

COMUNICACIÓN BIOMÉDICA

METANÁLISIS: UN ENFOQUE CUANTITATIVO PARA LA INTEGRACIÓN DE INVESTIGACIONES¹

Stephen B. Thacker²

El metanálisis es un procedimiento cada vez más usado en medicina clínica para mejorar los métodos tradicionales de revisión narrativa mediante acumulación sistemática de información y cuantificación del efecto. La combinación de datos de varios estudios mediante metanálisis puede incrementar la potencia estadística, apuntar alguna relación entre variables y hacer generalizables los resultados con mayor rigor que los métodos de revisión no cuantitativa. Como todos los métodos de revisión, el metanálisis puede resultar limitado por sesgos de selección, datos inadecuados o sesgos de interpretación de los datos. De todas formas, las ventajas mencionadas convierten al metanálisis en un método que vale la pena someter a prueba y evaluación empírica.

El objetivo de una revisión integradora de la literatura es resumir el conocimiento acumulado referente a un campo de interés y aclarar temas importantes que los investigadores dejaron sin resolver (1). El método tradicional de integración en la literatura médica ha sido la revisión narrativa. Un especialista en la materia revisa los estudios, decide cuáles son relevantes y muestra sus conclusiones en cuanto a resultados o, menos a menudo, en cuanto a metodología. También puede proponer temas para investigaciones futuras. Este tipo de revisión narrativa tiene dos debilidades básicas (2, 3). En primer lugar, no hay norma alguna sobre cómo conseguir los datos primarios o cómo integrar los resultados; lo que prima es el criterio subjetivo del revisor. Como consecuencia, no existe un estándar explícito para evaluar la calidad de una revisión. En segundo lugar, el revisor narrativo no sintetiza cuantitativamente los datos hallados en las distintas publicaciones. Así, a medida que aumenta el número de estudios en cualquier disciplina, se incrementa igualmente la probabilidad de que se llegue a conclusiones erróneas en una revisión narrativa (4).

¹ Esta traducción del artículo "Meta-analysis: A quantitative approach to research integration" (*Journal of the American Medical Association*, 1988;259(11):1685-1689) se publica con autorización de la revista y del autor.

² Centro para la Salud Ambiental y el Control de Lesiones, Centros para el Control de Enfermedades, Atlanta. Las separatas de la versión en inglés pueden solicitarse a: S. B. Thacker, M.D., MSc., Center for Environmental Health and Injury Control, Centers for Disease Control, Atlanta, GA 30333, Estados Unidos de América.

La investigación científica se basa en la integración y en la repetición de resultados. Con la posible excepción de un descubrimiento nuevo, un estudio aislado casi nunca hace una contribución decisiva al avance del conocimiento (5). En este artículo se resumen los problemas que afligen a los revisores de la literatura médica y se revisan los métodos alternativos para sintetizar estudios científicos. Examinó en particular el metanálisis, procedimiento cuantitativo para combinar datos, y también muestro un ejemplo de cómo se aplica en la literatura médica. Se describen luego las ventajas y las desventajas del metanálisis, los temas de investigación actual relacionados con el metanálisis y las posibles direcciones de la investigación ulterior.

LIMITACIONES DE LAS REVISIONES DE LA LITERATURA

Cualquier enfoque para la revisión de la literatura tiene limitaciones que pueden resumirse en lo siguiente (6): a) sesgo de selección debido a las costumbres de presentación y publicación de informes; b) falta en los estudios publicados de datos específicos requeridos para la revisión; c) sesgo de exclusión debido a las preferencias del investigador; d) calidad heterogénea de los datos primarios; y e) interpretación sesgada de los resultados. Estos problemas son aplicables a cualquier tipo de revisión de la literatura.

El revisor ha de preocuparse por dos tipos de sesgo en la literatura publicada. En primer lugar, cualquier revisión limitada a estudios publicados tiende a inflar el tamaño del efecto, ya que los autores y directores de revistas son proclives a dar cuenta sobre todo de resultados estadísticamente significativos. Cincuenta y ocho investigadores indicaron en una encuesta que habían llevado a cabo 921 ensayos clínicos aleatorios controlados de los que 96 (21,3%) no se habían publicado. La publicación fue significativamente más probable en el caso de ensayos que habían dado resultado positivo (77% frente a 42%, $P < 0,001$) (7). Además, no se debe dar por sentado que los métodos son mejores en los trabajos publicados, ya que la calidad de estos es muy variable (8). En segundo lugar, otro sesgo de publicación, el sesgo confirmatorio, se refiere a la tendencia de todo investigador a dar crédito y llamar la atención de los demás sobre las experiencias que dan apoyo a su punto de vista, dejando a un lado o desacreditando incluso a las que no lo apoyan. Los resultados de un estudio de 75 revisores científicos de revistas a los que se les pidió evaluar procedimientos experimentales idénticos revelaron escasa coincidencia entre los evaluadores y un sesgo en contra de los resultados contrarios a su propia perspectiva teórica (9). En consecuencia, cuando los datos constituyen una novedad o son impopulares, tienden a ser subnotificados en la literatura publicada.

Los datos disponibles procedentes de estudios primarios de investigación pueden ser inadecuados a juicio del que revisa la literatura. El revisor a menudo encuentra una notificación selectiva de los datos primarios, un análisis incorrecto de los mismos o descripciones inadecuadas de los estudios originales (10). En un estudio sobre las consecuencias de la psicoterapia, la omisión de datos impidió calcular el efecto en 26% de los estudios, lo cual es similar a informes publicados con anterioridad (11).

Además de localizar los estudios, el investigador debe decidir qué trabajos incluir en la revisión. Una posibilidad es usar todos los datos disponibles y maximizar así la representatividad de las conclusiones. No obstante, mediante este procedimiento se disminuirá la validez estadística de la síntesis de los datos por la inclusión de estudios menos rigurosos. Por otro lado, la exclusión de estudios por razones metodológicas incrementará la validez estadística, pero disminuirá el tamaño

del volumen total de datos acumulados y puede sacrificar así la posibilidad de generalizar los resultados.

La limitación más crítica para el revisor es probablemente la calidad variable de los datos. El efecto de la calidad de los datos se vio en un estudio de calidad de vida tras anastomosis coronaria (injerto quirúrgico) en el que los investigadores observaron que las estimaciones de beneficio eran en los ensayos aleatorios controlados 15% menores que en los ensayos con controles apareados (12). De igual modo, los resultados de estudios de atención médica tienden a mostrar una razón de posibilidades (*odds ratio*) cada vez menor cuando se incrementa el rigor de la investigación (8). No obstante, en un gran estudio sobre psicoterapia se vio que el efecto aumentaba cuando se imponían restricciones más estrictas (11). En las revisiones cuantitativas pueden usarse métodos estadísticos incluso análisis estratificados y métodos de múltiples variables para medir el efecto de la variación de la calidad de los estudios en los resultados (8, 13, 14).

Estas limitaciones se conocían ya hace tiempo, pero los intentos más recientes de integración de las investigaciones han llevado a una reconsideración de estos problemas.

ENFOQUES CUANTITATIVOS PARA COMPENDIAR RESULTADOS

Durante los últimos años han surgido varios enfoques diferentes para compendiar datos cuantitativos procedentes de diferentes investigaciones de problemas idénticos o similares. El enfoque más simple para la integración cuantitativa de investigaciones es el recuento de votos. Con este método los resultados de los estudios en consideración se clasifican en tres categorías: a) datos estadísticamente significativos en una dirección; b) datos estadísticamente significativos en dirección opuesta; c) datos sin diferencia estadísticamente significativa. La categoría que obtenga más votos es considerada la más próxima a la verdad (15). Aunque este procedimiento es sencillo, no tiene en cuenta la magnitud del efecto o el tamaño muestral. Además, este enfoque no considera los problemas ya mencionados inherentes a las revisiones tradicionales, tales como metodologías inadecuadas o calidad desigual de los datos.

En 1971, Light y Smith (15) propusieron un método alternativo al de votación. El método propuesto por estos autores aprovecha las agregaciones o grupos naturales de la población. El investigador estudia un problema en varios grupos, por ejemplo barrios o aulas escolares, y busca una explicación de las diferencias entre grupos. Si las diferencias son explicables, los datos pueden combinarse y puede describirse la variabilidad estadística.

Un tercer método para combinar datos de la literatura es el de agregación o acumulación (*pooling*), método por el cual se agregan datos de múltiples estudios de un solo tema, por ejemplo, el uso de β -bloqueantes tras infarto miocárdico (16). Este método está limitado por la disponibilidad de datos en bruto, por la variación en métodos de estudio, poblaciones y variables medidas y por consideraciones estadísticas (17, 18).

En 1976, en un estudio sobre la eficacia de la psicoterapia, Glass (19) acuñó el término *metanálisis*, "análisis estadístico de una serie numerosa de datos aislados de la literatura, realizado con el propósito de integrar los resultados". También puede considerarse *metanálisis* cualquier método sistemático que use análisis estadísticos para combinar datos de estudios independientes en la obtención de una estimación numérica del efecto global de un procedimiento o de una variable particular sobre un resultado específico (20).

Ha habido diversos enfoques del metanálisis, pero en general los componentes del mismo pueden considerarse los siguientes: a) definición del problema y de los criterios para la admisión de investigaciones; b) localización de las investigaciones; c) clasificación y codificación de las características de los estudios; d) cuantificación de las características de los estudios en una escala común; e) agrupación de los datos de las distintas investigaciones y determinación de relaciones entre las características de los estudios (análisis e interpretación); y f) presentación de los resultados (21, 22).

La definición del problema implica una definición explícita de los resultados considerados y de potenciales variables de confusión. Esto permite, cuando se hace cuidadosamente, que el investigador se centre en las magnitudes relevantes de los estudios considerados y pueda determinar métodos relevantes para clasificar y codificar las características de los estudios.

La revisión de la literatura implica un método sistemático para localizar estudios (1). En primer lugar se obtiene información de los colegas y especialistas conocidos en el tema de estudio. En segundo lugar se investigan índices sistemáticos (tales como el *Index Medicus* y el *Social Science Citation Index*), colecciones de *abstracts* (como el *International Pharmaceutical Abstracts*) y bases de datos computadorizadas (como MEDLINE y TOXLINE) para detectar artículos de investigación y fuentes de datos publicados o no publicados. En tercer lugar, se localizan otras referencias a partir de los trabajos publicados encontrados en las dos fases anteriores. A partir de investigadores académicos, privados o de instituciones públicas, pueden localizarse trabajos no citados y datos no publicados, lo cual puede reducir aun más los sesgos de presentación y publicación selectivas.

Para compendiar los resultados de los distintos estudios se utilizan diversos métodos (3, 23). La medida más usada en ciencias sociales es el tamaño del efecto, indicador tanto de la dirección como de la magnitud del efecto de un procedimiento sometido a estudio (19). Glass y sus colegas (24) desarrollaron este método en una evaluación de la eficacia de la psicoterapia a partir de datos de estudios controlados. La diferencia entre las medias de los dos grupos dividida por la desviación estándar del grupo control (o sea, $(X_t - X_c)/S_c$, siendo X_t la media del grupo experimental o expuesto a tratamiento, X_c la del grupo control o no expuesto y S_c la desviación estándar del grupo control) es un indicador del tamaño del efecto para datos cuantitativos. El tamaño del efecto expresa las diferencias en unidades de desviación estándar (DE), de manera que, por ejemplo, si el tamaño del efecto de un estudio son 0,2 unidades de DE, en conjunto, el tamaño del efecto es la mitad del tamaño de otro estudio que tenga un efecto de 0,4 unidades de DE. La medida apropiada del efecto en distintos estudios varía según la naturaleza del problema que se evalúa y la disponibilidad de los datos publicados (7, 25). Por ejemplo, la agregación de datos procedentes de ensayos clínicos controlados ha sido un método muy usado en la literatura médica (16, 26).

El tamaño del efecto para proporciones se ha calculado en la literatura sobre estudios de cohortes bien como diferencia, $P_t - P_c$, bien como razón, P_t/P_c (3). Esta última tiene la ventaja de indicar el cambio en relación al grupo control y en los estudios epidemiológicos es analíticamente equivalente a la razón de riesgos.

Sea cual sea el estadístico que usemos, un procedimiento cuantitativo sistemático para acumular resultados de diversos estudios debe incluir lo siguiente (27): a) los estadísticos descriptivos de los distintos estudios y los promedios adecuados; b) la varianza de un estadístico en los distintos estudios (es decir, pruebas de heterogeneidad); c) la corrección de la varianza restándole el error de muestreo; d) la

corrección de la media y la varianza para tener en cuenta artefactos del estudio no relacionados con el muestreo, por ejemplo, errores de medida; y e) la comparación entre la DE corregida y la media para evaluar el tamaño de la variación potencial en los distintos estudios. En trabajos de metodología estadística recién publicados se abordan los problemas de cálculo del tamaño del efecto o de las pruebas de significación en meta-análisis (28, 29).

VENTAJAS DEL METANÁLISIS

El metanálisis obliga a pensar sistemáticamente acerca de los métodos, resultados, categorizaciones, poblaciones e intervenciones al acumular datos diversos. Además ofrece un mecanismo para estimar la magnitud del efecto en cuanto a tamaño del efecto estadísticamente significativo o razón de posibilidades acumulada (*pooled odds ratio*). Además, la combinación de datos de varios estudios hace los resultados más generalizables y puede incrementar la potencia estadística, lo cual permite una evaluación más completa del procedimiento o de la variable (30). Los métodos cuantitativos aplicados a una gama de estudios también pueden ilustrar el carácter de las relaciones entre las variables, dando también un mecanismo para detectar y explorar las contradicciones aparentes en los resultados. Finalmente, los usuarios del metanálisis han expresado su esperanza de que este enfoque sistemático sea menos subjetivo y reduzca los sesgos atribuibles al investigador.

APLICACIONES DEL METANÁLISIS EN CIENCIAS DE LA SALUD

El interés en las aplicaciones clínicas del metanálisis ha aumentado vertiginosamente en los últimos años (31, 32). Se ha usado el metanálisis en muchos campos no relacionados con la salud mental ni con las ciencias de la educación; por ejemplo, en estudios sobre quimioterapia en el cáncer de mama (33), educación del paciente en medicina clínica (34), manipulación vertebral (35), efecto del ejercicio sobre las concentraciones séricas de lípidos (36) y tratamiento de la úlcera duodenal (37). También se han discutido las aplicaciones potenciales del metanálisis en salud pública (38). Una interesante aplicación del metanálisis fue un intento de cuantificar el efecto —en cuanto a supervivencia y seguridad de las intervenciones— de una amplia gama de innovaciones quirúrgicas y anestesiológicas (39). Más típico es el uso para extraer resultados globales de datos acumulados procedentes de un número reducido de estudios, generalmente ensayos clínicos controlados (26, 40–47). Las técnicas de agregación también se han aplicado a datos de ensayos no aleatorios tanto para estudiar temas cuya investigación ha sido incompleta, como para aumentar la representatividad (25, 48, 49).

ESTUDIO DE UN CASO: MONITOREO ELECTRÓNICO FETAL

El primer intento de evaluar la eficacia y la seguridad del uso sistemático del monitoreo electrónico fetal (MEF) intraparto fue una revisión realizada por Banta y Thacker en 1979 (50). La variable independiente fue la aplicación clínica de cualquier forma de MEF a embarazadas de alto o bajo riesgo; el resultado se midió mediante diversos índices de morbilidad materno-fetal y práctica de cesárea. También se tuvieron en cuenta los costos.

La revisión de la literatura comenzó con el intercambio de información con colegas obstetras, pediatras, epidemiólogos y especialistas en evaluación

tecnológica y economía. A partir del MEDLINE y el *Index Medicus* se obtuvieron referencias de los artículos de investigación publicados, a las que se añadieron las obtenidas a partir de los artículos ya localizados. También se intentaron localizar informes no publicados y *abstracts* de congresos profesionales. La revisión fue sistemática y amplia y se buscó experiencia comparable en los distintos trabajos, pero el análisis cuantitativo de los estudios se limitó a estadísticas descriptivas.

En 1987 se realizó un metanálisis del mismo tema. El estudio se centró en los ensayos clínicos controlados y en la revisión anterior de la literatura suplementada con información procedente de la Base de Datos de Oxford sobre Ensayos Clínicos Perinatales (*Oxford Database of Perinatal Trials*) y correspondencia directa con algunos investigadores (51). Se codificaron las variables y se las hizo comparables cuando fue posible. Por ejemplo, las determinaciones publicadas de la puntuación Apgar variaban en cuanto a momento de observación (1, 2 y 5 minutos) y clasificación (la puntuación 7 se consideraba anormal en unos trabajos y normal en otros); se pidió a los autores que dieran la puntuación Apgar correspondiente al minuto 1 en los casos en los que la puntuación de 7 se consideraba normal.

A continuación los datos primarios se organizaron en cuadros descriptivos, en los que se listaron los datos de cada estudio en lo referente a resultados —puntuaciones Apgar bajas, mortalidad perinatal y práctica de cesárea—, y medidas de precisión diagnóstica —sensibilidad, especificidad y valor predictivo (cuadro 1) (50). Los resultados de los ensayos clínicos controlados se evaluaron en cuanto a comparabilidad y luego se agregaron (cuadro 2), y el análisis de los agregados se estratificó se-

CUADRO 1. Exactitud del monitoreo electrónico fetal valorado mediante puntuaciones Apgar como índice del resultado^a

Investigador y año	No. de pacientes	VPP	VPN	Sensibilidad (%)	Especificidad (%)
Bissonette (69), 1975	714	80	94	57	82
Gaber y Stenchever (70), 1973	749	66	91	84	80
Schiffrin y Dame (71), 1972	307	43	93	54	90
Saldana <i>et al.</i> (72), 1976	620	23	86	71	44
Tipton y Shelley (73), 1971	100	81	93	82	93

^a Se consideró anormal la puntuación Apgar de menos de 7 al minuto (excepto Gabert, que consideró anormales los valores de menos de 6 al minuto). VPP indica valor predictivo positivo y VPN, valor predictivo negativo.

CUADRO 2. Datos agregados de seis ensayos clínicos controlados en los que se valoró la eficacia del monitoreo electrónico fetal sistemático durante el parto

	Razón de posibilidades acumulada ^a	Intervalo de confianza de 95%	Prueba de heterogeneidad ^b
Puntuación Apgar			
< 7	1,07	0,88–1,30	$\chi^2_6 = 1,28$
< 4	0,87	0,56–1,30	$\chi^2_6 = 1,84$
Convulsiones neonatales	0,39	0,14–1,08	$\chi^2_6 = 3,75$
Ingreso en la UCIN ^c	1,01	0,84–1,22	$\chi^2_6 = 12,15$
Muerte perinatal	1,73	0,53–5,64	$\chi^2_6 = 5,51$

^a Pooled odds ratio en inglés (*N. del t.*).

^b Todas las pruebas fueron no significativas. Los subíndices indican grados de libertad.

^c UCIN = unidad de cuidados intensivos neonatales.

gún la calidad de los datos (51). Finalmente se presentaron los resultados del análisis agregado, se dedujeron conclusiones y se hicieron las recomendaciones oportunas.

El estudio de 1979 concluyó que los datos no apoyaban el uso sistemático del MEF y recomendó realizar otros ensayos aleatorios controlados y restringir la técnica a los embarazos de alto riesgo (50). El informe de 1987 incluyó los ensayos aleatorios controlados citados en el estudio original y otros tres ensayos más (en total siete ensayos aleatorios controlados procedentes de cinco países). No se excluyó ningún ensayo clínico conocido, pero se analizó aparte el ensayo clínico de mayor tamaño (52), cuya muestra era mayor que el conjunto de las muestras de los otros seis estudios, y sus resultados se compararon con los resultados acumulados de los otros seis.

El análisis de diferentes subgrupos de estos estudios según diferencias de diseño (por ejemplo, la extracción de la muestra sanguínea del cuero cabelludo fetal) y calidad del estudio detectó pequeñas variaciones de los resultados, pero no hubo cambios en los hallazgos básicos. Sin embargo, los datos del ensayo aleatorio controlado que obtuvo la máxima puntuación en cuanto a calidad de diseño y realización indicaron que el MEF combinado con la toma de muestras de sangre del cuello cabelludo fetal podía usarse para detectar casos de riesgo de convulsiones neonatales (53). La proporción acumulada de nacimientos por cesárea fue dos veces mayor en el grupo sometido a MEF. Estos dos informes muestran las ventajas de un enfoque sistemático y abarcador en una revisión de la literatura, pero los métodos de metaanálisis usados en el informe de 1987 ilustran el incremento de la potencia estadística obtenido de la acumulación de datos y también la información generada por la estratificación de los estudios. Ensayos clínicos ulteriores tuvieron resultados compatibles con los de este metaanálisis (54, 55).

CRÍTICAS DEL METANÁLISIS

Cuando el metaanálisis comenzó a utilizarse en psicología no encontró aceptación universal. Fue calificado de "ejercicio práctico de mega-idiotez" y de "integración abusiva de las investigaciones" (56, 57). Además de las limitaciones ya mencionadas de toda revisión de la literatura, el metaanálisis afronta nuevos problemas en la medida que sintetiza cuantitativamente datos de diversos estudios.

La aplicación de pruebas de significación estadística familiares para la mayoría de los clínicos se basa en la suposición de que los datos están seleccionados al azar a partir de una población bien definida. Esta suposición puede verse violada de dos maneras: por la selección no aleatoria de estudios o por la realización de múltiples pruebas con los mismos datos, bien a partir de la publicación repetida de bases de datos parciales o completas, bien mediante el uso de más de un resultado de cada persona. Sin embargo, la estadística paramétrica que es la que se utiliza habitualmente se considera suficientemente robusta para que pueda ser usable en metaanálisis (58).

El uso actual de métodos estadísticos paramétricos en el metaanálisis exige más estudios teóricos (29). Otros temas metodológicos que han de preocupar a los metanalistas son los sesgos (59), la variabilidad entre estudios (60) y el desarrollo de modelos para medir la variabilidad en distintos estudios (61). La investigación estadística también debe echar luz sobre el efecto de las observaciones extremas aisladas (*outliers*) en el metaanálisis y la ayuda potencial que puede prestar esta técnica para resolver una cuestión de investigación (28). El desarrollo de métodos estadísticos válidos para combinar datos de distintos estudios, incluso datos de estudios de casos y

controles, permitirá al metanálisis maximizar el valor de la síntesis de datos que le es consustancial.

Una preocupación importante referente a la revisión cuantitativa de la literatura es que, aunque el metanálisis es más explícito, puede no ser más objetivo que una revisión narrativa de la literatura (62). Tanto los críticos como los partidarios del metanálisis reconocen que de la cuantificación puede derivarse una falsa sensación de validez científica, en vez de una comprensión real (63, 64). En otras palabras, el metanálisis, como método estadístico "moderno", no mejorará los datos de mala calidad, pero sí puede llevar a una complacencia indebida respecto a las conclusiones que se obtuvieron (65).

EVALUACIÓN DEL METANÁLISIS

La evaluación de una revisión de la literatura, al igual que su realización, ha de ser sistemática y cuantitativa. En la evaluación de un metanálisis ha de exigirse al menos lo siguiente: 1) una clara identificación de los problemas sometidos a investigación; 2) un esfuerzo activo para tener en cuenta todos los estudios disponibles; 3) una valoración del sesgo de publicación; 4) una identificación de los datos usados; 5) selección y codificación basadas en consideraciones teóricas, no en conveniencia; 6) consignación detallada de la codificación; 7) uso de varios evaluadores para valorar la codificación, incluyendo una evaluación de la concordancia entre evaluadores; 8) evaluación de la comparabilidad de los casos, controles y circunstancias en los estudios analizados; 9) consideración en la sección de discusión de otras posibles explicaciones; 10) relación entre las características de los estudios y los problemas que se están revisando; 11) estricta limitación de las generalizaciones al campo de la revisión de la literatura; 12) notificación con detalle suficiente para permitir la replicación del análisis por otro revisor; y 13) orientaciones para investigaciones ulteriores (3, 66).

COMENTARIO

El metanálisis intenta mejorar los métodos tradicionales de revisión narrativa agregando sistemáticamente información y cuantificando su efecto. El metanálisis surgió para tratar el problema de sintetizar gran cantidad de información sobre un tema particular, problema que se ha exacerbado por el gran volumen de investigaciones publicadas en los últimos 20 años. Sin embargo, se trata solamente de un paso en el proceso de perfeccionamiento de metodologías para cuantificar información procedente de múltiples estudios. El metanálisis no debe considerarse la palabra final en la revisión cuantitativa, ni tampoco debe ser despreciado por los problemas y críticas citados. No cabe duda de que es útil obtener, a partir de la revisión sistemática y rigurosa de la información disponible, incrementos en la potencia estadística y capacidad de generalización, llegar a una mejor comprensión de temas complejos o identificar correlaciones entre variables o "lagunas" que haya que cubrir mediante los estudios adecuados.

Al criticar el metanálisis debe distinguirse entre los problemas inherentes a cualquier revisión de la literatura y los que son específicos de este procedimiento. Por ejemplo, la calidad de los datos, los sesgos de muestreo y la localización de estudios pueden ser problemáticos en cualquier revisión de la literatura. De igual manera, la interpretación de los resultados en cada estudio puede verse influida por los diversos métodos de síntesis de los hallazgos de investigación, pero los sesgos no son exclusivos del metanalista. Por otra parte, hay que precaverse del peso inadecuado que se le da a un procedimiento tan solo porque es cuantitativo, particularmente cuando es

usado por quienes no entienden las limitaciones de los métodos estadísticos empleados. Por último, los críticos deberían evaluar empíricamente el efecto de sus críticas, de forma que contribuyera a llevar al metanálisis o a otros métodos de síntesis cuantitativa a un nivel más alto de utilidad.

Se ha sugerido que los investigadores deben combinar revisiones cuanti y cualitativas de los datos para permitir a quienes están en la práctica aplicar los resultados a problemas de pacientes individuales o de programas (67). De esta manera, los investigadores podrían estudiar temas que son importantes pero difíciles de cuantificar. La información no cuantitativa, como la opinión de un especialista y las pruebas anecdóticas, tienen de hecho importantes repercusiones en las prácticas habituales. Una preocupación que hay que señalar finalmente es que, aunque el mejor metanálisis sintetice todos los ensayos disponibles y estudios relevantes disponibles, puede no representar la práctica clínica por el cómo y el dónde se realizó la investigación (63).

Varios métodos pueden ser útiles para probar el metanálisis y mejorar los métodos de revisión cuantitativa. En primer lugar, pueden compararse los resultados de metanálisis con los de revisiones narrativas, para detectar diferencias de interpretación o conclusiones. En un estudio en el que se comparó un procedimiento estadístico para sintetizar hallazgos de investigación con revisiones narrativas, se vio que el revisor estadístico era más proclive a apoyar la hipótesis tanto en cuanto a dirección como en cuanto a magnitud, aunque las recomendaciones básicas no resultaron distintas en ambos grupos (68). Un segundo campo de investigación es la metodología estadística. Se requiere tanto una investigación teórica de los supuestos de métodos alternativos como investigación empírica que demuestre la exactitud y la eficiencia de estos métodos. En tercer lugar, los métodos para valorar metanálisis han de evaluarse y refinarse (66). Finalmente, al valorar el metanálisis, hay que ser cuidadoso y limitarse a extrapolar las conclusiones al campo de estudio cubierto por la revisión de la literatura. Esto vale para cualquier revisión acumulativa, pero los límites de la revisión deben ser cuidadosamente definidos y la interpretación debe circunscribirse a esos límites.

En resumen, la revisión sistemática y cuantitativa de la experiencia acumulada en un campo de investigación es fundamental para la buena práctica científica. El metanálisis es una metodología que puede someterse a prueba y evaluación empírica. Ningún científico que actúe racionalmente podrá evitar la necesidad de usar la información disponible de manera óptima. El marco particular en el que se realiza la revisión —el metanálisis o cualquier otro enfoque— debe considerarse un aspecto científico considerable. La importancia de su estudio ha de ser evidente en un campo como la medicina, en el que la información científica se incrementa de forma exponencial y el potencial para la aplicación de estos hallazgos de investigación no tiene precedentes.

REFERENCIAS

1. Cooper HM. Scientific guidelines for conducting integrative research reviews. *Rev Educ Res.* 1982;52:291–302.
2. Jackson GB. Methods for integrative reviews. *Rev Educ Res.* 1980;50:438–460.
3. Light RJ, Pillemer DB. *Summing up the science of reviewing research.* Cambridge, Mass: Harvard University Press; 1984.
4. Levine RL, Hunter JE. Regression methodology: correlation, meta-analysis, confidence intervals, and reliability. *J Leisure Res.* 1983;15:323–343.

5. Flake DW. The meta-analytic revolution in outcome research. *J Consult Clin Psychol.* 1983;51:65–70.
6. Hattle JA, Hansford BC. Meta-analysis: a reflection on problems. *Aust J Psychol.* 1984;36:239–254.
7. Chan SS, Sacks HS, Chalmers TC. The epidemiology of unpublished randomized control trials, abstracted. *Clin Res.* 1982;30:234A.
8. 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.
9. Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognitive Ther Res.* 1977;1:161–175.
10. Searles JS. A methodological and empirical critique of psychotherapy outcome meta-analysis. *Behav Res Ther.* 1985;23:453–463.
11. Landman JT, Dawes RM. Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. *Am Psychol.* 1982;37:504–516.
12. Wortman PM, Yeaton WH. Cumulating quality of life results in controlled trials of coronary artery bypass graft surgery. *Controlled Clin Trials.* 1985;6:288–305.
13. Chalmers TC, Celano P, Sacks HS, et al. Bias in treatment assignment in controlled clinical trials. *N Engl J Med.* 1983;309:1358–1361.
14. Yeaton WH, Wortman PM. Evaluation issues in medical research synthesis. En: Yeaton WH, Wortman PM, eds. *Issues in data synthesis.* San Francisco: Jossey-Bass Inc Publications; 1984:43–56.
15. Light RJ, Smith PV. Accumulating evidence: procedures for resolving contradictions among different research studies. *Harvard Educ Rev.* 1971;41:429–471.
16. Yusef S, Peto R, Lewis J, et al. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis.* 1985;27:335–371.
17. Elinarson TR, McGhain WF, Bootman JL, et al. Meta-analysis: quantitative integration of independent research results. *Am J Hosp Pharm.* 1985;42:1957–1964.
18. Bryant FB, Wortman PM. Methodological issues in the meta-analysis of quasi-experiments. En: Yeaton WH, Wortman PM, eds. *Issues in data synthesis.* San Francisco: Jossey-Bass Inc Publications; 1984:15–24.
19. Glass GV. Primary, secondary, and meta-analysis of research. *Educ Researcher.* 1976;5:3–8.
20. Leviton LC, Cook TD. What differentiates meta-analysis from other forms of review? *J Pers.* 1981;49:231–236.
21. Ottenbacher KJ, Petersen P. Quantitative reviewing of medical literature: An approach to synthesizing research results in clinical procedures. *Clin Pediatr.* 1983;28:423–427.
22. Curlette WL, Cannella KS. Going beyond the narrative summarization of research findings: the meta-analysis approach. *Res Nurs Health.* 1985;8:293–301.
23. Rosenthal R. Combining results of independent studies. *Psychol Bull.* 1978;85:188–193.
24. Glass GV, McGraw B, Smith ML. *Meta-analysis in social research.* Beverly Hills, Calif: Sage Publications Inc; 1981.
25. Goldsmith JR, Beeser S. Strategies for pooling data in occupational epidemiological studies. *Ann Acad Med Singapore.* 1984;13(suppl 2):297–307.
26. 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.
27. Hunter JE, Schmidt FL, Jackson GB. *Meta-analysis: cumulating research findings across studies.* Beverly Hills, Calif: Sage Publications Inc; 1982.
28. Kiltgaard RF. Identifying exceptional performers. *Policy Analysis.* 1978;4:529–547.
29. Hedges LV, Olkin J. *Statistical methods for meta-analysis.* Orlando, Fla: Academic Press Inc; 1985.

30. Pillemer DB, Light RJ. Synthesizing extremes: how to use research evidence from many studies. *Harvard Educ Rev.* 1980;50:176–195.
31. Proceedings of the Workshop on Methodologic Issues in Overviews of Randomized Clinical Trials. *Stat Med.* 1987;6:217–409.
32. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med.* 1987;107:224–233.
33. Himel HN, Liberati A, Gelber RD, et al. Adjuvant chemotherapy for breast cancer: a pooled estimate based on published randomized control trials. *JAMA.* 1986;256:1148–1159.
34. Mumford E, Schlesinger HJ, Glass GV: The effects of psychological intervention on recovery from surgery and heart attacks: an analysis of the literature. *Am J Public Health.* 1982;72:141–151.
35. Ottenbacher K, Difabio RP. Efficacy of spinal manipulation/mobilization therapy: a meta-analysis. *Spine.* 1985;10:833–837.
36. Tran ZV, Weltman A. Differential effects of exercise on serum lipid and lipoprotein levels seen with changes in body weight: a meta-analysis. *JAMA.* 1985;254:919–924.
37. Christensen E, Juhl E, Tygstrup N. Treatment of duodenal ulcer: randomized clinical trials of a decade (1964 to 1974). *Gastroenterology.* 1977;73:1170–1178.
38. Louis TA, Fineberg HV, Mosteller F. Findings for public health from meta-analysis. *Annu Rev Public Health.* 1985;6:1–20.
39. 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; 1977:124–169.
40. Chalmers TC, Matta RJ, Smith H Jr, et al. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction, *N Engl J Med.* 1977;297:1091–1096.
41. Stewart AL, Reynolds EOR, Lipscomb AP. Outcome for infants of very low birthweight: survey of world literature. *Lancet.* 1981;1:1038–1041.
42. Baum ML, Anish DS, Chalmers TC, et al. 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.
43. Blackburn BA, Smith H, Chalmers TC. The inadequate evidence for short hospital stay after hernia for varicose vein stripping surgery. *Mt Sinai J Med.* 1982;49:383–390.
44. Stamfer MJ, Goldhaber SZ, Yusuf S, et al. Effect of intravenous streptokinase on acute myocardial infarction. *N Engl J Med.* 1982;307:1180–1182.
45. Toth PJ, Horwitz RI. Conflicting clinical trials and the uncertainty of treating mild hypertension. *Am J Med.* 1983;75:482–488.
46. Thacker SB. Quality of controlled clinical trials: the case of imaging ultrasound in obstetrics: a review. *Br J Obstet Gynaecol.* 1985;92:437–444.
47. Colditz GA, Tuden RL, Oster G. Rates of venous thrombosis after general surgery: combined results of randomized clinical trials. *Lancet.* 1986;2:143–146.
48. Eddy DM: The use of confidence profiles to assess tissue-type plasminogen activator. En: Califf RM, Wagner GS, eds. *Acute coronary care 1987.* Boston: Martinus Nijhoff Publishing Co; 1986, cap. 7.
49. Schneider AP. II, Breast milk jaundice in the newborn: a real entity. *JAMA.* 1986;255:3270–3274.
50. Banta HD, Thacker SB. Assessing the costs and benefits of electronic fetal monitoring. *Obstet Gynecol Suru.* 1979;34:627–642.
51. Thacker SB. The efficacy of intrapartum electronic fetal monitoring. *Am J Obstet Gynecol.* 1987;156:24–30.
52. MacDonald D, Grant A, Sheridan-Pereira M, et al. The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring. *Am J Obstet Gynecol.* 1985;152:524–539.

53. Chalmers I. Randomized controlled trials of intrapartum monitoring. En: Thalhammer O, Baumgarten KV, Pollack A, eds. *Perinatal medicine*. Stuttgart: Georg Thieme Verlag; 1979:260–265.
54. Leveno KJ, Cunningham FG, Nelson S, et al. A prospective comparison of selective and universal electronic fetal monitoring in 34 995 pregnancies. *N Engl J Med*. 1986;315:615–619.
55. Luthy DA, Shy KK, Van Bell G, et al. A randomized trial of electronic fetal monitoring in premature labor. *Obstet Gynecol*. 1987;69:687–695.
56. Eysenck HJ. An exercise in mega-silliness. *Am Psychol*. 1978;33:517.
57. Eysenck HJ. Meta-analysis: an abuse of research integration. *J Spec Educ*. 1984;18:41–59.
58. Taveggia TC. Resolving research controversy through empirical cumulation. *Sociol Methods Res*. 1974;2:395–407.
59. Chalmers TC, Levin H, Sacks HS, et al. Meta-analysis of clinical trials as a scientific discipline: I, Control of bias and comparison with large cooperative trials. *Stat Med*. 1987;6:315–325.
60. Chalmers TC, Berrier J, Sacks HS, et al. Meta-analysis of clinical trials as a scientific discipline: II, Replicate variability and comparison of studies that agree and disagree. *Stat Med*. 1987;6:733–744.
61. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials*. 1986;7:177–188.
62. Mintz J. Integrating research evidence: a comentary on meta-analysis. *J Consult Clin Psychol*. 1983;51:71–73.
63. Cook TD, Leviton LC. Reviewing the literature: a comparison of traditional methods with meta-analysis. *J Pers*. 1980;48:449–472.
64. Cooper HM, Arkin RM. On quantitative reviewing. *J Pers*. 1981;49:225–230.
65. Wilson GT, Rachman SJ. Meta-analysis and the evaluation of psychotherapy outcome: limitations and liabilities. *J Consult Clin Psychol*. 1983;51:54–64.
66. Sacks HS, Berrier J, Reitman D, et al. Meta-analysis of randomized controlled trials. *N Engl J Med*. 1987;316:450–455.
67. Light PJ, Pillemer DB. Numbers and narrative: combining their strengths in research reviews. *Harvard Educ Rev*. 1982;52:1–26.
68. Cooper HM, Rosenthal R. Statistical versus traditional procedures for summarizing research findings. *Psychol Bull*. 1980;87:442–449.
69. Bissonnette JM. Relationship between continuous fetal heart rate patterns and Apgar score in the newborn. *Br J Obstet Gynaecol*. 1975;82:24–28.
70. Gabert HA, Stenchever MA. Continuous electronic monitoring of fetal heart rate during labor. *Am J Obstet Gynecol*. 1973;115:919–923.
71. Schifrin BS, Dame L. Fetal heart rate patterns, prediction of Apgar score. *JAMA*. 1972;219:1322–1325.
72. Saldana LR, Schulman H, Yang W. Electronic fetal monitoring during labor. *Obstet Gynecol*. 1976;47:706–710.
73. Tipton R, Shelley T. An index of fetal welfare in labour. *J Obstet Gynecol Br Commonw*. 1971;78:702–706.