

997

INDEXED

## COMUNICACIÓN BIOMÉDICA

# El metanálisis en la investigación clínica<sup>1</sup>

Kristan A. L'Abbé,<sup>2</sup> Allan S. Detsky<sup>2</sup> y Keith O'Rourke<sup>2</sup>

*El metanálisis es un proceso de combinación de resultados de estudios mediante el cual pueden obtenerse conclusiones sobre la efectividad terapéutica u orientaciones para planificar nuevas investigaciones. Aquí se revisan algunos aspectos del diseño y temas estadísticos importantes de ese proceso. Los temas de diseño son la elaboración del protocolo, los objetivos, la búsqueda de publicaciones, el sesgo de publicación, la medición de los resultados de los estudios y la calidad de los datos. Los temas estadísticos son la coherencia (homogeneidad) de los resultados de los estudios y las técnicas para acumular resultados de varios estudios. A partir de lo discutido respecto al diseño y los aspectos estadísticos, se dan orientaciones para evaluar la calidad del metanálisis. Se discuten las limitaciones y la posible orientación de los avances ulteriores; los investigadores deberían alcanzar un acuerdo general sobre cómo han de llevarse a cabo los metanálisis. Como estrategia explícita para compendiar resultados, el metanálisis podría ayudar a los que trabajan en el medio clínico o en la investigación a entender mejor los resultados de los estudios clínicos.*

Un investigador desea saber si la estreptoquinasa intravenosa es una terapéutica efectiva para los pacientes en fase aguda de infarto de miocardio. De los ocho estudios publicados en los que consta la mortalidad a los 45 días, tres dieron resultados estadísticamente significativos a favor de la estreptoquinasa ( $P < 0,05$ ), dos dieron resultados a favor de la estreptoquinasa, pero no estadísticamente significativos, y los resultados de los otros tres apoyaron la terapéutica de control con resultados estadísticamente no significativos. Además, los grupos de pacientes incluidos en los estudios eran de distintas características, al igual que algunos aspectos de los métodos de estudio.

<sup>1</sup> Esta versión al español del artículo "Meta-analysis in clinical research" (*Annals of Internal Medicine* 1987; 107:224-233) se publica con autorización del American College of Physicians, la revista *Annals of Internal Medicine* y los autores, ninguno de los cuales se hace responsable de la exactitud de la traducción.

<sup>2</sup> Departamento de Estudios de Posgrado en Salud Comunitaria y los Departamentos de Administración Sanitaria y Medicina, Universidad de Toronto, y la División de Medicina Interna General y Epidemiología Clínica, Hospital General de Toronto, Ontario, Canadá.

Las separatas de la versión original en inglés deben pedirse a Kristan J. Aronson, Ph.D.; Department of Preventive Medicine and Biostatistics, McMurrich Building, 4th floor, University of Toronto, Ontario, Canadá M5S 1A8.

¿Cómo puede decidir el clínico si la estreptoquinasa es una terapéutica efectiva a partir de los resultados contradictorios de estos estudios? ¿Cómo puede combinarse la información derivada de esos estudios para llegar a una conclusión general? El metanálisis es una estrategia de investigación ideada con la finalidad de responder a ese tipo de preguntas. Mediante este procedimiento, Stampfer et al. (1) combinaron los resultados de los ocho ensayos clínicos y mostraron que la estreptoquinasa reducía la mortalidad en 20%, lo cual era estadísticamente significativo ( $P < 0,01$ ).

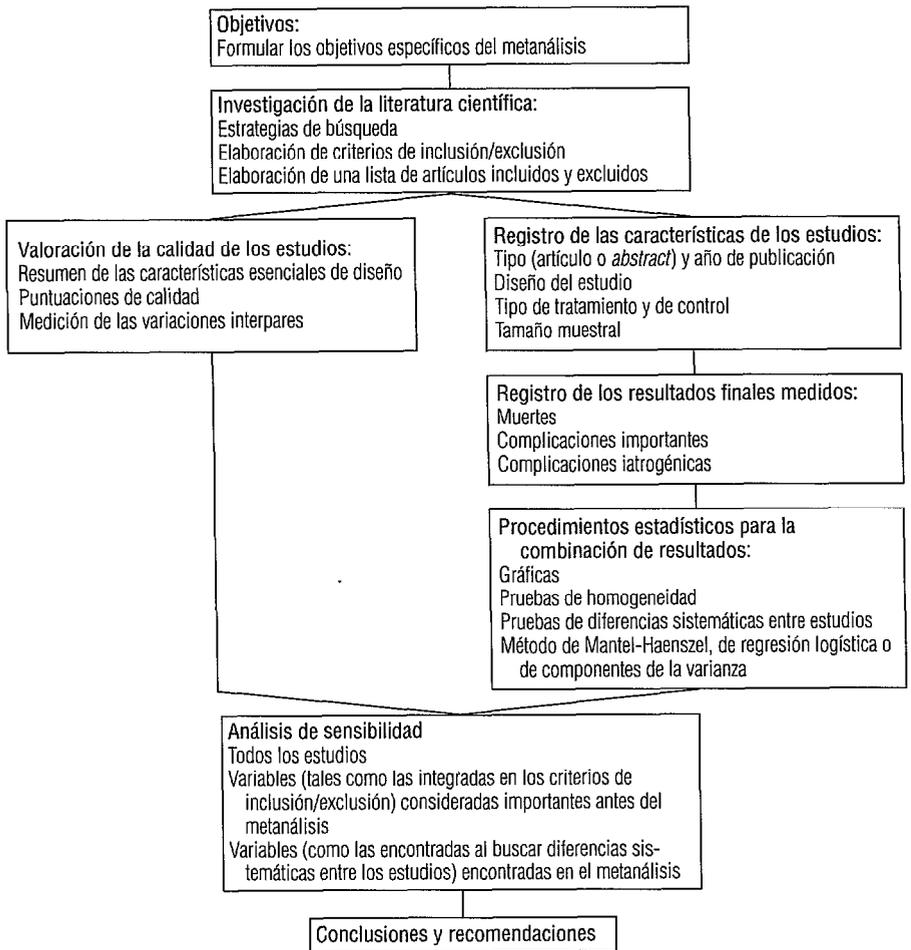
El metanálisis es un proceso de combinación de resultados de estudios para extraer conclusiones sobre la efectividad terapéutica o planificar investigaciones ulteriores. El producto final tiene elementos cuantitativos y cualitativos, ya que no solo incorpora los resultados numéricos y el tamaño muestral de cada estudio, sino también aspectos más subjetivos tales como la calidad, la magnitud del sesgo y la consistencia del diseño de los estudios. El metanálisis es una estrategia sistemática de revisión para resolver interrogantes de la investigación. Es especialmente útil cuando los resultados de varios estudios son contradictorios en cuanto a magnitud o dirección del efecto, cuando los tamaños muestrales individuales son demasiado pequeños para detectar un efecto y calificarlo de estadísticamente significativo, o cuando una investigación a gran escala exige demasiado tiempo o dinero.

Para evaluar tratamientos médicos y planificar nuevos estudios se requiere una mejor comprensión de los resultados de estudios clínicos anteriores. Los investigadores suelen basarse en revisiones de los trabajos publicados para delimitar los conocimientos actuales. El metanálisis puede ser una respuesta apropiada al volumen siempre creciente de investigaciones publicadas y no publicadas. Esta técnica enfoca la revisión de las publicaciones de una manera más estructurada que la revisión narrativa tradicional y de esta forma puede ser más útil para evaluar científicamente la acumulación de datos a favor o en contra de algo. Las conclusiones cualitativas o cuantitativas de un metanálisis pueden actualizarse a medida que se dispone de resultados de nuevos estudios. En nuestra opinión, el metanálisis es un enfoque explícito de revisión de la literatura científica, que complementa la revisión narrativa, pero que no se opone a ella.

Este trabajo da una visión general del metanálisis, desarrollado principalmente en psicología y ciencias de la educación, y discute sus aplicaciones a la investigación clínica. También examina aspectos de diseño y temas estadísticos. Como ayuda para los lectores hemos añadido un algoritmo o diagrama de flujo del metanálisis (figura 1). A nuestro juicio, el metanálisis puede tener muchas aplicaciones en la investigación clínica. Aquí se comentan los aspectos metodológicos de esta técnica y se señalan algunas posibles líneas de desarrollo de este enfoque integrativo de los resultados de la investigación.

Para los usuarios de informes de metanálisis presentamos una introducción a la técnica y un formulario que permite evaluar la calidad de los metanálisis publicados. Para quienes elaboran ese tipo de informes proporcionamos una pequeña introducción de nuestro enfoque del metanálisis, basado en nuestra aplicación de la técnica al trabajo clínico propio. Este enfoque debe mucho al trabajo de Sacks y Chalmers (2). Es de esperar que se sigan utilizando diversos enfoques de metanálisis, al menos hasta que haya algún consenso en lo referente a su uso en la investigación clínica. Por el momento, no hay un método "correcto" para llevar a cabo el metanálisis. La metodología para aplicar esta estrategia de investigación a las cuestiones clínicas está aún en evolución.

**FIGURA 1. Algoritmo de metanálisis**



## ASPECTOS DE DISEÑO

### Objetivos

En la investigación clínica, el objetivo principal del metanálisis es llegar a una conclusión general; por ejemplo, determinar cuán efectiva es una terapéutica para tratar una enfermedad. Casi cualquier interrogante clínico o cualquier controversia puede someterse a metanálisis. Es importante que los objetivos del metanálisis se especifiquen claramente antes de realizar el análisis. Como en cualquier trabajo de investigación, el desarrollo de un protocolo de trabajo formaliza las decisiones tomadas durante la fase de diseño para conseguir los objetivos. Mediante metanálisis se han estudiado temas básicos de la investigación clínica como los progresos en cirugía y anestesia (3), la efectividad de la profilaxis antibiótica en cirugía del colon (4) o los resultados terapéuticos, en la nefritis del lupus, de los inmunosupresores

con prednisona comparados con la prednisona sola (5). Para cada objetivo general, los investigadores pueden plantearse objetivos secundarios, por ejemplo, determinar a qué edades puede ser más efectivo el tratamiento. En la evaluación de fármacos para la nefritis lúpica, los objetivos secundarios fueron determinar los efectos relativos de los dos regímenes terapéuticos respecto a varios posibles resultados finales: deterioro renal, fallecimiento por nefropatía, muerte de origen extrarrenal y mortalidad global (5). El metanálisis también se usó para determinar el tamaño muestral necesario para futuros ensayos clínicos sobre la nefritis lúpica (5). También se han publicado modelos económicos en los que los resultados del metanálisis se usan para planificar futuros ensayos clínicos (6–8).

## Investigación de la literatura científica

Para recopilar todos los trabajos de interés publicados es posible hacer una búsqueda sistemática mediante computador o por medios convencionales. No obstante, las búsquedas computadorizadas no siempre son específicas y coherentes. Varios artículos en publicaciones clínicas han revisado los problemas que se plantean en ese campo (9–13), que rebasan el marco de este trabajo. Para conseguir mejores resultados debe consultarse a un bibliotecario especializado en investigación bibliográfica de temas clínicos.

La investigación bibliográfica no sirve para encontrar trabajos inéditos. Es posible que los estudios publicados sean sistemáticamente diferentes de los no publicados, de manera que el metanálisis basado únicamente en una búsqueda bibliográfica puede tener “sesgo de publicación”. Para minimizar este sesgo potencial pueden investigarse las referencias de los estudios publicados o las bases de datos de materiales no publicados (disponibles en servicios tales como BRS Saunders), pedir información a los investigadores con experiencia en la materia o estudiar el material de los Institutos Nacionales de Salud [de los Estados Unidos]. Cuando existen registros o inventarios de ensayos clínicos, son útiles para al menos tener un denominador apropiado para calcular la razón del número de estudios publicados respecto al total llevado a cabo. Un ejemplo de tales inventarios es la base de datos de investigaciones perinatales (11). A menudo se supone que los ensayos clínicos que no muestran diferencias significativas tienen menor probabilidad de ser publicados que los estudios que sí las muestran (14–20), pero no está en absoluto demostrado que el sesgo de publicación obre siempre en esa dirección.

Tampoco existen reglas aceptadas para manejar los estudios inéditos publicados solo como resúmenes (*abstracts*). Probablemente algunos lectores discreparán, pero, en nuestra opinión, el análisis inicial solo debe incluir estudios publicados en forma de artículos. A veces el desacuerdo puede resolverse fácilmente mediante el análisis de sensibilidad.

## Criterios de inclusión

Los estudios se integran o no en el metanálisis basándose en criterios de inclusión o de exclusión. Lo ideal es especificar los criterios de exclusión al diseñar el protocolo del metanálisis, en función de sus objetivos específicos. Es necesario explicar las razones que justifican los criterios elegidos. En la práctica no es raro que haya que modificar los criterios amplios de inclusión tras revisar todos los estudios encontrados en la búsqueda bibliográfica. Como base para los criterios de inclusión pueden escogerse aspectos tales como el diseño del estudio, el tamaño muestral, el

tipo de terapéutica experimental y de control, el que el estudio esté o no publicado y el resultado final de interés. Por ejemplo, los investigadores clínicos pueden considerar apropiado restringir el análisis a los estudios aleatorizados publicados, excluyendo así los estudios de cohorte (no aleatorizados), los estudios de casos y controles y los trabajos no publicados. Otros criterios de inclusión pueden ser, por ejemplo, un tamaño muestral de más de 10 por grupo, una dosis concreta del fármaco experimental comparado con la terapéutica de control o la medición de la mortalidad o de otro resultado final clínicamente relevante.

Si se establecen criterios de inclusión liberales, como parte del análisis se puede estudiar el efecto de la ampliación o restricción de criterios, subdividiendo los estudios según una o más características. Los ensayos pueden reagruparse por características como la asignación aleatoria o no aleatoria de los pacientes a las terapéuticas o las dosis de la terapéutica experimental, con objeto de determinar el efecto de estas características en los resultados del metanálisis. Este procedimiento se denomina análisis de sensibilidad. En un análisis de sensibilidad los estudios no aleatorizados mostraron constantemente mayor efectividad de la cirugía anastomótica coronaria frente al tratamiento médico (no quirúrgico) que los estudios aleatorizados (21). Correlacionar las diversas características de diseño de los estudios con la magnitud de las medidas individuales y combinadas de efecto mostrará explícitamente el efecto de ampliar o estrechar los criterios de inclusión y puede así fortalecer el análisis.

No conocemos criterios estándar para la inclusión de estudios en el metanálisis. No obstante, no hay criterios universales apropiados, ya que el metanálisis puede aplicarse a una amplia gama de temas. Los criterios de inclusión y exclusión pueden variar en función de diferentes objetivos. Por ejemplo, si el objetivo del análisis es ver el efecto de alguna intervención sobre la calidad de la vida, los únicos estudios que podrán incluirse en el metanálisis serán aquellos en los que se haya medido dicha calidad. De igual forma, si el objetivo es evaluar los efectos colaterales (la seguridad) de una técnica nueva, solo podrán incluirse los estudios en los que se hayan tenido en cuenta los efectos colaterales. Pensamos que a menudo los investigadores estarán en desacuerdo respecto a los criterios de inclusión y exclusión para un metanálisis concreto. Por ello deben especificarse los criterios y su fundamento y hacerse constar todos los estudios encontrados, incluso los excluidos.

## Recogida de datos

Una vez recopilados y seleccionados los estudios según los criterios de inclusión y exclusión, se revisan y se resume la información descriptiva. Solo debe recogerse información correspondiente a grupos separados de pacientes. A veces, los resultados de una investigación aparecen en más de una publicación, de manera que uno de los informes corresponde a los mismos pacientes que el otro, o bien a un subgrupo de aquellos. Al llevar a cabo un metanálisis el investigador ha de tener cuidado para no incluir datos del mismo grupo de pacientes más de una vez.

Pueden diseñarse formularios para recoger datos resumidos, por ejemplo, el número de pacientes por grupo, el efecto medido en cada grupo de tratamiento y otras características descriptivas importantes. El efecto de interés puede que se haya medido mediante una variable continua tal como la tensión arterial o un índice de calidad de la vida; o una variable categórica, como la mortalidad o las complicaciones; o una variable ordinal (una serie de categorías ordenadas), como el estadio tumoral; o variables de tipo temporal expresadas en tablas de mortalidad, por

ejemplo la supervivencia sin enfermedad. Las mediciones del efecto terapéutico (diferencias en resultado final entre los grupos de tratamiento) corresponden a estos tipos de datos. A los efectos que aquí interesan nos centraremos en los resultados finales de tipo categórico, ya que muchas entidades clínicas se definen de esa manera. El apéndice 1 presenta cuatro medidas del efecto del tratamiento para la comparación de resultados finales categóricos. Para resumir los datos lo mejor es registrar los números en bruto (10 fallecimientos entre los 20 pacientes de un grupo de tratamiento), no en forma de proporciones (0,5) o diferencias de proporciones entre grupos. La figura 2 es un ejemplo de formulario para el resumen de datos.

En esta fase pueden surgir dos problemas. En primer lugar, los datos relevantes pueden no constar en el informe publicado. Se puede intentar localizar a los investigadores para obtener los datos que faltan, pero eso puede no ser posible por las limitaciones económicas o temporales del metanálisis o por el número de estudios en los que faltan datos. El segundo problema se refiere a la calidad del proceso de resumen de los datos, que puede ser origen de error y, de ahí, de sesgo potencial. Nuestra recomendación es que dos personas familiarizadas con el asunto clínico en cuestión extraigan un resumen independientemente y sometan luego sus resultados a cotejo, para prevenir esta forma de error.

**FIGURA 2. Formulario-resumen de datos de estudios para un proyecto de metanálisis**

Categoría del estudio \_\_\_\_\_ Revisor \_\_\_\_\_

Título \_\_\_\_\_

Autores \_\_\_\_\_

1. TIPO DE TERAPÉUTICA DE TRATAMIENTO

Dosis \_\_\_\_\_

Edad de los pacientes \_\_\_\_\_

Duración del tratamiento \_\_\_\_\_

Otras variables clínicas \_\_\_\_\_

2 TIPO DE TERAPÉUTICA DE CONTROL

Dosis \_\_\_\_\_

Edad de los pacientes \_\_\_\_\_

Duración del tratamiento \_\_\_\_\_

Otras variables clínicas \_\_\_\_\_

3. RESULTADOS

	$P_t$ Tratamiento (x/n y %)	$P_c$ Control (x/n y %)	Valor P
Diferencia en mortalidad a los ( <i>período especificado</i> )	_____	_____	_____
Diferencia en complicaciones a los ( <i>período especificado</i> )	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
Resultados finales no clínicos _____	_____	_____	_____
_____	_____	_____	_____

Información:— tal como consta en el estudio \_\_\_\_\_  
 o — estimada a partir de la información del estudio \_\_\_\_\_  
 o — proporcionada por el investigador \_\_\_\_\_

## Evaluación de la calidad del estudio

Antes de evaluar la calidad de los estudios, recomendamos formar un grupo de seis a ocho colegas para desarrollar un protocolo de evaluación de la calidad y un formulario para los datos. Nosotros hemos modificado un sistema de evaluación de la calidad que idearon Chalmers et al. (22). Para evitar un sesgo potencial, este sistema exige que los revisores lean solo los métodos y los resultados de los informes publicados y que toda la información identificadora (autores, instituciones e intervenciones) sea eliminada de las copias de los informes entregadas a los evaluadores. Cada estudio se califica con una puntuación de 0,0 a 1,0. Estas calificaciones de calidad pueden usarse de varias formas. Una es incluir o excluir cada estudio del metanálisis en función de una calificación límite previamente establecida. Otra posibilidad es utilizar las puntuaciones de calidad como pesos para ponderar cada estudio en el proceso de combinación (apéndice 2). Por último, los resultados finales del estudio pueden correlacionarse con las puntuaciones de calidad para determinar el efecto de la calidad del estudio en los resultados demostrados.

Como la evaluación de la calidad es un proceso subjetivo, hay muchas posibilidades de error y de sesgo. Nosotros hicimos que dos revisores evaluaran la calidad de cada estudio y a continuación, comentaran sus desacuerdos en una "reunión de consenso". Tal como indica Fleiss (23), es conveniente estimar la variación de las calificaciones de calidad de una pareja de evaluadores en un estudio piloto realizado antes de llevar a cabo el metanálisis.

## ASPECTOS ESTADÍSTICOS

### Marco general

Tres aspectos configuran el marco general estadístico y lógico para el metanálisis. En primer lugar, ¿son las medidas de resultado final compatibles entre los distintos estudios? (Dicho de otra manera, ¿es posible combinar los estudios de manera razonable?) En segundo lugar, ¿las variaciones de los resultados de los estudios se correlacionan con las variaciones en las características del estudio? Y, por último, ¿cuál es la mejor estimación del valor y del intervalo de confianza para la medida combinada de resultado final (por ejemplo, la reducción de la mortalidad)?

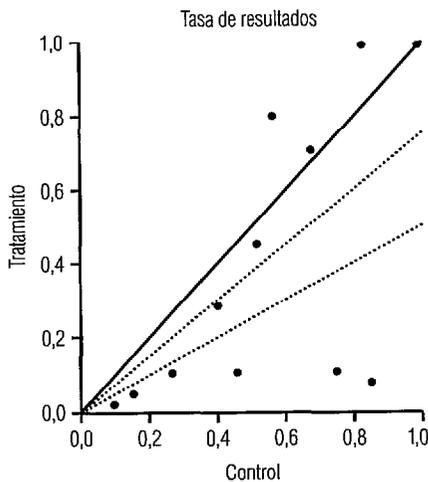
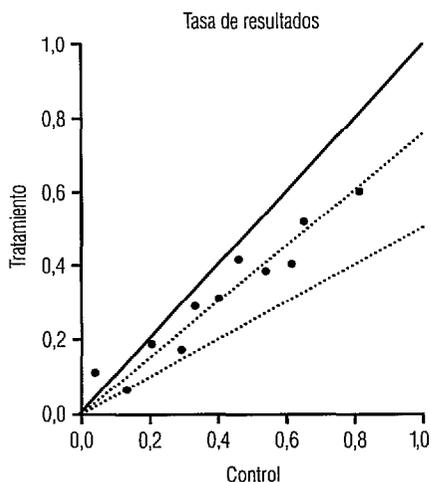
Las respuestas a estas preguntas configuran un marco general para combinar lo probado por varias tablas dos por dos procedentes de uno o varios estudios. De hecho, el metanálisis tiene una analogía directa con la combinación de los resultados en un estudio multicéntrico, donde también surgen problemas tales como las diferencias de protocolos y de poblaciones estudiadas, aunque estos problemas suelen ser mucho más importantes en el metanálisis. La combinación de resultados puede generar una mayor potencia estadística para resolver un interrogante clínico, suponiendo que los estudios tengan valor. En el metanálisis las estadísticas tienen el objetivo habitual de describir los resultados y cuantificar la incertidumbre, una vez que el buen juicio y el sentido común (los elementos más importantes) han clarificado los temas cualitativos.

## Homogeneidad y fuentes de variación

El supuesto que subyace a la combinación de resultados de estudios individuales en una estadística descriptiva es que sus diferencias son solo casuales (variación de muestreo), y que, por tanto, los resultados de todos los estudios son homogéneos, es decir, que reflejan el mismo efecto "verdadero". Dicho de otra manera, que cuando los resultados se combinan, los errores aleatorios se compensan y  $n$  resultados resultan mejor que uno. De todas formas, este supuesto debe cuestionarse. Si las variaciones no parecen solamente fortuitas, la combinación de resultados será más compleja y en ciertos casos puede no ser aconsejable. Como primer paso, puede ser útil una representación gráfica de los resultados de los diversos estudios. Hay una técnica que muestra la variación observada en la reducción del riesgo (definida en el apéndice 1), representando la tasa de eventos en los grupos terapéuticos en el eje vertical y la de los grupos de control en el eje horizontal. (Otras veces puede ser más apropiado representar gráficamente las diferencias absolutas de tasas de eventos.) La figura 3 muestra los resultados finales de los estudios según una simulación de computador diseñada para mostrar efectos terapéuticos coherentes (homogéneos), medidos por la reducción proporcional de riesgo. La figura 4 muestra un conjunto simulado de resultados finales en el que no existe coherencia de efectos del tratamiento (es decir, muestra heterogeneidad).

**FIGURA 3.** Ejemplo de coherencia de efectos de los tratamientos (homogeneidad), medidos por la reducción proporcional del riesgo. Tasas de eventos representadas en el eje vertical para el grupo de tratamiento y en el horizontal para el grupo control. La diagonal continua indica tasas iguales. La línea de puntos superior muestra una reducción del riesgo de 25% ( $(P_c - P_t)/P_c = 0,25$ ); la inferior, de 50%

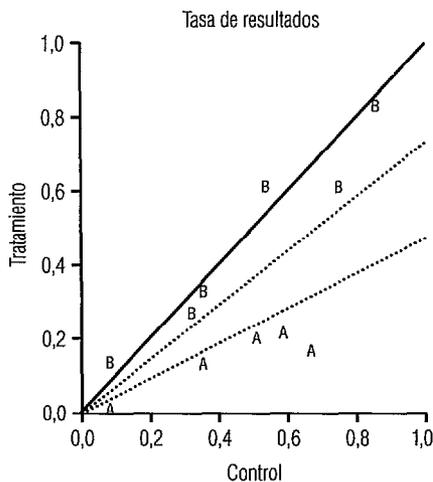
**FIGURA 4.** Ejemplo de efectos incoherentes de los tratamientos (heterogeneidad), medidos por la reducción proporcional del riesgo. Tasas de eventos representadas en el eje vertical para el grupo de tratamiento y en el horizontal para el grupo control. La diagonal continua indica tasas iguales. La línea de puntos superior muestra una reducción del riesgo de 25% ( $(P_c - P_t)/P_c = 0,25$ ); la inferior, de 50%



Hay enfoques estadísticos más formales para investigar la variabilidad de los resultados finales de los estudios, comparando las variaciones observadas a las que se esperaría debidas simplemente al muestreo. Para proporciones, por ejemplo, puede utilizarse una prueba de ji cuadrado ( $\chi^2$ ) de coherencia u homogeneidad, como la que proponen Mantel y Haenszel (24). También pueden usarse técnicas de regresión para investigar si los resultados finales medidos son coherentes o varían (25–27). Sin embargo, como la potencia de estas pruebas estadísticas de homogeneidad a menudo es baja porque la mayor parte de los metanálisis combinan los resultados de un número muy limitado de estudios individuales, los investigadores deben valerse del juicio subjetivo bien fundado y examinar una gráfica de los resultados finales de los estudios para evaluar su coherencia cuando las pruebas estadísticas formales no rechazan la hipótesis de homogeneidad. Recomendamos no confiar en pruebas estadísticas sin más.

La falta de uniformidad de los resultados medidos en los estudios puede reflejar el hecho de que el efecto del tratamiento varía en función de características particulares de los estudios, por ejemplo, las dosis. Si las diferencias de resultados no se deben a variabilidad aleatoria de muestreo alrededor de un solo parámetro poblacional, la representación gráfica podría indicar la influencia de las características del estudio en las diferencias demostradas entre tratamientos. Por ejemplo, la figura 5 muestra que la variabilidad de los resultados finales de los estudios se debe probablemente a que en unos estudios se usó una dosis de tratamiento alta y en otros

**FIGURA 5. Ejemplo de efectos incoherentes de los tratamientos (heterogeneidad), debido a que en unos estudios se usaron dosis altas (A) y en otros, dosis bajas (B). Tasas de eventos representadas en el eje vertical para el grupo de tratamiento y en el eje horizontal para el grupo control. La diagonal continua indica tasas iguales. La línea de puntos superior muestra una reducción del riesgo de 25% ( $[P_c - P_t]/P_c = 0,25$ ); la inferior, de 50%**



una dosis baja. Otro ejemplo puede ser el de los estudios que investigan el efecto en pacientes jóvenes, frente a aquellos en los que se investiga en pacientes mayores. La descripción del efecto a diferentes edades es así una meta importante del metanálisis, que puede alcanzarse basándose en subgrupos de estudios similares. No solo es importante saber si una terapéutica hace más bien que mal; también es importante la cuestión de en qué pacientes y bajo qué circunstancias el tratamiento funciona mejor.

## Combinación de resultados

Cuando se juzga que hay una relativa homogeneidad o constancia de la medida común de resultado final en los distintos estudios, puede obtenerse una medida sumaria del resultado final combinando los resultados. Las técnicas se discuten en el apéndice 2.

Los lectores de informes de metanálisis publicados deben confiar en el revisor estadístico de la revista para evaluar si es o no apropiado el procedimiento de combinación contenido en esos informes. Sin embargo, todos los lectores deberían buscar en la publicación una discusión del tema de la homogeneidad (mediante pruebas gráficas y estadísticas, por ejemplo) previa al uso de procedimientos de combinación que la dan por supuesta. Además, los autores podrían mostrar qué pasa cuando se estudia el efecto de las características de diseño de los estudios sobre las diferencias demostradas en el conjunto de aquellos. Si los criterios de inclusión y exclusión son tan estrechos que no hay variaciones en las características de diseño de los estudios (como en una investigación multicéntrica), esas técnicas no serán necesarias.

Lo que aquí se ha presentado es un marco conceptual para el proceso de análisis estadístico, aplicable a todos los tipos de datos que se encuentran en las investigaciones médicas. Recomendamos a quienes desean realizar metanálisis que consideren los aspectos mencionados y los que se detallan en el apéndice 2, así como consultar con un estadístico la planificación y las etapas de análisis del metanálisis.

## Análisis de la sensibilidad

Siempre habrá opiniones distintas sobre cuál es el método más apropiado para hacer un metanálisis particular, de manera que el investigador debe preguntarse en todos los casos en qué medida sus resultados metanalíticos son sensibles a la forma en que se hizo el metanálisis. Este proceso se denomina análisis de la sensibilidad. Por ejemplo, podemos querer ver cómo varía el resultado combinado cuando se incluyen en el metanálisis estudios aleatorizados o no, en vez de estudios aleatorizados solamente. Esto implica simplemente un segundo análisis llevado a cabo de la manera ya indicada, incluyendo la segunda vez los datos de los estudios no aleatorizados. La comparación de los resultados de los dos metanálisis puede mostrar que el tipo de estudios tiene poca influencia en los resultados del metanálisis, en cuyo caso dejan de tener importancia las discrepancias acerca de los criterios de inclusión; o puede ser que el tipo de estudio sí sea importante a la hora de sacar conclusiones del análisis. El análisis de la sensibilidad utilizado como herramienta en el marco del metanálisis puede contribuir a la resolución de polémicas clínicas, por

ejemplo, la habida sobre la asociación entre la terapéutica con corticosteroides suprarrenales y las úlceras pépticas (28–36). También podría utilizarse, por ejemplo, para mostrar por qué de dos metanálisis sobre ese tema se obtuvieron conclusiones diferentes (28, 37).

En el proceso de metanálisis también pueden descubrirse factores básicos para el análisis de sensibilidad, por ejemplo, las diferencias entre estudios con pacientes jóvenes y estudios con pacientes viejos. El efecto de estos factores puede apreciarse fácilmente en la figura 5. La valoración del efecto de los elementos más subjetivos del metanálisis, como la elección de criterios de inclusión y de exclusión, puede fortalecer las conclusiones metanalíticas.

Un tema que tiene que ver con lo anterior es la medida en que los resultados del metanálisis están potencialmente distorsionados por el sesgo de publicación. Si suponemos que los estudios negativos o los estudios que no muestran diferencia son más difíciles de publicar que los estudios positivos, y puede que esto no sea verdad en todos los casos, hay que ver si este tema tiene importancia cuando el metanálisis muestra diferencias significativas entre los grupos de tratamiento. Rosenthal (38) ha indicado una fórmula para estimar el número de estudios no publicados necesario para convertir una diferencia combinada estadísticamente significativa en una diferencia no significativa. Si ese número es pequeño en relación con el de estudios publicados que se han combinado en el metanálisis, hay razones para preocuparse por la influencia de un sesgo de publicación. Si el número es grande, es posible preocuparse menos del potencial sesgo de publicación. Desgraciadamente, el método descrito por Rosenthal no especifica el tamaño de los estudios no publicados, es decir, el número de pacientes en cada grupo de tratamiento.

Otro método para estimar el sesgo de publicación potencial es el ideado por Begg (39), que requiere estimar el tamaño de la población de la que se han extraído los pacientes de los estudios publicados. Mediante esa información y el tamaño muestral, puede calcularse el potencial máximo de sesgo de publicación para cada estudio; Begg proporciona una tabla para facilitar los cálculos. La limitación del método de Begg es que exige conocer la incidencia de enfermedad en una población dada y la proporción de personas elegibles y de quienes consentirían participar en un estudio hipotético.

Para valorar el posible sesgo de publicación nosotros hemos ideado un método más sencillo en el que se supone que existe un solo estudio aleatorizado negativo [no publicado] y se hace variar el tamaño de los grupos de tratamiento hasta que el valor  $P$  excede de 0,05 (o sea, hasta que no hay significación estadística) al añadir ese estudio a la serie de resultados combinados en el metanálisis. De manera similar, pueden suponerse varios estudios relativamente pequeños (por ejemplo, con 20 pacientes en cada grupo de tratamiento) para determinar cuántos estudios no publicados que no muestren diferencia en las tasas de letalidad son necesarios para que el valor  $P$  combinado exceda de 0,05. Si se necesita un estudio muy grande (como de unos 200 pacientes por grupo) o muchos pequeños, es mejor no darle mucha importancia al posible sesgo de publicación.

Que nosotros sepamos, no se ha discutido nunca lo opuesto al sesgo de publicación, es decir, la publicación selectiva de estudios negativos. Sin embargo, los resultados de un estudio publicado en este mismo número de *Annals of Internal Medicine* (40) muestran que los estudios publicados como *abstracts* que no fueron posteriormente publicados como artículos mostraron diferencias mayores que los estudios publicados como artículos.

## Ejemplos de metanálisis

*Meta-Analysis in Social Research*, de Glass, McGaw y Smith (41), presenta el metanálisis como un instrumento valioso de investigación, subrayando su aplicabilidad en ciencias sociales. En *Evaluation Studies Review Annual*, Vol 8 (42) y en *Meta-Analytic Procedures for Social Research* (43) se encuentran temas metodológicos y varios ejemplos de metanálisis aplicados principalmente a la psicología y ciencias de la educación. *Summing Up*, de Light y Pillemer (44), proporciona una perspectiva general del proceso metanalítico y un formulario de evaluación muy útil en el que constan preguntas de revisión tales como "¿cuál es el objetivo concreto de la revisión?" o "¿cómo fueron seleccionados los estudios?". Louis et al. (45) han recopilado los resultados relevantes derivados de metanálisis en temas de salud pública. En farmacología se ha publicado una revisión reciente de metanálisis, con un ejemplo ilustrativo de su uso para valorar la efectividad de la quimopapaína (46).

En las revistas clínicas, además de los metanálisis ya citados se han publicado otros sobre el tratamiento adelgazante y sus efectos sobre la hipertensión esencial (47), la terapéutica con betabloqueantes a largo plazo tras el infarto de miocardio (48) y el contraste entre los efectos de la estreptoquinasa y la nitroglicerina intracoronarias en pacientes con infarto de miocardio agudo (49). En metanálisis recientes se ha evaluado la mortalidad, los infartos repetidos y los efectos colaterales en pacientes con infarto de miocardio en fase aguda tratados con fibrinolíticos intravenosos e intracoronarios (50); se ha estudiado la prevención de la preeclampsia y sus secuelas con diuréticos (51); se han medido los cambios de los lípidos séricos y las lipoproteínas en el entrenamiento físico (52); se han examinado los efectos del tratamiento con antihistamínicos H<sub>2</sub> en la hemorragia aguda gastroduodenal (53), y se han estimado los resultados en cuanto a calidad de la vida en los ensayos clínicos de cirugía anastomótica de injerto coronario (54).

Sacks y Chalmers (2) han presentado un metametanálisis de más de 30 informes de metanálisis que combinaron los resultados de cientos de estudios aleatorizados. Estos autores propugnan una mejor metodología en los metanálisis y señalan los defectos en protocolos, búsqueda de trabajos publicados, evaluación de la calidad de los estudios y técnicas de síntesis de resultados. Basándonos en nuestra discusión del trabajo de Sacks y Chalmers (2), proponemos una lista de revisión de siete puntos (cuadro 1) para evaluar la calidad del metanálisis.

### CUADRO 1. Puntos para evaluar la calidad de un metanálisis

- 
- ¿Hay constancia de que se usó un protocolo de trabajo?
  - ¿Se describen explícitamente los métodos para buscar investigaciones publicadas?
  - ¿Se especifican los criterios de inclusión y de exclusión, los artículos incluidos y excluidos y las razones para la exclusión?
  - ¿Hay gráficos y pruebas de homogeneidad?
  - ¿Se han empleado las estadísticas y los análisis de sensibilidad apropiados?
  - Si el análisis combinado muestra diferencias significativas, ¿se discute el tema del sesgo de publicación?
  - ¿Se extraen conclusiones respecto a recomendaciones sobre el tratamiento (beneficioso, dudoso, perjudicial) y orientaciones para futuros estudios?
-

## DISCUSIÓN

Se han indicado algunos temas de diseño y aspectos estadísticos del proceso de combinación de resultados de estudios, es decir, del instrumento de investigación denominado metanálisis. Este tipo de análisis es útil para los investigadores clínicos que desean evaluar estrategias terapéuticas basadas en varios estudios previos, así como para quienes planifican ensayos clínicos. Los métodos aquí propuestos representan una estrategia explícita y reproducible para sintetizar la información, frente al enfoque narrativo tradicional.

Diversos aspectos del metanálisis han sido objeto de críticas, provenientes principalmente de los campos de la psicología y las ciencias de la educación, donde ha sido más utilizado. Se ha dicho que el metanálisis no tiene en cuenta la calidad de los estudios, que solo deberían combinarse los estudios "buenos" (55, 56). Aquí se ha mostrado cómo, mediante el cambio de los criterios de inclusión y exclusión, el establecimiento de una correlación entre la tasa de efectividad mostrada por los estudios individuales y las puntuaciones de calidad, o el uso de una regla de ponderación al combinar los resultados, el metanálisis puede incorporar la valoración de la calidad de los estudios.

Una segunda crítica importante se refiere a que es ilógico combinar los resultados de estudios con diferentes procedimientos de medición y realizados en distintos momentos en diferentes tipos de pacientes (55, 57-59). Este problema se presenta en cualquier tipo de revisión de la literatura científica, como ya hemos señalado. El metanálisis ofrece la oportunidad de explorar los efectos de estas diferencias a través de la comparación de la magnitud del efecto del tratamiento en los distintos estudios y en el resultado general del metanálisis. Mediante el análisis de sensibilidad, los estudios con diferentes características de diseño pueden incluirse o excluirse del metanálisis y puede examinarse el efecto de combinar resultados de estudios potencialmente distintos.

Un tercer problema es el potencial sesgo de selección que plantean las investigaciones publicadas. Este también es un problema de cualquier revisión y aquí se han enunciado diversos métodos aún en perfeccionamiento para estimar el sesgo de publicación.

Además de estas críticas publicadas, hay otras limitaciones potenciales del metanálisis. En primer lugar, la validez del metanálisis depende de que la información en los artículos publicados sea completa y exacta. Los estudios de las publicaciones muestran que a menudo no se notifica información importante (22, 60, 61), lo cual impide responder a las preguntas planteadas en el protocolo metanalítico (54). Esto no es un problema del metanálisis en sí, pero apunta a la necesidad de que las publicaciones clínicas proporcionen información más estandarizada y más completa (22). En segundo lugar, aunque la combinación de los resultados incrementa la potencia estadística respecto a cada estudio original, la validez relativa de las distintas técnicas estadísticas disponibles para la combinación de resultados no está bien estudiada y es necesario prestar más atención a esos temas de investigación. En un metanálisis reciente en el que se usaron varias de las técnicas previamente citadas como adecuadas para la combinación de resultados, se obtuvieron estimaciones cuantitativas similares y conclusiones cualitativas idénticas (54).

El metanálisis puede ser menos convincente que un solo estudio aleatorizado, de gran tamaño y de "alta calidad". Sin embargo, en la práctica, puede que no se haya hecho tal tipo de estudio, lo cual obliga a evaluar las publicaciones exis-

tentes. Además, el tamaño muestral o el tiempo necesario para llevar a cabo un estudio aleatorizado a gran escala puede ser prohibitivo, especialmente cuando hay que tomar decisiones reglamentarias a pesar de los defectos de las investigaciones publicadas. Al valorar la posible causalidad en una asociación (62), los resultados similares en poblaciones distintas bajo circunstancias distintas que puede mostrar el metanálisis pueden reforzar la idea de la relación causal. En nuestra opinión, el perfeccionamiento del metanálisis convencerá a los escépticos de que se trata de un instrumento valioso para la investigación y de que su uso se justifica porque ahorra tiempo y dinero. Sin embargo, sigue en pie la cuestión de cuándo los resultados del metanálisis son suficientes para resolver una controversia clínica y cuándo se necesita un estudio prospectivo a gran escala para obtener una respuesta definitiva. Incluso cuando este último resulta necesario, el metanálisis de los estudios previos puede ser útil para planificarlo.

A medida que hay más y más estudios publicados y no publicados, aumenta la necesidad de síntesis sistemática de los resultados de los estudios. El metanálisis puede proporcionar una integración de los resultados de la investigación con una metodología explícita, cuantitativa y más rigurosa que el método tradicional de revisión narrativa. Los investigadores deberían llegar a un acuerdo general sobre cómo llevar a cabo los metanálisis y mejorar el desarrollo de protocolos para el metanálisis mediante una conferencia de consenso sobre métodos de revisión de la literatura. La aplicación del metanálisis a la investigación clínica debe producir una metodología diferente de la que se ha empleado en los ramos aplicados de la educación o la psicología. Consideramos también necesario un registro de ensayos clínicos en progreso, gestionado por una entidad oficial tal como los Institutos Nacionales de Salud o la Organización Mundial de la Salud, que no solo serviría como red informal para el intercambio de noticias sobre la investigación sino que proporcionaría también medios para estimar el sesgo de publicación. Las instituciones podrían exigir el registro de todos los ensayos clínicos antes de dar a los investigadores su aprobación de estudios con sujetos humanos.

El mejoramiento de la revisión de investigaciones puede llevar a avances del conocimiento. El perfeccionamiento y la evolución ulterior de la aplicación del metanálisis a la combinación de resultados de estudios clínicos puede ser valiosa en ese proceso.

## APÉNDICE 1

En el caso más simple, el resultado que se evalúa solo tiene dos categorías (dicotómico), por ejemplo, vivo o muerto. Usaremos  $P_c$  para representar la proporción de eventos en el grupo control de terapéutica estándar (número de fallecimientos dividido por el total de personas en el grupo, por ejemplo) y  $P_t$  para representar la proporción de eventos en el grupo terapéutico experimental. Se muestran como ejemplo los cálculos para  $P_c = 50/100 = 0,5$  y  $P_t = 40/100 = 0,4$ .

La primera forma de expresar las diferencias entre terapéuticas (efectividad) es mediante la diferencia absoluta de frecuencias de cierto resultado final en los dos grupos de pacientes. En la notación propuesta, la diferencia de riesgos para un resultado final categórico será  $P_c - P_t$ . En el ejemplo mencionado,  $0,5 - 0,4 = 0,1$ .

La segunda medida se basa en la razón de las frecuencias del resultado final en los dos grupos de pacientes. La razón de riesgos (o riesgo relativo) será  $P_t/P_c = 0,4/0,5 = 0,8$ .

De donde se deduce que el riesgo del resultado final no buscado (el fallecimiento) fue en el grupo experimental del 80% respecto al grupo control.

Una tercera medida es la reducción de riesgo (proporcional),

$$\frac{P_c - P_t}{P_c} = \frac{0,5 - 0,4}{0,5} = 0,2.$$

De donde deducimos que la terapéutica experimental reduce el riesgo de sufrir un resultado no deseado en un 20% respecto a la terapéutica de control. A partir del riesgo relativo puede calcularse la reducción de riesgo, mediante la fórmula:

$$1 - 1/(\text{riesgo relativo}) = \text{reducción del riesgo}.$$

Una cuarta medida es el valor inverso de la diferencia de riesgo  $P_c - P_t$ , es decir

$$\frac{1}{0,5 - 0,4} = 10$$

que es el número de pacientes que habrían de ser tratados con la terapéutica experimental para evitar un resultado no deseado, el "número a tratar" (63).

El método para expresar diferencias combinadas referentes a datos categóricos puede venir dado por la elección del procedimiento estadístico de combinación de resultados (por ejemplo, la estimación de Mantel-Haenszel o la de máxima verosimilitud dan una razón de posibilidades (*odds ratio*) que constituye una buena aproximación al riesgo relativo, mientras que los modelos de componentes de la varianza proporcionan una estimación de la diferencia de riesgos.

## APÉNDICE 2

En las publicaciones sobre metanálisis en ciencias sociales se han sugerido muchas técnicas para combinar resultados. De ellas, la denominada votación da una idea de la dirección, pero no del tamaño muestral ni de la magnitud del efecto combinado (64, 65). Mediante este método se compara el número de estudios con efectos positivos estadísticamente significativos con el número de estudios con efectos negativos estadísticamente significativos. También se ha sugerido combinar los valores  $P$  (15, 66), pero en ese caso no se consideran ni el tamaño muestral ni la magnitud del efecto. Los procedimientos de cálculo de promedios (41) se han usado para estimar la magnitud del efecto, pero estas técnicas han sido criticadas tanto desde el punto de vista estadístico como conceptual (67).

Cuando se dispone de datos categóricos en bruto de un pequeño número de estudios que se consideran homogéneos, puede obtenerse un resultado global mediante la estimación del riesgo relativo combinado de Mantel-Haenszel (24). Esta técnica no plantea dificultades de cálculo y es aplicable a muchos campos de la investigación clínica. Se suman las tasas de ocurrencia de cierto resultado en los distintos estudios, como si fueran los resultados de los distintos centros en un estudio multicéntrico. Puede usarse el estadístico  $\chi^2$  con un grado de libertad para juzgar si es significativa la medida descriptiva global del efecto del tratamiento; el error estándar de la estimación permite calcular un intervalo de confianza para el riesgo relativo. No obstante, basándonos en los avances en métodos estadísticos y en tecnología computacional, nosotros hemos obtenido mediante una regresión lo-

gística una estimación de máxima verosimilitud de la razón de posibilidades combinada, que es un valor aproximado del riesgo relativo (25–27).

Respecto al método de Mantel-Haenszel, las ventajas de la regresión logística representan la base teórica para esta técnica, que permite controlar simultáneamente la influencia de características del diseño del estudio tales como las dosis terapéuticas o la edad de los pacientes, que hipotéticamente podrían tener relación con las diferencias manifestadas entre los tratamientos. Es decir, que la regresión logística permite incluir estas características en un modelo, como variables independientes en la ecuación de regresión con la que se estiman los efectos ajustados de los tratamientos. (Estas variables se denominan a veces covariables.<sup>3</sup>) La técnica de Mantel-Haenszel no permite la inclusión de varias covariables. Así, cuando el supuesto de homogeneidad se rechaza, la regresión logística puede usarse para buscar diferencias sistemáticas entre estudios. Si el supuesto de homogeneidad se rechaza y los modelos logísticos no producen resultados convincentes para explicar la heterogeneidad basada en diferencias sistemáticas en las variables de diseño de los estudios, en nuestra opinión debe acudir a un análisis de componentes de la varianza (68, 69), como hicieron Clark et al. (70) para evaluar la efectividad de los agentes preventivos de la caries.

La estimación del efecto del tratamiento puede complementarse mediante sistemas de ponderación basados en la precisión de las estimaciones, en la importancia relativa o la calidad de los estudios o en una población de referencia que se utiliza para estandarizar los resultados. Los sistemas de ponderación pueden usarse aislados o combinados. Por ejemplo, la estimación de Mantel-Haenszel del riesgo relativo combinado usa un enfoque basado en la precisión que tiene en cuenta el tamaño muestral de cada estudio. Stampfer et al. (1) usaron este sistema de ponderación en su estudio de la estreptoquinasa intravenosa en la fase aguda del infarto de miocardio. Estos autores calcularon un promedio ponderado de razones de riesgo, asignando a cada estudio un peso proporcional a su precisión (valor inverso de la varianza de la razón de riesgos). En la regresión logística, el peso basado en los tamaños muestrales se incorpora automáticamente. Klein et al. (71) usaron un conjunto de puntuaciones de calidad para ponderar los resultados de cada estudio en un metanálisis de la nutrición parenteral total en ensayos clínicos oncológicos. A nuestro parecer, en los metanálisis futuros probablemente será más habitual el empleo de sistemas de ponderación como los mencionados.

Hay algunos procedimientos estadísticos que no deben usarse en el metanálisis. Por ejemplo, no suele ser adecuado sumar los valores de  $ji$  cuadrado individuales y luego sustraer el valor de  $ji$  cuadrado global (23). Puede ser problemático usar ciertos métodos estadísticos por su falta de sensibilidad a la acumulación de indicios de asociación, cuando los tamaños muestrales diferentes de cada estudio no se tienen adecuadamente en cuenta, es decir, cuando se da demasiado peso a las desviaciones individuales de estudios de pequeño tamaño muestral (72, 73).

## AGRADECIMIENTO

Los autores agradecen a John McLaughlin, David Andrews y Harvey Risch sus valiosos comentarios; a Desiree Chanderbhan la preparación del manuscrito, y a los revisores los comentarios que mejoraron una versión anterior del artículo.

<sup>3</sup> En el original inglés *covariates*, término para el que no existe una traducción estándar aceptada. A veces se encuentra traducido por "covariados" o "covariantes" (N. del t.).

Trabajo parcialmente financiado con el Fondo HS-05374-01 del Centro Nacional para la Investigación en Servicios de Salud y Evaluación de la Tecnología Sanitaria y por el Programa de Investigación y Desarrollo del Servicio Nacional de Salud (Canadá) a través de una beca doctoral otorgada al Dr. L'Abbé y un subsidio académico de investigación al Dr. Detsky.

## REFERENCIAS

1. Stampfer MJ, Goldhaber SZ, Yusuf S, Peto R, Hennekens CH. Effect of intravenous streptokinase on acute myocardial infarction: pooled results from randomized trials. *N Engl J Med* 1982; 307:1180–1182.
2. Sacks HS, Chalmers TC. Meta-analysis of randomized control trials (RCT's). *Clin Res* 1985;33:525A.
3. Gilbert JP, McPeck B, Mosteller F. Progress in surgery and anesthesia: benefits and risks of innovative therapy. En: Bunker JP, Barnes BA, Mosteller F, eds. *Costs, risks and benefits of surgery*. New York: Oxford University Press; 1977.
4. Baum ML, Anish DS, Chalmers TC, Sacks HS, Smith H Jr, Fagerstrom RM. A survey of clinical trials of antibiotic prophylaxis in colon surgery: evidence against further use of no-treatment controls. *N Engl J Med* 1981;305:795–799.
5. Felson DT, Anderson J. Evidence for the superiority of immunosuppressive drugs and prednisone over prednisone alone in lupus nephritis: results of a pooled analysis. *N Engl J Med* 1984;311:1528–1533.
6. Thompson MS. Decision analytic determination of study size: the case of electronic fetal monitoring. *Med Decis Making* 1981;1:165–179.
7. Weinstein MC. Cost-effective priorities for cancer prevention. *Science* 1983;221:17–23.
8. Detsky AS. Using economic analysis to determine the resource consequences of choices made in planning clinical trials. *J Chron Dis* 1985;38:753–765.
9. Hewitt P, Chalmers TC. Using MEDLINE to peruse the literature. *Controlled Clin Trials* 1985;6: 75–83.
10. Hewitt P, Chalmers TC. Perusing the literature: methods of accessing MEDLINE and related data bases. *Controlled Clin Trials* 1985;6:168–177.
11. Haynes RB, McKibbin KA, Walker CJ, et al. Computer searching of the medical literature: an evaluation of MEDLINE search systems. *Ann Intern Med* 1985;103:812–816.
12. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. *Controlled Clin Trials* 1985;6:306–317.
13. Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature: a comparison of MEDLARS and manual methods. *Controlled Clin Trials* 1985;6: 271–279.
14. Greenwald A. Consequences of prejudices against the null hypothesis. *Psychol Bull* 1975; 82:1–20.
15. Rosenthal R. Combining results of independent studies. *Psychol Bull* 1978;85:185–193.
16. Staquet MJ, Rozenzweig M, Von Hoff DD, Muggia FM. The delta and epsilon errors in the assessment of cancer clinical trials. *Cancer Treat Rep* 1979;63:1917–1921.
17. Smith ML. Publication bias and meta-analysis. *Eval Educ* 1980;4:22–24.
18. Simon R. Randomized clinical trials and research strategy. *Cancer Treat Rep* 1982;66:1083–1087.
19. Zelen M. Strategy and alternate randomized designs in cancer clinical trials. *Cancer Treat Rep* 1982;66:1095–1100.
20. Fleming TR. Historical controls, data banks, and randomized clinical trials in clinical research: a review. *Cancer Treat Rep* 1982;66:1101–1105.

21. Wortman PM, Yeaton WH. Synthesis of results in controlled clinical trials of coronary bypass graft surgery. En: Light RJ, ed. *Evaluation Studies Review Annual, Volume 8*. Beverly Hills, California: Sage Publications; 1983.
22. Chalmers TC, Smith H Jr, Blackburn B, et al. A method for assessing the quality of a randomized control trial. *Controlled Clin Trials* 1981;2:31–49.
23. Fleiss JL. *Statistical methods for rates and proportions*. 2ª ed. New York: John Wiley & Sons; 1981.
24. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22:719–748.
25. Davis W, Breslow NE, Day NE. *Statistical methods in cancer research. Vol. I: The analysis of case-control studies*. Lyon, France: IARC Scientific Publications No. 32; 1980.
26. Cox DR. *Analysis of binary data*. London: Chapman and Hall Ltd.; 1970.
27. McCullagh P, Nelder JA. *Generalized linear models*. London: Chapman and Hall Ltd.; 1983.
28. Conn HO, Blitzer BL. Nonassociation of adrenocorticosteroid therapy and peptic ulcer. *N Engl J Med* 1976;294:473–479.
29. Green SB, Gail MH, Byar DP. Steroids and peptic ulcer [Carta]. *N Engl J Med* 1976;294:1291.
30. Long WB, Shiau Y. Steroids and peptic ulcer [Carta]. *N Engl J Med* 1976;294:1291.
31. Grass J. Steroids and peptic ulcer [Carta]. *N Engl J Med* 1976;294:1292.
32. Sturdevant RA. Steroids and peptic ulcer [Carta]. *N Engl J Med* 1976;294:1292.
33. Lubin A. Steroids and peptic ulcer [Carta]. *N Engl J Med* 1976;294:1292.
34. McDonald CJ. Steroids and peptic ulcer [Carta]. *N Engl J Med* 1976;294:1292.
35. Cooke AR. Steroids and peptic ulcer [Carta]. *N Engl J Med* 1976;294:1292–1293.
36. Conn HO, Blitzer BL. Steroids and peptic ulcer [Carta]. *N Engl J Med* 1976;294:1293.
37. Messer J, Reitman D, Sacks HS, Smith H Jr, Chalmers TC. Association of adrenocorticosteroid therapy and peptic-ulcer disease. *N Engl J Med* 1983;309:21–24.
38. Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979;86:638–641.
39. Begg CB. A measure to aid the interpretation of published clinical trials. *Stat Med* 1985;4:1–9.
40. Detsky AS, Baker JP, O'Rourke K, Goel V. A meta-analysis of perioperative parental nutrition. *Ann Intern Med* 1987;107:195–203.
41. Glass GV, McGaw B, Smith ML. *Meta-analysis in social research*. Beverly Hills, California: Sage Publications; 1981.
42. Light RJ, ed. *Evaluation Studies Review Annual, Volume 8*. Beverly Hills, California: Sage Publications; 1983.
43. Rosenthal R. *Meta-analytic procedures for social research*. Beverly Hills, California: Sage Publications; 1984.
44. Light RJ, Pillemer DB. *Summing up: the science of reviewing research*. Cambridge, Massachusetts: Harvard University Press; 1984.
45. Louis TA, Fineberg HV, Mosteller F. Findings for public health from meta-analyses. *Annu Rev Public Health* 1985;6:1–20.
46. Einaron TR, McGhan WF, Bootman JL, Sabers DL. Meta-analysis: quantitative integration of independent research results. *Am J Hosp Pharm* 1985;42:1957–1964.
47. Hovell MF. The experimental evidence for weight-loss treatment of essential hypertension: a critical review. *Am J Public Health* 1982;72:359–368.
48. Bassan MM, Shalev O, Eliakim A. Improved prognosis during longterm treatment with beta-blockers after myocardial infarction: analysis of randomized trials and pooling of results. *Heart Lung* 1984;13:164–168.

49. Rentrop KP, Feit F, Blanke H, et al. Effects of intracoronary streptokinase and intracoronary nitroglycerin infusion on coronary angiographic patterns and mortality in patients with acute myocardial infarction. *N Engl J Med* 1984;311:1457–1463.
50. Yusuf S, Collins R, Peto R, et al. Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results on mortality, reinfarction, and side-effects from 33 randomized controlled trials. *Eur Heart J* 1985;6:556–585.
51. Collins R, Yusuf S, Peto R. Overview of randomized trials of diuretics in pregnancy. *Br Med J* 1985;290:17–23.
52. Vutran Z, Weltman A. Differential effects of exercise on serum lipid and lipoprotein levels seen with changes in body weight. *JAMA* 1985;254:919–924.
53. Collins R, Langman M. Treatment with histamine H2 antagonists in acute upper gastrointestinal hemorrhage. *N Engl J Med* 1985;313:660–666.
54. Wortman PM, Yeaton WH. Cumulating quality of life results in controlled trials of coronary artery bypass graft surgery. *Controlled Clin Trials* 1985;6:289–305.
55. Eysenck JH. An exercise in mega-silliness [Carta]. *Psychol Bull* 1978;85:517.
56. Eysenck JH. Correspondence. *Bull Br Psychol Soc* 1978;31:56.
57. Gallo PS. Meta-analysis—a mixed meta-phor. *Am Psychol* 1978;33:515–517.
58. Presby S. Overly broad categories observe important differences between therapies. *Am Psychol* 1978;33:514–515.
59. Goldman L, Feinstein AR. Anticoagulants and myocardial infarction: the problems of pooling, drowning, and floating. *Ann Intern Med* 1979;90:92–94.
60. DerSimonian R, Charlotte J, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;306:1332–1337.
61. Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. *Controlled Clin Trials* 1984;5:328–347.
62. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
63. Laupacis A, Sackett DL. The number to treat (NTT): an alternative measure of efficacy [Resumen]. *Clin Res* 1986;34:373A.
64. Mosteller F. Assessing unknown numbers: order of magnitude estimation. En: Fairly WB, Mosteller F, eds. *Statistics and public policy*. Reading, Massachusetts: Addison-Wesley Publishing Co.; 1977.
65. Hedged LV, Olkin I. Vote-counting methods in research synthesis. *Psychol Bull* 1980;88:359–369.
66. Jones LV, Fiske DW. Models for testing the significance of combined results. *Psychol Bull* 1953;50:375–382.
67. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. London: Academic Press Inc.; 1985.
68. Harville DA. Maximum likelihood approaches to variance components estimation and to related problems. *J Am Stat Assoc* 1977;72:320–338.
69. DerSimonian R, Laird N. Combining evidence in clinical trials. *Controlled Clin Trials* 1986;7:177–188.
70. Clark DC, Hanley JA, Stamm JW, Weinstein PL. An empirically based system to estimate the effectiveness of caries-preventive agents: a comparison of the effectiveness estimates of APF gels and solutions, and fluoride varnishes. *Caries Res* 1985;19:83–95.
71. Klein S, Simes J, Blackburn GL. Total parenteral nutrition and cancer clinical trials. *Cancer* 1986;58:1378–1386.
72. Pasternack BS, Mantel N. A deficiency in the summation of chi procedure. *Biometrics* 1966;22:407–409.
73. Mantel N, Brown C, Byar DP. Tests for homogeneity of effect in an epidemiologic investigation. *Am J Epidemiol* 1977;106:125–129.