

# GUÍA PARA LA LECTURA DE ARTÍCULOS DE REVISIÓN<sup>1</sup>

Andrew D. Oxman<sup>2</sup> y Gordon H. Guyatt<sup>3</sup>

*Una manera de enfrentar la masa inmensa de publicaciones médicas es confiar en las revisiones de la literatura. El método es eficiente, pero una revisión que no satisface los criterios científicos puede descarriar al lector. Los criterios que se dan en este artículo pretenden ayudar a los lectores a valorar la calidad científica de los trabajos de revisión. Los criterios tienen que ver con la definición del tema, la exhaustividad de la estrategia de investigación, los procedimientos para seleccionar y valorar los estudios primarios y los métodos para combinar los resultados y obtener conclusiones apropiadas. La aplicación de estos criterios permite que los clínicos inviertan su valioso tiempo de lectura en material de alta calidad, y que puedan juzgar la validez de las conclusiones de un autor.*

Los clínicos que intentan mantenerse actualizados deben encontrar maneras de enfrentarse a un volumen de publicaciones que crece exponencialmente. Se han publicado métodos eficientes para hallar y conservar los estudios interesantes (1-6) y desechar los estudios inválidos o inaplicables (7-12). Sin embargo, revisar toda la literatura para responder a una cuestión planteada en el ejercicio de la medicina lleva mucho tiempo, y no es factible que los clínicos lean toda la literatura primaria relacionada con cada una de las muchas dificultades que tienen que afrontar en la práctica día tras día.

Una solución a este problema es el artículo de revisión o de repaso general, en el que se examinan y resumen los estudios primarios relevantes para una cuestión clínica. Sin embargo, las revisiones, igual que los estudios primarios, deben ser leídas de forma crítica y selectiva. Una metodología defectuosa en una investigación sobre procedimientos diagnósticos o terapéuticos puede invalidar los resultados; lo mismo ocurre cuando se extraen conclusiones incorrectas de una revisión científica de la literatura. Los autores de revisiones recogen y analizan datos de investigaciones primarias, pero a veces lo hacen de manera subjetiva y subconsciente. La diferencia fundamental entre una revisión y un estudio primario es la unidad de análisis, no los principios científicos que se aplican.

Lo que sigue son cinco recomendaciones contradictorias para el tratamiento de la hipertensión leve, todas ellas publicadas.

<sup>1</sup> Esta traducción del artículo "Guidelines for reading literature reviews" (*Canadian Medical Association Journal*, 1988;138:697-703) se publica con autorización de los autores y de la Canadian Medical Association.

<sup>2</sup> Departamento de Epidemiología Clínica y Bioestadística y Departamento de Medicina, McMaster University, Hamilton, Ontario, Canadá.

<sup>3</sup> Ministerio de Salud de Ontario, Canadá. Las separatas de la versión original en inglés de este artículo deben pedirse a: Dr. Gordon H. Guyatt, McMaster University Health Sciences Centre, 3H7-1200 Main St. W, Hamilton, Ont. L8N 3Z5, Canadá.

□ Los datos disponibles ... llevaron al revisor a concluir que el tratamiento de la hipertensión leve [90 a 104 mm Hg] hasta conseguir tensiones diastólicas de menos de 90 mm Hg es la política apropiada de salud pública, basada en las pruebas más actuales (13).

□ La mayoría de los pacientes con tensión arterial diastólica entre 90 y 104 mm Hg deben ser tratados, a menos que haya contraindicaciones para la medicación... En algunos pacientes debe intentarse una modificación drástica dietética y conductual antes de comenzar un tratamiento farmacológico, o como medida complementaria de este (14).

□ El tratamiento no farmacológico con frecuencia es efectivo en la hipertensión leve. La elección inicial entre tiazidas y betabloqueantes adrenérgicos a menudo depende de la preferencia personal del médico... Con precaución, los riesgos del tratamiento hipotensor son considerablemente menores que los beneficios (15).

□ Los beneficios del tratamiento farmacológico para pacientes con hipertensión leve [tensión arterial diastólica entre 90 y 105 mm Hg] siguen sin demostrarse. El tratamiento no medicamentoso tampoco ha sido suficientemente investigado (16).

□ Por el momento, en relación con los tratamientos a base de diuréticos que son los que se han estudiado en ensayos clínicos, el tratamiento de la hipertensión leve o moderada [tensión arterial diastólica de menos de 115 mm Hg] solo tiene valor directamente demostrado cuando la tasa de accidentes cerebrovasculares es lo suficientemente elevada (quizá por la edad o por la existencia de enfermedad cerebrovascular) para que reducirla a la mitad justifique el costo y la molestia del tratamiento... Es posible que los hipotensores que no afectan a los lípidos tengan más efecto sobre los infartos de miocardio que sobre los accidentes cerebrovasculares. Pero en los ensayos clínicos revisados la reducción de la tasa de infartos de miocardio es incierta [Rory Collins, observaciones inéditas, 1987].

Si no se tiene algún criterio para valorar las revisiones de donde se han tomado estas recomendaciones, decidir qué revisión hay que creer es como decidir qué dentífrico usar. Es una cuestión de gusto más que de ciencia.

No hay que ir muy lejos para encontrar otros ejemplos de revisiones recientes sobre cuestiones clínicas importantes, en las que se ha llegado a diferentes conclusiones: ¿Debe evitarse la administración de corticosteroides por la posibilidad de osteoporosis clínicamente importante? (17, 18) ¿Qué ventajas tiene la cateterización cardíaca derecha en pacientes en estado crítico? (19, 20) ¿Hay que tratar activamente la hipopotasemia leve? (21, 22).

Está claro que la competencia científica del autor no es criterio suficiente para la credibilidad de una revisión, ya que es frecuente que distintos expertos lleguen a distintas conclusiones al revisar el mismo tema. Tampoco es criterio suficiente el prestigio de la revista que publica la revisión. Estudios recientes de la literatura médica han mostrado que la calidad científica de la mayor parte de las revisiones que se publican, incluso en las revistas más consideradas, es bastante baja (23-27).

En este artículo se presenta una serie de criterios para que los lectores valoren los artículos de revisión. Se han sugerido anteriormente criterios similares, en particular en la literatura de la psicología y las ciencias sociales (28-30). Nos cen-

traremos especialmente en cómo puede decidir el lector si vale la pena leer una revisión y si es apropiado creer sus conclusiones. Nuestras orientaciones también pueden ser útiles para quienes estén proyectando escribir un artículo de revisión.

## Criterios

Hemos concretado nuestros criterios en una serie de preguntas (cuadro 1). Antes de ver cada criterio en detalle, vale la pena hacer algunos comentarios generales. En primer lugar, las preguntas están pensadas para valorar las revisiones generales de estudios primarios sobre cuestiones pragmáticas. En segundo lugar, el término "estudios primarios" se refiere a los informes de investigación que contienen la información original sobre la que se basa la revisión. En tercer lugar, la intención de estos criterios es promover un uso eficiente de la literatura médica y un escepticismo sano; no se trata de apoyar el nihilismo. Los lectores que apliquen estos criterios podrán comprobar que la mayor parte de los trabajos de revisión publicados tienen deficiencias científicas importantes (23–27). Claro está que cuando se ha estudiado la calidad científica de los informes de investigaciones médicas se ha visto que la mayor parte de los estudios primarios también tienen deficiencias científicas considerables.

Es necesario mejorar el diseño, la realización y la redacción de los informes, tanto de trabajos de revisión como de estudios primarios. Sin embargo, existen enormes cantidades de información valiosa y, para tomar decisiones juiciosas en la práctica clínica, hay que basarse en las investigaciones disponibles. La mayor parte de las revisiones publicadas no dan demasiado apoyo a sus conclusiones, pero los lectores con capacidad crítica pueden discernir qué información es útil y hacer sus propias inferencias, que serán o no las mismas que las de los autores.

### ¿Se exponen claramente los interrogantes y los métodos?

Ante un artículo de revisión lo primero que debe decidir el lector es si el trabajo trata un tema relevante para su práctica clínica o si es de su interés. Por lo tanto, es necesario que se exponga claramente de qué trata el artículo.

Cualquier interrogante causal tiene tres elementos fundamentales: la población, la exposición o la intervención y el resultado. El cuadro 2 presenta ejemplos de estos elementos en cinco campos clave de la investigación clínica. Para que el lector pueda decidir cuanto antes si el artículo le interesa, es necesaria una exposición clara de la pregunta considerada, con especificación de los tres elementos. Si la pregunta o preguntas que se pretenden contestar no están claramente especificadas, quizá el lector haga bien en dejar de leer. Las preguntas vagas suelen llevar a respuestas vagas.

#### CUADRO 1. Criterios para evaluar artículos de revisión

- 
- ¿Se especificaron claramente las preguntas a las que se pretende responder y los métodos del estudio?
  - ¿Se utilizaron métodos sistemáticos para localizar los artículos relevantes?
  - ¿Se especificaron los métodos utilizados para decidir qué artículos se incluirían en la revisión?
  - ¿Se evaluó la validez de los estudios primarios?
  - ¿Fue reproducible y exento de sesgos el método de evaluación de los estudios primarios?
  - ¿Se analizó la variabilidad de los resultados de los estudios primarios?
  - ¿Se combinaron debidamente los resultados de los estudios primarios?
  - ¿Se apoyan en los datos citados las conclusiones de los autores de la revisión?
-

**CUADRO 2. Ejemplos de los elementos de una pregunta causal**

Tipo de pregunta	Población	Exposición o intervención	Resultado
Etiología	Varones homosexuales	Virus de la inmunodeficiencia humana	Síndrome de inmunodeficiencia adquirida
Diagnóstico	Pacientes con traumatismo craneal	Tomografía computadorizada	Hemorragia
Pronóstico	Pacientes con colitis ulcerosa	Colitis ulcerosa	Cáncer de colon
Terapéutica	Pacientes con enfermedad de Alzheimer	Fármacos colinérgicos	Estado funcional
Prevención	Mujeres posmenopáusicas	Suplemento dietético de calcio	Fractura de cadera

Muchos trabajos de revisión tratan de varios temas; por ejemplo, en un capítulo de un libro o un artículo sobre el síndrome de inmunodeficiencia adquirida se pueden revisar los conocimientos actuales sobre etiología, diagnóstico, pronóstico, tratamiento y prevención de la enfermedad. Este tipo de revisión puede ser muy útil para los lectores que buscan una visión general del tema. Sin embargo, las conclusiones de estas revisiones generales suelen estar poco o nada fundamentadas. Lo habitual es que se presenten las inferencias como juicios de hecho seguidos de una o más referencias. En ese caso, el lector no tiene base sobre la cual juzgar la calidad o la validez de las inferencias si no lee los artículos que se citan. Los lectores que buscan respuestas a preguntas clínicas específicas no deben fiarse de estos trabajos de revisión que tratan un tema muy amplio y abarcan muchas cuestiones.

Además, la exposición explícita de los métodos usados para revisar las investigaciones es necesaria para que el lector pueda evaluar con conocimiento de causa el rigor científico de la revisión y la solidez de los datos en los que se apoyan las inferencias. Por desgracia, esta información a menudo brilla por su ausencia. En general, cuando en una revisión no se menciona cómo se hizo algo —por ejemplo, cómo se seleccionaron los estudios primarios que se incluyeron—, es razonable suponer que no se hizo con rigor, lo que lleva a pensar en alguna posibilidad de invalidez de la revisión. Los lectores que necesiten respuestas a problemas clínicos específicos deben buscar revisiones que indiquen claramente los métodos utilizados. Sin conocer los métodos del autor, el lector no puede distinguir las afirmaciones basadas en datos de aquellas apoyadas únicamente en la opinión de quien escribió la revisión.

### ¿Se usaron métodos sistemáticos para localizar los estudios relevantes?

Es sorprendentemente difícil localizar todas las investigaciones publicadas sobre un tema, incluso cuando se trata de un aspecto relativamente concreto (31–33). Según Dickersin et al. (33), una búsqueda en el MEDLINE fue capaz de detectar solamente 29% de los ensayos clínicos relevantes para la prevención y el tratamiento de la hiperbilirrubinemia perinatal.

El problema es agravado por el hecho de que algunos de los datos relevantes pueden no estar ni siquiera publicados. Además, los estudios inéditos pueden ser sistemáticamente distintos de los que han aparecido en revistas sometidas a arbitraje por expertos, no porque sus métodos sean deficientes, sino porque sus resultados pueden ser "negativos". Las investigaciones al respecto han mostrado que, de dos artículos que usan los mismos métodos para investigar una cuestión, el estudio que rinde resultados positivos tiene más posibilidades de ser publicado que el de resultados negativos (33-37). Un estudio llevado a cabo por una entidad que haya invertido en el tratamiento investigado (por ejemplo, una compañía farmacéutica que trabaja sobre un producto nuevo) y que dé resultados negativos, puede que ni siquiera sea enviado a publicación. Corresponde, pues, al autor determinar en qué medida hay "sesgo de publicación" en el tema de estudio.

Las estrategias de investigación varían mucho de unos autores a otros y los expertos no suelen ser más sistemáticos que los no expertos a la hora de investigar (38). Cuanto más selectivo o anárquico sea el método de búsqueda de publicaciones, más probable será que haya sesgos en la revisión. Por ejemplo, lo más probable es que los autores presten mayor atención a los artículos que apoyan sus ideas previas.

El lector necesita tener la seguridad de que todas las publicaciones pertinentes e importantes se han incluido en la revisión. Cuanto más completa sea la búsqueda llevada a cabo por los autores, más probable es que hayan sido encontrados todos los artículos importantes para el tema. El lector debe buscar una mención explícita de los métodos de investigación utilizados. Lo ideal es que dichos métodos incluyan una búsqueda en una o más bases de datos bibliográficos (con especificación de las palabras clave y otros aspectos de las estrategias de investigación utilizadas [39]), una investigación de los informes que sean citados en artículos importantes localizados mediante una base de datos tal como el *Science Citation Index*, las referencias de todos los artículos pertinentes encontrados y la comunicación personal con investigadores u organizaciones activas en el tema que se está revisando (para garantizar que no se omiten artículos importantes publicados y en particular para detectar estudios metodológicamente adecuados que no hayan sido publicados).

### **¿Se usaron métodos explícitos para determinar qué artículos incluir en la revisión?**

Una revisión sistemática de la literatura generará un conjunto de artículos de los que muchos no serán pertinentes al tema que se estudia; otros serán tan deficientes desde el punto de vista metodológico que no darán información válida. Por lo tanto, los autores han de seleccionar los artículos válidos para la revisión. Cuando este proceso es asistemático, lo cual es muy común, surgen muchas oportunidades para que haya sesgos. Así, es frecuente encontrar dos revisiones del mismo tema en las que se han incluido distintos estudios primarios cuya selección ha generado diferentes conclusiones en cada una de las revisiones. Por ejemplo, en dos revisiones cuidadosas y sólidas desde el punto de vista metodológico, en las que se investigó si los corticosteroides se asocian con úlcera péptica, los dos grupos de autores usaron distintos criterios para seleccionar los estudios que incluirían en la revisión (40, 41). Esta diferencia fue la razón principal para el llamativo resultado de las dos revisiones: conclusiones diametralmente opuestas respecto a si la asociación existe o no.

Los autores deben especificar cómo seleccionaron los artículos haciendo referencia a los tres elementos básicos de un estudio primario: población, exposición o intervención y resultado. Por ejemplo, al evaluar los efectos de los fármacos

colinérgicos en pacientes con demencia, los autores podrían especificar los criterios siguientes.

- Población: pacientes con demencia senil en los que se excluyeron causas distintas a la enfermedad de Alzheimer.
- Intervención: administración de agentes colinérgicos por vía bucal.
- Resultado: indicadores de memoria y estado funcional.

Pueden usarse otros criterios metodológicos para seleccionar los estudios primarios que se incluirán en la revisión. En este ejemplo, los autores pueden considerar solo los estudios en los que los pacientes fueron seleccionados al azar para recibir terapéutica medicamentosa o placebo, y en los que tanto los investigadores como el paciente ignoraban el grupo al que había sido asignado el paciente.

### **¿Se tuvo en cuenta la validez de los estudios primarios?**

Los autores llegarán a conclusiones adecuadas solo si valoran con mucha meticulosidad la validez de los estudios primarios en los que se basa la revisión. Si todos los estudios tienen deficiencias básicas, sus conclusiones pueden ser cuestionables, incluso cuando sus resultados son comparables. Por ejemplo, si la literatura sobre cirugía derivativa extra-intracraneal para el riesgo de accidente cerebrovascular hubiera sido revisada antes de la publicación de los resultados de un ensayo clínico aleatorizado (42), se habría encontrado gran número de estudios con resultados positivos pero con diseños muy deficientes y, por tanto, muy sujetos a sesgo. La conclusión apropiada hubiera sido que la efectividad del procedimiento aún estaba en duda, a pesar del volumen de estudios con resultados positivos; claro está que el ensayo clínico posterior no mostró ventaja alguna de la cirugía respecto al tratamiento médico.

Existen guías metodológicas para las investigaciones sobre etiología (10, 43), diagnóstico (8), y tratamiento (11, 44). En un estudio sobre tratamiento lo importante es si la asignación fue o no aleatoria, si los sujetos y los investigadores desconocían o no el grupo al que había sido asignado el paciente, y si todos los resultados relevantes fueron registrados. Los aspectos importantes del diseño y de la realización de cada estudio primario deben criticarse y los estándares usados en estas críticas deben ser explícitos. Las críticas deben recogerse en detalle suficiente para que los lectores puedan juzgar la calidad metodológica de los estudios primarios. Una crítica estudio por estudio puede ser tediosa; la presentación de la evaluación metodológica en un cuadro permite una evaluación rápida de la validez. Los lectores deben considerar con cautela los resultados de cualquier revisión que enfoque los resultados de los estudios sin discutir a fondo los métodos mediante los cuales se obtuvieron a tales resultados.

Cuando la información sobre los métodos o los resultados se ha omitido de un informe publicado, los autores de una revisión pueden ponerse en contacto con los autores de la investigación para obtener la información requerida. Dice mucho a favor de la calidad de una revisión el hecho de que los autores discutan las consecuencias de la información que falta y hayan intentado obtener los datos relevantes.

## ¿Es reproducible y objetiva la valoración de los estudios primarios?

La evaluación por expertos de los estudios primarios de investigación normalmente genera un grado de desacuerdo extraordinario y preocupante. Por ejemplo, cuando se correlacionan las recomendaciones de publicar o no publicar los estudios de investigación primaria casi siempre se obtienen resultados de menos de 0,5, con un promedio de 0,3 (28, 45, 46); o sea, no mucho más de lo que esperaríamos encontrar por azar.

A las valoraciones no solo les falta reproducibilidad, sino que muchas veces están sesgadas. En un estudio llevado a cabo por Peters y Ceci (47) se reenviaron a diversas revistas artículos previamente publicados, procedentes de instituciones de mucho prestigio, pero sustituyendo los nombres de los autores y de las instituciones con nombres ficticios. Mahoney (35) envió un artículo a distintos árbitros, cambiando los resultados sin cambiar los métodos. Estos estudios mostraron que los trabajos procedentes de instituciones respetadas y que exponen resultados positivos tienen más posibilidades de ser aceptados. Además, en el estudio de Peters y Ceci muchos de los artículos fueron rechazados por "deficiencias metodológicas importantes", y en el estudio de Mahoney los métodos del estudio fueron peor considerados cuando los resultados fueron negativos.

Es posible también que los autores discrepen sobre los resultados de un estudio. Hay constancia de muchas revisiones en las que un autor que prefería un tratamiento particular clasificó el estudio primario como positivo mientras que otro autor al que no le gustaba ese tratamiento lo consideró negativo. Por ejemplo, Miller (48) encontró cinco artículos de revisión en los que se comparaba el tratamiento farmacológico asociado a psicoterapia con la psicofarmacoterapia aislada para tratar a pacientes psiquiátricos. De los 11 estudios citados en dos o más de las revisiones, los resultados de seis eran interpretados como positivos al menos en una revisión y como negativos al menos en otra.

Los problemas de reproducibilidad y sesgo pueden afectar a dos etapas del proceso de revisión: la decisión sobre los estudios que se incluirán y la valoración de la calidad de los estudios incluidos. Esos problemas pueden minimizarse si se observan criterios explícitos. Sin embargo, muchos de los criterios exigen una gran capacidad crítica del autor de una revisión. En uno de los ejemplos mencionados, la definición de la población como pacientes con demencia senil en los que se habían excluido causas distintas de la enfermedad de Alzheimer era uno de los criterios para inclusión en una revisión del tratamiento de la enfermedad de Alzheimer con colinérgicos. ¿Es suficiente que aparezca en el texto una afirmación tal como "se usaron los métodos habituales para diagnosticar la enfermedad de Alzheimer", o se requieren detalles de cómo se descartaron otras causas de demencia?

Los criterios explícitos ofrecen pocas ventajas si no son reproducibles por otros autores. Lo ideal sería que todos los estudios primarios potencialmente apropiados fueran evaluados para su inclusión al menos por dos autores, cada uno desconocedor de la decisión del otro. El grado de acuerdo debería hacerse constar. La reproducibilidad debería cuantificarse mediante una medida estadística como el coeficiente de correlación intragrupal (49) o el estadístico kappa ( $\kappa$ ) (50), que cuantifique el grado de acuerdo en exceso o defecto de lo que pudiera esperarse por casualidad. También debería usarse un método similar para evaluar la reproducibilidad de los criterios con los que se determinó la validez de los estudios primarios.

Aun cuando los criterios para la inclusión en el estudio o la evaluación de la validez pueden reproducirse, no queda garantizado que no se ha introducido algún sesgo. Por ejemplo, si los autores creen que un nuevo tratamiento es eficaz, puede

que empleen criterios de inclusión por los cuales los estudios con resultados negativos son sistemáticamente excluidos; puede que también juzguen con mayor rigor la validez de los resultados negativos que incluyan en el estudio. ¿Qué puede hacerse para evitar esa clase de sesgo?

En los ensayos clínicos controlados y aleatorizados, el sesgo se evita cuando tanto el paciente como el investigador desconocen si el paciente está tomando el medicamento activo o un placebo. En una valoración de estudios primarios las causas principales de sesgo se relacionan con los autores, sus instituciones y los resultados del estudio. Pero puede evaluarse el contenido y la calidad de un estudio a partir de sus métodos, sin conocer lo demás. Las otras secciones del artículo podrían dejarse aparte, de manera que los investigadores desconocieran las instituciones de los autores y los resultados. Lo ideal sería que las decisiones sobre inclusión y validez se tomaran en estas condiciones. Esta precaución adicional reforzaría las conclusiones de la revisión.

### **¿Se analizó la variación de los resultados de los estudios considerados?**

Los autores de revisiones con toda seguridad encontrarán variaciones en los resultados de los estudios en los que se investiga la cuestión de interés. Claro que si todos los resultados de los estudios primarios fueran iguales, probablemente no sería necesario un artículo de revisión. La tarea del revisor es precisamente explicar dicha variabilidad.

Las posibles fuentes de variación son el diseño de la investigación, la pura casualidad y las diferencias en los tres aspectos básicos del estudio (población, exposición o intervención y resultado medido) (51). Cuando en una revisión se incluyen ensayos clínicos controlados y aleatorizados, estudios de comparación pre y postintervención y estudios con un grupo de comparación histórico, y resulta que los ensayos clínicos controlados y aleatorizados muestran resultados sistemáticamente distintos a los de los otros estudios, la diferencia de diseños posiblemente explica la diferencia de resultados. Por ejemplo, Sacks et al. (52) vieron que los ensayos clínicos aleatorizados y controlados mostraban siempre un menor efecto que los estudios en los que se había usado un grupo de comparación histórico.

Una segunda explicación de las diferencias de resultados de los estudios es la mera casualidad. Incluso cuando dos investigaciones tienen métodos comparables y el verdadero valor de los efectos es idéntico, el efecto del azar llevará a diferencias aparentes en la magnitud de los resultados. Si las muestras son pequeñas, el azar por sí mismo puede dar lugar a diferencias aparentemente grandes en la dimensión de los efectos. Algunos ensayos clínicos con aspirina (ácido acetilsalicílico, AAS) en pacientes con ataques isquémicos transitorios han mostrado una tendencia a favor del placebo; otros ensayos han mostrado reducciones del riesgo de hasta 50% con el uso de AAS (53). Sin embargo, los intervalos de confianza que representan los extremos superior e inferior de magnitud del efecto compatible con los resultados observados se superponen. Por lo tanto, aunque los resultados aparentemente discrepantes quizá apunten a hipótesis por verificar en estudios ulteriores, todos ellos son compatibles con una reducción del riesgo de 15 a 30% por el tratamiento con AAS.

Cuando las diferencias de los resultados de los estudios son tan grandes que no pueden ser explicadas por el azar, los autores de la revisión deben buscar diferencias en cuanto a la población, la exposición o intervención, o los resultados

medidos. En nuestro ejemplo de tratamiento de pacientes de enfermedad de Alzheimer con colinérgicos, los estudios que dieron resultados negativos podrían haber incluido mayor número de pacientes graves que los estudios que dieron resultados positivos. Pudiera ser, entonces, que la intervención solo tuviera efecto en pacientes con enfermedad leve. Pero también podría ser que la intervención hubiera sido distinta, es decir, que se hubieran dado dosis mayores o fármacos distintos en los estudios que tuvieron resultados positivos. Por último, también pueden haber sido distintas las pruebas para cuantificar la memoria y el estado funcional. Algunas pruebas son más sensibles al estado del paciente que otras. Horwitz (51) ha dado muchos datos de cómo las diferencias metodológicas de los ensayos clínicos aleatorizados y controlados pueden llevar a resultados diferentes.

Los lectores de una revisión deben verificar si se han considerado estas cinco explicaciones de la diferencia de resultados en los estudios y deben ser escépticos cuando las diferencias se atribuyen solo a una explicación sin dar la debida consideración a las demás.

### **¿Se combinaron adecuadamente los resultados de los estudios primarios?**

El metanálisis (uso de diversas técnicas estadísticas para combinar los resultados de diferentes estudios) es cada vez más popular, especialmente como método para combinar los estudios de ensayos clínicos aleatorizados. Sin embargo, sigue siendo muy controvertido y no cabe esperar que en la práctica clínica se juzguen los méritos de una técnica estadística particular utilizada por los autores de un metanálisis. Sin embargo, hay temas que sí pueden valorar los lectores del ámbito clínico.

La forma más basta de metanálisis, en la que se compara el número de estudios con resultado positivo y con resultado negativo, no es satisfactoria. Este "recuento de votos" no tiene en cuenta la magnitud del efecto del tratamiento y el tamaño muestral de cada estudio. Los metanálisis más satisfactorios proporcionan dos trozos de información: la magnitud del efecto global debido al tratamiento y la verosimilitud de que ese efecto haya ocurrido por casualidad si el verdadero efecto fuera cero. Lo primero puede expresarse como una reducción porcentual del riesgo, lo último como un valor  $P$ .

La ventaja inicial del metanálisis es que los resultados de distintos estudios pueden combinarse con exactitud y fiabilidad para estimar mejor la magnitud media de los efectos de la exposición o tratamiento de interés. De todas formas, antes de combinar los resultados hay que considerar si es apropiado agregar todos los estudios. Los diseños o los tres aspectos básicos de los estudios a veces difieren en tal medida que combinar estadísticamente los resultados no tiene sentido. El metanálisis puede usarse para analizar la variación de los resultados de los estudios y generar o evaluar hipótesis sobre el origen de las diferencias. Sin embargo, el máximo rendimiento se obtiene cuando los métodos de los estudios primarios son similares y las diferencias de los resultados pueden explicarse por el azar.

Las revisiones en las que los resultados no se combinan mediante métodos estadísticos deben establecer explícitamente la base para las conclusiones y deberían explicar de alguna manera los resultados contradictorios. Hay que precaverse contra las revisiones que concluyen la inexistencia de un efecto sin haber considerado la potencia estadística de los estudios para detectar un efecto clínicamente importante. Cuando varios estudios no muestran una diferencia significativa, los revisores que no han usado el metanálisis tienden a concluir que no hay efecto, cuando la agregación mediante un

método estadístico demuestra justamente lo contrario. Esto fue mostrado experimentalmente por Cooper y Rosenthal (54), que asignaron revisores al azar a usar o no metanálisis para combinar los resultados de varios estudios, incluidos algunos que no mostraban resultados estadísticamente significativos. Otro investigador hizo la misma observación cuando encuestó a los investigadores que habían llevado a cabo un ensayo clínico del citrato de tamoxifeno, como terapéutica coadyuvante en el cáncer de mama (Rory Collins, comunicación personal, 1987). La mayoría de los investigadores concluyeron de la información disponible que el tamoxifeno no prolongó el intervalo exento de enfermedad; sin embargo, la agregación estadística de todos los resultados disponibles mostró un efecto estadísticamente significativo y clínicamente importante.

Es conveniente recordar que los demás criterios ya mencionados siguen siendo aplicables aun cuando los autores de una revisión hayan usado metanálisis.

### ¿Se fundamentan en los datos citados las conclusiones de la revisión?

Se haya usado o no el metanálisis, los resultados de los estudios primarios individuales deben indicarse con detalle suficiente para que los lectores sean capaces de evaluar críticamente el fundamento de las conclusiones a las que llegan los autores de la revisión. El método de presentación de resúmenes de los estudios individuales dependerá de la cuestión estudiada. Para cuestiones de efectividad del tratamiento o de prevención, la magnitud del efecto y su intervalo de confianza proporcionan la información clave. En las revisiones de métodos diagnósticos deben indicarse las sensibilidades, especificidades y razones de verosimilitudes (con sus intervalos de confianza) (8). Las curvas de supervivencia pueden servir eficientemente para indicar los resultados principales de estudios de pronóstico.

En temas de etiología y causación en los que no se dispone de estudios controlados y aleatorizados, los autores de la revisión pueden evaluar los datos probatorios mediante los criterios de inferencia causal. Estos criterios han sido presentados de distinta manera por diversos investigadores (10, 44, 55, 56), pero los aspectos comunes son la magnitud y la consistencia de la asociación entre el agente causal y el resultado, y la necesidad de demostrar la relación temporal apropiada. Nuestra versión de estos criterios consta en el cuadro 3. Por supuesto, los comentarios de la revisión sobre cada uno de estos criterios deben referirse directamente a los datos de los estudios primarios citados.

#### CUADRO 3. Criterios para valorar la solidez de una inferencia causal

- 
- ¿Es adecuada la relación temporal? (Es imprescindible una respuesta afirmativa, pero dicha respuesta no dice nada a favor de la inferencia causal.)
  - ¿Son importantes los datos a favor?
  - ¿Es intensa la asociación?
  - ¿Son compatibles los distintos estudios?
  - ¿Hay una relación dosis-respuesta?
  - ¿Hay datos indirectos que apoyen la inferencia, es decir, datos que pongan en relación resultados intermedios, o datos de investigaciones en poblaciones diferentes (incluso animales) o de relaciones análogas (o sea, exposiciones o intervenciones similares)?
  - ¿Se han desechado otras hipótesis plausibles que pudieran plantearse?
-

## Conclusión

Una revisión de la literatura es una tarea científica y, como en otros ámbitos de la ciencia, existen determinadas normas para llevarla a cabo de tal modo que se alcancen conclusiones válidas. Los lectores de la literatura clínica que no son capaces de evaluar críticamente los métodos de los estudios primarios pueden deducir conclusiones erróneas e, igualmente, los lectores que no son capaces de valorar la calidad científica de una revisión pueden dejarse llevar a conclusiones equivocadas. En este artículo hemos propuesto ocho criterios a lectores interesados en hallar respuestas a cuestiones clínicas de interés en su práctica diaria. La aplicación de esos criterios permitirá a los lectores descartar rápidamente los artículos de revisión que son de poco interés, o científicamente dudosos, detectar fuentes potenciales de error y poder confiar en las conclusiones obtenidas a partir de una evaluación sistemática de las investigaciones publicadas.

## AGRADECIMIENTO

Damos las gracias a los doctores Geoff Norman, David Streiner, David L. Sackett y Brian Hutchinson, y al profesor Mike Gent por su contribución a la elaboración de estos criterios.

Este trabajo fue financiado en parte por el Ministerio de Salud de Ontario.

## REFERENCIAS

1. Haynes RB, McKibbon KA, Fitzgerald D, et al. How to keep up with the medical literature: I. Why try to keep up and how to get started. *Ann Intern Med.* 1986;105:149-153.
2. Idem. How to keep up with the medical literature: II. Deciding which journals to read regularly. *Ibid:* 309-312.
3. Idem. How to keep up with the medical literature: III. Expanding the number of journals you read regularly. *Ibid:* 474-478.
4. Idem. How to keep up with the medical literature: IV. Using the literature to solve clinical problems. *Ibid:* 636-640.
5. Idem. How to keep up with the medical literature: V. Access by personal computer to the medical literature. *Ibid:* 810-824.
6. Idem. How to keep up with the medical literature: VI. How to store and retrieve articles worth keeping. *Ibid:* 978-984.
7. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals: I. Why to read them and how to start reading them critically. *Can Med Assoc J.* 1981;124:555-558.
8. Idem. How to read clinical journals: II. To learn about a diagnostic test. *Ibid:* 703-710.
9. Idem. How to read clinical journals: III. To learn the clinical course and prognosis of disease. *Ibid:* 869-872.
10. Idem. How to read clinical journals: IV. To determine etiology or causation. *Ibid:* 985-990.
11. Idem. How to read clinical journals: V. To distinguish useful from useless or even harmful therapy. *Ibid:* 1156-1162.
12. Idem. How to read clinical journals: VI. To learn about the quality of clinical care. *Can Med Assoc J.* 1984;130:377-382.

13. Labarthe DR. Mild hypertension: the question of treatment. *Ann Rev Public Health*. 1986;7:193–215.
14. Haber E, Slater EE. High blood pressure. En Rubenstein M, Federman DD, eds. *Scientific American Medicine*. 9ª ed., New York: Sci Am; 1986: sect. 1, VII1–VII29.
15. Risks of antihypertensive therapy [E]. *Lancet*. 1986;2:1075–1076.
16. Sacks HS, Chalmers TC, Berk AA, et al. Should mild hypertension be treated? An attempted meta-analysis of the clinical trials. *Mt Sinai J Med*. 1985;52:265–270.
17. Guyatt GH, Webber CE, Mewa AA, et al. Determining causation—a case study: adrenocorticosteroids and osteoporosis. *J Chronic Dis*. 1984;37:343–352.
18. Baylink DJ. Glucocorticoid-induced osteoporosis. *N Engl J Med*. 1983;309:306–308.
19. Swan HJC, Ganz W. Hemodynamic measurements in clinical practice: a decade in review. *J Am Coll Cardiol*. 1983;1:103–113.
20. Robin ED. The cult of the Swan-Ganz catheter: overuse and abuse of pulmonary flow catheters. *Ann Intern Med*. 1985;103:445–449.
21. Harrington JT, Isner JM, Kassirer JP. Our national obsession with potassium. *Am J Med*. 1982;73:155–159.
22. Kaplan NM. Our appropriate concern about hypokalemia. *Am J Med*. 1984;77:1–4.
23. Mulrow CD. The medical review article: state of the science. *Ann Intern Med*. 1987;106:485–488.
24. Sacks HS, Berner J, Reitman D, et al. Meta-analyses of randomized controlled trials. *N Engl J Med*. 1987;316:450–455.
25. Williamson JW, Goldschmidt PG, Colton T. The quality of medical literature. An analysis of validation assessments. En: Bailar JC III, Mosteller F, eds. *Medical Uses of Statistics*. Waltham, Mass: NEJM Bks; 1986;370–391.
26. Halvorsen KT. Combining results from independent investigations: meta-analysis in medical research. *Ibid*: 392–416.
27. Oxman AD. *A Methodological Framework for Research Overviews*. [Tesis de maestría] McMaster U., Hamilton, Ont., 1987:23–31, 98–105.
28. Light RJ, Pillemer DB. *Summing Up: the Science of Reviewing Research*. Cambridge, Mass: Harvard University Press; 1984.
29. Jackson GB. Methods for integrative reviews. *Rev Educ Res*. 1980;50:438–460.
30. Cooper HM. *The Integrative Research Review: a Systematic Approach*. Beverly Hills, Calif: Sage; 1981.
31. Glass GV, McGaw B, Smith MI. *Meta Analysis in Social Research*. Beverly Hills, Calif: Sage; 1981.
32. Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature. *Controlled Clin Trials*. 1985;6:271–279.
33. Dickersin K, Hewitt P, Mutch I, et al. Perusing the literature comparison of MEDLINE searching with a perinatal trials database. *Ibid*: 306–317.

34. Simes RJ. Publication bias. The case for an international registry of clinical trials. *J Clin Oncol.* 1986;4:1529–1541.
35. Mahoney MJ. Publication prejudices an experimental study of confirmatory bias in the peer review system. *Cognit ther Res.* 1977;1:161–175.
36. Devine EC, Cook TD. Effects of psycho educational intervention on length of hospital stay: a meta analytic review of 34 studies. En: Light RJ, ed. *Evaluation Studies Review Annual.* 8a ed. Beverly Hills, Calif: Sage; 1983:417–432.
37. Simes RJ: Confronting publication bias: a cohort design for meta-analysis. *Stat Med.* 1987;6:11–29.
38. Cooper HM: Literature searching strategies of integrative research reviewers: a first survey. *Knowledge.* 1986;8:372–383.
39. Huith EJ. Needed: review articles with more scientific rigor. *Ann Intern Med.* 1987;106:470–471.
40. Messer J, Reitman D, Sacks HS, et al. Association of adrenocorticosteroid therapy and peptic-ulcer disease. *N Engl J Med.* 1983;309:21–24.
41. Conn HO, Blitzer BL. Nonassociation of adrenocorticosteroid therapy and peptic ulcer. *N Engl J Med.* 1976;294:473–479.
42. EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. *N Engl J Med.* 1985;313:1191–1200.
43. Hill AB. *Principles of Medical Statistics.* 9<sup>a</sup> ed. London: Lancet; 1971:312–320.
44. Chalmers TC, Smith H, Blackburn B, et al. A method for assessing the quality of a randomized controlled trial. *Controlled Clin Trials.* 1981;2:31–49.
45. Bailar JC III, Patterson K. Journal peer review: the need for a research agenda. En: Bailar JC III, Mosteller F, eds. *Medical Uses of Statistics.* Waltham, Mass: NEJM Bks; 1986:349–369.
46. Marsh HW, Ball S. Interjudgmental reliability of reviews for the *Journal of Educational Psychology.* *J Educ Psychol.* 1981;73:872–880.
47. Peters DP, Cesi SJ. Peer-review practices of psychological journals: the fate of published articles, submitted again. *Behav Brain Sci.* 1982;5:187–255.
48. Miller TI. *The Effects of Drug Therapy on Psychological Disorders* [tesis doctoral]. Boulder: U of Colorado; 1977.
49. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86: 420–428.
50. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37–46.
51. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med.* 1987;82:498–510.
52. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. *Am J Med.* 1982;72:233–240.
53. Sze PC, Pincus M, Sacks HS, et al. Antiplatelet agents in secondary stroke prevention. A meta-analysis of the available randomized control trials [resumen]. *Clin Res.* 1986;34:385A.
54. Cooper HM, Rosenthal R. Statistical versus traditional procedures for summarizing research findings. *Psychol Bull.* 1980;87:442–449.
55. Susser M. Reviews and commentary: the logic of Sir Karl Popper and the practice of epidemiology. *Am J Epidemiol.* 1986;124:711–718.
56. Guyatt CH, Newhouse MT. Are active and passive smoking harmful? Determining causation. *Chest.* 1985;88:445–451.