

INCLUSIÓN DE LA CALIDAD DE UNA INVESTIGACIÓN EN EL METANÁLISIS*

Miguel Delgado-Rodríguez¹ / María Sillero Arenas²

¹ Cátedra de Medicina Preventiva y Salud Pública. Facultad de Medicina, Universidad de Cantabria. ² Delegación Provincial de la Salud, Jaén

Resumen

El objetivo de la presente revisión es la forma de considerar la calidad de un estudio en la síntesis de información procedente de varias investigaciones. Se ha criticado al metanálisis el que no valora de manera adecuada la calidad de una investigación, lo que puede resultar en que los estudios pobres desplacen a los buenos por el peso de los números. La inclusión de la calidad de una investigación en el metanálisis cuantitativo es un aspecto debatido en la actualidad. Está claro que es necesario determinar la calidad de un estudio. Sin embargo, y aunque existe un cierto grado de consenso en lo que es un buen estudio y cuáles son las características que debe reunir, no hay un protocolo de evaluación unánimemente aceptado. Otros problemas añadidos son la falta de validación de los cuestionarios existentes y la no comprobación de la reproducibilidad de sus resultados. Existen varios procedimientos para incorporar la calidad en el metanálisis: a) es aceptado de manera general que un estudio haya de reunir una calidad mínima y esté libre de sesgos importantes; b) se tiende a rechazar la incorporación de la calidad como criterio de ponderación; c) existen dudas acerca de los análisis gráficos o de regresión en los que se relacione la magnitud del efecto con la calidad global. Puede ser superfluo o erróneo.

Palabras clave: Metanálisis. Calidad.

INCLUSION OF INVESTIGATION QUALITY IN META-ANALYSIS

Summary

The objective of this review is how study quality is taken into account in the synthesis of information from several studies. Meta-analysis has been criticized for failing into account differential quality of evidence in primary studies, allowing poor studies to drive out good science by weight of numbers. Notwithstanding there is presently no agreement on the inclusion of a study quality score in quantitative meta-analysis; on the other hand, agreement does exist on knowing study quality. However, and although there is a certain level of consensus about what a good study is and on what characteristics it should have, no evaluation protocol on study quality is uniformly accepted. Furthermore, there are no data on the reliability and validity of quality questionnaires. There are several procedures to include study quality in meta-analysis: a) it is generally accepted that bias-free studies should be pooled only; b) most researches reject to include quality scores in the weight; and c) no agreement is reached on the use the quality scores as a covariate in graphical methods and regression analysis. It may be superfluous or misleading.

Key words: Meta-analysis; Quality.

Introducción

Se ha criticado al metanálisis el que no tenga en cuenta la calidad diferente de los estudios primarios y que permita que la investigación pobre pueda desbancar los estudios buenos por el simple peso de los números¹. Si los resultados dependen en cierta manera de la

calidad con la que han sido obtenidos, de ello podría sugerirse un sistema de ponderación que de alguna manera tuviera en cuenta la calidad de los estudios que van a ser combinados. La inclusión de la calidad como un parámetro cuantitativo más en el metanálisis constituye uno de los aspectos más debatidos en el momento presente²⁻⁴; no existen, sin embargo, tantas dudas en la utilidad potencial de indicadores de

Correspondencia: M. Delgado-Rodríguez. Cátedra de Medicina Preventiva. Facultad de Medicina, Universidad de Cantabria. Avenida Cardenal Herrera Oria, s/n; 39011 Santander.

* Presentado en parte a la XII Reunión Anual de la Sociedad Española de Epidemiología, Alicante, 1994.

Este artículo fue recibido el 18 de enero de 1995 y fue aceptado, tras revisión, el 17 de mayo de 1995.

calidad en otros aspectos (valoración de la heterogeneidad, por ejemplo). El objetivo del presente trabajo es analizar la valoración de la calidad de una investigación en el contexto de la síntesis de información de diferentes investigaciones, poniendo un énfasis especial en los estudios observacionales, sobre los que existe una gran polémica acerca de si son o no un substrato adecuado de las técnicas de metanálisis^{5,6}.

Antes de plantearse la posibilidad de incluir la calidad de una investigación en el metanálisis de varios estudios, parece lógico, siguiendo el esquema de Detsky y cols.⁷ plantearse dos preguntas: ¿hasta qué punto importa? y ¿cómo se puede medir? A continuación se intentará responder a la forma de incorporarla en las estimaciones de la magnitud del efecto.

¿Hasta qué punto importa la calidad de una investigación?

No es necesario plantearse si la calidad de una investigación es relevante; es algo fuera de toda duda. Dentro de la calidad serían distinguibles dos aspectos: la calidad máxima alcanzable por un diseño y la propia del mismo. No todos los diseños son capaces de proporcionar el mismo grado de fiabilidad en sus resultados. La *US Preventive Services Task Force* clasifica el nivel de validez de las evidencias de los distintos procedimientos utilizados para valorar una maniobra diagnóstica o terapéutica en: I) ensayos clínicos aleatorizados; II-1) ensayos clínicos no aleatorizados; II-2) estudios de cohortes o de casos y controles; II-3) series temporales; y III) estudios descriptivos, opiniones⁸. Esta clasificación no detalla todos los diseños, pero sí sitúa al mismo nivel a los estudios de cohortes y a los de casos y controles. La ordenación que hace Mclure con arreglo a una serie de ejes taxonómicos de los distintos estudios observacionales y experimentales establece un orden más detallado⁹. Coloca por detrás de los estudios de cohortes a los estudios de casos y controles, algo que, según se ha sugerido, *a priori* no debiera ser así si se aplican los principios de causalidad de Hill¹⁰.

Lo anterior tiene gran interés en el metanálisis, dado que con frecuencia se analizan situaciones en las que diferentes aproximaciones metodológicas se utilizan para estudiar la misma asociación. Sirva de ejemplo el estudio de la asociación entre la terapia hormonal sustitutiva en la menopausia y el riesgo de cáncer de mama, en la que hasta 1991 se

habían empleado 23 estudios de casos y controles, 14 de cohortes y un ensayo clínico¹¹. Si se presume, como se deduce de lo comentado en el párrafo anterior, que la calidad de las evidencias no es la misma según el tipo de diseño, ¿es adecuado combinar distintas metodologías y sintetizarlas en un solo resultado? Si se cree que el metanálisis sólo es utilizable en estudios experimentales y no en otro tipo de diseños^{5,12-14}, el problema se reduce. La realidad es, sin embargo, que el metanálisis se aplica también a la combinación de diseños observacionales, y con frecuencia el investigador se plantea la necesidad de combinar los dos diseños más utilizados en epidemiología observacional, los estudios de cohortes y los estudios de casos y controles. La distancia metodológica entre ambos diseños, clásicamente asumida¹⁵, se ve considerablemente reducida cuando se aplican los planteamientos de Olli S. Miettinen¹⁶⁻¹⁸, que hoy parecen bastante claros^{19,20}. Una forma de evitar la respuesta a la pregunta planteada podría ser una argumentación en base a la heterogeneidad: si los resultados son heterogéneos entre los distintos diseños epidemiológicos no podrán ser combinables, no así si los resultados son coincidentes. La realidad es que el tipo de abordaje de una misma población base por diferentes tipos de diseños (cohortes, casos y controles, etc.) produce con frecuencia resultados discordantes^{11,21}. Tampoco ayuda la comprobación del alto número de inconsistencias entre los resultados de los estudios de casos y controles que analizan el mismo tema, por ejemplo, el 36,4% -56 de 154- tras una revisión en Medline durante cinco años²².

La calidad de un estudio afecta a la existencia de sesgos y a la precisión del estimador⁷. Los estudios de alta calidad reducen la probabilidad de que un sesgo justifique los resultados y contribuyen a una mayor precisión en el estimador final del efecto, simplemente por la supresión de la variabilidad adicional introducida por los errores sistemáticos. Los metanálisis que combinan estudios de calidad variable (más frecuentes en epidemiología observacional) pueden incluir investigaciones afectadas por uno o varios errores sistemáticos y aumentar la probabilidad de producción de un error tipo I (concluir que algo funciona cuando en realidad no) o tipo II (concluir que algo no funciona cuando en realidad sí). La falta de precisión, al aumentar la variabilidad de los resultados (mayor varianza), disminuye la potencia estadística (aumento del error tipo II)⁷. Todo esto se traduce en la que la inclusión de estudios de diferente calidad produce una mayor presencia de heterogeneidad del efecto²⁴.

Tabla 1. Protocolo general de valoración de estudios

Diseño del estudio

- ¿El objetivo del estudio está lo suficientemente descrito?
- ¿El diseño del estudio está lo suficientemente descrito?
- ¿El diseño del estudio es adecuado al objetivo?
- ¿El procedimiento de selección de los sujetos está claramente descrito?
- ¿La muestra de sujetos fue apropiada con respecto a la población a la que se pretenden generalizar los resultados?
- ¿El tamaño de la muestra se basó en consideraciones de potencia estadística previas al estudio?
- ¿El diseño del estudio es aceptable?

Realización del estudio

- ¿Se obtuvo una tasa de respuesta satisfactoria (alta)?

Presentación y análisis

- ¿Hay un párrafo en el que se describan de manera adecuada o se den referencias de los métodos estadísticos empleados?
- ¿Son los procedimientos estadísticos empleados apropiados a los datos?
- ¿Fueron usados correctamente?
- ¿La presentación del material estadístico (tablas, gráficos, números) es satisfactoria?
- ¿Se presentan suficientes análisis?
- ¿Se dan intervalos de confianza para los resultados principales?

Valoración global

- ¿Las conclusiones sacadas se justifican de los análisis realizados?
- ¿El artículo es aceptable estadísticamente?

Fuente: Altman²⁶.

¿Cómo se puede medir la calidad de una investigación?

Procedimiento de valoración

Es necesaria la confección, dentro del protocolo de estudio, de un cuestionario de evaluación de la investigación^{1,24}. Sea cual fuere el protocolo que se aplique, es necesario seguir en su valoración una estrategia cuidadosa que evite el sesgo del observador, y que podría acabar en un sesgo de inclusión o de puntuación de la calidad^{1,25}. La estrategia podría ser la siguiente, recomendada por el grupo de Chalmers²⁶. Los aspectos más relevantes en el proceso de recogida de información son garantizar la ausencia de sesgo y el que haya consistencia²⁴. El cuestionario debe ser aplicado por investigadores enmascarados frente a ciertas características del estudio a evaluar, y así no se influyan por ciertos aspectos del mismo. Los datos que no se deben conocer son: la institución ejecutora y el nombre de los autores (su prestigio puede ejercer un cierto grado de presión), y el sentido de la asociación (si hay relación positiva, negativa o nula entre la exposición y el efecto), ya que su conocimiento podría influenciar también la selección. Esto se consigue si

se suprime por otros investigadores esos datos identificadores. Normalmente se fotocopian de manera diferencial (para distorsionar los formatos clásicos de algunas revistas de gran prestigio que son fácilmente identificables) las secciones de Material y Métodos y la de Resultados, tachando en esta última los epígrafes indicativos de los sujetos pertenecientes a los grupos índice y de referencia. Una alternativa interesante en el momento presente sería el uso de los dispositivos de introducción de datos de ordenador mediante barrido óptico (los conocidos *scanners*). Su aplicación haría muy fácil, y sin exigir mucho tiempo, la transformación de los formatos reconocibles de las revistas en una única forma de presentación, que facilitaría además la estandarización en la evaluación.

Se recomienda que los evaluadores no estén involucrados en el proceso de recogida de la información referente a la magnitud del efecto (por ejemplo, riesgo relativo de la asociación entre anti-conceptivos orales y riesgo de cáncer de mama). La familiarización con los resultados, junto a una posible opinión propia sobre la asociación (que puede ir formándose conforme se lee la información), pueden contribuir a la aparición de sesgos del observador. Se considera que es suficiente que sean dos los evaluadores, pero que realicen su tarea con independencia. Una vez concluida, examinarán sus conclusiones de manera conjunta. Si existen discrepancias, deberán intentar solucionarlas entre ellos, y si se mantienen en sus posturas iniciales, la decisión final será tomada por un tercer investigador o consultor^{23,26}.

El procedimiento comentado supone la existencia de un material escrito sobre el que se realiza la evaluación: ya sea en base a la información (reducida) publicada de un estudio y/o una información (potencialmente grande) proporcionada por los autores (especialmente si no ha sido publicado). La variabilidad en la calidad, si se incorporan estudios no publicados, es presumiblemente mayor. La mayoría de los metanálisis se realizan sobre estudios aparecidos en la prensa científica médica. Lo que aparece como artículo en una revista científica es el resultado de un proceso de edición, sin excesiva estandarización²⁷. Primero, los autores detallan aquellos aspectos metodológicos que consideran más relevantes (que pueden no ser todos los que con posterioridad interesen). Segundo, el comité editorial y los evaluadores de la revista tienen la oportunidad de solicitar información adicional sobre las circunstancias de realización del estudio; esta información puede o no ser incluida en el manuscrito finalmente aceptado. Por último, unos evaluadores utilizarán la información contenida en el artículo, una vez editado, para valorar la calidad. En este proceso

Tabla 2. Criterios de evaluación de los principales diseños epidemiológicos observacionales

	Casos y controles	Cohortes
1. Método predeterminado para la selección de los grupos de estudio	+	+
2. Definición clara de la exposición	+	+
3. Recogida de datos no sesgada	+	+
4. Igualdad anamnésica	+	NA
5. Utilización de criterios de exclusión que no provoquen sesgo	+	NA
6. Igual examen diagnóstico	+	+
7. Igual vigilancia médica	+	+
8. Control de factores de confusión		
Igual susceptibilidad demográfica	+	+
Igual susceptibilidad clínica	+	+
9. Evitar sesgo de Berkson	+	NA
10. Cohortes de estreno o inceptión	NA	+
11. Uso de casos incidentes	+	NA

Fuente: Realini y Goldzieher³¹, modificado.

hay demasiados elementos de subjetividad para garantizar el punto central de la evaluación de estudios publicados: que cualquier estudio que analice el mismo tema adquiera un formato similar cuando aparezca en la prensa científica y así ser capaz de satisfacer todas las preguntas de manera equivalente.

Cuestionarios de evaluación

La elaboración de un cuestionario de evaluación tropieza con varias dificultades: tratar de responder a la variabilidad existente en el diseño epidemiológico, la reproducibilidad, y su validez. Existen directrices generales de evaluación aplicables a cualquier tipo de diseño y a cualquier situación. Sirvan de ejemplo las enumeradas por Altman (Tabla 1)²⁸. Este breve protocolo admite una fácil cuantificación (0. no, 1. dudoso, 2. sí). No obstante, el tipo de diseño condiciona que las cuestiones pertinentes para un ensayo clínico, no lo sean para un estudio observacional, y viceversa. Ello ha motivado que se diseñen protocolos de evaluación diferentes, al menos para la epidemiología experimental y la observacional.

Dentro de la epidemiología experimental, el diseño más estandarizable es el ensayo clínico, donde es posible mantener una mayor homogeneidad. Ello ha permitido la elaboración de protocolos de variación. En la prensa científica se han publicado algunos cuestionarios para la evaluación de ensayos clínicos controlados; a pesar de ello no existe un gran con-

senso en lo que constituye un buen ensayo clínico. El cuestionario que ha alcanzado más difusión ha sido el de Chalmers y cols.²⁹. Desde su publicación han aparecido muchos otros.

La mayoría de los cuestionarios elaborados no fueron desarrollados específicamente para su uso en el metanálisis. Su finalidad principal fue proporcionar una lista de preguntas fundamentales que los realizadores de ensayos clínicos y, especialmente, los lectores debían tener en cuenta para hacerse una idea de la calidad del trabajo. En su desarrollo no se aplicaron análisis de correspondencias, factorial o de componentes principales, que hubieran ayudado a dilucidar los apartados con capacidad de discriminación. Esto justificaría los resultados que encontraron Detsky y cols.⁷, de alta correlación entre cuestionarios cortos y largos.

En epidemiología observacional no existe ningún cuestionario maestro, ya que éste debe ser adaptado al problema a estudiar, con una valoración particular de los errores que amenazan a la asociación en estudio, de las mediciones de la exposición y del efecto. No obstante, en la prensa científica existen algunas muestras de los principios a seguir para la evaluación de estudios observacionales. El texto de Elwood³⁰ ofrece de manera detallada las etapas a seguir en esta evaluación. De manera más sucinta, el metanálisis de Realini y Goldzieher³¹ da un ejemplo de valoración cualitativa de estudios observacionales y resume los criterios desarrollados por Feinstein para evaluar este tipo de estudios (Tabla 2). Algo similar sucede con el artículo de Longnecker y cols.³², que ofrece una valoración de estudios observacionales, dentro de un metanálisis de consumo de alcohol y cáncer de mama. Friedenreich y cols.³³ ofrecen un cuestionario cuantificado para valorar estudios de casos y controles que valoran dieta, tras el desarrollo de unos criterios generales de evaluación de estudios de casos y controles y de cohortes³⁴. Siguiendo estas ideas, diseñamos un cuestionario específico y cuantificable para el estudio de la asociación de los anticonceptivos orales y los cánceres de mama y cuello uterino^{35,36}.

La reproducibilidad de un cuestionario de evaluación es el segundo problema. La concordancia que se alcanza entre evaluadores es variable. Detsky y cols.⁷ han encontrado que la correlación intraclase fue de 0,92 en el cuestionario de Chalmers y cols., valorando cuatro revisores cada estudio. Estos resultados son bastante mejores que los obtenidos por el propio grupo de Chalmers en el análisis de fármacos antiinflamatorios no esteroideos (usando el cuestionario diseñado por ellos): hubo un 19% de desacuerdo en la decisión final acerca de la acep-

Tabla 3. Riesgo relativo de cáncer invasivo de cérvix tras consumo de anticonceptivos orales

Tipo de diseño	Todos los estudios (n= 16)		Sólo metodológicamente aceptables (n= 10)	
	RR (IC 95%)	Heterogeneidad (valor p)	RR (IC 95%)	Heterogeneidad (valor p)
Todos	1,1 (0,9-1,3)	<0,001	1,2 (1,1-1,4)	0,056
Corte	0,5 (0,3-0,8)	0,392	-	-
Cohortes	2,1 (1,2-3,8)	0,301	2,1 (1,2-3,8)	0,301
Casos y controles	1,2 (1,0-1,3)	0,016	1,2 (1,0-1,3)	0,174

Fuente: Delgado y cols.²¹ y Sillero Delgado (resultados no publicados).

tación o rechazo en el apartado de Métodos y un 9% en el de Resultados²⁶.

El tercer problema que presenta el desarrollo de un cuestionario de evaluación es su validación. ¿Son falsos los resultados de un estudio valorado como defectuoso metodológicamente? No siempre. Esto sitúa el problema de la validación de un cuestionario de calidad dentro de la validez de escalas en las que no hay un criterio de referencia. La validez del contenido de las preguntas que componen la escala (*content validity*) no suele plantear problemas, todas ellas suelen hacer referencias a aspectos relevantes del diseño relacionados con una mejor recogida de información. Sí existen problemas en la validez de la construcción de la escala (*construct validity*)³⁷. Muchos de los cuestionarios que se han comentado ofrecen una puntuación final de la calidad del estudio. Esto implica que cada apartado que compone el cuestionario ha de ser puntuado. Esto supone para Greenland² la mezcla de hechos objetivos (diseño del estudio, métodos de reunión de información, control de factores de confusión, etc.) con otros subjetivos y arbitrarios, como es la propia puntuación. No se han establecido hasta ahora cuáles deben ser los pesos de las distintas variables que reflejan la calidad de una investigación. Lo anterior motiva que los efectos de variables importantes de calidad pueden quedar confundidos dentro del parámetro global de calidad, hasta tal punto que no se encuentre ninguna relación entre calidad y efecto².

Funciones del cuestionario de evaluación

El cuestionario de calidad puede tener una cuádruple función. En primer lugar, una cualitativa, al discriminar si el resultado final de un estudio está probablemente sesgado o no. Esto es, *determinar un nivel de calidad mínima o umbral para que el estudio sea incluido* en la ponderación. Con ello nos aproximamos a la táctica recomendada por Slavin³⁸, y que Spitzer adopta^{6,39}, de valorar la inferencia en función de la mejor evidencia disponible. Aunque existen algunas

dudas al respecto⁴⁰, creemos que los estudios con sesgos no deberían ser incluidos. Por ejemplo, en un metanálisis de la asociación entre anticonceptivos orales y cáncer de cuello uterino, ¿se debería incluir un estudio en el que no se ajustó en el análisis por ninguna variable relacionada con la conducta sexual, cuando ese cáncer se considera en la actualidad una enfermedad de transmisión sexual y las estimaciones ajustadas difieren notablemente de las crudas? Creemos que no y la aplicación de un protocolo redujo los 51 estudios recogidos sobre la asociación a 18 considerados metodológicamente aceptables^{21,41}. Esta función puede ser esencial en el análisis, ya que ayuda a discriminar la heterogeneidad entre estudios, al condicionar la inclusión a que se cumplan unos requisitos de diseño y análisis por parte de los estudios^{35,42}. En la tabla 3 se puede apreciar un ejemplo de la reducción de heterogeneidad, cuando se suprimen los estudios con probable sesgo, en la asociación entre anticonceptivos orales y cáncer de cuello invasivo. Indirectamente también lo apoya el hecho constatado por Easterbrook y cols.⁴³, en su análisis de sesgo de publicación en Oxford, en el que los estudios más pobres metodológicamente tendían a obtener resultados más significativos. Goodwin y Boyd⁴⁴ comprueban, en su análisis de la relación entre patrón mamográfico y el riesgo de cáncer de mama, que los estudios que cumplen la mayor parte de los requisitos de calidad establecidos obtienen unos valores de riesgo relativos similares (mayor homogeneidad), que no existe cuando se incluyen estudios con diferentes patrones de calidad. El cuestionario de evaluación de la calidad se transforma en estos casos en un filtro; si los criterios utilizados no son objetivos y sugerentes de que indican la "verdad", se puede introducir un sesgo de selección que altere los resultados del estudio.

En segundo lugar, el cuestionario permite *detectar los errores más frecuentes y elaborar las bases de una investigación correcta sobre el tema en estudio*. La mayor parte de los errores se describen en el curso de las distintas investigaciones; cada estudio nuevo intenta superar limitaciones de los ya existen-

tes. Por ejemplo, en dos metanálisis sobre consumo de anticonceptivos orales y los cánceres de mama y cuello uterino, nuestro grupo realizó un análisis de los sesgos más frecuentes^{41,45}, que desembocaron, por ejemplo, en el cáncer de mama, con unas recomendaciones generales metodológicas para realizar en ese terreno una investigación⁴⁶.

En tercer lugar, el cuestionario de calidad permite la *extracción de información referente a las características del diseño*. Estos datos pueden ser utilizados en el análisis con la finalidad de justificar la heterogeneidad que se encuentre entre las distintas estimaciones individuales del efecto. Son muy numerosos los metanálisis que han encontrado que la heterogeneidad puede ser justificada por el empleo de diferentes procedimientos. Por ejemplo, Friedenreich y cols.³³ valoraron la calidad de trece estudios de casos y controles de cáncer de colon y recto y consumo de fibra en la dieta. Globalmente no apreciaron una relación entre calidad y magnitud de la asociación. Sí comprobaron que el uso de un cuestionario validado, junto al uso de datos cualitativos en la estimación de nutrientes, produjo estimaciones más protectoras en la asociación entre fibra en la dieta y riesgo de cáncer colorrectal. Ambos hechos, sugeridos por el metanálisis, están de acuerdo con el nivel de conocimientos existente⁴⁷: son indicaciones de mejor recogida de información, producen un menor error de mala clasificación no diferencial, con lo que el valor de la fuerza de la asociación, al corregirse, se aleja del valor nulo (recuérdese que el sesgo de mala clasificación no diferencial produce un sesgo hacia el nulo).

Si la metodología es ordenable en función de la calidad, esto podría sugerir dos conclusiones: a) el abandono de ciertos procedimientos en el futuro; y b) en el metanálisis del presente, la inclusión de la calidad como criterio de ajuste, siempre y cuando la calidad no esté condicionada por variables que a su vez influyen en el resultado final (el parámetro combinable) y el criterio de ponderación (normalmente el inverso de la precisión del estimador).

Por último, tiene una *función cuantitativa* al poder resumir la calidad general de un estudio en un número (generalmente como un tanto por ciento, siendo el máximo 100). Es preciso recordar aquí que un cuestionario debe estar formado por preguntas con capacidad discriminativa, que no sean redundantes.

¿Cómo incorporar la calidad en las estimaciones?

Se puede realizar de varias formas. En primer lugar, la selección de estudios probablemente no sesgados,

a través de un *umbral mínimo de calidad*, es una manera de introducir la calidad en el metanálisis cuantitativo. Ya se ha comentado en el epígrafe anterior.

En segundo lugar, a través de la puntuación del cuestionario de calidad, se puede *incorporar la calidad como criterio de ponderación*. Existen serias dudas al respecto. En principio podría parecer lógico, ya que se daría un mayor peso o influencia a las investigaciones más correctas metodológicamente. El proceso de revisión adquiere así una importancia crucial, ya que su cuantificación será uno de los parámetros que se tendrán en cuenta en el análisis cuantitativo. La incorporación de la calidad ofrece la ventaja de que aumenta el intervalo de confianza⁵, una cualidad nada desdeñable dado que el metanálisis ejerce un efecto acumulativo sobre la varianza, sea cual sea la dirección del efecto. Esta ventaja potencial puede ser también alcanzada en el análisis cuantitativo con un modelo de efectos aleatorios.

Si se han eliminado los estudios de baja calidad, como parece razonable, ¿sería bueno seguir considerando la ponderación por calidad? Si la calidad mantuviera relación directa con la proximidad al valor "verdadero" del efecto, la respuesta sería afirmativa. El problema radica en que partimos de la calidad (muchas veces subjetivamente valorada) para considerar un resultado como válido, cuando el proceso debiera realizarse al contrario. ¿Existe relación entre calidad y "verdad" por encima de un umbral? Probablemente no. Uno de los pocos estudios en el campo de la salud que han incluido la calidad (como tanto por uno) en la ponderación de cada estudio individual ha sido el de Longnecker y cols.³² y ha sido objeto de una fuerte crítica⁵.

Aparte de la determinación de un umbral de calidad (aceptable), de la inclusión de la calidad como criterio de ponderación (polémico), Detsky y cols.⁷ describen otros dos procedimientos para tener en cuenta la calidad de las investigaciones: la *representación gráfica de la magnitud del efecto frente a la puntuación de calidad* y la combinación secuencial de ensayos clínicos según la puntuación de calidad. La representación gráfica de la magnitud del efecto frente a la calidad global no constituye en sí un procedimiento; más bien es un paso previo a la incorporación de la calidad como criterio de ponderación, o a un análisis de regresión que intente justificar la heterogeneidad existente en función de variables específicas ligadas a la calidad de la investigación. Así lo hicieron, por ejemplo, Berlin y Colditz⁴⁸ al estudiar la actividad física en la prevención de la enfermedad coronaria. Encontraron que los estudios metodológicamente más correctos tendían a mostrar un mayor beneficio de la actividad física. Algo similar encontraron Boyd y cols.⁴⁹, en su

análisis de la dieta grasa y cáncer de mama, en el que los estudios mejor diseñados y ejecutados encontraron una mayor fuerza de la asociación (riesgo relativo de 1,15 frente a 1,06). Por el contrario, otros autores no han encontrado relación entre la calidad y la magnitud de la asociación³².

Olkin³ prefiere incluir la calidad global de un estudio dentro de un análisis general de sensibilidad como una covariable. Greenland⁴ prefiere no calcular indicadores globales de calidad y relacionarlos con el efecto, porque o es superfluo o es erróneo. Si existe variación en el criterio de calidad, el analista debería investigar la razón y ello le conduciría a un análisis de los componentes de la escala de calidad; la valoración, global de la calidad sería innecesaria en estos casos. Si no hay variación, puede ser debida a tres razones: no influye la calidad, un pequeño efecto de los componentes de calidad se diluye en el indicador global, o un gran efecto contrario entre componentes de la calidad se neutralizan entre sí. En el primer caso, la valoración de la calidad global es superflua; en los dos restantes, si no se analizan los componentes de la calidad, pueden sacarse conclusiones erróneas. Por todo esto, Greenland⁴ sustenta que es mejor utilizar indicadores de características específicas del diseño como justificadores de la heterogeneidad.

Por último, la *combinación secuencial de ensayos clínicos según la calidad* empieza por el estudio de mayor calidad y gradualmente va incorporando a la ponderación a los estudios por orden decreciente de calidad. Esto se considera en el momento presente una técnica habitual del metanálisis acumulado⁵⁰, y su finalidad principal es valorar heterogeneidad del resultado en función del nivel de calidad del estudio.

Conclusiones

En resumen, con respecto a la valoración de la calidad sería recomendable una mayor estandarización en la publicación de artículos científicos. La difusión de directrices metodológicas y su aceptación por parte de editores y autores debería suponer un mayor grado de normalización en la redacción final de los manuscritos. Esto facilitaría el proceso de evaluación de los mismos.

En el momento presente existe acuerdo en la necesidad de valorar la calidad de una investigación (determinación de errores, explicar heterogeneidad, etc.), pero no así en el cómputo de indicadores globales de calidad. Los defensores del análisis individualizado de los componentes de la calidad de un estudio no ven, por lo tanto, la necesidad de desarrollar mejores cuestionarios de calidad en el terreno del metanálisis. Por el contrario, los partidarios de que la calidad global puede ser una covariable importante para explicar los resultados, insisten en que se deben desarrollar mejores cuestionarios de evaluación. En ellos sería importante delimitar cuáles son variables que tiene poder discriminatorio (información independiente), y si la adición de información específica sobre el tema del metanálisis está o no correlacionada con los indicadores generales de la calidad de un estudio.

Por último, parece sensato que se excluyan los estudios probablemente sesgados; en el momento presente no es recomendable incluir a la calidad como un componente más del peso de una investigación.

Bibliografía

1. Wachter KW. Disturbed by meta-analysis? *Science* 1988; 241: 1407-8.
2. Greenland S. Invited commentary: a critical look at some popular meta-analytical methods. *Am J Epidemiol* 1994; 140: 290-6.
3. Olkin I. Invited commentary: Re: "A critical look at some popular meta-analytical methods". *Am J Epidemiol* 1994; 140: 297-300.
4. Greenland S. Quality scores are useless and potentially misleading. Reply to "Re: "A critical look at some popular meta-analytical methods". *Am J Epidemiol* 1994; 140:300-1.
5. Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994; 140:771-8.
6. Spitzer WO. The challenge of meta-analysis. *J Clin Epidemiol* 1995; 48: 1-4.
7. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbé KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992; 45: 255-65.
8. US Preventive Services Task Force. *Guide to Clinical Preventive Services. An Assessment of the Effectiveness of 169 Interventions*. Baltimore: Williams & Wilkins, 1989.
9. Maclure M. Taxonomic axes of epidemiologic study designs: a refutationist perspective. *J Clin Epidemiol* 1991; 44: 1045-53.
10. Austin H, Hill HA, Flanders WD, Greenberg RS. Limitations in the application of case-control methodology. *Epidemiol Rev* 1994; 16: 65-76.
11. Sillero-Arenas M, Delgado-Rodríguez M, Rodrigues-Canteras R, Bueno-Cavanillas A, Gálvez-Vargas R. Menopausal hormone replacement therapy and breast cancer. A meta-analysis. *Obstet Gynecol* 1992; 79: 286-94.
12. Bulpitt CJ. Meta-analysis. *Lancet* 1988; ii: 93-4.
13. Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control bias and comparison with large co-operative trials. *Stat Med* 1987; 6: 315-26.
14. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316: 150-5.
15. Schlesselman JJ. *Case-control studies. Design, conduct, analysis*. New York: Oxford University Press, 1982.
16. Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol* 1976; 103: 226-35.

17. Miettinen OS. *Theoretical Epidemiology. Principles of Occurrence Research in Medicine*. New York: Wiley, 1985.
18. Miettinen OS. The "case-control" study: valid selection of subjects. *J Chron Dis* 1985; 38: 543-8.
19. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol* 1992; 135: 1019-28.
20. Lasky T, Stolley PD. Selection of cases and controles. *Epidemiol Rev* 1994; 16: 6-17.
21. Delgado-Rodríguez M, Sillero-Arenas M, Martín-Moreno JM, Gálvez-Vargas R. Oral contraceptives and cancer of the cervix uteri. A meta-analysis. *Acta Obstet Gynecol Scand* 1992; 71: 368-76.
22. Mayes LC, Horwitz RI, Feinstein AR. A collection of 56 topics with contradictory results in case-control research. *Int J Epidemiol* 1988; 17: 680-5.
23. Delgado Rodríguez M, Sillero Arenas M, Gálvez Vargas R. Metaanálisis en epidemiología. I. Características generales. *Gac Sanit* 1991; 5:265-72.
24. Pettiti DB. *Meta-analysis. Decision analysis and cost-effectiveness analysis*. New York: Oxford University Press, 1994.
25. Felson DT. Bias in meta-analytic research. *J Clin Epidemiol* 1992; 45: 885-92.
26. Chalmers TC. Problems induced by meta-analyses. *Stat Med* 1991; 10: 971-80.
27. Trilla A. Por qué y cómo se aceptan o rechazan artículos para su publicación en las revistas biomédicas. *Med Clin (Barc)* 1990; 95: 732-4.
28. Altman DG. *Practical statistics for medical research*. Londres: Chapman & Hall, 1991: 477-504.
29. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; 2: 31-49.
30. Elwood JM. *Causal relationships in medicine: a practical appraisal*. Oxford: Oxford University Press, 1988.
31. Realini JP, Goldzieher JW. Oral contraceptives and cardiovascular disease: a critique of the epidemiologic studies. *Am J Obstet Gynecol* 1985; 152: 729-98.
32. Longnecker MP, Berlin JA, Orza MJ, Chalmers TC. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 1988; 260: 652-6.
33. Friedenreich CM, Brandt RF, Riboli E. Influence of methodologic factors in a pooled analysis of 13 case-control studies of colorectal cancer and dietary fiber. *Epidemiology* 1994; 5: 66-79.
34. Friedenreich CM. Methods for pooled analyses of epidemiologic studies. *Epidemiology* 1993; 4: 295-302.
35. Sillero Arenas M. *Metaanálisis de la relación entre los contraceptivos orales y los cánceres de mama y cuello uterino* [Tesis Doctoral]. Granada: Universidad de Granada, 1989.
36. Delgado Rodríguez M, Sillero Arenas M, Rodríguez R, Bueno Cavanillas A, Martín Moreno JM, Gálvez Vargas R. Anticonceptivos orales y cáncer de mama: análisis de la potencia estadística de asociación. *Gac Sanit* 1991; 5: 59-67.
37. Feinstein AR. *Clinimetrics*. New Haven: Yale University Press, 1987.
38. Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. *J Clin Epidemiol* 1995; 48: 9-18.
39. Spitzer WO. Meta-meta-analysis: unanswered questions about aggregating data. *J Clin Epidemiol* 1991; 44: 103-7.
40. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol* 1991; 44: 127-39.
41. Sillero-Arenas M, Rodríguez R, Delgado Rodríguez M, Bueno Cavanillas A, Gálvez Vargas R. Patterns of reserach. Oral contraceptives and cancer of the cervix uteri. *Acta Obstet Gynecol Scand* 1991; 70: 143-8.
42. Delgado-Rodríguez M, Sillero-Arenas M, Rodríguez R, López R, Gálvez-Vargas R. Oral contraceptives and breast cancer. A meta-analysis. *Rev Epidém Santé Publ* 1991; 39: 165-81.
43. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; 337: 867-72.
44. Goodwin PJ, Boyd NF. Mammographic parenchymal pattern and breast cancer risk: a critica appraisal of the evidence. *Am J Epidemiol* 1988; 127: 1097-108.
45. Delgado Rodríguez M, Rodríguez R, Sillero Arenas M, Lardelli Claret P, Bueno Cavanillas A, Gálvez Vargas R. Contraceptivos orales y cánceres de mama y cuello uterino. Metaanálisis cualitativo. *Toko-Gin Práct* 1990; 46: 336-44.
46. Delgado Rodríguez M, Sillero Arenas M, Rodríguez R, Gálvez Vargas R. Pertinencia y normas para la realización de un estudio sobre anticonceptivos orales y cáncer de mama en España. *Toko-Gin Práct* 1991; 47: 19-24.
47. Willett WW. *Nutritional Epidemiology*. New York: Oxford University Press, 1990.
48. Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol* 1990; 132: 612-28.
49. Boyd NF, Martin LJ, Noffel M, Lockwood GA, Trichler DL. A meta-analysis of studies of dietary fat and breast cancer risk. *Br J Cancer* 1993; 68: 627-36.
50. Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *J Clin Epidemiol* 1995; 48: 45-57.

