

## Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica<sup>1</sup>

Segunda edición

Richard K. Riegelman y Robert P. Hirsch

PARTE III:

Capítulo 8. Diseño del estudio  
Capítulo 9. Resumen: el estudio de un estudio  
Capítulo 10. Ejercicios para detectar errores:  
estudios observacionales

---

<sup>1</sup>Título original: *Studying a Study and Testing a Test. How to Read the Medical Literature* Second edition. © Richard K. Riegelman, Robert P. Hirsch. Publicado por Little, Brown and Company, Boston, Massachusetts 02108, Estados Unidos de América. Los pedidos del libro en inglés deben dirigirse a esta dirección.

Versión en español autorizada por Little, Brown and Company; se publica simultáneamente en forma de libro (Publicación Científica 531) y como serie en el *Boletín de la Oficina Sanitaria Panamericana* Traducción de José María Borrás, revisada por el Servicio Editorial de la Organización Panamericana de la Salud.

© Little, Brown and Company, 1989. Todos los derechos reservados. Ninguna parte de esta publicación puede ser reproducida ni transmitida en ninguna forma ni por ningún medio de carácter mecánico o electrónico, incluidos fotocopia y grabación, ni tampoco mediante sistemas de almacenamiento y recuperación de información, a menos que se cuente con la autorización por escrito de Little, Brown and Company.

# DISEÑO DEL ESTUDIO

Una vez revisados los requisitos de la aplicación correcta de los componentes del marco uniforme, volvamos al principio para formular algunas preguntas básicas.

1. ¿Estaban definidos adecuadamente los objetivos del estudio?
2. ¿Cuál es el tipo de estudio? ¿Es apropiado para responder a las preguntas planteadas?
3. ¿Cuál es el tamaño de la muestra? ¿Es suficiente para responder a las preguntas del estudio?

Las respuestas a estos interrogantes le dirán al lector si los investigadores escogieron un diseño de estudio apropiado; esto es, aquel que define y puede responder a las preguntas planteadas.

## OBJETIVO DEL ESTUDIO

Supongamos que unos investigadores desean estudiar los efectos orgánicos de la hipertensión arterial. La imposibilidad de realizar biopsias renales y angiografías cerebrales puede obligarlos a explorar detenidamente el fondo del ojo. Supongamos que otros desean investigar los efectos a largo plazo de un nuevo fármaco para prevenir la osteoporosis y que el tiempo, el dinero y el deseo de publicar limitan su investigación a sus efectos a corto plazo sobre el metabolismo y la densidad ósea. Estos ejemplos ilustran la importancia de que los investigadores y el lector distingan entre lo que idealmente desearían estudiar aquellos y lo que de hecho estudian.

Al definir los objetivos del estudio es esencial formular una hipótesis específica. Cuando se estudia el daño orgánico producido por la hipertensión, los investigadores pueden formular la hipótesis de que el grado de daño orgánico está asociado con el grado de hipertensión. Sin embargo, esta hipótesis no es suficientemente concreta para ser contrastada. Por el contrario, para ello, es preciso formular una hipótesis específica como: el aumento del estrechamiento de las arterias de la retina, medido mediante fotografías sucesivas tomadas durante tres años de observación, está asociado con un aumento de la tensión arterial diastólica, utilizando como medida la media de tres mediciones realizadas al inicio del estudio. Esta última afirmación constituye una hipótesis de estudio específica que se puede abordar por medio de una investigación.

La incapacidad para definir claramente las hipótesis que se desean contrastar dificulta al investigador y al lector la selección y la valoración del diseño del estudio, respectivamente. También hace más difícil determinar si se alcanzaron los objetivos del estudio. En última instancia, como se señaló al presentar las pruebas de significación estadística, las pruebas de significación habituales no se pueden aplicar si no se define un resultado o desenlace específico que se pueda valorar.

## EVALUACIÓN DEL TIPO DE ESTUDIO

Una vez definidas las hipótesis específicas del estudio, el lector está preparado para identificar el tipo de estudio realizado y evaluar su idoneidad. Rara es la ocasión en la que solo hay un tipo de diseño apropiado para responder a la pregunta del estudio. A veces, las desventajas de un tipo de diseño pueden obstaculizar notablemente el cumplimiento de los objetivos del estudio. Para ayudar al lector a juzgar la idoneidad del diseño escogido, esbozaremos las ventajas y desventajas de los tipos básicos de estudio.

Los estudios de casos y controles o retrospectivos presentan la ventaja distintiva de que permiten estudiar enfermedades muy poco frecuentes. Si la enfermedad es rara, con los estudios de casos y controles se pueden detectar diferencias entre los grupos empleando muchos menos individuos de los que se necesitarían con otros diseños. El tiempo necesario para realizar un estudio de casos y controles es mucho menor, porque la enfermedad ya se ha manifestado. Este tipo de estudio permite a los investigadores examinar simultáneamente asociaciones entre varios factores y una enfermedad. Por ejemplo, es posible examinar diversas variables que puedan estar asociadas con el cáncer de colon. En el mismo estudio, se podrían investigar la dieta anterior, la cirugía, la colitis ulcerosa, los pólipos, el alcohol, los cigarrillos, los antecedentes familiares y muchas otras variables.

La mayor objeción a estos estudios es su tendencia a presentar una serie de errores metodológicos y sesgos, que ya se indicaron en los estudios hipotéticos de capítulos anteriores. Muchos sesgos, como el de declaración y el de recuerdo, comprometen la exactitud de los datos referentes a las características previas. Sin embargo, el estudio de casos y controles puede ser el método adecuado para revelar la existencia de una asociación previa, especialmente cuando no hay razones para creer que el conocimiento del investigador o de los sujetos estudiados sobre la presencia de la enfermedad influye en la valoración de los datos del pasado.

La ventaja principal de los estudios de cohortes es que ofrecen más garantías de que la característica estudiada precede al desenlace estudiado. Esta es una distinción fundamental cuando se valora una relación de causa-efecto. Los *estudios de cohortes concurrentes*, en los que se sigue la evolución de los pacientes durante largos períodos, son caros y requieren mucho tiempo. No obstante, es posible realizar un estudio de cohortes sin un período de seguimiento tan largo. Cuando existen datos fiables de épocas anteriores sobre la presencia o ausencia de la característica estudiada, estos se pueden utilizar para realizar un *estudio de cohortes no concurrentes*. En un estudio de cohortes no concurrentes la asignación de los individuos a los grupos se lleva a cabo a partir de los datos del pasado. Después de la asignación, el investigador puede investigar si la enfermedad se desarrolló posteriormente.

Por ejemplo, si conociéramos las concentraciones de colesterol de un grupo de adultos jóvenes medidas 15 años antes del inicio del estudio actual, podríamos seguir prospectivamente a los pacientes que todavía no han desarrollado la consecuencia clínica de la hipercolesterolemia para valorar el desarrollo ulterior de enfermedad coronaria, accidentes vasculares cerebrales u otras consecuencias que po-

drían aparecer poco tiempo después de iniciar el estudio. El elemento fundamental que caracteriza a todos los estudios de cohortes es la identificación de los individuos del grupo de estudio y del grupo control sin conocer si se ha desarrollado la enfermedad estudiada.<sup>1</sup>

Los estudios de cohortes permiten delimitar diversas consecuencias que pueden estar asociadas con un único factor de riesgo. Los investigadores pueden estudiar simultáneamente la relación entre la hipertensión y el accidente vascular cerebral, el infarto de miocardio, la insuficiencia cardíaca o la enfermedad renal. Los estudios de cohortes pueden ayudar a comprender con más detalle el efecto de un factor etiológico sobre varios desenlaces. No obstante, la posibilidad de que con estos estudios se descubran nuevos factores etiológicos es menor que con los de casos y controles.

Ambos tipos de estudios son observacionales; esto es, en ellos las características y los desenlaces de los individuos no se imponen, sino que se observan. Los ensayos clínicos aleatorios se distinguen de los estudios observacionales en que el investigador interviene asignando al azar a los individuos al grupo de control y al de estudio. La capacidad de asignar a los individuos contribuye a asegurar que la característica estudiada, y no alguna predisposición subyacente, produce los resultados del estudio. Cuando se realizan adecuadamente, los ensayos clínicos aleatorios pueden cumplir los tres criterios de causa contribuyente.

En los capítulos 11 y 12 examinaremos en profundidad las ventajas e inconvenientes de los ensayos clínicos controlados.

Puede ser útil examinar una posible secuencia de estudios realizados para comprobar la existencia de una causa contribuyente. Muchas veces, los investigadores inician una investigación con un estudio de casos y controles con objeto de indagar la existencia de posibles causas.<sup>2</sup> Estos estudios ofrecen la ventaja de la rapidez, el bajo costo y la capacidad de investigar numerosas causas a la vez. Además, tienen por objeto demostrar la existencia de asociaciones o relaciones entre factores. A veces, pueden ser fiables para garantizar que la causa precede al efecto, si bien pueden dejar algunas dudas sobre cuál precede a cuál.

Una vez que se ha comprobado la existencia de una asociación en uno o más estudios de casos y controles, los investigadores llevan a cabo frecuentemente un estudio de cohortes concurrentes. A pesar de la necesidad de interpretar los datos cuidadosamente —como se demostró en el ejemplo del abandono del tabaco—, con los estudios de cohortes concurrentes a menudo es posible comprobar que la causa precede al efecto.

Después de demostrar que una posible causa precede al efecto, los investigadores pueden utilizar un estudio de intervención, por ejemplo, un ensayo clínico aleatorio, para comprobar que la modificación de la causa altera el efecto. En este estudio, los individuos se asignan al azar y a ciegas al grupo de estudio y al de control. Solo el grupo de estudio es expuesto a la posible causa o al tratamiento propuesto. El ensayo clínico aleatorio cumple idealmente con los tres criterios de causa contribuyente y, por ello, es un instrumento potente para demostrar que una determinada causa es contribuyente.

---

<sup>1</sup> Los estudios de cohortes se realizan cada vez con más frecuencia utilizando bases de datos que se han completado antes de iniciar el estudio. Esta situación representa el caso extremo de los estudios no concurrentes, a veces denominados *estudios de cohortes retrospectivos*. El elemento clave que transforma a estos estudios en estudios de cohortes es el hecho de que la identificación de los sujetos para su inclusión en el estudio se realiza sin saber si han desarrollado la enfermedad.

<sup>2</sup> Con la creciente disponibilidad de grandes bases de datos, los investigadores pueden empezar realizando un estudio de cohortes no concurrentes, que también puede llevarse a cabo rápidamente y a bajo costo

En teoría, esta secuencia de estudios funcionaría de la siguiente manera: para comprobar que los estrógenos sin progesterona son una causa contribuyente del cáncer de útero, un investigador podría utilizar, en primer lugar, un estudio de casos y controles con el que se examinarían diversas variables, incluyendo la asociación hipotética entre los estrógenos y el cáncer de útero. Si se encontrara una asociación, se podría realizar un estudio de cohortes concurrentes para establecer con más firmeza que la toma de estrógenos sin progesterona precede al desarrollo de un cáncer de útero. Los investigadores desearían estar seguros de que los estrógenos no se están administrando para tratar una pérdida de sangre que pudiera ser un signo de cáncer de útero. En un estudio de cohortes concurrente se seleccionarían grupos similares de mujeres que han tomado estrógenos y de mujeres que no los han tomado; se seguiría a ambos grupos durante un período de tiempo y se investigaría si las mujeres que toman estrógenos desarrollan cáncer de útero con más frecuencia que las que no los toman. Este estudio de cohortes concurrente puede demostrar más firmemente que la toma de estrógenos precede al desarrollo de un cáncer de útero.

En teoría, la investigación proseguiría con un ensayo clínico controlado en el cual las mujeres se asignarían al azar al grupo de las que toman estrógenos sin progesterona o al de las que toman placebo. Sin embargo, después de obtener pruebas de que los estrógenos son peligrosos, no sería ético o podría ser imposible realizar un ensayo clínico controlado sobre la relación entre los estrógenos y el cáncer de útero. En este caso, los investigadores podrían realizar un *experimento natural*, para respaldar la idea de que los estrógenos son una causa contribuyente del cáncer de útero. Este experimento natural se podría efectuar en el caso de que un grupo de mujeres dejase de tomar estrógenos como resultado de la publicidad generada por los estudios. Si la tasa de cáncer de útero del grupo de mujeres que dejan de tomar estrógenos disminuyera y no lo hiciera la de las mujeres que continúan tomándolos, este experimento aportaría la prueba más convincente disponible de que la modificación de la causa altera el efecto.

Los tipos básicos de estudios presentados en este libro no son los únicos que se encuentran en la literatura médica. Muchas veces se llevan a cabo estudios transversales. En estas investigaciones, la característica estudiada y el desenlace se miden en el mismo momento; en otras palabras, la asignación y la valoración se realizan en el mismo momento. Los estudios transversales son relativamente baratos y rápidos. Son útiles cuando se espera que es improbable que la exposición cambie con el tiempo o que el tiempo entre la exposición y el desarrollo de la enfermedad sea muy corto. Cuando se quiere estudiar la relación entre la tromboflebitis y la toma de píldoras anticonceptivas se puede usar un estudio transversal. Uno podría desear estudiar si es más probable que las mujeres con tromboflebitis estén tomando píldoras anticonceptivas en el momento en que aparece la tromboflebitis.

## TAMAÑO DE LA MUESTRA

Una vez que se han valorado los objetivos y el tipo de estudio, el lector debe concentrarse en el tamaño de la muestra de individuos seleccionada. Además, ha de preguntarse si el número de pacientes incluidos en el estudio es suficiente para demostrar con una probabilidad razonable la existencia de una diferencia estadís-

ticamente significativa entre las muestras del estudio y si dicha diferencia existe realmente en la población de la cual se han extraído las muestras.

Cuando nos preguntamos por la idoneidad del tamaño de la muestra, es preciso distinguir entre los estudios de casos y controles, por un lado, y los ensayos clínicos aleatorios y los estudios de cohortes, por el otro. Recuerde que en los estudios de casos y controles el desenlace es una característica del paciente, mientras que en los estudios de cohortes y en los ensayos clínicos aleatorios el desenlace es una enfermedad. Por consiguiente, en el estudio de casos y controles sobre la tromboflebitis, estaríamos interesados en conocer la magnitud de la verdadera diferencia en la toma de píldoras anticonceptivas que, dado el tamaño de la muestra utilizado en el estudio, es probable demostrar como estadísticamente significativa. Al realizar un estudio de cohortes o un ensayo clínico aleatorio, uno está interesado en saber cuán pequeña debe ser una verdadera diferencia en la probabilidad de desarrollar una enfermedad como la tromboflebitis para que sea probable demostrar que es estadísticamente significativa, dado el tamaño de la muestra empleado en el estudio.

Las respuestas a estas cuestiones dependen de la magnitud de los errores de tipo I y de tipo II que el lector y el investigador estén dispuestos a tolerar. Recuerde que el error de tipo II es la probabilidad de no demostrar una diferencia estadísticamente significativa cuando realmente existe una diferencia en la población de la que se ha extraído la muestra del grupo de estudio y del de control.

El error de tipo I aceptado habitualmente es 5%. El error de tipo II aceptado está abierto a discusión. La mayoría de los investigadores desearían que la probabilidad de no demostrar una diferencia estadísticamente significativa cuando realmente existe una verdadera diferencia fuese 10% o menor. Si aceptamos un error de tipo I de 5% y uno de tipo II de 10% y utilizamos las tablas estadísticas estándar, se pueden extraer las siguientes conclusiones sobre el tamaño de la muestra.<sup>3</sup>

1. Si el grupo de estudio y el de control están formados por 100 individuos cada uno, el estudio tiene potencia estadística para detectar una diferencia estadísticamente significativa, si la frecuencia real de un desenlace, como la muerte en una población, es de 20% o más alta en una población y de 5% o menor en la otra.
2. Si tanto el grupo de estudio como el de control están formados por 250 individuos cada uno, la investigación tiene una potencia estadística para detectar una diferencia estadísticamente significativa, si la frecuencia real de un desenlace en una población es de 20% o más alta y en la otra población, de 10% o menor.
3. Si el grupo de estudio y el de control están formados por 500 individuos cada uno, la investigación tiene una potencia estadística para detectar una diferencia estadísticamente significativa, si la verdadera frecuencia de un desenlace en una población es de 10% o más alta y en la otra, de 5% o menor.
4. Cuando la frecuencia de ambos desenlaces es baja y la diferencia entre los porcentajes de los desenlaces es pequeña, se necesitan muestras grandes para detectar una diferencia significativa. Por ejemplo, para detectar que una verdadera diferencia entre dos poblaciones es estadísticamente significativa, cuando la frecuencia del desenlace en un grupo es de 2% y en el otro, de 1%, se necesitarían más de 3 500 individuos en cada grupo. Cuando utilice estas orientaciones, recuerde que, incluso con una potencia estadística alta, una muestra

<sup>3</sup> Fleiss JL. *Statistical methods for rates and proportions* 2a ed. Nueva York: Wiley; 1981, pp 260–280.

concreta extraída de poblaciones en las que existen verdaderas diferencias todavía puede ser insuficiente para detectar una diferencia estadísticamente significativa.

Estas estimaciones son útiles para el lector de la literatura médica, porque le permiten estimar si el estudio tiene una posibilidad real de demostrar una significación estadística a partir del tamaño de su muestra.

Ahora aplicaremos estos principios para demostrar por qué los estudios de casos y controles son útiles para estudiar enfermedades raras que afectan a un número relativamente bajo de individuos. No olvide que el término *desenlace* se refiere a una característica del paciente en los estudios de casos y controles y a la enfermedad misma en los de cohortes y en los ensayos clínicos aleatorios. El siguiente ejemplo hipotético muestra la dificultad de demostrar una diferencia estadísticamente significativa cuando se emplea un estudio de cohortes con el fin de investigar una enfermedad muy poco frecuente.

Los investigadores deseaban estudiar si la toma de píldoras anti-conceptivas está asociada con la infrecuente aparición de accidentes vasculares cerebrales (AVC) en la mujer joven. Para ello, siguieron durante 10 años a 20 000 mujeres que tomaban píldoras anticonceptivas y a 20 000 que utilizaban otros métodos de planificación familiar. Después de gastar varios millones de dólares en el seguimiento, encontraron 2 casos de accidente vascular cerebral entre las usuarias de las píldoras y uno entre las no usuarias. La diferencia no fue estadísticamente significativa.

Cuando una enfermedad es muy rara, como los AVC en las mujeres jóvenes, muchas veces es preciso estudiar a un número muy alto de individuos para detectar una diferencia estadísticamente significativa, si se utiliza un estudio de cohortes. Suponga, por ejemplo, que la proporción de accidentes vasculares cerebrales en las mujeres jóvenes que no toman la píldora es 1 por 100 000 ó 0,001%. Supongamos también que la píldora aumenta 10 veces el riesgo de padecer la enfermedad, es decir, hasta 1 por 10 000 ó 0,01%. La diferencia en el desenlace es de 0,01% a 0,001% ó 0,009%. El uso de un estudio de cohortes para demostrar una diferencia estadísticamente significativa, existiendo una diferencia verdadera tan pequeña, puede requerir más de 100 000 mujeres en cada grupo.

Por otro lado, si se realiza un estudio de casos y controles en mujeres jóvenes con un AVC como grupo de estudio y mujeres jóvenes sin AVC como grupo de control, el desenlace que se medirá será la toma de píldoras anticonceptivas, en lugar de los AVC. La inclusión de 100 mujeres en cada grupo sería suficiente para detectar una diferencia estadísticamente significativa si existiera una diferencia real en la toma de píldoras anticonceptivas de 20% entre las que padecen AVC y de 5% entre el grupo sin AVC. En este ejemplo, es factible realizar un estudio de casos y controles sobre relación entre las píldoras anticonceptivas y los AVC, utilizando solo una pequeña proporción de los individuos requeridos para estudiar la misma cuestión con un estudio de cohortes. Por lo tanto, aquel estudio de cohortes estaba condenado al fracaso desde el principio; un estudio de casos y controles habría sido mucho más apropiado. Siempre que en una investigación no se logre detectar una diferencia estadísticamente significativa, el lector se debe preguntar si el tamaño de la muestra del estudio era suficiente para detectarlas.

En el capítulo 11 exploraremos en mayor profundidad las implicaciones del tamaño de la muestra. La evaluación del diseño de un estudio exige que el lector valore sus objetivos, la idoneidad del tipo de estudio utilizado y la suficiencia del tamaño de la muestra. El lector capaz de comprender estos problemas básicos puede evaluar los resultados de un estudio de forma más inteligente.

# RESUMEN: EL ESTUDIO DE UN ESTUDIO

## DISEÑO DEL ESTUDIO

Analizar si un estudio se diseñó adecuadamente para responder a las preguntas planteadas, el revisor debe determinar, en primer lugar, si los objetivos del estudio se definieron con suficiente precisión y si la hipótesis se formuló de forma clara. A continuación, debe preguntarse si el tamaño de la muestra fue suficiente para reponder a la pregunta planteada en el estudio.

El lector de la literatura también debe decidir si el diseño empleado fue el apropiado para contestar a la cuestión planteada, teniendo en cuenta las ventajas y desventajas de cada tipo de estudio.

## ASIGNACIÓN

Los investigadores intentan formar grupos de estudio y de control que sean semejantes en todas las características excepto en la estudiada. Los estudios de casos y controles y los de cohortes pueden contener un *sesgo de selección*. Este sesgo se produce cuando el grupo de estudio y el de control se escogen de tal forma que las frecuencias de un factor de riesgo o pronóstico que influya en el resultado de la investigación son distintas en ambos grupos. El sesgo de selección es un tipo especial de variable de confusión producida por diferencias aleatorias entre el grupo de estudio y el de control que están relacionadas con el desenlace estudiado. Cuando aparecen variables de confusión potenciales es importante identificarlas para poder incluirlas en el análisis.

## VALORACIÓN

Para valorar el desenlace de un estudio, el lector debe considerar si se han cumplido los criterios de una valoración válida. Los investigadores deben demostrar que escogieron una medida adecuada del desenlace, aquella que mide lo que se propone medir. Deben haber realizado una valoración exacta; esto es, aquella medición que se aproxima a la medida verdadera del fenómeno. La medición de un desenlace debe ser completa. Por último, deben haber considerado si el proceso de la observación influyó en el desenlace valorado.

## ANÁLISIS

El análisis implica el uso de métodos estadísticos para investigar el efecto al azar y el de los sesgos, así como para realizar estimaciones puntuales sobre los datos de la muestra. Es posible que el sesgo o el azar produzcan variables de confusión

que pueden prevenirse al inicio del estudio apareando a los grupos de estudio y de control o bien apareando a los individuos de cada grupo. Las pruebas de significación estadística son métodos de contraste de hipótesis que sirven para valorar los efectos del azar en los resultados de una investigación. Estas pruebas suponen una hipótesis y conllevan errores de tipo I y de tipo II. Son un método de prueba por eliminación. En los estudios clínicos, la razón de productos cruzados y el riesgo relativo son las medidas básicas de la fuerza de una asociación. Los intervalos de confianza de 95% están sustituyendo paulatinamente a las pruebas de significación o se dan como información adicional. Estos intervalos proporcionan el valor numérico observado o estimación puntual del valor de la población, así como el intervalo de valores que contiene el verdadero valor poblacional (parámetro) con un nivel de confianza de 95%. Las pruebas de significación estadística y los intervalos de confianza se calculan con los mismos métodos estadísticos. A veces, el lector puede usar rápidamente los intervalos de confianza para realizar una prueba de significación estadística.

## INTERPRETACIÓN

Los autores de un estudio deben preguntarse qué significan los resultados para las personas incluidas en la investigación. Deben cuestionar también si la magnitud de las diferencias o la fuerza de la asociación es tal que los resultados son clínicamente útiles o importantes. Asimismo, han de plantearse si se han cumplido los criterios de una relación de causa-efecto.

Es preciso, además, que los autores y el lector apliquen el concepto clínico de causa contribuyente. La causa contribuyente requiere que la supuesta causa esté asociada con el efecto y lo preceda y, por añadidura, que la modificación de la causa altere el efecto. No se exige que la causa sea necesaria o suficiente para producir el efecto. Cuando no se consigue cumplir los criterios definidos, los criterios auxiliares, accesorios o de apoyo pueden ayudar a respaldar la observación de la existencia de una relación de causa-efecto. Estos criterios son la fuerza de la asociación, la consistencia, la plausibilidad biológica y la relación dosis-respuesta.

## EXTRAPOLACIÓN

Finalmente, el lector debe preguntarse qué significan los resultados del estudio para los individuos no incluidos en el mismo. Al extrapolar los resultados a un individuo, es preciso que el lector distinga entre el riesgo relativo y el absoluto. Cuando se extrapola a nuevos grupos de sujetos con el factor de riesgo, el número de pacientes que es preciso tratar ofrece una medida de síntesis útil sobre el número de individuos que es necesario tratar para obtener un desenlace negativo menos o uno positivo más. También es preciso considerar el riesgo atribuible porcentual. Cuando se extrapola a poblaciones compuestas por individuos con y sin el factor de riesgo, se debe considerar el riesgo atribuible poblacional porcentual. Es importante reconocer el peligro que supone la extrapolación lineal más allá del intervalo de los datos observados. También hay que tener en cuenta cómo las distintas características de una nueva población objetivo pueden influir en la capacidad de extrapolar los resultados.

Pocas investigaciones pueden zafarse de estos errores; no obstante, su presencia no invalida automáticamente una investigación. Es responsabilidad del lector atento identificar estos errores y tenerlos en cuenta cuando se aplican los resultados del estudio.

## PREGUNTAS ACERCA DEL ESTUDIO DE UN ESTUDIO

Ahora reuniremos el material precedente y veremos si usted puede aplicar lo que ha aprendido a varios artículos de investigación simulados. El método crítico para evaluar un estudio de investigación se perfila en la siguiente lista de preguntas que uno debe formularse cuando está estudiando un estudio.

1. Diseño del estudio: ¿estaba diseñado adecuadamente?
  - a. Los objetivos del estudio, ¿estaban definidos correctamente? Las hipótesis del estudio, ¿estaban formuladas con claridad?
  - b. ¿Cuál era el tipo de estudio? ¿Era el adecuado para responder a las preguntas planteadas?
  - c. ¿Cuál fue el tamaño de los grupos de estudio? ¿Era suficiente para contestar las preguntas del estudio?
  
2. Asignación: ¿se asignaron adecuadamente los pacientes al grupo de estudio y al de control?
  - a. Si el estudio fue de casos y controles o de cohortes, ¿pudo existir un sesgo de selección?
  - b. Si el estudio fue un ensayo clínico aleatorio, ¿se mantuvo la asignación al azar y a ciegas?
  - c. Sin tener en cuenta el tipo de estudio, ¿los grupos de estudio y de control fueron comparables respecto a características distintas del factor estudiado o pudo haber influido en los resultados una variable de confusión?
  
3. Valoración: ¿se valoró el desenlace adecuadamente en los grupos de estudio y de control?
  - a. La medida del desenlace, ¿era apropiada para los objetivos del estudio?
  - b. La medida del desenlace, ¿fue exacta, reflejando entonces el verdadero valor del fenómeno?
  - c. La medida del desenlace, ¿fue completa?
  - d. ¿Afectó el proceso de observación al desenlace?
  
4. Análisis: ¿comparó correctamente el desenlace en los grupos de estudio y de control en el análisis?
  - a. ¿Se ajustaron los resultados para tener en cuenta el efecto de posibles variables de confusión?
  - b. La prueba de significación estadística, ¿fue realizada correctamente para valorar la probabilidad de que la diferencia o la asociación observada fuese debida al azar si la hipótesis nula fuera verdadera?
  - c. ¿Se proporcionó la estimación puntual del valor de la población (parámetro) y su intervalo de confianza de 95%?
  - d. ¿Se consideró el número de hipótesis formuladas? Usando el enfoque bayesiano, ¿se asignó a cada hipótesis la probabilidad previa antes de empezar el estudio para poder calcular la probabilidad de la hipótesis después de obtener los datos?
  - e. ¿Podría el error de tipo I o el de tipo II explicar los resultados?

5. Interpretación: ¿se llegó a conclusiones válidas sobre el significado de la investigación para los sujetos incluidos en el estudio?
  - a. ¿Es la magnitud de la diferencia o de la fuerza de la asociación lo suficientemente grande como para ser clínicamente importante o útil?
  - b. ¿Se cumplieron los tres criterios de causa contribuyente?
  - c. ¿Los investigadores distinguieron entre causa contribuyente y causa necesaria y suficiente?
  - d. Si no se cumplieron los tres criterios de causa contribuyente, ¿se cumplieron los criterios auxiliares?
6. Extrapolación: ¿se realizaron correctamente las extrapolaciones a los individuos y situaciones no incluidos en el estudio?
  - a. ¿Consideraron los investigadores tanto el riesgo relativo como el absoluto al extrapolar los resultados a los individuos?
  - b. Cuando se extrapoló a nuevos grupos con el factor de riesgo, ¿los investigadores tomaron en consideración el riesgo atribuible porcentual?
  - c. Cuando se extrapoló a nuevos grupos formados por individuos con y sin el factor de riesgo, ¿los autores tuvieron en cuenta el riesgo atribuible poblacional porcentual?
  - d. ¿Los autores realizaron una extrapolación más allá del intervalo de los datos?
  - e. ¿Los autores consideraron las diferencias entre el grupo de estudio y la población objetivo?

# EJERCICIOS PARA DETECTAR ERRORES: ESTUDIOS OBSERVACIONALES

Los siguientes estudios hipotéticos incluyen errores del tipo ejemplificado en cada uno de los componentes del marco básico. Estos ejercicios para detectar errores se han diseñado con el fin de comprobar su capacidad para aplicar el marco básico al examen crítico de un estudio. Se presentan ejemplos de estudios de casos y controles y de cohortes. Por favor, lea los ejercicios y escriba una crítica de cada estudio. Al final de cada ejercicio encontrará una crítica en la que se señalan los errores más importantes.

Observe que el último ejercicio es el mismo que leyó en el primer capítulo. Compare su crítica actual de este ejercicio con la que escribió previamente para ver el progreso realizado.

## EJERCICIO No. 1: ESTUDIO DE CASOS Y CONTROLES

Se llevó a cabo un estudio de casos y controles para estudiar los factores asociados con el desarrollo en el feto de enfermedades cardíacas congénitas (ECC). El grupo de estudio estaba formado por 200 mujeres que habían tenido abortos espontáneos durante el primer trimestre, en los que se detectaron malformaciones cardíacas congénitas. El grupo control estaba compuesto por 200 mujeres con abortos inducidos en el primer trimestre y en los que no se hallaron esas malformaciones.

Se intentó entrevistar a todas las mujeres durante el primer mes posterior al aborto, para determinar qué factores del embarazo podrían estar asociados con una ECC. Se estudiaron 100 variables. Los encuestadores consiguieron que participaran 120 de las 200 mujeres del grupo de estudio y 80 de las 200 del grupo de control. El resto de mujeres rehusaron participar.

Los investigadores encontraron las siguientes diferencias entre las mujeres cuyos fetos tenían ECC y aquellas cuyos fetos no la tenían.

1. La *ventaja* de tomar medicamentos contra la náusea de las mujeres con fetos que presentaban ECC fue tres veces más elevada que la de las mujeres con fetos sin ECC. Esta diferencia fue estadísticamente significativa.
2. No se observaron diferencias en el uso de tranquilizantes entre el grupo de estudio y el de control.
3. La media de la edad de las mujeres cuyos fetos presentaron ECC fue de 23 años y la de las mujeres del grupo control, 18. Los resultados fueron estadísticamente significativos.
4. Las mujeres del grupo de estudio bebían una media de 3,7 tazas de café diarias, mientras que las mujeres con fetos sin ECC bebían una media de 3,5 tazas. Esa diferencia también fue estadísticamente significativa.
5. Entre las 96 variables restantes, los autores observaron que la *ventaja* de tener el pelo rubio y de medir más de 167 centímetros era el doble en las mujeres que dieron a

luz fetos con ECC. Ambas diferencias fueron estadísticamente significativas empleando los métodos estadísticos habituales.

Los autores llegaron a las siguientes conclusiones.

1. La medicación contra la náusea causa ECC, porque las mujeres que dan a luz fetos con ECC la toman con frecuencia.
2. Los tranquilizantes se pueden usar con seguridad en las mujeres embarazadas, ya que no están asociadas con un aumento del riesgo de ECC.
3. Dado que es más probable que las mujeres de 20 años de edad tengan fetos con ECC, se debe animar a las mujeres a que tengan sus hijos antes de los 20 años.
4. Como el café aumenta el riesgo de ECC, su consumo se debe eliminar completamente durante el embarazo, lo cual eliminaría en gran parte el riesgo de ECC.
5. A pesar de que no se haya formulado la hipótesis de que el pelo rubio y la talla pueden ser factores de riesgo de ECC, se ha demostrado que pueden ser factores predictivos importantes de la enfermedad.

## CRÍTICA: EJERCICIO No. 1

### Diseño del estudio

Los investigadores no formularon claramente los objetivos de su estudio. ¿Estaban interesados en un tipo específico de ECC? Las enfermedades congénitas del corazón son una serie de enfermedades que afectan a las válvulas, el septo y los vasos sanguíneos. Al reunir todas las enfermedades bajo el encabezamiento de ECC, estaban suponiendo la existencia de una etiología común para todas ellas. Además, no quedó clara la hipótesis concreta que se pretendía contrastar en el estudio. Los grupos escogidos consistían en uno de estudio, cuyas integrantes tuvieron un aborto espontáneo, y uno de control, a cuyas participantes se les indujo un aborto con su consentimiento. Es de esperar que estos grupos difirieran en varios aspectos. Hubiera sido preferible escoger grupos más comparables de mujeres, por ejemplo, aquellas que habían tenido un aborto inducido con y sin ECC o aquellas que habían tenido un aborto espontáneo con y sin ECC.

Con este diseño de estudio debe recordarse que solo se podían estudiar las ECC que eran suficientemente graves como para producir un aborto espontáneo. Aunque ello puede proporcionar información importante, los factores que causan ECC suficientemente grave como para producir un aborto pueden ser distintos de los que causan ECC en los recién nacidos a término.

### Asignación

Para determinar si existió un sesgo de selección, primero debemos preguntarnos si el grupo de estudio y el de control difirieron en algunos aspectos. Segundo, si estas diferencias influyeron en los resultados. Es probable que las experiencias de las mujeres que padecieron un aborto espontáneo difirieran en múltiples aspectos.

tos de las que lo tuvieron por inducción. Es probable también que las actitudes de las mujeres acerca de sus embarazos fueran distintas y que estas pudieran influir en la toma de medicamentos durante el embarazo. Tales diferencias entre el grupo de estudio y el de control pudieron influir en el resultado. Por consiguiente, en este estudio se pudo haber introducido un sesgo de selección.

## Valoración

La elevada tasa de pérdidas en el seguimiento de las participantes sugiere la posibilidad de que las mujeres a las que no se pudo seguir tuvieran características diferentes. Una tasa elevada de pérdidas en el seguimiento debilita las conclusiones que es posible extraer a partir de cualquier diferencia observada. Una posibilidad es el sesgo de recuerdo por parte de las participantes, especialmente cuando estas mujeres experimentaron una experiencia traumática, como el de tener un aborto con ECC, y se les solicitó varias veces que recordaran sucesos subjetivamente como el uso de medicamentos o el consumo de café. La notificación retrospectiva de la toma de fármacos, por ejemplo, puede estar influida por las emociones que provoca la pérdida del feto en las mujeres que experimentan un aborto espontáneo inesperado. El resultado puede ser un escrutinio más detallado de la memoria que conduzca a un recuerdo más preciso del uso de medicamentos.

## Análisis, interpretación y extrapolación

1. Aunque se supusiera que la inferencia sobre relación entre la medicación contra la náusea y la ECC es correcta, con ello no se demostraría la existencia de una relación de causa-efecto. Con los estudios de casos y controles no se pueden afirmar definitivamente qué factor es la “causa” y cuál es el “efecto”. Es posible que las mujeres que dieron a luz fetos con ECC tuvieran más náuseas y, por lo tanto, que tomaran más antieméticos. Antes de establecer una relación causal en el sentido clínico de causalidad (causa contribuyente), los investigadores deben demostrar que la causa postulada precede al efecto y que su modificación lo modifica. Los autores de este estudio hicieron una interpretación que no está necesariamente justificada por los datos. El ajuste de los resultados de los grupos de estudio y de control según la diferencia en la frecuencia de náusea como parte del análisis sería un método para evaluar más a fondo la posible relación entre los antieméticos y la ECC, aunque todavía se usara un estudio de casos y controles.

2. La ausencia de una diferencia entre los grupos en términos del uso de tranquilizantes no garantiza necesariamente la seguridad de estos medicamentos. Las muestras pueden ser demasiado pequeñas para examinar completamente el riesgo de que los tranquilizantes causen ECC. Un pequeño aumento del riesgo requiere muchos más sujetos de estudio antes de que la investigación tenga la potencia estadística suficiente y con un alto grado de certeza para demostrar una diferencia entre los grupos. Por esta razón, se pudo haber cometido un error de tipo II. Incluso aunque no existiera el riesgo de que los tranquilizantes causen ECC, no tendríamos garantías de que estos medicamentos produzcan otros efectos adversos sobre el feto que los conviertan en inseguros durante el embarazo. Por consiguiente, los investigadores extrapolaron los resultados bastante más allá de sus datos.

3. La diferencia de edad entre los dos grupos de mujeres puede estar relacionada tanto con el tipo de aborto como con la presencia o ausencia de ECC. Por ello, la edad puede constituir un sesgo de selección, si es más probable que las mujeres tengan un aborto inducido en la adolescencia que en edades más avanzadas; esta relación puede explicar, por sí sola, las diferencias de edad observadas entre ambos grupos. Aunque el estudio haya mostrado que el riesgo de ECC era menor para las embarazadas adolescentes, los riesgos médicos y sociales pueden superar este beneficio. La mera presencia de una diferencia estadísticamente significativa no significa que se haya alcanzado una conclusión clínicamente importante.

4. La diferencia en el consumo de café fue estadísticamente significativa, aunque no muy grande. Un resultado estadísticamente significativo es aquel cuya probabilidad de observarse por azar es baja, si no existen verdaderas diferencias en la población de la que se han extraído los datos de la muestra. Sin embargo, es clínicamente improbable que una reducción tan pequeña tenga un efecto notable sobre el riesgo de ECC. La significación estadística debe distinguirse de la importancia clínica y de la causa contribuyente. Beber café puede tener un efecto, pero con diferencias tan pequeñas como las observadas, uno debe tener cuidado en extraer demasiadas conclusiones.

5. No es sorprendente que, al analizar 100 variables, se encuentren algunas asociaciones estadísticamente significativas. Cuando se utilizan muchas variables, uno no puede aplicar el nivel de significación estadística habitual para rechazar la hipótesis nula de no asociación. El nivel de 5% habitual supone que se ha formulado una hipótesis antes de iniciar el estudio. Por ello, los autores no pueden concluir con seguridad que la estatura y el color del pelo sean factores de riesgo de la ECC.

Con un enfoque bayesiano se puede decir que la probabilidad de que el color del cabello y la estatura estén asociados con la ECC es menor. Por lo tanto, el observar una asociación puede representar un error de tipo I, dado que la probabilidad de detectar una asociación después de obtener los resultados del estudio todavía es relativamente baja.

## **EJERCICIO No. 2: ESTUDIO DE COHORTES**

Con el fin de estudiar los efectos de una unidad coronaria (UC) que funcione adecuadamente sobre los infartos de miocardio (IM), varios investigadores realizaron un estudio de cohortes concurrentes de los efectos de una nueva UC.

Durante el primer año de funcionamiento de la UC, se ingresaron 100 pacientes del grupo de estudio remitidos por sus médicos con el diagnóstico de sospecha de IM. En salas hospitalarias fuera de la UCC, se ingresaron 100 pacientes del grupo de control en los que se había descartado el diagnóstico de IM.

A los pacientes de la UC se les administró lidocaína si sus enzimas cardíacas eran positivas para IM a las 24 horas de su ingreso. Además, se les administraron tratamientos invasivos para valorar y tratar las oclusiones de sus arterias coronarias. Las complicaciones de los pacientes de sala se monitorearon y se trataron cuando se presentaron.

Al comparar a los pacientes de sala con los de la UC, los investigadores observaron que la media de la edad de los de la UC era de 58 años y la de los controles, 68. Una cuarta parte de los pacientes de la UC y un veinteaño de los de sala desarrollaron hipotensión. El 80% de los pacientes de la UC y 20% de los de sala presentaron arritmias ventriculares. Los investigadores siguieron la evolución de los pa-

cientes durante su hospitalización y el año posterior, y recogieron los siguientes datos sobre los desenlaces.

1. En 36% de los pacientes de la UC y en 30 de los de sala se encontraron finalmente pruebas enzimáticas o electrocardiográficas definitivas de IM.
2. Ocho pacientes de la UC y cuatro de los de sala fallecieron en el hospital. Estas diferencias no fueron estadísticamente significativas.
3. Los pacientes de la UC permanecieron hospitalizados durante una media de 12 días y los de la sala, durante 15 días. Las diferencias fueron estadísticamente significativas.
4. Ninguno de los pacientes que recibieron lidocaína en la UC falleció.
5. Un año después del alta médica, los antiguos pacientes de la UC eran capaces de hacer, en promedio, 20% más ejercicio que los de la sala.

Los autores llegaron a las siguientes conclusiones:

1. La atención de la UC aumenta la tasa de desarrollo de IM entre los pacientes ingresados en el hospital con dolor de pecho.
2. Dado que las diferencias entre las tasas de mortalidad no fueron estadísticamente significativas, las tasas de mortalidad fueron idénticas en ambos grupos.
3. Como las diferencias en la duración de la hospitalización fueron estadísticamente significativas, los investigadores concluyeron que, mediante la creación de la UC, habían demostrado un importante ahorro en los costos.
4. Habida cuenta de que la lidocaína previno todas las muertes, si se empleaba después del diagnóstico definitivo de IM, el uso de este medicamento en el momento del ingreso eliminaría toda la mortalidad debida al IM.
5. Ya que los pacientes de la UC toleraron mejor el ejercicio durante el año posterior al alta, la UC causa una mejora de la supervivencia a largo plazo.

## CRÍTICA: EJERCICIO No. 2

### Diseño del estudio

Los investigadores intentaban estudiar el efecto de una UC que funcionara bien y, para ello, decidieron realizar su estudio en una UC nueva. Sin embargo, las nuevas instalaciones no pueden operar a pleno rendimiento en su primer año. De este modo, los investigadores no seleccionaron las mejores condiciones para estudiar los efectos de una UC que funcionara bien. Además, no formularon sus hipótesis específicas antes de iniciar el estudio.

### Asignación

El sesgo de selección puede estar presente en este estudio, si los individuos con mal pronóstico fueron ingresados en la UC por sus médicos. Este factor puede ser importante, si los médicos ingresaron selectivamente en la UC a los pacientes más enfermos. En este caso, sería de esperar que el sesgo de selección influyese en los resultados.

## Valoración

Los investigadores encontraron una tasa más elevada de arritmias entre los pacientes de la UC, lo que podría ser el resultado del método empleado para valorarlas en dicha unidad. Si los pacientes de la UC fueron monitoreados continuamente —al contrario que los de la sala—, es posible que dada la intensidad con que fueron observados, se descubriera un porcentaje más alto de las arritmias desarrolladas.

## Análisis

Los investigadores también observaron que la media de la edad de los pacientes de la UC era menor que la de los de sala. Este factor pudo haber aparecido por azar o ser resultado del deseo del médico de ofrecer una asistencia más intensiva a los pacientes jóvenes. Es probable que la edad de los pacientes se asocie a desenlaces tales como la tolerancia al ejercicio después del IM, porque los hombres jóvenes toleran mejor el ejercicio. Este hecho podría explicar la diferencia observada, sin tener en cuenta si la diferencia ocurrió a causa de un sesgo o por azar. Las diferencias en la edad son una variable de confusión potencial que debe ser motivo de ajuste en el análisis.

## Interpretación y extrapolación

1. Los investigadores encontraron que un porcentaje menor de pacientes de la sala presentaron finalmente pruebas de IM y concluyeron que la tasa más elevada de IM en la UC había sido causada por la atención médica prestada en ella. El primer requisito para establecer una relación de causa-efecto es que la causa preceda al efecto. En esta situación, es probable que los pacientes ya hubiesen padecido o estuviesen padeciendo el IM cuando ingresaron en el hospital. Por eso, en muchos casos, el efecto (IM) puede haber precedido a la causa (ingreso en la UC). Hay pocas pruebas que apoyen la interpretación de que la UC esté asociada con una tasa de IM más elevada.

2. Los autores concluyeron que las tasas de mortalidad se debían considerar similares, porque no se observó una diferencia estadísticamente significativa entre dichas tasas. No demostrar una diferencia estadísticamente significativa no implica que esta diferencia no exista. Cuando el número de individuos incluidos en una muestra es muy bajo, se necesita una diferencia muy grande para demostrar que es estadísticamente significativa. Los autores no pensaron en la posibilidad de cometer un error de tipo II. Es posible que los pacientes más graves fueran ingresados en la UC y que, por este motivo, se esperara una tasa de mortalidad más alta entre ellos. Cuando el número de individuos estudiados es tan bajo, como en este estudio, es preferible presentar los resultados sin aplicar ninguna prueba de significación estadística. En este estudio existía una diferencia; aunque no fue estadísticamente significativa, el número de muertes observado no puede considerarse idéntico.

3. La estancia hospitalaria de los pacientes de la UC fue menos prolongada que la de los de la sala. Los resultados fueron estadísticamente significativos, lo que indica que no es probable que las diferencias fueran debidas al azar. El que estas diferencias sean importantes desde el punto de vista de los costos es otra cuestión, porque los costos adicionales de la atención en la UC pueden superar las pequeñas diferencias en la duración de la estancia hospitalaria. Esta consideración puede ilustrar la distinción que debe hacerse entre una diferencia estadísticamente significativa y una clínicamente importante.

4. La lidocaína se administró a los pacientes de la UC solamente después de establecer un diagnóstico enzimático definitivo, y el IM estaba presente desde hacía 24 horas. En ese momento, el riesgo de morir de IM había descendido en gran parte, especialmente el riesgo de arritmia. Es probable, por consiguiente, que la administración de lidocaína tuviera poco que ver con el hecho de que no se produjeran muertes entre aquellos a los que se administró. Los autores fueron más allá de los datos al extrapolar sus resultados a todos los pacientes de la UC y al no observar que el grupo de pacientes que recibieron lidocaína era diferente del de los pacientes recién ingresados en la UC.

5. Los autores concluyeron que la UC había mejorado la perspectiva de recuperación, dado que los pacientes ingresados en ella toleraban mejor el ejercicio un año más tarde. Existen pocas pruebas para establecer una relación de causa-efecto. Como los pacientes de la UC eran más jóvenes que los de la sala, era previsible que toleraran mejor el ejercicio. Además, en la UC murieron más pacientes. La capacidad de sobrevivir a un IM pudo haber seleccionado a un grupo de pacientes con una mayor tolerancia al ejercicio. Finalmente, no se aportaron pruebas para respaldar que la tolerancia al ejercicio un año después del IM estuviese realmente asociada con una supervivencia a largo plazo. Los autores debieron haber tenido más cuidado al relacionar un mejor pronóstico con la atención prestada en la UC; al hacerlo, extrapolaron bastante más allá de los datos que observaron.

## **EJERCICIO DE REPASO: ESTUDIO DEL TAMIZAJE MÉDICO EN UNA POBLACIÓN MILITAR**

En el primer año del servicio militar se ofreció la posibilidad a 10 000 soldados de 18 años de edad de participar en un examen médico de salud anual constituido por una historia clínica, un examen físico y diversas pruebas de laboratorio. El primer año participaron 5 000 y los 5 000 restantes no lo hicieron. Los 5 000 que participaron fueron escogidos como grupo de estudio y los 5 000 que no participaron formaron el grupo de control. A los que participaron el primer año se les practicaron exámenes médicos anuales durante su servicio militar.

Al finalizar el servicio, tanto a los 5 000 del grupo de estudio como a los 5 000 del grupo control se les hizo una amplia evaluación de su historia clínica y se les practicó un examen físico y una evaluación de laboratorio para determinar si las visitas anuales habían producido alguna diferencia en su salud y en sus estilos de vida.

Los investigadores obtuvieron la siguiente información:

1. Según el consumo de alcohol declarado, la tasa de alcoholismo de los no participantes fue dos veces más alta que la de los participantes.
2. Se establecieron el doble de diagnósticos en los participantes que en los no participantes.
3. Los participantes tuvieron un promedio de ascensos dos veces más alto que los no participantes.
4. No se observaron diferencias estadísticamente significativas entre las tasas de infarto de miocardio (IM) de ambos grupos.
5. No se encontraron diferencias entre los grupos respecto a la tasa de aparición de cáncer de testículo o de enfermedad de Hodgkin, que son los dos tipos de cáncer más frecuentes en la gente joven.

Después, los autores extrajeron las siguientes conclusiones:

1. El tamizaje anual puede reducir a la mitad la tasa de alcoholismo en la población en el servicio militar.
2. Dado que el número de enfermedades diagnosticadas en los participantes fue el doble que en los no participantes, sus enfermedades se estaban diagnosticando en un estadio temprano del proceso patológico, momento en el cual el tratamiento es más beneficioso.
3. Como a los participantes se les concedieron el doble de ascensos que a los no participantes, el programa de tamizaje tuvo que haber contribuido a la calidad de su trabajo.
4. Habida cuenta de que no se observaron diferencias entre las tasas de IM de ambos grupos, el tamizaje y la intervención sobre los factores de riesgo de la enfermedad coronaria no se deben incluir en un futuro programa de tamizaje sanitario.
5. Dado que la frecuencia de la enfermedad de Hodgkin y del cáncer de testículo fue igual en ambos grupos, los futuros exámenes de salud no deben incluir esfuerzos para diagnosticar estas enfermedades.

## **CRÍTICA: EJERCICIO DE REPASO**

### **Diseño del estudio**

Los investigadores solo establecieron como objetivo general el estudio del valor anual de los exámenes de salud. No definieron la población a la que deseaban aplicar sus resultados, no formularon una hipótesis específica, ni identificaron claramente las preguntas específicas de su estudio.

Si el objetivo de los investigadores era estudiar los efectos de los exámenes anuales de salud, no cumplieron su objetivo, ya que no disponían de pruebas de que los participantes del primer año tomaran realmente parte en los exámenes siguientes.

Por añadidura, la elección de los participantes pudo no ser la más apropiada para responder a la pregunta formulada en la investigación. En el estudio se seleccionaron personas jóvenes que ya habían sido sometidas a un tamizaje de enfermedades crónicas en virtud de haber superado las pruebas físicas para entrar en el servicio militar. Siendo un grupo joven y sano, es posible que el grupo estudiado no fuera la población adecuada para probar la utilidad de un examen de salud en poblaciones de más edad o en alto riesgo, como las de militares de más edad, en las que es de esperar que la frecuencia de enfermedades sea más alta.

### **Asignación**

Los individuos de este estudio se seleccionaron a sí mismos; ellos decidieron si participaban o no. Por lo tanto, los participantes pueden considerarse voluntarios. Los investigadores no presentaron ninguna prueba indicativa de que los que decidieron participar se diferenciasen en algún aspecto de los que no participaron. Es probable que aquellos tuvieran hábitos de salud distintos de los no participantes. Estas diferencias pudieron haber contribuido a las del desenlace. Dado que no se dispone de una evaluación inicial del grupo de control, no se sabe si sus integrantes eran diferentes y en qué forma de los del grupo de estudio. Por este motivo desconocemos si el grupo de control y el de estudio eran comparables.

Los individuos de ambos grupos se asignaron a sí mismos a partir de su participación en los exámenes de salud del primer año. Como esos exámenes se llevaron a cabo anualmente, los que habían participado al principio se pudieron haber retirado del estudio. Por ello, el estado de los individuos del grupo de estudio y del de control puede no reflejar de forma válida su participación real en el tamizaje.

## Valoración

La valoración del desenlace se realizó solo en aquellos que fueron licenciados del servicio militar. No se incluyó a los que permanecieron en el ejército. Los individuos que murieron durante el servicio militar no se hubieran incluido en la valoración de los que dejaron el servicio. Estos pudieron haber sido los sujetos más adecuados para valorar los beneficios potenciales conseguidos con el tamizaje.

Los que participaron en los múltiples exámenes de salud estaban sometidos a una observación más intensa que los que no lo hicieron. Esta diferente intensidad podría explicar que en ellos se establecieran más diagnósticos durante su servicio militar. Si bien los no participantes podían haber tenido el mismo número de enfermedades, no todas resultaron diagnosticadas.

## Análisis, interpretación y extrapolación

1. La tasa de alcoholismo de los participantes fue más baja que la de los no participantes, quizá a causa de las diferencias entre los grupos antes de su entrada en el estudio. Si fuese menos probable que los grandes bebedores participaran en el tamizaje, este hecho solo hubiera modificado la frecuencia de alcoholismo. En el análisis no se incluyeron datos comparativos de los participantes a su entrada en el estudio, ni se ajustaron según las diferencias. Además, la validez del método utilizado para valorar el consumo de alcohol es cuestionable. Como no existe un criterio uniforme de diagnóstico, es posible que existieran diferencias en el recuerdo y en la declaración. Aunque no se hubiera cometido ninguno de estos errores potenciales, no hay pruebas en el estudio de que el tamizaje por sí mismo fuera el factor causante de una tasa de alcoholismo más baja. La extrapolación a los militares en general excedió el intervalo de los datos.

2. Si el nivel de motivación más elevado estuviera asociado con la participación en el estudio y con los ascensos en el ejército, la motivación sería una variable de confusión, al estar relacionada con la participación y con el desenlace. Sin ajustar los datos según esta variable de confusión potencial, no se puede llegar a ninguna conclusión sobre la relación entre la participación y los ascensos.

3. Muchos de los que tuvieron IM podrían haber muerto y, de ese modo, estar excluidos de la valoración. Además, uno esperaría que la tasa de IM fuera más baja en una población joven. Incluso con el elevado número de participantes estudiados, el tamaño de la muestra pudo haber sido insuficiente para detectar diferencias estadísticamente

significativas para diferencias reales pero pequeñas entre los grupos. Si se supone que las modificaciones de los factores de riesgo del IM también alteran el pronóstico, en este estudio no existe una indicación de que los que participaron tuvieran más factores de riesgo identificados o bien más factores de riesgo alterados. Es posible que los efectos de las posibles modificaciones de los factores de riesgo no se hagan aparentes hasta años después de que los participantes hayan abandonado el servicio militar. Por lo tanto, con este estudio no se puede determinar si el tamizaje de los factores de riesgo coronarios modifica el pronóstico de la enfermedad.

4. La ausencia de diferencias entre las tasas de aparición de cáncer de testículo y de la enfermedad de Hodgkin no se puede valorar a partir de los que dejaron el servicio con vida. Aunque esas tasas fueran idénticas, dicen poco sobre el fracaso o el éxito del programa de tamizaje. Un programa de tamizaje del cáncer pretende detectar la enfermedad en un estadio temprano, pero no intenta prevenirla. Por eso, la tasa de aparición de cáncer no se puede utilizar para evaluar el éxito de un programa de tamizaje. Uno esperaría encontrar tasas idénticas de desarrollo de ambos tipos de cáncer. El estadio de la enfermedad en el momento del diagnóstico y el pronóstico de los que desarrollaron una de las dos enfermedades serían medidas más apropiadas para valorar el éxito de dicho programa. Como no se presentan estos datos, no es posible hacer interpretaciones.

Una vez criticados estos ejercicios de detección de errores, el lector puede sentirse desanimado, pero sepa que la mayor parte de los estudios de investigación contienen bastantes menos errores que los ejercicios que acabamos de presentar. Sin embargo, puede ser de ayuda para el lector recordar que algunos errores son inevitables y que su detección no es sinónimo de invalidez de la investigación.

La práctica de la medicina clínica exige que los clínicos actúen sobre la base de probabilidades y la lectura crítica de la literatura médica les ayuda a definir con más exactitud esas probabilidades. El arte de la lectura de la literatura médica consiste en la capacidad de extraer conclusiones útiles a partir de datos inciertos. Aprender a detectar errores no solo ayuda al clínico a identificar las limitaciones de un estudio concreto, sino también a moderar la tendencia natural a poner en práctica automáticamente los resultados más recientes de la investigación.

## Corrección

En la sección de COMUNICACIÓN BIOMÉDICA del mes de julio, Vol. 111, No. 1, 1991, página 68, primer párrafo, líneas 5 y 6, donde dice "Para obtener una tasa de error tipo II . . .", debe decir "Para obtener un valor *P* . . .".