. *¿Superan los beneficios a los riesgos y costes?*
 - balance del beneficio y del daño en un paciente concreto
 - preferencias y valores del paciente
-

1.3. A continuación, debemos verificar cuáles son las características de los pacientes y de las intervenciones estudiadas. Se trata de los cuatro componentes básicos de la mayoría de las investigaciones clínicas, según las siglas PICO, de Pacientes, Intervención, Comparación y *Outcome* —desenlace o resultado—:

1.3.1. Pacientes: criterios de selección, es decir, de inclusión (el diagnóstico clínico, aunque puede estar matizado por el estadio del cuadro) y de exclusión (por ejemplo, criterios de gravedad).

1.3.2. Intervención: tipo de intervención, terapéutica, diagnóstica, exposición potencialmente nociva, o la característica pronóstica de interés, para los grupos experimentales, con posible especificación del momento y modo de administración y de su duración, si procede.

1.3.3 Comparación: en muchos estudios se compara un grupo sometido a la intervención con un grupo control, que sirve de comparación. En este grupo se deben detallar las mismas características antes referidas a la intervención, pero referidas al grupo control. En particular, tipo de comparación (placebo, tratamiento activo en estudios de tratamiento, otras pruebas diagnósticas o el patrón de referencia en los estudios diagnósticos,...).

1.3.4. Desenlace o resultado: el suceso o enfermedad que la intervención del ensayo se propone tratar, mejorar, retrasar o prevenir; o el diagnóstico definitivo en el caso de una prueba diagnóstica; o la enfermedad que es presunta consecuencia de la exposición; o el desenlace final de un estudio pronóstico.

Deben indicarse las variables de resultado primarias y secundarias; la primaria es la que posee mayor importancia en la revisión.

Los criterios de definición de todas estas características deben explicarse con suficiente claridad.

La exposición previa de los criterios de inclusión reduce el sesgo de selección de citas que se produce en las revisiones narrativas y en la sección de discusión en los artículos originales, en donde la evidencia muestra que los autores tienden a citar preferentemente a aquellos autores que corroboran sus propias conclusiones.

En este punto, se debe analizar si las características anteriormente mencionadas (selección de pacientes, intervención, comparación y variables de resultado) son las adecuadas para responder a la pregunta formulada en los objetivos y si los estudios seleccionados contienen esos elementos con un nivel de homogeneidad clínica suficiente para permitir que la respuesta sustentada en el conjunto de resultados (y, en su caso, en el análisis de la agregación de los mismos mediante un metanálisis) tenga rigor y, por tanto, validez.

Por otra parte, el clínico que encuentra una revisión sistemática se debe preguntar si esa revisión plantea la misma cuestión que quiere resolver del paciente que tiene a su cargo o, por el contrario, plantea una cuestión distinta.

2. *¿Fueron adecuados los criterios empleados para la selección de artículos?*

Antes que nada, los artículos seleccionados tienen que abordar la pregunta objeto de la revisión.

Además, es necesario escoger el diseño apropiado de los artículos seleccionados, que depende del tipo de problema que aborde la revisión (tratamiento, diagnóstico, etiología o pronóstico):

2.1. Si la revisión aborda una cuestión de tratamiento, los artículos de la selección han de ser ensayos controlados con asignación aleatoria con un análisis por intención de tratar.

2.2. Si la revisión aborda una cuestión de diagnóstico, los artículos de la selección deben comparar la prueba diagnóstica con una prueba estándar de referencia («patrón oro») y la comparación entre los resultados de las dos pruebas debe ser enmascarada e independiente. Además, los pacientes incluidos en el estudio deben abarcar el espectro completo del tipo de pacientes en quienes se aplicaría la prueba en la práctica clínica.

2.3. Si la revisión aborda una cuestión de etiología o efectos nocivos, los artículos deben determinar con claridad el grupo de casos y el grupo de controles así como las exposiciones y los resultados, que deben obtenerse de la misma forma en los dos grupos.

2.4. Si la revisión aborda una cuestión de pronóstico, los artículos deben estudiar a pacientes en un momento similar en el transcurso de su enfermedad, el tiempo de seguimiento debe ser suficientemente prolongado y la verificación del seguimiento debe ser bastante completa.

Por tanto, una revisión debe establecer de antemano tanto los criterios de inclusión y de exclusión de los artículos que va a seleccionar.

A.2. Preguntas de detalle

3. *¿Se incluyeron todos los estudios relevantes e importantes?*

Los autores tuvieron que efectuar una búsqueda exhaustiva de los estudios que cumplen los criterios

de inclusión. Esta búsqueda exhaustiva incluye lo siguiente:

3.1. Una búsqueda bibliográfica en los distintos repertorios: MEDLINE y EMBASE, con sus correspondientes estrategias de búsqueda orientadas a recuperar el máximo de referencias con objeto de no perder ninguna; si lo que se buscan son ensayos controlados aleatorizados hay que buscar en el *Cochrane Central Register of Controlled Trials* (CENTRAL), que contiene casi 500.000 ensayos; hay, además, repertorios específicos para enfermería CINAHL, psicología (PSYCLIT, PSYCINFO, PSYINDEX), literatura gris en Europa (SIGLE), para ámbitos geográficos (LILACS, literatura en Ciencias de la Salud publicada en los países de América Latina y el Caribe o el Índice Médico Español —IME—, de autores españoles). Si el objetivo de la revisión lo requiere, los investigadores deberían haber buscado en alguna de estas bases de datos.

3.2. La búsqueda bibliográfica debe complementarse con la búsqueda manual de las referencias de los artículos seleccionados en los repertorios bibliográficos, así como la búsqueda manual de las revistas en donde es más probable que se hayan publicado estudios similares y la búsqueda en los libros de resúmenes de los congresos en donde se hayan podido presentar comunicaciones.

3.3. Contacto con expertos en el campo de estudio para rastrear trabajos no incluidos en las búsquedas anteriores, bien por su relevancia o por estar todavía en fase de publicación o bien para identificar posibles estudios no publicados.

3.4. Decidir si incluir estudios no publicados. La no inclusión de un estudio no publicado puede ocasionar lo que se llama «sesgo de publicación», que se produce cuando los resultados de los estudios publicados difieren de los de los estudios no publicados, por lo general porque los estudios publicados suelen tener resultados positivos con mayor frecuencia que los no publicados. Al no tener en cuenta este hecho, la combinación de los resultados puede sobreestimar las medidas de efecto. Por otra parte, los estudios no publicados suelen ser más pequeños y tienden a ser de peor calidad que los publicados, por lo que su inclusión puede incorporar sesgos no deseados. Si se incluyen estudios no publicados se debe obtener el máximo de información sobre ellos. Hay técnicas estadísticas gráficas que ayudan a explorar la presencia de sesgo de publicación.

3.5. Otro aspecto importante en la búsqueda de artículos es el idioma de su publicación. Por ejemplo, en el campo de los ensayos clínicos, se sabe que los ensayos clínicos publicados en idiomas distintos al inglés tienden a tener resultados más favorables a la intervención que los ensayos publicados en inglés, quizá porque su calidad y su tamaño de muestra es menor. No se sabe si este fenómeno es suficientemente importante para influir sobre las medidas finales del metanálisis.

3.6. Se deben conocer otras restricciones habidas en las búsquedas, como los años de búsqueda. El MEDLINE está en vigor desde 1966, pero muchas búsquedas no necesitan remontarse tan lejos, pues se centran en intervenciones más recientes.

4. ¿Se evaluó la validez de los estudios incluidos?

Esta pregunta tiene que ver con la calidad de los estudios incluidos. El problema que plantean los artículos que componen una revisión sistemática es su validez. La inclusión de un artículo de calidad cuestionable puede afectar al conjunto de la revisión, sobre todo si se efectúa un metanálisis.

En este sentido, en el campo de los ensayos clínicos hay pruebas de que la inclusión de ensayos de baja calidad, por ejemplo aquellos sin aleatorización oculta, puede sobreestimar la medida de efecto de un metanálisis, ya que los ensayos de peor calidad tienden a tener resultados más favorables a las intervenciones en estudio.

No hay acuerdo sobre cómo evaluar la calidad de los artículos originales en una revisión sistemática: se puede hacer mediante puntuaciones, listas-guía o mediante elementos aislados. En una revisión sistemática de ensayos clínicos se emplea la escala de Jadad, listas-guía o el cumplimiento de tres requisitos básicos (ocultación de la distribución aleatoria, análisis por intención de tratar y enmascaramiento de la intervención).

Para el caso de los ensayos controlados aleatorizados, la Colaboración Cochrane recomienda tomar en cuenta cuatro posibles sesgos: el sesgo de selección, cuando las características iniciales de los participantes de los grupos de estudio son distintas, que se previene mediante la ocultación de la distribución aleatoria, primer criterio de calidad; el sesgo de ejecución, que se produce cuando hay diferencias sistemáticas en la atención prestada a los participantes del grupo control con respecto al grupo intervención, que se previene mediante el enmascaramiento de los proveedores de la asistencia sanitaria y de los pacientes, segundo criterio de calidad;

el sesgo de desgaste, que se produce cuando los abandonos son distintos en el grupo control que en el grupo intervención, que se corrige cuando se puede efectuar un análisis estadístico por tratamiento asignado o por intención de tratar, tercer criterio de calidad; y el sesgo de detección, cuando se producen diferencias en la evaluación del desenlace en los dos grupos, que se previene con el enmascaramiento del evaluador, cuarto criterio de calidad. Con arreglo a la presencia de estos elementos, se clasifican los ensayos en uno de los siguientes tres grupos: A) riesgo bajo de sesgo, cuando se cumplen todos los criterios mencionados; B) riesgo moderado de sesgo, cuando se cumplen parcialmente uno o varios criterios; C) riesgo elevado de sesgo, cuando no se cumple uno o más criterios. Los autores de la revisión deberían obtener información adicional de los autores de los ensayos si no queda claro el cumplimiento de alguno de estos criterios.

Sea cual sea el método de evaluación de calidad, los autores deben comparar los resultados de aquellos estudios con calidad considerada alta con los de los estudios con calidad considerada baja, con objeto de apuntar posibles diferencias en las medidas de efecto que puedan deberse a los sesgos de los estudios con calidad inferior.

5. *¿Fueron reproducibles las evaluaciones de los estudios?*

La selección de estudios, la evaluación de su calidad y la extracción de datos son procesos sometidos a subjetividad. Por tanto, los instrumentos empleados para ello deben ser lo más válidos posibles. En ausencia de criterios objetivos para establecer su validez, se suele recurrir a que dos evaluadores hagan cada uno el trabajo por su cuenta para después contrastar sus resultados y resolver sus diferencias, ya sea por acuerdo entre ellos o por el concurso de un tercer observador. En ese sentido, la medida del acuerdo entre los dos observadores puede suponer una medida indirecta de la validez.

6. *¿Fueron similares los resultados en los distintos estudios de la revisión?*

Las revisiones reúnen estudios que, aunque tengan en común una serie de características pertinentes al objetivo del estudio, difieren en otras muchas, como pueden ser la gravedad o comorbilidad de los pacientes, su ámbito geográfico y social, la intensidad de la intervención en duración, dosis o medicamento concreto, y otros factores. Algunas de esas

diferencias pueden ser lo suficientemente grandes para hacer poco razonable un metanálisis, la combinación de los resultados de todos los estudios en una medida sumaria.

Si se quiere combinar los resultados de los distintos estudios, estos se deben expresar gráficamente, mediante el diagrama del bosque (*forest plot*), que muestra la medida de efecto de cada estudio, con sus intervalos de confianza. Mediante ese gráfico se puede ver a grandes rasgos si los resultados concuerdan o difieren entre sí.

Hay procedimientos estadísticos para estudiar el grado de concordancia de los resultados de los estudios, u homogeneidad, en términos estadísticos. En concreto, se emplean dos: la prueba de la homogeneidad, que aporta un valor de P, que indica la probabilidad de encontrar las diferencias halladas de los resultados de los estudios o diferencias más extremas si todos los estudios fueran muestras aleatorias de una misma distribución; y el índice I^2 , más apropiado que la prueba de homogeneidad, porque su valor no depende del número de estudios y porque es una prueba más sensible, pues la prueba de homogeneidad puede dar un valor de P mayor de 0,05 cuando los estudios no son homogéneos.

Cuando hay heterogeneidad se plantea la duda de si tiene sentido efectuar un metanálisis cuando los estudios dan resultados distintos. Algunos autores desaconsejan por completo el metanálisis cuando hay heterogeneidad pues, argumentan, se infringe su presunción estadística básica, es decir, que, empleando términos estadísticos, todos los estudios constituyen variables aleatorias de una distribución común; si hay heterogeneidad, no hay distribución común y no se puede buscar ninguna medida común. Otros autores argumentan que la presunción básica no se cumple nunca, que siempre hay cierto grado de heterogeneidad y proponen métodos estadísticos que tienen en cuenta una cierta heterogeneidad.

En cualquier caso, siempre que se encuentra una cierta discordancia entre los resultados de los estudios, los autores deben intentar explorar sus causas, qué diferencias en las características de los pacientes, de la intervención o del desenlace pueden explicar las diferencias, con objeto de formular hipótesis para futuros estudios.

7. *¿Quién financió la revisión?, ¿se describió el papel del promotor en su diseño, ejecución, análisis y publicación?*

Los intereses económicos de las compañías con interés comercial, al poder colisionar con la necesi-

dad científica de divulgar resultados que pueden ser perjudiciales para sus productos, puede dar lugar a sesgos en el proceso de desarrollo de una revisión sistemática. Dentro de la lectura de una revisión se necesita disponer de información acerca de:

- quién fue el promotor que financió la revisión;
- los conflictos de intereses de los autores de la publicación;
- el grado de independencia de los investigadores con respecto al promotor en las distintas fases de la revisión: diseño, ejecución y recogida de datos, análisis estadístico y publicación de resultados.

Algunas revistas médicas exigen información detallada sobre estos apartados a los autores de los manuscritos que se aceptan para publicación.

B. ¿CUÁLES SON LOS RESULTADOS?

8. ¿Cuáles fueron los resultados globales de la revisión?

8.1. Antes de conocer los resultados globales de la revisión, se deben conocer las características de los estudios que componen la revisión: tipo de diseño, características de los participantes (número, variables de interés como edad, sexo o enfermedades asociadas), pormenores de la intervención y de la comparación (dosis, duración), tiempo de seguimiento, definiciones de los desenlaces, y otras características que sean relevantes.

8.2. Es útil un diagrama de flujo que explique cuántos estudios se identificaron inicialmente, cuántos se seleccionaron, cuántos se excluyeron y por qué, cuántos se emplearon en el análisis.

8.3. Si los autores decidieron efectuar un metanálisis, éste aporta una medida de efecto sumaria de los estudios, que depende del tipo de revisión. Para las revisiones de tratamiento, la medida de efecto puede ser un riesgo relativo, una razón de posibilidades (*odds ratio*) o la diferencia de riesgos si las variables de resultado son dicotómicas o una diferencia de medias ponderada o estandarizada si la variable de resultado es una variable cuantitativa. Para las revisiones diagnósticas, las medidas de efecto son la razón de verosimilitudes, una curva ROC sumaria, la sensibilidad, especificidad y *odds ratio* diagnóstica. Para un estudio etiológico, la medida de efecto es la razón de posibilidades. Para un estudio pronóstico, la medida de efecto es una razón de riesgos o una razón de posibilidades.

Estas medidas de efecto se ponderan según la importancia de cada estudio, cuyo peso, por lo general, viene dado por el inverso de la variancia, que tiene que ver con el tamaño de muestra, de modo que aquellos estudios con un tamaño de muestra mayor son los que aportan una contribución mayor en la ponderación final. A diferencia del metanálisis, en las revisiones narrativas se tiende a clasificar los estudios en dos grupos, los que dan resultados favorables y los que los dan desfavorables, sin ninguna consideración a su importancia relativa.

8.4. Hay dos procedimientos estadísticos para hacer metanálisis: el modelo de efectos fijos y el modelo de efectos aleatorios. Aquél parte de la presunción estadística básica del metanálisis de que todos los estudios son variables aleatorias de una distribución común. El modelo de efectos aleatorios rechaza esa presunción y supone que hay cierta heterogeneidad entre los estudios. En la práctica, si no hay heterogeneidad llamativa entre los estudios los dos modelos arrojan resultados comparables. Muchas revisiones describen los resultados de los dos modelos.

8.5. En la mayoría de las revisiones se efectúa un «análisis de sensibilidad», es decir, análisis separados de estudios según compartan determinadas características de calidad, pacientes, intervención o desenlace. El propósito de estos análisis es explorar posibles diferencias en los resultados que pudieran ser atribuibles a las diferencias en dichos factores.

8.6. Los análisis de subgrupos, muy tentadores para conocer los resultados específicos para determinadas características, deberían estar planificados de antemano. Son análisis que están sujetos a sesgos y sus resultados muchas veces no tienen más mérito que el de formular hipótesis para futuros estudios. Por tanto, los análisis de subgrupos *post-hoc* o *a posteriori* tienen escaso crédito y pueden ser engañosos. En ningún caso se pueden hacer subgrupos según el riesgo de los pacientes de presentar los sucesos, pues se producen sesgos importantes; únicamente se deben hacer según características iniciales de los pacientes. Los metanálisis que emplean datos de pacientes individuales son más fructíferos en los análisis de subgrupos predefinidos. Otra técnica que se emplea para explorar diferencias entre subgrupos es la meta-regresión, en donde se evalúan los desenlaces con respecto a determinadas características de los pacientes con técnicas de regresión lineal.

8.7. Si hay sospecha de «sesgo de publicación», la revisión debería explicar los procedimientos empleados para su evaluación.

8.8. Los resultados deben incluir una descripción de los acontecimientos adversos por grupo, con especial énfasis en los acontecimientos adversos graves.

8.9. Al igual que en los ensayos clínicos se intenta obtener el número necesario para tratar (NNT) para tener una estimación del beneficio (o perjuicio) potencial de una intervención para un paciente determinado, en los metanálisis de tratamiento también se puede intentar obtener una estimación del NNT, ya sea para obtener un beneficio (NNTB), o para obtener un daño o perjuicio (NNTD, de *harm*). El NNT promedio se puede calcular a partir del riesgo relativo o de la razón de posibilidades descritos en los resultados (ver calculadora de NNT en un metanálisis, http://www.senefrobe.org/herramientas_publico.asp). Sin embargo, no se puede tomar este NNT promedio como el NNT aplicable a nuestro paciente, porque, a diferencia de las medidas relativas como el riesgo relativo o la razón de posibilidades, que suelen ser invariables con respecto al riesgo previo, el NNT depende del riesgo previo de presentar el suceso que tiene el paciente y ese riesgo puede diferir según los estudios incluidos en la revisión. A mayor riesgo previo, menor es el NNT y, por tanto, mayor beneficio relativo. Por el contrario, cuando el riesgo previo es pequeño, el NNT aumenta y el beneficio se reduce.

¿Cómo se calcula el riesgo de nuestro enfermo? Se puede aplicar uno de los siguientes tres procedimientos:

- en primer lugar, si hay un análisis de subgrupos, el clínico puede extrapolar a su paciente el riesgo del subgrupo más afín. Los sesgos potenciales de los análisis de subgrupos desaconsejan este enfoque.

- en segundo lugar, el clínico puede obtener el riesgo a partir de análisis multivariantes de estudios de cohorte en pacientes similares al suyo, en estudios pronósticos, cuando existen.

- por último, el clínico puede emplear su juicio y dividir el NNT promedio obtenido en el metanálisis por un factor f que relaciona el riesgo de nuestro paciente en relación con el riesgo promedio del metanálisis. Dicho factor f se expresa como una cifra con un decimal, de modo que si el paciente tiene un riesgo menor que el promedio $f < 1$ y si el riesgo es mayor, $f > 1$. Por ejemplo, si el NNT promedio es de 34 y consideramos que nuestro paciente tiene el doble de riesgo, $f = 2$ y $NNT = 17$; si consideramos que nuestro paciente tiene la mitad de riesgo, $f = 0,5$ y $NNT = 68$.

9. ¿Fueron precisos los resultados?

9.1. La precisión de los resultados la aportan sus intervalos de confianza, que informan del grado de

incertidumbre de los resultados. Por ejemplo, el intervalo de confianza del 95% da el recorrido de incertidumbre que incluiría el valor verdadero en 95 veces entre 100 estudios similares. Las medidas de efecto de todas las variables de resultado deberían ir acompañadas de sus intervalos de confianza.

9.2. Razones de la imprecisión de los resultados: cuando los intervalos de confianza son anchos, se debe reflexionar acerca de las razones de ello. El principal motivo en un metanálisis es su escasa potencia estadística, que se produce cuando el número de estudios es escaso o el tamaño de muestra de sus estudios es pequeño. También puede ocurrir cuando el tiempo de seguimiento es reducido o cuando la medida de la variable de resultado es imprecisa *per se*.

9.3. Los resultados de un metanálisis se pueden expresar también mediante la significación estadística, expresada mediante los valores de probabilidad o valores de P. El valor de P indica la probabilidad de que los datos observados, o un resultado más extremo, aparezcan por azar cuando las intervenciones no difieren entre sí. Sin embargo, los valores de P no expresan la magnitud de efecto y dependen poderosamente del tamaño de la muestra. Aisladamente, los valores de P no reflejan importancia o falta de importancia clínica. Aunque hay una correspondencia entre la significación estadística de los valores de P y los límites del intervalo de confianza del 95%, el intervalo de confianza aporta una información clínicamente más relevante y más fácil de interpretar. Muchos metanálisis omiten la descripción de los valores de P y sólo los presentan para las pruebas de homogeneidad.

C. ¿PUEDEN AYUDARNOS ESTOS RESULTADOS?

10. ¿Se pueden aplicar estos resultados en nuestro medio?

Esta cuestión tiene que ver con la validez externa de la revisión. Determinada información recogida en el estudio, como las características de elegibilidad de los pacientes, el ámbito y lugar donde se recogieron los datos, las características de la intervención, la definición y evaluación de las variables de resultado, la representatividad de los casos reclutados sobre el total de casos potenciales, la duración del tratamiento, o el tiempo de seguimiento, ayudan a ver si los resultados de la revisión serían aplicables a nuestros pacientes. Cuando nuestros pacientes o nuestro ámbito no coinciden en características

demográficas, de comorbilidad u otras, la pregunta es: ¿se pueden extender los resultados de esta revisión a nuestra población a pesar de las diferencias? ¿se pueden extender los resultados de esta revisión a situaciones con mayor tiempo de seguimiento o mayor duración del tratamiento?

Otro aspecto importante, que se aplica más a intervenciones terapéuticas y diagnósticas, es la viabilidad de la intervención en nuestro ámbito, en términos de disponibilidad, coste, formación y recursos.

11. ¿Se tuvieron en cuenta todos los resultados de importancia clínica?

Los resultados de importancia clínica abarcan desde las variables clínicas importantes que el estudio pretende prevenir, hasta acontecimientos adversos, junto con otros aspectos como la calidad de vida relacionada con la salud, la satisfacción del paciente o los costes económicos. La cuestión que se plantea es: en caso de que no se hayan tenido en cuenta todos los resultados de importancia clínica, ¿en qué medida afecta a la aplicación hipotética de los resultados a nuestra práctica clínica?

12. ¿Superan los beneficios a los riesgos y costes?

Aunque el propósito de una revisión es establecer exhaustivamente las pruebas existentes sobre una determinada cuestión su interpretación también necesita conocer qué otros estudios previos han abordado esa cuestión u otras directamente relacionadas.

En un paciente particular los resultados de una revisión se deben expresar en términos que incluyen su riesgo inicial y los efectos terapéuticos, que quedan mejor expresados mediante la medida del número necesario para tratar (NNT), tanto para obtener un beneficio (NNTB) como para obtener un perjuicio o daño (NNTH). La mayoría de las intervenciones que aportan un beneficio conllevan un perjuicio añadido en forma de un aumento de la incidencia de determinados acontecimientos adversos, de modo que, en muchas ocasiones, se pueden calcular los NNTB y NNTH y sopesarlos, tanto en general, como en pacientes particulares.

En el balance final de beneficios y riesgos, antes de tomar una decisión se debe tener en cuenta otro aspecto importante, la incorporación a la decisión de las preferencias y los valores del propio paciente. Hay pruebas de que la participación activa del paciente en la toma de decisiones mejora su calidad de vida y sus resultados. La aplicación a la práctica de este aspecto es un campo de investigación actual que va a dar muchos frutos en el futuro.

REFERENCIAS

- Oxman AD, Cook DJ, Guyatt GH: Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA* 272: 1367-1371, 1994.
- Egger M, Davey Smith D, Altman DG (eds): Systematic reviews in health care. Meta-analysis in context. 2ª edición. Londres: BMJ books, 2001.
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB: Medicina basada en la evidencia: cómo ejercer y enseñar la MBE. Madrid: Churchill Livingstone, 1997.
- García López FJ, Gutiérrez Bezoán S, Galende Domínguez I, Avendaño Solá C: Evaluación de calidad de los ensayos clínicos: justificación, utilidad e inconvenientes. *Medicina Clínica* (Barcelona) 112 (Supl. 1): 35-42, 1999.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF: Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 354: 1896-1900, 1999.
- Guyatt G, Rennie D. Users' guides to the medical literature: essentials of evidence-based clinical practice. Chicago: AMA Press, 2002.
- Guyatt G, Rennie D. Users' guides to the medical literature: a manual for evidence-based clinical practice. Chicago: AMA Press, 2002.
- Furukawa TA, Guyatt GH, Griffith LE: Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol* 31: 72-76, 2002.
- Higgins JP, Thompson SG: Quantifying heterogeneity in a meta-analysis. *Stat Med* 21: 1539-1558, 2002.
- McQuay HJ, Moore RA: Using numerical results from systematic reviews in clinical practice. *Ann Intern Med* 126: 712-720, 1997.

REFERENCIAS GENERALES

- La declaración QUORUM. Improving the quality of reports of meta-analyses or randomized controlled trials: the QUORUM statement checklist. <http://www.consort-statement.org/QUOROM.pdf>
- Manual Cochrane de Revisiones Sistemáticas de Intervenciones. <http://www.cochrane.es/?q=es/handbook>
- CASPe (*Critical Appraisal Skills Program* en España): programa de habilidades en lectura crítica, España. <http://www.redcaspe.org/herramientas/index.htm>
- Centre for Evidence-Based Medicine, Oxford-Centre for Evidence Based Medicine. The EBM toolbox. <http://www.cebm.net/toolbox.asp>
- Centre for Evidence-Based Medicine, Oxford-Centre for Evidence Based Medicine. Doing EBM. http://www.cebm.net/using_ebm.asp
- User's guides interactive: an online tool to guide clinicians in the appraisal and application of evidence into their everyday practice. <http://www.usersguides.org/>
- Modelo para la realización del resumen estructurado de una revisión sistemática. http://www.senefrobe.org/herramientas_publico.asp
- Calculadora de NNT en un metanálisis. http://www.senefrobe.org/herramientas_publico.asp
- Enseñar MBE: herramientas y calculadoras interactivas. <http://www.infodoctor.org/rafabravo/herramientas.htm>
- Cucherat M. Méthodologie e interprétation des essais thérapeutiques. <http://www.spc.univ-lyon1.fr/polycop/>