

Evaluación de la eficacia de tratamientos y programas mediante estudios de casos y controles¹

Joe V. Selby²

En vista del uso cada vez más frecuente de los estudios de casos y controles para evaluar la eficacia de vacunas (1) y pruebas de detección (2), tiene interés revisar la aplicación de este método eficiente y relativamente económico al estudio de la eficacia de otras medidas preventivas, como los programas de salud pública o los tratamientos médicos. Los estudios observacionales, ya sean diseños de casos y controles o de cohorte, tienen serias limitaciones para investigar la eficacia de tratamientos. No obstante, en algunos casos puede que no sean factibles los ensayos aleatorizados. En esta presentación se comparan los méritos de los estudios observacionales y experimentales de tratamientos, se contrasta la elección de métodos de casos y controles o de cohorte con ese propósito, y se brindan ejemplos de la aplicación del método de casos y controles a la evaluación de efectos intencionados y no intencionados de tratamientos y a la evaluación de programas y procedimientos de control de calidad.

MÉTODOS OBSERVACIONALES PARA EVALUAR LA EFICACIA DE TRATAMIENTOS O DE PROGRAMAS

Desventajas de los métodos observacionales para estudiar tratamientos

La desventaja principal de los estudios no aleatorizados para evaluar la eficacia de un tratamiento es la alta probabilidad de que, sin aleatorización, se seleccionen para recibir un tratamiento o participar en un programa personas que difieren, en cuanto a riesgo de presentar cierto resultado, de las que se seleccionan o asignan a otro tratamiento alternativo o a no ser tratadas. Estas diferencias pueden proceder de decisiones tomadas por los médicos, por los directores de programas o por los propios sujetos. En cualquier caso, son factores de confusión cuando se estudian los efectos de tratamientos.

Confusión por indicación. A diferencia de las evaluaciones de vacunas o de pruebas de detección, cuyo enfoque general y uniforme son poblaciones o grupos de edad en bloque, la mayor parte de los estudios de tratamientos médicos se llevan a cabo con pacientes que sufren un trastorno concreto. Los tratamientos se seleccionan sobre la base de una evaluación clínica del riesgo que tiene cada paciente de que la enfermedad progrese o se complique. Este riesgo a su vez se relaciona con la gravedad de la enfermedad, el estado de salud general del paciente y la presencia de otros factores de riesgo o enfer-

¹ Este artículo se publicó en inglés en *Epidemiologic Reviews* (1994;16:90-101) con el título "Case-Control Evaluations of Treatment and Program Efficacy". Traducción de Luis Carlos Silva Ayçaguer revisada por la redacción del *Boletín de la OSP*. Copyright © 1994, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, MD, Estados Unidos de América. Véanse otros artículos de esta serie en pp. 228-242 y en números anteriores y posteriores del *Boletín de la OSP*.

² Dirección postal: Kaiser Permanente Medical Care Program of Northern California, Division of Research, 3505 Broadway, Oakland CA 94611. (Las solicitudes de separatas del original deben enviarse a esta dirección.)

medades asociadas. A menor escala, los programas de prevención a veces intentan alcanzar su efectividad máxima enfocando poblaciones o subgrupos de pacientes que se sabe que son de alto riesgo. En ambos casos, el pronóstico del grupo tratado y el del no tratado son distintos por razones no relacionadas con el tratamiento. Este sesgo se ha denominado confusión por indicación (3, 4), sesgo de susceptibilidad (5) y sesgo de encauzamiento (*channelling bias*) (6).

En un estudio de casos y controles sobre factores de riesgo de amputación de una extremidad inferior, anidado en el estudio de una cohorte numerosa de pacientes diabéticos (cuadro 1), buscamos indicios de que la asistencia a clases de educación para diabéticos pudiera disminuir el riesgo de amputación. Se midió la asistencia durante cinco intervalos de tiempo antes de la amputación. Al contrario de lo que esperábamos, las visitas a clases de educación fueron significativamente más frecuentes en los casos que en los controles en cada intervalo, incluido el período de más de 10 años que precedió a la amputación. Tuvieron poco efecto sobre estas asociaciones positivas los ajustes según diversas medidas de gravedad de la enfermedad, que incluían el tipo y la duración de la diabetes, la clase de tratamiento, el control promedio de la glucemia a lo largo del tiempo, la tensión

arterial basal, el colesterol y el hábito de fumar. Las asociaciones se atribuyeron por tanto a la capacidad del paciente o del clínico para percibir la susceptibilidad del paciente mucho antes de que se realizaran las amputaciones y asignar o seleccionar para clases de educación a aquellos pacientes en mayor riesgo.

Es posible que la capacidad de los clínicos y de los propios pacientes para evaluar las diferencias de riesgo no pueda captarse ni controlarse con simples medidas de la gravedad de la enfermedad y de los trastornos asociados (7, 8). Mossey y Shapiro (7) describieron un índice de autovaloración de la salud que consiste en una sola pregunta y que, cuando se administra a pacientes de edad avanzada predice diferencias de casi tres a uno en las tasas de mortalidad durante los 7 años siguientes, independientemente de las medidas objetivas de gravedad de la enfermedad y comorbilidad. Esta intensa capacidad predictiva ha de reflejar la percepción que tiene el paciente de los síntomas de enfermedad preclínica o de efectos no medidos de factores relacionados con el estilo de vida en lo que hace al estado de salud y el pronóstico. Sea como fuere, el índice subraya la importancia del estado de salud como predictor adicional de riesgo. Como es probable que las medidas del estado de salud

CUADRO 1. Comparación de casos y controles en cuanto a exposición a clases de educación sobre diabetes: estudio de casos y controles anidado de amputación de una extremidad inferior en una cohorte de pacientes diabéticos, Programa Kaiser Permanente de Atención Médica

Intervalo (en años) antes de la amputación	Media del número de clases		Razón de posibilidades apareada*	Intervalo de confianza de 95%
	Casos (n = 149)	Controles (n = 264)		
0 a 1	2,4	0,6	1,41	1,25 a 1,58
1 a 2	0,9	0,5	1,25	1,07 a 1,45
2 a 5	2,6	1,3	1,25	1,14 a 1,37
5 a 10	3,3	2,3	1,12	1,05 a 1,20
>10	3,6	2,3	1,13	1,06 a 1,21

* *Matched odds ratio* en el original inglés (*N. del editor*). Se refiere al cambio en el riesgo de amputación asociado con una clase adicional de educación sobre diabetes.

(estado funcional, calidad de vida) también influyan en la elección de tratamientos por parte de pacientes y clínicos, esas medidas pueden actuar como factores de confusión en los estudios no aleatorizados de la eficacia de tratamientos.

Confusión por autoselección. La proclividad general de una persona a acatar los consejos médicos y a practicar un estilo de vida saludable también puede influir en el pronóstico y la selección de tratamientos. En las intervenciones que dependen en gran parte de la decisión del paciente (como la asistencia a clases de educación para la salud, la elección de tratamiento quirúrgico en lugar de médico y la decisión de tomar o no estrógenos tras la menopausia), las diferencias entre los pacientes que escogen o aceptan un tratamiento y los que no lo hacen (factores de autoselección) pueden distorsionar las asociaciones entre tratamientos específicos y resultados. Asimismo, los clínicos pueden seleccionar tratamientos distintos basándose en cómo evalúan cuán probable es que el paciente cumpla las indicaciones. Hay un ejemplo bien conocido de sesgo de autoselección que se encontró en el Proyecto sobre Medicación Coronaria (9). En el grupo control tratado con placebo, los sujetos que tomaron al menos 80% de las tabletas de placebo tuvieron una tasa de mortalidad 36% menor que la de los que no las tomaron, después de ajustar el resultado según 40 posibles factores de confusión. Una asociación similar se observó entre el cumplimiento del tratamiento y la supervivencia en el grupo que tomó clofibrato. Por lo tanto, un estudio observacional de este proyecto realizado solamente en el grupo tratado habría encontrado un efecto reductor de la mortalidad, fuerte pero espurio, del uso regular del clofibrato, efecto que no se encontró en el ensayo aleatorizado.

Sesgo de detección. En un ensayo aleatorizado se establecen criterios estrictos y uniformes para monitorear la aparición del resultado de interés tanto en los sujetos tra-

tados como en los controles, lo cual asegura la misma probabilidad de detectarlo en ambos grupos. Sin esos criterios, en los estudios observacionales existe el riesgo de que haber recibido cierto tratamiento pueda alterar la probabilidad de ser diagnosticado. Por ejemplo, la decisión de prescribir o aceptar un tratamiento podría acompañarse de un cambio de la frecuencia con que se busca el resultado o afectar a la percepción del clínico o del paciente, de manera que se altere el umbral para hacer un diagnóstico. En cualquier caso, entre los tratados y los no tratados hay una diferencia en la incidencia aparente del resultado. Las asociaciones de los estrógenos administrados en la posmenopausia con el cáncer de mama (10) y de los anticonceptivos orales con el cáncer de cérvix (11) se han atribuido a una supervisión más intensa de las mujeres que los usan. Estos sesgos son mayores en diagnósticos más discrecionales, lo cual puede explicar por qué, en los estudios prospectivos, los estrógenos prescritos para la posmenopausia se asocian con una alta incidencia de cáncer de mama sin un aumento de mortalidad por este cáncer (10, 12).

Problemas adicionales de los estudios observacionales. Al planificar un ensayo controlado, generalmente se tiene sumo cuidado de estandarizar todos los aspectos de la intervención. En las situaciones observacionales, la falta de estandarización y las incongruencias de atención pueden menoscabar la capacidad para identificar grupos similarmente expuestos o para formular interrogantes sobre equivalencias en la calidad del tratamiento. Además, la dificultad para cuantificar el cumplimiento puede complicar aun más la cuantificación del tratamiento recibido. En general, estas limitaciones contribuyen a subestimar el verdadero efecto beneficioso. Sin embargo, si la selección de un tratamiento específico por parte del clínico se relaciona con la calidad general de la atención que presta, las asociaciones del tratamiento con los resultados de la enfermedad pueden inflarse espuriamente en cualquier dirección.

Estos problemas, particularmente la confusión por indicación y autoselección, han llevado a algunos autores (13, 14) a desalentar casi totalmente el uso de métodos observacionales para evaluar los beneficios de tratamientos o de intervenciones programáticas.

Ventajas potenciales de los métodos observacionales sobre los ensayos aleatorizados

Los estudios observacionales suelen ser menos costosos que los ensayos aleatorizados, porque ahorran los gastos de reclutamiento, aleatorización y seguimiento. En un ensayo aleatorizado, la insuficiencia del tamaño de la muestra o del tiempo de seguimiento puede hacer que pase inadvertido un efecto real; mientras más raro es o más se demora el resultado de interés, más aumenta ese riesgo. Los análisis observacionales retrospectivos, sean de casos y controles o de cohorte, pueden realizarse más rápidamente y a menor costo en muestras mayores de pacientes.

Los ensayos aleatorizados pueden no ser prácticos o incluso éticos cuando se trata de un programa o tratamiento que ya se usa ampliamente o que se sospecha eficaz hasta tal punto que ni los pacientes ni los médicos aceptan ser asignados al grupo de control con un tratamiento alternativo o sin tratamiento alguno. Los ensayos aleatorizados se realizan, por necesidad, en personas seleccionadas que desean y pueden aceptar voluntariamente la aleatorización. A menudo se excluyen aquellas con enfermedades asociadas importantes o en situaciones extremas en cuanto a edad, nivel socioeconómico, gravedad de su enfermedad o incumplimiento. Los estudios observacionales pueden incluir a todos los pacientes potencialmente expuestos al tratamiento en cuestión. Esta capacidad para estudiar pacientes "reales" en condiciones clínicas "reales" hace que el estudio observacional sea un método muy atractivo, por lo menos como complemento de los ensayos aleatorizados.

Los ensayos aleatorizados usualmente se limitan a evaluar un tratamiento, a lo sumo dos. Sin embargo, en el entorno clínico a menudo hay bastantes más de dos posibles tratamientos (según intensidad y duración de la terapéutica y uso de medidas adicionales complementarias). Todas estas posibilidades pueden estudiarse con métodos observacionales si para cada tratamiento hay suficientes pacientes que lo hayan recibido. Las muestras de mayor tamaño permiten examinar interacciones entre tratamientos. Esta ventaja tiene un atractivo particular en el estudio de enfermedades que pueden requerir varios tratamientos o programas complejos con múltiples componentes. Por ejemplo, en investigaciones de control de calidad a menudo se comienza por tratar de comprender las diferencias observadas en resultados que son el producto final de situaciones clínicas complejas.

Debido a las limitaciones de los ensayos aleatorizados, persiste el interés en aplicar y mejorar los métodos observacionales para evaluar tratamientos e intervenciones programadas (4, 5, 15, 16). La Agencia para la Política e Investigación en Atención de Salud (*Agency for Health Care Policy and Research*)³ apoya varios estudios sobre efectividad médica que emplean métodos observacionales para comparar tanto beneficios como riesgos de los tratamientos (17). El movimiento de "medicina de resultados" (18, 19) deposita muchas esperanzas en las grandes bases clínicas de datos como fuente informativa para estudios observacionales de los méritos relativos de programas de salud o de atención médica; pero ¿podrán eliminarse los peligros de los estudios observacionales en estas ambiciosas empresas?

Cómo aumentar la validez de los estudios observacionales

La necesidad de controlar la confusión por indicación y por autoselección ha lleva-

³ Se trata de una institución de los Estados Unidos de América (N. del editor).

do a una búsqueda de medidas más exactas de los factores de confusión. Se han desarrollado medidas de gravedad específicas para distintas enfermedades (20), así como índices para evaluar la gravedad de ciertos diagnósticos en cuidados intensivos (21, 22), en el hospital general (23, 24) y en la consulta externa (25). Los índices de comorbilidad o de "mezclas de casos" procuran contemplar el impacto de otras enfermedades cuando se estudia un grupo particular de pacientes hospitalizados (26) o ambulatorios (27). Se han desarrollado medidas multidimensionales sintéticas del estado funcional y de salud así como instrumentos que miden la calidad de vida (28, 29). Esas medidas han mostrado ser predictoras de la mortalidad y de la utilización de atención médica. Estos instrumentos pueden administrarse a los sujetos de estudio en el proceso de recolección de los datos de la exposición y de los resultados. Sin embargo, esos mismos instrumentos se administran de ordinario a un número elevado de pacientes en muchas situaciones clínicas. Los resultados se incorporan a bases de datos clínicas automatizadas para emplearlos como "resultados" de garantía de la calidad, realizar ajustes según "mezclas de casos" en estudios de mortalidad y de otros resultados finales, y proporcionar información a los clínicos. Cuando es evidente que estas medidas predicen un resultado específico de interés, pueden ofrecer un medio conveniente de ajuste posterior o de estratificación para estudios observacionales más detallados sobre tratamiento, particularmente cuando la información se recolecta en el momento en que se decide el tratamiento o poco después.

Como alternativa a la medición metódica de los factores, puede intentarse reducir la confusión aprovechando las variaciones de la práctica médica según las distintas zonas geográficas o, con menos frecuencia, dentro de una misma población. Un ejemplo de esa variación es el de las tasas de intervención de revascularización coronaria en los residentes de New Haven, Connecticut, que duplica a la de Boston, Massachusetts; complementariamente, la

tasa de endarteriectomía carotídea es en Boston el doble que en New Haven (30).

Esta variación puede reflejar incertidumbre sobre la eficacia de un tratamiento o diferencias de capacitación profesional o diseminación de información entre los clínicos. El punto importante es que esa variación no obedece a diferencias en el pronóstico de los pacientes. Strom et al. (31) revisaron los 100 medicamentos aprobados más recientemente por la Administración de Alimentos y Medicamentos (FDA) desde 1978 y concluyeron que 72% de las indicaciones terapéuticas posibles de estos medicamentos podrían someterse a estudios observacionales de eficacia relativa poscomercialización, principalmente por el conocimiento de variaciones en la práctica no relacionadas con las indicaciones. En esencia, dichos estudios sacan partido de experimentos naturales.

Varios autores (5, 16, 32, 33) han debatido la aplicación de los principios de los ensayos aleatorizados a los estudios observacionales. Las estrategias incluyen medición detallada y estratificación de variables relacionadas con el pronóstico, uso de criterios de inclusión y exclusión comparables con los que se habrían usado en un ensayo aleatorizado y protocolos para proceder al entrecruzamiento (*crossover*) de tratamientos.

También puede ser útil examinar la especificidad del efecto de un tratamiento para el resultado buscado a fin de evaluar la confusión residual (33). Si un beneficio aparente representa un efecto causal de un tratamiento, no debe estar presente entre los resultados no relacionados con dicho tratamiento. Tal especificidad del efecto se demostró en un estudio de casos y controles sobre métodos de detección mediante sigmoidoscopia rígida (34). En ese estudio se encontró una reducción de 60 a 70% de la tasa de mortalidad por tumores malignos que se encontraban al alcance del sigmoidoscopia, aunque no se comprobó beneficio alguno en cuanto a tumores que estaban fuera de su alcance.

Es de importancia decisiva tener en cuenta la confusión potencial por indicación

y autoselección; pero esa precaución debe equilibrarse con una apreciación de la fuerza de asociación que se necesita para que los factores de confusión expliquen las asociaciones sustanciales que se observen (o para que oscurezcan las asociaciones sustanciales verdaderas). Se ha demostrado (35-38) que es poco probable que los factores de confusión expliquen totalmente asociaciones observadas con tamaños del efecto mucho mayores de 2 (o menores que su inverso, 0,5). Por ejemplo, para crear espuriamente una eficacia aparente de un tratamiento de 50% (es decir, una reducción de 50% de un resultado indeseable), un factor de confusión dicotómico tendría que estar causalmente asociado con el resultado con un tamaño de efecto de 3,4 y presentarse, como mínimo, 3,4 veces menos frecuentemente en los expuestos al tratamiento que en los no expuestos (37). Por tanto, cuando se esperan efectos verdaderamente marcados del tratamiento, incluso asociaciones fuertes como las descritas entre factores de confusión y el resultado difícilmente podrán enmascarar el efecto en su totalidad. A la inversa, cuando se observa un efecto sustancial, es poco probable que los factores de confusión puedan explicar por completo esa asociación.

ESTUDIOS DE CASOS Y CONTROLES FRENTE A ESTUDIOS DE COHORTE EN LA INVESTIGACIÓN DE EFECTOS DE TRATAMIENTO

El cuadro 2 resume las ventajas y desventajas del método de casos y controles comparado con el de cohortes (sea retrospectivo o prospectivo). La única razón para tener en cuenta los estudios de casos y controles es su superioridad en eficiencia para estudiar resultados raros. Es conveniente intentar aumentar la eficiencia (reduciendo el tamaño de muestra) siempre que un estudio requiera una costosa recolección de datos adicionales. Estos datos pueden ser necesarios para verificar resultados, cuan-

tificar las exposiciones al tratamiento o medir la gravedad de la enfermedad u otros factores de confusión potenciales. En la rara circunstancia de que todos los datos necesarios sobre exposición al tratamiento, resultados y factores de confusión de una cohorte entera de pacientes estén disponibles de forma automatizada (como ocurre a veces con grandes bases de datos clínicas), no ofrece ventaja alguna analizar los datos mediante el método de casos y controles por más raro que sea el resultado. Un análisis de cohortes retrospectivo con el mismo número de casos brinda estimaciones ligeramente más precisas y evita los problemas propios del enfoque de casos y controles.

Una desventaja importante del estudio de casos y controles al evaluar los efectos de un tratamiento es que se limita a un solo resultado (el que define la condición de ser caso o control). Esto no suele constituir un problema en la investigación etiológica, donde el mayor interés radica en explicar la aparición de un acontecimiento específico. Incluso en estudios de eficacia de vacunas y

CUADRO 2. Ventajas y desventajas de los estudios de casos y controles comparados con los de cohorte en la evaluación de tratamientos médicos o programas

Ventajas
<ul style="list-style-type: none"> • Diseño muy eficiente cuando se estudian resultados poco frecuentes • Adecuados para resultados con periodos de latencia largos
Desventajas
<ul style="list-style-type: none"> • Solo se puede estudiar un resultado cada vez • No pueden obtenerse directamente tasas de incidencia • Las estimaciones de razón de posibilidades no son atractivas desde el punto de vista intuitivo • Los efectos temporales en las asociaciones pueden ser difíciles de discernir • Los sesgos de selección y de recuerdo constituyen problemas potenciales • Por lo general no son adecuados para estudiar tratamientos introducidos recientemente

procedimientos de detección, lo que más interesa es precisamente la detección de un resultado único. En cambio, cuando se estudia un tratamiento puede haber varios posibles resultados de interés. Por ejemplo, al investigar un tratamiento médico de la arteriopatía coronaria en comparación a una intervención quirúrgica, también conviene estudiar aspectos como la incidencia posterior de infarto de miocardio, la mortalidad, la rehospitalización, los síntomas y la calidad de vida, así como los efectos adversos inesperados. El enfoque analítico de casos-cohorte (39), en el cual se examina una subcohorte elegida al azar y no un grupo de control apareado, retiene buena parte de la eficiencia de los estudios de casos y controles al tiempo que permite analizar dos o más resultados (40).

Una segunda desventaja estriba en que las tasas de incidencia no pueden obtenerse directamente a partir del análisis de casos y controles y no es posible calcular directamente las razones de riesgo. Aunque la mayoría de los epidemiólogos se han familiarizado con la estimación que facilita la razón de posibilidades (*odds ratio*), no puede decirse lo mismo de la comunidad médica. La desconfianza general que prevalece en torno a los estudios "retrospectivos" y a los métodos de casos y controles en particular puede dificultar la diseminación de los resultados, aunque sean convincentes.

Dos escollos citados a menudo en los estudios de casos y controles, los riesgos de sesgos de selección y de recuerdo, pueden superarse anidando el estudio en una cohorte de pacientes ya configurada. Hoy día la identificación de cohortes enteras de pacientes es fácil utilizando bases de datos clínicas computadorizadas tales como las de egresos hospitalarios, las de farmacia o los registros de diagnósticos de pacientes ambulatorios de organizaciones de mantenimiento de la salud, de la Administración de Veteranos⁴ o de registros poblacionales. La verificación ulterior de todos los casos que aparecen en la cohorte y la elección de controles de la misma

cohorte elimina los problemas de sesgo de selección. También se elimina el problema del sesgo de recuerdo sobre tratamientos que se registran en las historias clínicas o en las bases de datos automatizadas antes de que se conozca la condición de caso o control de un sujeto.

El diseño de casos y controles presenta algunos inconvenientes de tipo temporal. Como no pueden construirse y compararse curvas de supervivencia de los grupos de tratamiento, las interacciones entre el tipo de tratamiento y su duración (tales como el entrecruzamiento del riesgo acumulado durante períodos cortos frente a períodos largos de seguimiento) pueden ser menos evidentes. También puede ser más difícil definir retrospectivamente el estado basal o punto en el que se asigna el tratamiento inicial. Algunas veces hay pacientes que comienzan recibiendo tratamiento médico y cuando este fracasa se les ofrece cirugía: el análisis de casos y controles se sesgaría si esos fracasos del tratamiento médico en situaciones más complejas se clasificaran erróneamente como tratamiento quirúrgico. La disponibilidad de cohortes de pacientes puede aminorar las consecuencias de este problema siempre que los pacientes sean clasificados en un punto previo a la decisión sobre el tratamiento inicial. En este caso, se puede fijar una fecha de partida para cada paciente, las decisiones iniciales sobre el tratamiento pueden anotarse y analizarse, y los entrecruzamientos de tratamientos se pueden identificar y analizar.

⁴ Las organizaciones de mantenimiento de la salud son programas para la provisión de servicios de salud en ciertas zonas geográficas. Estas organizaciones estadounidenses brindan atención de salud de carácter básico y suplementario a un grupo de personas voluntariamente suscritas, a cambio del pago periódico de una cantidad determinada que no tiene relación con los servicios realmente recibidos por el suscriptor. La Administración de Veteranos es una institución oficial que proporciona beneficios a quienes han servido en las fuerzas armadas (N. del editor).

EJEMPLOS DE LA APLICACIÓN DE MÉTODOS DE CASOS Y CONTROLES

Beneficios esperados del tratamiento

La confusión por indicación se suele producir con mayor frecuencia al estudiar los resultados que se intentan prevenir con el tratamiento. Horwitz y Feinstein (41, 42) aplicaron los principios del ensayo controlado aleatorizado a los estudios de casos y controles en evaluaciones del tratamiento anticoagulante (41) y la profilaxis con lidocaína (42) para prevenir la mortalidad después de un infarto de miocardio. Los dos estudios fueron anidados en la misma cohorte de pacientes hospitalizados por infarto agudo de miocardio durante 5 años.

Los autores revisaron los informes de seis ensayos aleatorizados, de los cuales solo uno mostraba que el tratamiento anticoagulante era beneficioso después del infarto de miocardio. Sin embargo, los estudios restantes eran relativamente pequeños y podían haber fracasado en la detección de un beneficio verdadero. En relación con la lidocaína, un ensayo previo mostró reducción de la aparición de fibrilación ventricular, pero no de la mortalidad. Los autores concluyeron que la mortalidad relativamente baja esperada con cada tratamiento en los pacientes que sobrevivirían para ser aleatorizados haría excesivos los tamaños de muestra necesarios para futuros ensayos aleatorizados sobre mortalidad.

Para semejar un ensayo aleatorizado, se excluyeron los pacientes que habrían resultado no elegibles para aleatorización por fallecer tempranamente (es decir, antes de que la aleatorización fuera posible) o por fuerte indicación o contraindicación del tratamiento estudiado. Se definió un momento base o "tiempo focal" en la admisión a la unidad de cuidados especiales, y las órdenes de admisión se utilizaron para determinar el grupo de tratamiento. Es interesante que los autores decidieran no realizar el análisis según la "intención de tratar," argumen-

tando que en los pacientes que "se entrecruzaron" al grupo tratado (desde el no tratado) en algún momento posterior a la decisión sobre el tratamiento inicial debía haber existido una fuerte indicación del tratamiento, lo cual los convertía en no elegibles para un ensayo. A partir de las historias clínicas se construyeron índices de gravedad de la enfermedad y de comorbilidad para la estratificación pronóstica.

En el estudio del tratamiento anticoagulante (41), el beneficio aparente derivado de su administración se redujo marcadamente cuando se aplicaron los criterios de exclusión del ensayo aleatorizado. La estratificación posterior según la gravedad sugirió un fuerte efecto protector en los pacientes con insuficiencia cardíaca o choque cardiogénico (razón de posibilidades = 2,6; $P < 0,025$) y un efecto mínimo, estadísticamente no significativo, en los casos más leves sin signos de insuficiencia (razón de posibilidades = 1,15). Este análisis estratificado puede explicar los efectos más débiles, no significativos, que predominaban entre los ensayos aleatorizados, ya que la mayor parte de los infartos de miocardio incluidos en los ensayos son leves.

En el estudio de la profilaxis con lidocaína (42), se observó que el tratamiento no tenía efecto sobre la mortalidad por fallo de bomba u otras causas no asociadas con arritmia. Sin embargo, la razón de posibilidades fue 3,17 ($P < 0,05$) cuando las muertes debidas a arritmia se compararon con los supervivientes, lo cual sugería una reducción del riesgo de muerte por arritmia ventricular asociada con el tratamiento.

Psaty et al. (43) llevaron a cabo un estudio de casos y controles anidado en una cohorte de pacientes con hipertensión con objeto de examinar la eficacia del tratamiento con betabloqueantes, comparado con otros tratamientos antihipertensivos, en la prevención de las primeras manifestaciones de las coronariopatías (angina con hospitalización, infarto de miocardio o muerte súbita). De los cuatro primeros ensayos aleatorizados examinados, solo uno puso de manifiesto un beneficio significativo de los

betabloqueantes con respecto al tratamiento con diuréticos. Dos de los otros tres ensayos mostraron beneficios menores no significativos, pero en ellos se constataron altas tasas de abandono y de entrecruzamiento.

Ambos, casos y controles, se seleccionaron de los que recibían medicación antihipertensiva y sus nombres se registraron en una base de datos de farmacia automatizada por el Grupo de Salud Cooperativo de Seattle, Washington. Los casos incidentes de coronariopatía se escogieron a partir de los registros automatizados de diagnóstico en el momento del alta y de los certificados de defunción del hospital. Se excluyeron las personas con historia de enfermedad coronaria comprobada o sospechada en cualquier momento anterior a la hospitalización, porque esos pacientes no habrían sido elegibles para un ensayo de prevención. Esta exclusión también disminuyó la posibilidad de confusión por indicación, ya que los betabloqueantes se podían haber prescrito preferentemente a los pacientes sospechados de padecer enfermedad coronaria.

La información sobre uso actual de medicación antihipertensiva, tensión arterial, otros factores de riesgo coronario y enfermedades asociadas de los casos se obtuvo de las historias clínicas inmediatamente antes de la fecha de admisión, y, en los controles, en una fecha seleccionada al azar dentro del mismo año de la prescripción que la identificaba. La razón de posibilidades ajustada según manifestaciones no mortales sugirió que el riesgo asociado con el uso actual de betabloqueantes se redujo 38% (intervalo de confianza de 95%: 1% a 61%). Además, se observó un efecto dosis-respuesta: las dosis altas actuales se asociaban con un riesgo sustancialmente menor. Ni el consumo anterior ni su duración total difirieron significativamente entre casos y controles. Los autores concluyeron que si la confusión residual por indicación estaba presente, sería más probable que indujera a subestimar la eficacia en el sentido de que los pacientes en mayor riesgo de enfermedad coronaria y aquellos con hipertensión más difícil de con-

trolar tendrían una mayor probabilidad de recibir betabloqueantes.

Al menos dos estudios de casos y controles han medido la eficacia de la profilaxis con antibióticos previa a tratamiento odontológico para prevenir la endocarditis bacteriana en pacientes con valvulopatía conocida (44, 45). En el estudio más amplio (44), 48 casos de endocarditis bacteriana aguda que se produjeron durante los 180 días siguientes a un tratamiento para el cual la profilaxis estaba indicada, se compararon con 200 controles que también padecían valvulopatía conocida y que se habían sometido a un tratamiento dental durante el mismo período. La razón de posibilidades del uso de profilaxis en los tratamientos que ocurrieron durante los primeros 30 días antes del diagnóstico fue 0,51, lo que sugirió una eficacia de 49%. El estudio más reducido de pacientes con riesgo muy alto (45) sugirió un beneficio mayor (eficacia de 91%).

Beneficios inesperados del tratamiento

La posibilidad de confusión por indicación debe ser más baja cuando no se espera un efecto potencial al administrar el tratamiento (13). No obstante, debe considerarse y evaluarse la posibilidad de que el tratamiento se haya podido asignar o seleccionar sobre la base de una variable causalmente relacionada con el resultado. Además, la confusión por factores de autoselección no se reduce necesariamente, porque la perspectiva mejorada de los pacientes que siguen el tratamiento puede influir en diversos resultados.

Mediante el método de casos y controles se ha estudiado la reducción del número de fracturas y del riesgo de enfermedad coronaria, dos beneficios inicialmente inesperados del tratamiento con estrógenos. En varios de esos estudios (46-48), se observó una disminución de 50 a 60% del riesgo de fracturas de cadera y de antebrazo en consumidores de estrógenos. La protección disminuyó rápidamente cuando se suspen-

dió su consumo. La prescripción preferencial de estrógenos durante la posmenopausia a mujeres presuntamente en mayor riesgo de fracturas (confusión por indicación) habría sesgado los beneficios observados hacia la hipótesis nula. Por lo tanto, en un estudio se excluyeron las pacientes que habían recibido estrógenos por esa indicación (47). Parece improbable que la confusión por autoselección pudiera producir el beneficio sustancial observado en estos estudios o ser responsable de la restricción del efecto al uso en ese momento.

Stampfer y Colditz (49) revisaron los datos de los estudios observacionales sobre el efecto protector de los estrógenos respecto del riesgo de enfermedad coronaria. En seis estudios de casos y controles de base poblacional sobre el infarto de miocardio o la muerte súbita, las razones de posibilidades estimadas se encontraron congruentemente por debajo de 1, con un recorrido de 0,34 a 0,97 y un promedio ponderado común de 0,76 (intervalo de confianza de 95%: 0,61 a 0,94). Aunque se ha mencionado la posible existencia de confusión por factores de autoselección (50), los datos disponibles sugieren que las usuarias de estrógenos en la posmenopausia difieren leve e incongruentemente de las no usuarias con respecto a una gama de factores de riesgo coronario. No obstante, en vista del pequeño tamaño del efecto para la enfermedad coronaria, la posibilidad de confusión residual por factores de autoselección no medidos no puede descartarse con tanta certidumbre como en los estudios de prevención de fracturas.

Por lo menos tres estudios de casos y controles (51–53) evaluaron la eficacia de los suplementos vitamínicos periconcepcionales (particularmente el ácido fólico) para prevenir el desarrollo de defectos del tubo neural. Los ensayos aleatorizados habían demostrado que los suplementos del ácido fólico eran eficaces para prevenir la recurrencia de un defecto del tubo neural en mujeres con alto riesgo por haber tenido un hijo anterior con este tipo de defecto. Sin embargo, en mujeres que no tenían esa his-

toria anterior, la probabilidad conjunta de concebir o tener un niño con un defecto del tubo neural se consideró tan baja que un ensayo aleatorizado no hubiera sido viable. En dos de tres estudios de casos y controles se encontró una reducción de 60 a 70% del riesgo asociado con un alto consumo dietético de ácido fólico o suplementos vitamínicos con dicho ácido antes o inmediatamente después de la concepción. En el tercero, un estudio de base poblacional llevado a cabo cuidadosamente en mujeres de California e Illinois (52), no se encontró efecto alguno. Como los datos sobre el consumo de folato y vitaminas en los tres estudios se obtuvieron después de haber detectado las anomalías, se ha sugerido la posibilidad de que en los resultados hayan influido sesgos de recuerdo y se ha puesto en tela de juicio la exactitud de la evaluación. Más recientemente, se comunicaron desde Hungría los resultados de un ensayo aleatorizado sobre el suplemento con ácido fólico para prevenir la primera aparición de un defecto del tubo neural (54). Este estudio también sugirió un efecto protector de las multivitaminas que contienen folato, pero a pesar de la inclusión y aleatorización de 7540 mujeres en las que posteriormente se contabilizaron 4753 embarazos, los hallazgos se basaron solo en 6 defectos del tubo neural (todos en el grupo de control). Sobre la base de las pruebas que sugieren un beneficio derivado del ácido fólico para la prevención primaria, se han propuesto recomendaciones de suplementar la alimentación con folatos (55). Es posible que el único estudio de casos y controles con resultados negativos refleje una población con un riesgo especialmente bajo, de manera que el suplemento adicional de folato podría no haber conferido protección adicional alguna.

Eficacia y efectividad de los programas de prevención

Baker y Curbow (56) han analizado la utilidad potencial de los estudios de casos y controles para evaluar programas, aunque muy pocos de esos estudios se han notifica-

do. Hasta la fecha, buena parte de las investigaciones han estimado indirectamente la eficacia de los programas, comparando los niveles de cumplimiento de sus objetivos como exposición de interés, tal como ocurre en los programas que promueven el uso de cascos por ciclistas (57) o el empleo de sales de rehidratación oral (58). Una evaluación directa de la eficacia del programa se obtiene comparando la exposición o participación de casos y controles en el programa, como ocurre en la evaluación de servicios de urgencia que se describe más adelante (59). En teoría, los estudios de casos y controles también podrían emplearse para estimar la efectividad del programa, es decir, la magnitud del impacto de la intervención en la población a la cual se dirige. Esto requeriría disponer de una población con un riesgo comparable pero no involucrada en el programa, tal como se hace en un estudio de campo. Ser miembro del grupo objetivo reemplaza entonces la participación en el programa como exposición de interés. Centrando la atención en la colectividad de los miembros de la población antes que en la participación individual se evitan los problemas de autoselección e incorporación de individuos con riesgo más alto. La aplicación de ajustes según factores asociados con la autoselección, tales como el nivel socioeconómico o educativo, ayudaría a asegurar que las diferencias no sospechadas en cuanto al riesgo de la población no actúen como factores de confusión de asociaciones.

Thompson et al. (57) estudiaron la eficacia del uso de cascos por ciclistas en la prevención de lesiones cefálicas causadas por accidentes. Durante un año, todos los casos fueron ciclistas que buscaron atención médica en uno de cinco grandes servicios urbanos de urgencia como consecuencia de heridas en la cabeza producidas en accidentes de tráfico. Se eligieron grupos separados de controles de accidentados que fueron atendidos en los mismos lugares y que tenían heridas no localizadas en la cabeza, y en miembros de una gran organización de mantenimiento de la salud que respondie-

ron a un cuestionario por correo y manifestaron haber tenido un accidente de bicicleta el año anterior. Después de ajustar los datos según diversos factores de confusión potenciales, incluida la gravedad de la lesión, la razón de posibilidades de 0,15 que se halló sugirió una eficacia de 85% del casco para prevenir lesiones cefálicas en accidentes de bicicleta. La razón de posibilidades asociada con lesiones cerebrales graves en 99 casos y los respectivos controles fue ligeramente menor (0,12, lo que corresponde a una eficacia de 88%).

Pulungsih et al. (58) evaluaron la eficacia de la terapia de rehidratación oral en Indonesia. Recabando datos de la literatura sobre evaluación de la eficacia de vacunas mediante estudios de casos y controles, los autores calcularon las posibilidades de que los pacientes que ingresaron en el hospital con diarreas líquidas hubiesen usado terapia de rehidratación oral. Se consideraron casos las personas con deshidratación grave. La razón de posibilidades de 0,28 sugirió que la eficacia de la preparación y uso adecuados fue 72%, mientras que una preparación inadecuada se asoció con una reducción del riesgo de deshidratación de 63%.

Sampalis et al. (59) evaluaron los servicios de urgencia de traumatología en Montreal, Quebec, Canadá, con un estudio de casos y controles. Los sujetos fueron víctimas de un traumatismo que consiguieron llegar vivos al hospital. Los casos fueron los que murieron en los 6 días siguientes; los controles se eligieron entre los que sobrevivieron más de 6 días. Para ajustar los datos según distintas gravedades, se utilizó un índice de gravedad de las heridas. Después de ajustar los resultados según la gravedad de las heridas, se observó que una demora de más de una hora en llegar al hospital triplicaba el riesgo de morir. La calidad de los servicios hospitalarios de traumatología se relacionó inversamente con el riesgo de morir, al margen de la gravedad de la lesión. La presencia de un médico en el lugar en que se dieron los primeros auxilios no se asoció con un riesgo menor. Sin embargo, como un médico solo acudía al lugar del

accidente en caso de ser requerido por una enfermera, la probabilidad de que las personas con heridas leves fuesen vistas por un médico fue menor. Esta confusión por indicación puede haber oscurecido el verdadero beneficio a pesar de que se haya usado el índice de gravedad de la herida.

Greenland et al. (60) aplicaron el método de casos y controles para investigar la garantía de la calidad en un estudio en que se intentaba explicar cierta diferencia observada entre las tasas de mortalidad de dos unidades de cuidados intensivos neonatales. El método de casos y controles constituyó un medio eficiente de estudiar diversos factores de confusión potencialmente explicativos, como las diferencias en la distribución del peso al nacer. Este estudio reveló que el exceso de mortalidad se circunscribía a los niños con bajo peso al nacer. El próximo paso que lógicamente procedía dar, no emprendido por los autores, hubiera sido usar el método de casos y controles para investigar detalladamente las prácticas específicas de enfermería como posibles explicaciones de las diferencias de mortalidad observadas.

CONCLUSIÓN

Como es probable que la asignación no aleatoria a tratamientos introduzca sesgos, los ensayos clínicos aleatorizados siguen siendo el paradigma metodológico para la mayor parte de la evaluaciones de la eficacia de tratamientos o programas. Sin embargo, si los estudios observacionales se diseñan cuidadosamente, pueden brindar información importante que no es posible recabar con ensayos aleatorizados, por ejemplo estimaciones de eficacia para subgrupos de pacientes no incorporados en ensayos aleatorizados y estimaciones de eficacia relativa para varias intensidades de tratamientos o dosis. Los estudios de beneficios no intencionados de tratamientos (así como los de efectos adversos) son adecuados para los enfoques no experimentales. Los estudios observacionales son apropiados para demostrar un efecto terapéutico buscado y que

se espera de tamaño considerable y para evaluar la eficacia cuando la variación de los tratamientos se debe básicamente a diferentes patrones prácticos más que a indicaciones diferentes de los tratamientos entre los pacientes.

El desarrollo continuado de bases de datos clínicos en grandes poblaciones de pacientes brindará nuevas oportunidades para llevar a cabo estudios tanto de cohorte como de casos y controles anidados en los que se evalúen tratamientos médicos o programas, y la calidad de atención. La inclusión de medidas de gravedad de la enfermedad y del estado de salud contribuirá a los esfuerzos por controlar o estratificar según el riesgo inicial. El método de casos y controles es preferible para estudiar resultados raros o que tardan en aparecer cuando se necesitan datos adicionales y el interés se centra en un resultado simple.

AGRADECIMIENTO

El autor agradece a los doctores Gary D. Friedman y Stephen van den Eeden la cuidadosa revisión y los constructivos comentarios que hicieron a un borrador inicial de este trabajo.

REFERENCIAS

1. Comstock GW. Evaluating vaccination effectiveness and vaccine efficacy by means of case-control studies. *Epidemiol Rev* 1994;16:77-89.
2. Weiss NS. Application of the case-control method in the evaluation of screening. *Epidemiol Rev* 1994;16:102-108.
3. Miettinen OS. Efficacy of intervention practice: will epidemiology provide the answers? En: Melmon KL, ed. *Drug therapeutics — concepts for physicians*. 2nd ed. New York, NY: Elsevier/North-Holland Publishing Company; 1981: 201-208.
4. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *Int J Epidemiol* 1980;9:361-367.
5. Feinstein AR. Epidemiologic analyses of causation: the unlearned scientific lessons of randomized trials. *J Clin Epidemiol* 1989;42: 481-489.

6. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 1991; 10:577–581.
7. Mossey JM, Shapiro E. Self-rated health: a predictor of mortality among the elderly. *Am J Public Health* 1982;72:800–808.
8. Charlson ME, Sax FL, MacKenzie CR, et al. Assessing illness severity: does clinical judgment work? *J Chronic Dis* 1986;39:439–452.
9. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *N Engl J Med* 1980;303:1038–1041.
10. Bergkvist L, Adami HO, Persson I, et al. The risk of breast cancer after estrogen and estrogen-progestin replacement. *N Engl J Med* 1989;321: 293–297.
11. Irwin KL, Rosero-Bixby L, Oberle MW, et al. Oral contraceptives and cervical cancer risk in Costa Rica: detection bias or causal association? *JAMA* 1988;259:59–64.
12. Yuen J, Persson I, Bergkvist L, et al. Hormone replacement therapy and breast cancer mortality in Swedish women: results after adjustment for 'healthy drug-user' effect. *Cancer Causes Control* 1993;4:369–374.
13. Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983;2: 267–271.
14. Byar DP. Problems with using observational databases to compare treatments. *Stat Med* 1991;10:663–666.
15. Psaty BM, Koepsell TD, Siscovick D, et al. An approach to several problems in using large databases for population-based case-control studies of the therapeutic efficacy and safety of antihypertensive medicines. *Stat Med* 1991;10: 653–662.
16. Hoes AW, Grobbee DE, Lubsen J. Primary prevention in hypertension: valid conclusions from observational studies. *Circulation* 1991;84(6 Suppl):VI78–183.
17. Salive ME, Mayfield JA, Weissman NW. Patient outcomes research teams and the Agency for Health Care Policy and Research. *Health Serv Res* 1990;25:697–708.
18. Ellwood PM. Shattuck lecture—Outcomes management: a technology of patient experience. *N Engl J Med* 1988;318:1549–1556.
19. Epstein AM. The outcomes movement—Will it get us where we want to go? *N Engl J Med* 1990;323:266–270.
20. Stein REK, Gortmaker SL, Perrin EC, et al. Severity of illness: concepts and measurements. *Lancet* 1987;2:1506–1509.
21. Knaus WA, Wagner DP, Zimmerman JE, et al. Variations in mortality and length of stay in intensive care units. *Ann Intern Med* 1993;118: 753–761.
22. Lemeshow S, Teres D, Klar J, et al. Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993;270:2478–2486.
23. Horn SD, Sharkey PD, Buckle JM, et al. The relationship between severity of illness and hospital length of stay and mortality. *Med Care* 1991;29:305–317.
24. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993;270:2957–2963.
25. Parkerson GR Jr, Broadhead WE, Tse CKJ. The Duke Severity of Illness checklist (DUSOI) for measurement of severity and comorbidity. *J Clin Epidemiol* 1993;46:379–393.
26. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–383.
27. Starfield B, Weiner J, Mumford L, et al. Ambulatory care groups: a categorization of diagnoses for research and management. *Health Serv Res* 1991;26:53–74.
28. Greenfield S, Nelson EC. Recent developments and future issues in the use of health status assessment measures in clinical settings. *Med Care* 1992;30(5 Suppl):MS23–MS41.
29. Deyo RA, Carter WB. Strategies for improving and expanding the application of health status measures in clinical settings: a researcher-developer viewpoint. *Med Care* 1992;30(5 Suppl):MS176–MS186.
30. Wennberg JE, Freeman JL, Culp WJ. Are hospital services rationed in New Haven or overutilised in Boston? *Lancet* 1987;1:1185–1189.
31. Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? *Am J Med* 1984;77:703–708.
32. Gray-Donald K, Kramer MS. Causality inference in observational vs experimental studies: an empirical comparison. *Am J Epidemiol* 1988;127: 885–892.
33. Weiss NS. Therapeutic efficacy: nonexperimental studies. En: Weiss NS. *Clinical epidemiology: the study of the outcome of illness*. New York, NY: Oxford University Press; 1986:72–89.
34. Selby JV, Friedman GD, Quesenberry CP Jr, et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* 1992;326:653–657.

35. Bross IDJ. Pertinency of an extraneous variable. *J Chronic Dis* 1967;20:487-495.
36. Bross IDJ. Spurious effects from an extraneous variable. *J Chronic Dis* 1966;19:637-647.
37. Winkelstein W Jr, Shillitoe EJ, Brand R, et al. Further comments on cancer of the uterine cervix, smoking, and herpesvirus infection. *Am J Epidemiol* 1984;119:1-8.
38. Breslow NE, Day NE. *Statistical methods in cancer research: Vol 1, The analysis of case-control studies*. Lyon, France: International Agency for Research on Cancer; 1980:96. (IARC scientific publication 32.)
39. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986;73:1-11.
40. Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* 1991;2:155-158.
41. Horwitz RI, Feinstein AR. The application of therapeutic-trial principles to improve the design of epidemiology research: a case-control study suggesting that anticoagulants reduce mortality in patients with myocardial infarction. *J Chronic Dis* 1981;34:575-583.
42. Horwitz RI, Feinstein AR. Improved observational method for studying therapeutic efficacy: suggestive evidence that lidocaine prophylaxis prevents death in acute myocardial infarction. *JAMA* 1981;246:2455-2459.
43. Psaty BM, Koepsell TD, LoGerfo JP, et al. Beta-blockers and primary prevention of coronary heart disease in patients with high blood pressure. *JAMA* 1989;261:2087-2094.
44. Van der Meer JT, Van Wijk W, Thompson J, et al. Efficacy of antibiotic prophylaxis for prevention of native-valve endocarditis. *Lancet* 1992;339:135-139.
45. Imperiale TF, Horwitz RI. Does prophylaxis prevent postdental infective endocarditis? A controlled evaluation of protective efficacy. *Am J Med* 1990;88:131-136.
46. Hutchinson TA, Polansky SM, Feinstein AR. Postmenopausal oestrogens protect against fractures of hip and distal radius: a case-control study. *Lancet* 1979;2:705-709.
47. Weiss NS, Ure CL, Ballard JH, et al. Decreased risk of fractures of the hip and lower forearm with postmenopausal use of estrogen. *N Engl J Med* 1980;303:1195-1198.
48. Paganini-Hill A, Ross RK, Gerkins VR, et al. Menopausal estrogen therapy and hip fractures. *Ann Intern Med* 1981;95:28-31.
49. Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 1991;20:47-63.
50. Barrett-Connor E. Postmenopausal estrogen and prevention bias. *Ann Intern Med* 1991;115:455-456.
51. Mulinare J, Cordero JF, Erickson JD, et al. Periconceptional use of multivitamins and the occurrence of neural tube defects. *JAMA* 1988;260:3141-3145.
52. Mills JL, Rhoads GG, Simpson JL, et al. The absence of a relation between the periconceptional use of vitamins and neural-tube defects. *N Engl J Med* 1989;321:430-435.
53. Werler MM, Shapiro S, Mitchell AA. Periconceptional folic acid exposure and risk of recurrent neural tube defects. *JAMA* 1993;269:1257-1261.
54. Czeizel AE, Dudas I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med* 1992;327:1832-1835.
55. Recommendations for use of folic acid to reduce number of spina bifida cases and other neural tube defects. *MMWR* 1992;41(RR-14):1-7.
56. Baker F, Curbow B. The case-control study in health program evaluation. *Eval Prog Planning* 1991;14:263-272.
57. Thompson RS, Rivara FP, Thompson DC. A case-control study of the effectiveness of bicycle safety helmets. *N Engl J Med* 1989;320:1361-1367.
58. Pulungsih SP, Ittiravivongs A, Sutoto, et al. Assessment of the effectiveness of oral rehydration therapy against severe diarrheal dehydration. *Southeast Asian J Trop Med Public Health* 1992;23:420-426.
59. Sampalis JS, Lavoie A, Williams JI, et al. Impact of on-site care, prehospital time, and level of in-hospital care on survival in severely injured patients. *J Trauma* 1993;34:252-261.
60. Greenland S, Watson E, Neutra RR. The case-control method in medical care evaluation. *Med Care* 1981;19:872-878.